

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
IVAN FRANKO NATIONAL UNIVERSITY OF LVIV
STATE UNIVERSITY OF INFORMATION
AND COMMUNICATION TECHNOLOGIES



SECURITY OF MODERN INFORMATION AND COMMUNICATION SYSTEMS

PROCEEDINGS

of the International Scientific and Technical Conference

16-18 October 2025
Lviv (Ukraine)

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ФРАНКА
ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ



БЕЗПЕКА СУЧАСНИХ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМ

МАТЕРІАЛИ

Міжнародної науково-технічної конференції

16-18 жовтня 2025
м. Львів (Україна)

УДК [003.26+004+519.816]:004.056:65(063)

SMICS: Безпека сучасних інформаційно-комунікаційних систем: матеріали Міжнар. наук.-техн. конф., м. Львів, 16-18 жовтня 2025 р.: ЛНУ ім. І. Франка, 2025, 468 с.

Збірник містить тексти наукових матеріалів доповідей та тез учасників міжнародної науково-технічної конференції «SMICS: Безпека сучасних інформаційно-комунікаційних систем». Основною метою конференції є ознайомлення з сучасними досягненнями та висвітлення результатів наукових досліджень з усіх аспектів кібербезпеки та захисту інформації.

РЕДАКЦІЙНА КОЛЕГІЯ

ДИЯК І. – декан факультету прикладної математики та інформатики Львівського національного університету імені Івана Франка, д.ф.-м.н., професор;

ВЕНГЕРСЬКИЙ П. – завідувач кафедри кібербезпеки Львівського національного університету імені Івана Франка, д.ф.-м.н., доцент;

ГУТІК О. – професор кафедри кібербезпеки Львівського національного університету імені Івана Франка, к.ф.-м.н., доцент;

КОКОВСЬКА Я. – доцент кафедри дискретного аналізу та інтелектуальних систем, Львівського національного університету імені Івана Франка, к.ф.-м.н.

ПРИГАРА М. – доцент каф. технології машинобудування Ужгородського національного університету, к.т.н., доцент.

УКЛАДАННЯ ТА ВЕРСТКА МАТЕРІАЛІВ

ТРУШЕВСЬКИЙ В. – доцент кафедри кібербезпеки Львівського національного університету імені Івана Франка, к.ф.-м.н., доцент.

ОРГАНІЗАТОРИ КОНФЕРЕНЦІЇ

1. Львівський національний університет імені Івана Франка
2. Державний університет інформаційно-комунікаційних технологій
3. ГО “Асоціація спеціалістів кібербезпеки”
4. Академія кіберзахисту (CDA)
5. Rowan University, USA
6. Halmstad University, Sweden
7. University of the National Education Commission, Poland
8. Opole University of Technology, Poland

ПАРТНЕРИ КОНФЕРЕНЦІЇ



softserve

intellias



kamula.tech

ОРГКОМІТЕТ КОНФЕРЕНЦІЇ

Голови	ГЛАДИШЕВСЬКИЙ Р. - ректор Львівського національного університету імені Івана Франка, академік НАН України, д.х.н., професор ШУЛЬГА В. - ректор Державного університету інформаційно-комунікаційних технологій, д.і.н., професор КОРЧЕНКО О. – президент Громадської організації «Асоціація спеціалістів кібербезпеки», чл.-кор. Національної Академії Наук України, д.т.н., професор
Заступники голів	ДИЯК І. – декан факультету прикладної математики та інформатики Львівського національного університету імені Івана Франка, д.ф.- м.н., професор ІВАНЧЕНКО Є - директор Навчально-наукового інституту кібербезпеки та захисту інформації Державного університету інформаційно-комунікаційних технологій, д.т.н., професор УДОВИК І. - декан факультету інформаційних технологій, Національний технічний університет «Дніпровська політехніка», к.т.н., професор ХОХЛАЧОВА Ю.Є. – вчений секретар Громадської організації «Асоціація спеціалістів кібербезпеки», к.т.н., професор
Програмний комітет	БУГРІЙ О. - професор кафедри математичної статистики та диференціальних рівнянь Львівського національного університету імені Івана Франка, д.ф.-м.н., професор ВИНОКУРОВА О. - професор кафедри кібербезпеки Львівського національного університету імені Івана Франка, д.т.н., професор HNATYSHIN V. - Department of Computer Science Head, Full Professor, Rowan University (USA) ДАВИДЕНКО А. – провідний науковий співробітник Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, д.т.н., професор КАЗАКОВА Н. – завідувач каф. інформаційних технологій Одеського національного університету імені І.І.Мечнікова, д.т.н., професор КИРИК М. - професор кафедри кібербезпеки Львівського національного університету імені Івана Франка, д.т.н., професор КЛЬОЦ Ю. - завідувач каф. кібербезпеки Хмельницького національного університету, к.т.н., доцент КОНДРАТЮК С. – професор кафедри публічного управління та адміністрування Державного університету інформаційно-комунікаційних технологій, к.п.н., професор КУЧИК О. – завідувач кафедри міжнародної безпеки та стратегічних студій Львівського національного університету імені Івана Франка, к.і.н., доцент МУЛЕСА П. - завідувач каф. кібернеттики та прикладної математики Ужгородського національного університету, д.п.н., доцент ОПІРСЬКИЙ І. – завідувач каф. безпеки інформаційних технологій НУ “Львівська політехніка”, д.т.н., професор ПАРХУЦЬ Л. – професор кафедри кібербезпеки Львівського національного університету імені Івана Франка, д.т.н., професор РІЗАК В. М. - завідувач каф. твердотільної електроніки та інформаційної безпеки Ужгородського національного університету, д.ф.-м.н., професор SEMENOV S. – Head of the Department of Computer Engineering and Cybersecurity, Full Professor, Krakow University of the National Education

Commission (Poland)

ТКАЧ Ю. - завідувач кафедри кібербезпеки та математичного моделювання
НУ "Чернігівська політехніка", к.т.н., д.п.н., професор

ТКАЧУК Р. – начальник кафедри управління інформаційною безпекою
Львівського державного університету безпеки життєдіяльності, д.т.н., професор

TORSTENSON O. – Programme director Master's Programme in Network
Forensics (Halmstad, Sweden)

ШЕЛЕСТ М. - професор кафедри кібербезпеки та математичного
моделювання НУ "Чернігівська політехніка", д.т.н., професор

ШУВАР Р. – завідувач кафедри системного проектування Львівського
національного університету імені Івана Франка, к.ф.-м.н., доцент

Співголови

ВЕНГЕРСЬКИЙ П. – завідувач кафедри . кібербезпеки Львівського
національного університету імені Івана Франка, д.ф.-м.н., доцент

ГАЙДУР Г. – завідувач кафедри систем та технологій кібербезпеки
Державного університету інформаційно-комунікаційних систем, д.т.н.,
професор

Організаційний комітет

БРИЧ Т.Б. – доцент кафедри кібербезпеки Львівського національного
університету імені Івана Франка, к.т.н.-м. н.

ВЛАСОВ В. – доцент кафедри математичної статистики і диференціальних
рівнянь Львівського національного університету імені Івана Франка, к.ф.-м. н.

КОКОВСЬКА Я. – доцент кафедри дискретного аналізу та інтелектуальних
систем, Львівського національного університету імені Івана Франка, к.ф.-м.н.

ПОПАДЮК О. – доцент кафедри кібербезпеки Львівського національного
університету імені Івана Франка, PhD

ПРИГАРА М. – доцент каф. технології машинобудування Ужгородського
національного університету, к.т.н., доцент

ТРУШЕВСЬКИЙ В. – доцент кафедри кібербезпеки Львівського
національного університету імені Івана Франка, к.ф.-м.н., доцент

ЩЕРБИНА М. – ст. викладач кафедри кібербезпеки Львівського
національного університету імені Івана Франка

В'ЯЧАЛО М. – асистент кафедри кібербезпеки Львівського національного
університету імені Івана Франка

ЗЛАТОУС С. – асистент кафедри кібербезпеки Львівського національного
університету імені Івана Франка

КУЗБИТ Ю. – асистент кафедри кібербезпеки Львівського національного
університету імені Івана Франка

ЛАЩІВСЬКА С. – інженер кафедри кібербезпеки Львівського національного
університету імені Івана Франка

Вчений секретар

ГУТІК О. – професор кафедри кібербезпеки Львівського національного
університету імені Івана Франка, к.ф.-м.н., доцент

ЗМІСТ

КІБЕРБЕЗПЕКА У ДЕРЖАВНІЙ ПОЛІТИЦІ ТА МІЖНАРОДНИХ ВІДНОСИНАХ.....	15
EVT-калібрування порогів IDS на основі VaR/CVaR у цифровій юстиції: методологія, впровадження та міжнародні стандарти.....	16
<i>Валерій Бушков</i>	16
Кібердипломатія ЄС в умовах російської агресії в Україні.....	20
<i>Сергій Кондратюк¹</i>	20
Стратегія постквантового переходу ЄС у контексті квантової переваги як геополітичного чинника безпеки кіберпростору.....	26
<i>Олександр Кучик¹, Данило Щербина¹</i>	26
Вітчизняний досвід формування системи професійно-кваліфікаційної стандартизації сфери кібербезпеки, ІКТ та критичної інфраструктури.....	32
<i>Сергій Мельник¹</i>	32
Кібербезпека зсередини: як вбудовані бекдори підривають довіру до державних і міжнародних систем.....	34
<i>Олександр Полевод¹, Михайло Шелест¹, Юлія Ткач¹</i>	34
Аналіз ефективності навчання кібербезпеки в Україні в сучасних умовах.....	41
<i>Михайло Пригара¹, Володимир Щербина²</i>	41
Формування кібердипломатичних критеріїв стратегічного оцінювання кібербезпеки держави.....	46
<i>Володимир Шульга¹, Олександр Корченко¹, Євгенія Іванченко¹, Світлана Казмірчук¹, Сергій Кондратюк¹</i>	46
The formation of a national strategy for cyber protection of critical infrastructure facilities.....	52
<i>Oleksii Hutsaliuk^{1,†}, Iuliia Bondar^{2,†} and Oksana Yakusheva^{3,†}</i>	52
Research on Models, Methods, and Systems for Assessing the Negative Consequences of Personal Data Breaches.....	57
<i>Iryna Lozova¹, Yuliia Khokhlovachova², Mykhailo Rizak³, Oleksandr Kotyk⁴</i>	57
Cybersecurity as a factor of national and european security: the experience of slovakia in the eu context.....	64
<i>Olena Plaksiuk¹, Oksana Yakusheva², Oleksandr Yakushev^{2,3}, Oleksandr Chernyshov⁴</i>	64
Training specialists in cybersecurity and information protection as a component of the national strategy under the condition of hybrid warfare.....	71
<i>Larysa Prodanova¹, Oksana Zakharova¹, Oksana Yakusheva¹ and Nelia Nagaichuk¹</i>	71

The analysis of cybersecurity audit standards in different parts of the world in the context of digital transformation.....	77
Olesya Voytovych, Vitalii Volynets.....	77
БЕЗПЕКА КІБЕРФІЗИЧНИХ СИСТЕМ ТА КРИТИЧНОЇ ІНФРАСТРУКТУРИ.....	82
Методи та моделі оцінювання стану кібербезпеки об'єктів критичної інфраструктури суб'єктів авіаційної галузі.....	83
Андрій Біскупський ¹ , Ігор Іванченко ²	83
Тестування згортової нейронної мережі при використанні в задачах забезпечення безпеки критичної інфраструктури.....	85
Олена Висоцька ¹ , Антон Герасименко ² , Анатолій Давиденко ³ , Олександр Корченко ²	85
Методика швидкого розгортання захищеного супутникового зв'язку у важкодоступних районах з гарантіями якості зв'язку для служб порятунку.....	90
Василь Волошин ¹ , Лариса Дакова ¹	90
Програмно апаратний комплекс обробки акустичних сигналів для виявлення та ідентифікації БПЛА в реальному часі.....	94
Михайло Євдокімов ¹ , Максим Шевляков ¹ , Любомир Боценюк ¹ , Василь Різак ¹	94
Кібербезпека виробничих систем як складова охорони праці: вплив кіберзагроз на безпеку працівників та способи їх мінімізації.....	97
Олена Крайнюк ¹ , Юрій Буц ² , Михайло Пікасов ¹	97
Використання ансамблевого класифікатора в системах виявлення вторгнень.....	102
Андрій Кулько ¹ , Сергій Толюпа ¹ , Олександр Бучик ¹	102
Створення багатоконтурної інтелектуальної системи кіберзахисту.....	108
Євген Меленті ¹ , Сергій Євсєєв ² , Станіслав Мілевський ² , Олена Ахієзер ² , Владислав Сокол ²	108
Метод оцінювання стану інформаційної безпеки елементів кіберфізичних систем.....	113
Наталія Петляк ¹ , Артем Барабаш ¹	113
Інтеграція 3D-моделювання, етичного хакінгу та технічного захисту для проактивного тестування стійкої критичної інфраструктури в умовах кіберфізичних загроз.....	117
Петро Поночовний ¹ , Ігор Аверічев ¹ , Артем Роженко ¹ , Євген Кихтенко ¹	117
Метод обчислення лишків в задачах шифрування інформації.....	122
Лідія Тимошенко ^{1,†} , Степан Івасьєв ^{2,*,†} , Юрій Грінівецький ¹	122
Моделі багаторівневого доступу для захисту даних при керуванні безпілотними летальними апаратами.....	126
Юлія Ткач ^{1,†} , Ігор Дюба ^{2,*,†} та Олександр Суліма ^{3,†}	126
Загрози безпеці при інтеграції КФС в глобальні мережі.....	129
Юрій Хлапонін ¹	129

Method of applying homomorphic encryption for protecting private data.....	133
<i>Yurii Baryshev¹, Vladyslava Lanova² and Viktoriia Klysh³</i>	133
Single-Layer Boundary Handling for SPH with Volume-Corrected Mass.....	137
<i>Ostap Hrytsyshyn, Valeriy Trushevskyy</i>	137
Modeling self-similar traffic of information and communication networks for attack detection systems.....	141
<i>Valerii Korniienko¹ and Valerii Magro¹</i>	141
Identification of self-similar traffic feature vectors in information and communication networks for attack detection systems.....	146
<i>Valerii Korniienko¹ and Valerii Magro¹</i>	146
Methodology for OSINT analysis of geopolitical risks in critical open-source infrastructure....	151
<i>Ivan Parkhomenko¹, Larysa Myrutenko², Yurii Sokyran³</i>	151
Practical Exploitation of HDD Vibrations for Acoustic Eavesdropping.....	156
<i>Artur Pelcharskyi¹, Maksym Buleshnyi^{1,*}, Mykhailo Buleshnyi^{1,*}, Yuliia Moliashcha^{1,*}, Oleg Farenjuk^{1,2}</i>	156
Real-Time Acoustic Impulse Response Generation Using Static Ray Tracing for Anomaly Detection.....	161
<i>Oleksandr Terletskyy, Valeriy Trushevskyy</i>	161
Phishing URL recognition based on fuzzy clustering with global optimization.....	167
<i>Iryna Udovych¹, Olha Shevtsova¹ and Maksim Fesenko¹</i>	167
Ensuring Data Security in Special-Purpose Information Systems.....	171
<i>Iryna Udovych¹, Serhii Matsiuk¹ and Illia Olishevskyy¹</i>	171
КІБЕРЗАХИСТ КОМУНІКАЦІЙНИХ СИСТЕМ І ХМАРНОЇ ІНФРАСТРУКТУРИ.....	175
Глибоке навчання для мережевої безпеки: Модель Attention-CNN-LSTM для точного виявлення вторгнень.....	176
<i>Віталій Дністровський</i>	176
Аналіз адаптивного кодування LT, Raptor і LDPC у мережах нового покоління.....	179
<i>Сергій Дунаєв¹, Ілля Букачич¹, Станіслав Мілевський¹ and Ірина Аксьонова¹</i>	179
Виявлення атак типу "блокування IP через NAT" в публічних мережах.....	184
<i>Юрій Кльоц¹, Сергій Мостовий¹, Назар Тростянецький¹</i>	184
Безсерверні та контейнеризовані архітектури: порівняльний аналіз підходів до безпеки.....	188
<i>Юрій Кулик¹, Богдан Скоринович¹ та Володимир Гавриляк¹</i>	188
Прогнозування повільних DDoS атак на основі кореляційного аналізу індивідуальних траєкторій трафіку.....	194

<i>Віталій Савченко¹, Галина Гайдур¹, Світлана Легомінова¹</i>	194
Недвійкові афінні перетворення як інструмент підвищення стійкості криптоалгоритмів.....	200
<i>Артем Соколов¹, Володимир Радущ²</i>	200
Метод захисту інформаційних систем на основі застосування міжмережевих екранів.....	204
<i>Максим Стасюк¹, Олександр Лаптев²</i>	204
Вибір стійких трансформант перетворення Уолша-Адамара для аудіостеганографічних методів.....	206
<i>Михайло Хименко¹, Артем Соколов²</i>	206
Ефективність використання інструментів MULI та UniGuardian для виявлення вразливостей LLM.....	210
<i>Ігор Хоменко¹</i>	210
Analysis of Tools for Detecting Vulnerabilities from CWE Top 25.....	215
<i>Yurii Baryshev¹, Ivan Kravchuk^{2*}</i>	215
Using the discrete cosine transform method to ensure the confidentiality and authenticity of audio data.....	219
<i>Alla Havrylova¹, Vladyslav Murzhak¹</i>	219
Process decomposition in cloud computing security techniques.....	222
<i>Alla Havrylova¹, Oleh Azarov¹</i>	222
Information Security Risk Management in Information Systems Based on the Method of Paired Comparisons.....	224
<i>Yurii Klots¹, Vira Titova¹ and Nataliia Petliak¹</i>	224
Enhancement of Security Mechanisms for an Open Information Platform Supporting the Publication of Scientific Research.....	230
<i>Roman Moroz¹</i>	230
Mathematical Method for Evaluating the State of Cybersecurity of Cloud Services for Information Infrastructure Objects.....	236
<i>Volodymyr Shulha¹, Yevhenii Pedchenko², Yevheniia Ivanchenko³, Ihor Ivanchenko⁴, Mari Petrovska⁵, Serhii Skvortsov⁶ and Maryna Pedchenko⁷</i>	236
ПРОГРАМНО-АПАРАТНІ ТА ТЕХНІЧНІ ЗАСОБИ ЗАХИСТУ ІНФОРМАЦІЇ.....	242
Удосконалення фізичних принципів збільшення дальності дії систем управління безпілотними апаратами.....	243
<i>Сергій Герасимов¹, Олег Чернявський¹, Алла Гаврилова¹</i>	243
Алгоритм потокового шифрування на основі.....	246
латинських квадратів.....	246
<i>Богдан Загирняк¹, Галина Крайнічук (Шелепало)¹</i>	246

Методологія забезпечення цілісності та доступності версій CRM-системи Salesforce.....	251
<i>Златоус Святослав¹, Петро Венгерський²</i>	251
Логічні схеми квазігруп за парастрофною симетрією для LW-криптографії.....	255
<i>Галина Крайнічук¹, Людмила Іванова¹ та Тетяна Кирилащук¹</i>	255
Метод побудови криптосистеми Мак-Еліса на основі m-кодів.....	260
<i>Олександр Лаптев¹, Вадим Стеценко²</i>	260
Метод визначення ймовірності негласного отримання інформації потенційним порушником.....	266
<i>Тетяна Лаптева¹, Сергій Лаптев², Євген Меленті³</i>	266
Кіберполігон кафедри ТЕІБ УжНУ: освітнє середовище для моделювання та запобігання кіберзагроз.....	272
<i>Богдан Маліцький¹, Олександр Черепов¹, Юрій Матювка¹, Василь Різак¹</i>	272
Рекурсивний генератор з латинських квадратів.....	274
<i>Богдан Микитченко¹, Галина Крайнічук (Шелепало)¹</i>	274
Радіостанція ISM діапазону із захистом каналу зв'язку від несанкціонованого доступу.....	279
<i>Андрій Садченко¹, Олег Кушніренко¹</i>	279
Алгоритм захисту аналогового відеоканалу розвідувального БПЛА.....	283
<i>Андрій Садченко¹, Олег Кушніренко¹, Олександр Дубіна¹, Євгеній Павлюченко¹</i>	283
Метод захисту кодової бази шляхом шифрування репозиторію.....	287
<i>Андрій Скін¹, Юрій Баришев²</i>	287
ECG for Multi-Factor Cybersecurity: Rhythm-Adaptive Authentication During Mild Physical Activity.....	291
<i>Serhii Lupenko^{1,2,3}, Roman Butsiy^{3,*}</i>	291
Audio Signal Steganography with Neural Networks for Data Integrity in Secure Communications.....	296
<i>Ostap-Sviatoslav Malets¹, Olha Smotr¹</i>	296
Graph-Based Analysis and Hidden Channel Discovery in SELinux Policies.....	298
<i>Vitaliy Kukharskyi¹, Mykola Shcherbyna¹</i>	298
ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ У КІБЕРБЕЗПЕЦІ.....	305
Адаптивна модель для виявлення шкідливих URL з використанням машинного навчання.....	306
<i>Петро Венгерський¹, Володимир Лесик¹</i>	306
Застосування ПЛІС для захисту каналу передачі відеоінформації з БпЛА з використанням штучного інтелекту.....	312
<i>Сергій Гільгурт, Іван Яблоков²</i>	312

Застосування штучного інтелекту для виявлення кіберзагроз.....	317
<i>Олексій Залізний, Владислав Токарев¹</i>	317
Мультирепрезентаційна GNN-модель з узгодженням і адаптивним злиттям для детекції спаму.....	320
<i>Ірина Замрій¹, Іван Шахматов²</i>	320
Інтегрована модель виявлення аномалій та семантичної атрибуції кіберінцидентів для об'єктів критичної інформаційної інфраструктури.....	326
<i>Євгенія Іванченко¹, Микола Рижаків²</i>	326
Штучний інтелект як нова загроза: deepfake та автоматизовані кібератаки.....	331
<i>Крістіна Козир¹, Белла Ісаєва¹</i>	331
Синергія машинного навчання з підкріпленням та великих мовних моделей у мультиагентному фреймворку для тестування на проникнення.....	334
<i>Андрій Притула¹ та Леонід Куперштейн¹</i>	334
Цифрові вакцини як нова парадигма кіберзахисту в умовах атакуючого ІІІ.....	338
<i>Дмитро Прокопович-Ткаченко¹, Максим Білан¹, Олександр Черкаський², Давид Черкаський¹, Данило Переметчик¹</i>	338
Штучний інтелект у реальному часі для виявлення аномальної активності в мережах.....	343
<i>Владислав Сущенко¹, Богдан Марченко¹</i>	343
Захист систем комунікації агентів в мультиагентних розподілених інформаційних системах.....	346
<i>Максим Хорканін¹</i>	346
GPT і Human-in-the-Loop криптоаналіз: дешифрування шифру УПА.....	350
<i>Михайло Шелест¹, Владислав Ворона¹, Анатолій Борода², Юлія Ткач¹</i>	350
Securing Interactive Learning Tools: CRDT/OT, Signed Events, and Analytics.....	356
<i>Tomash Hombosh and Pavlo Mulesa²</i>	356
Detection and Response as Code in the Secure SDLC for Cloud-Native Applications.....	361
<i>Oleksandr Vakhula¹</i>	361
AI-Driven Approaches to Safeguard Workers in Cyber-Physical Environments.....	366
<i>Maksym Krainiuk¹, Oleh Shcherbak¹ and Olena Krainiuk¹</i>	366
Secure and Energy-Efficient Object Recognition on Edge AI Devices.....	371
<i>Vlad Lubin</i>	371
Direct Preference Optimization Using Synthetic Data: Domain-Specific Model Training and Benchmarking for GPT models.....	375
<i>Bohdan Pavlyshenko¹, Ivan Bulka²</i>	375
Multilingual Knowledge Graphs for Thematic Text Analysis.....	380

<i>Bohdan Pavlyshenko¹ and Mykola Stasiuk¹</i>	380
A hybrid ML approach for anomaly detection in information networks.....	384
<i>Andriy Riy, Marian Kyryk, Taras Kolodiy</i>	384
Longitudinal Evaluation of Deep Learning Architectures for Detecting Algorithm-Generated Domains.....	388
<i>Vasyl Salyk¹ and Petro Venherskyi²</i>	388
AI-Generated and AI-Enabled Cyberthreats.....	397
<i>Olga Torstensson¹</i>	397
Intelligent Agents in Multi-Agent Information Security Systems.....	406
<i>Roman Yaroviychuk¹</i>	406
Lightweight Machine Learning Models for Securing Linux-Based IoT Devices.....	410
<i>Yurii Yashunin¹</i>	410
МЕТОДИ ПРОТИДІЇ ДЕЗІНФОРМАЦІЙНОМУ ВПЛИВУ ТА СОЦІАЛЬНОМУ ІНЖИНІРИНГУ.....	412
Розробка ансамблю моделей на основі ResNet50 для детекції дипфейків.....	413
<i>Дмитро Азарний</i>	413
Математичне моделювання поширення пропаганди та фейкової інформації у соціальних мережах: модель SEIPR.....	415
<i>Ярина Борисюк</i>	415
Тенденції інформаційного простору та їхній вплив на стратегічні комунікації сектору безпеки і оборони: виклики та можливості.....	419
<i>Ірина Макаренко</i>	419
A Model for Keyword Spotting in a Voice Signal to Counter Social Engineering and Disinformation.....	422
<i>Andrii Didus¹ and Ihor Tereikovskiy¹</i>	422
The method of cognitive reverse reframing under conditions of large-scale armed aggression of the russian federation against Ukraine.....	426
<i>Serhii Bazarnyi¹, Roman Zubko²</i>	426
БЕЗПЕКА І АНАЛІЗ ВЕЛИКИХ ДАНИХ.....	430
Композиційний шифр як послідовна дія діагонального та секвенціального шифрів.....	431
<i>Олег Гутік¹, Дмитро Дияк¹, Юрій Кузбик¹</i>	431
Інтеграція масштабованих ML-моделей валідації вразливостей у SIEM-системи для автоматизації та прискорення реагування на інциденти.....	437
<i>Владислав Денисюк¹</i>	437

Нейромережева модель для розпізнавання емоцій текстових фрагментів.....	441
<i>Олександр Коровий¹ та Ігор Терейковський¹</i>	441
Протидія атаці Вінера на RSA шифр.....	446
<i>Юлія Мисло¹, Михайло Пагіря¹</i>	446
Програмна система для реалізації методології криптографічного захисту інформації на основі системи залишкових класів.....	450
<i>Ігор Якименко¹, Михайло Касянчук¹, Василь Яцків¹, Людмила Бабала¹, Степан Івасьєв¹</i> ...	450
LLM-Based Methodology for Data Classification under SOC 2 Type 2.....	454
<i>Oleh Deineka¹, Oleh Harasymchuk¹</i>	454
Integrated Framework Deployment for Cybersecurity of Information and Communication Channels in On-Demand Printing.....	459
<i>Tetyana Neroda¹</i>	459
Application of Scallop neurosymbolic framework to defending against cyberattacks.....	465
<i>Vitaly Vlasov¹, Sofia Mikhailova¹</i>	465



КІБЕРБЕЗПЕКА У ДЕРЖАВНІЙ ПОЛІТИЦІ ТА МІЖНАРОДНИХ ВІДНОСИНАХ

EVT-калібрування порогів IDS на основі VaR/CVaR у цифровій юстиції: методологія, впровадження та міжнародні стандарти

Валерій Бушков

Державний торговельно-економічний університет, вул. Кіото, 19, 02156 м. Київ, Україна

Анотація

У тезах подано розгорнуту методологію застосування теорії екстремальних значень (EVT) та фінансових метрик Value at Risk (VaR) і Conditional Value at Risk (CVaR) для калібрування порогів систем виявлення вторгнень (IDS) у цифровій юстиції. Розглянуто проблематику «важких хвостів» подій у судових IT-сервісах, запропоновано інтегровану модель risk-based управління, показано узгодження з міжнародними стандартами (AI Act ЄС, NIST AI RMF, ISO/IEC 27001/27005, принципи OECD та UNESCO). Окреслено архітектурні рішення, економіко-правові аспекти та приклади практичних

Ключові слова

цифрова юстиція; EVT; IDS; VaR/CVaR¹

1. Вступ

Цифрова трансформація правосуддя призвела до стрімкого зростання операційних ризиків у таких сервісах, як електронна подача документів (e-filing), кваліфікований електронний підпис (КЕП/PKI), відеозасідання, сховище доказів та інтеграційні шини реєстрів. Основна проблема полягає в тому, що більшість подій є «нормальними», однак саме рідкісні екстремальні значення викликають непропорційні наслідки — зриви слухань, втрату доказів або компрометацію chain-of-custody. Традиційні статичні пороги IDS не враховують таку специфіку, що підвищує ймовірність як хибних позитивів, так і пропуску критичних інцидентів. Це створює потребу у впровадженні адаптивних порогів, побудованих на теорії екстремальних значень (EVT) та ризик-метриках Value at Risk (VaR) і Conditional Value at Risk (CVaR) [1–3].

2. Мета та завдання

Мета роботи — розробити інтегровану методологію EVT-калібрування порогів IDS для цифрових судів.

Основні завдання:

1. побудова моделі «поріг → ризик → політика» з урахуванням вартості інцидентів;
2. забезпечення пояснюваності та contestability-by-design [4];
3. узгодження з міжнародними стандартами кібербезпеки та управління ризиками (AI Act ЄС, NIST AI RMF, ISO/IEC 27001/27005, принципи OECD та UNESCO) [1–5];
4. демонстрація практичних кейсів для ключових судових сервісів (e-filing, КЕП/PKI, відеозасідання, сховище доказів, chain-of-custody).

¹Corresponding author

 0009-0005-5097-2689 (В.Г. Бушков)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

3. Методи

Цифрова трансформація правосуддя супроводжується зростанням операційних ризиків у таких сервісах, як електронна подача документів, кваліфікований електронний підпис, відеозасідання та управління цифровими доказами. Більшість подій у цих системах є рутинними, проте саме рідкісні екстремальні випадки здатні спричинити зриви слухань, втрату доказів або порушення цілісності ланцюга зберігання даних. Традиційні статичні пороги систем виявлення вторгнень не враховують такої специфіки, що зумовлює підвищення частки хибних спрацювань і водночас ризик пропуску критичних інцидентів. Це створює потребу у впровадженні адаптивних підходів до визначення порогів, заснованих на теорії екстремальних значень та фінансових ризик-метриках Value at Risk і Conditional Value at Risk, що дозволяють відобразити хвостові ризики у практичні керуючі пороги для цифрових судових сервісів

Теорія екстремальних значень (EVT).

У дослідженні використано підхід Peaks-Over-Threshold (POT), який дозволяє моделювати рідкісні події у «хвостах» розподілу. Цей метод орієнтований на аналіз саме тих аномалій, що у цифровій юстиції мають найбільший вплив — зриви слухань, втрата доказів чи компрометація процесу chain-of-custody [1–2].

Використання ризик-метрик VaR і CVaR.

Для калібрування порогів IDS застосовано фінансові метрики Value at Risk (VaR) і Conditional Value at Risk (CVaR). Їх використання дає змогу перекласти хвостові ризики у практичні порогові значення, що враховують не лише частоту, але й очікувану вартість інцидентів. Такий підхід дозволяє відокремлювати події з незначними наслідками від критичних, які вимагають негайних заходів реагування [3–4].

Стрімінгові квантільні ескізи.

Для обчислення порогових значень у реальному часі, без необхідності зберігати повні масиви даних, використано методи компактних квантільних структур: t-digest та KLL-скетч. Вони забезпечують ефективну оцінку квантилів у потокових даних і підвищують точність саме у зонах рідкісних подій, що критично для цифрових судових сервісів [5–6].

Регуляризація та моніторинг дрейфу.

Для забезпечення стабільності оцінок при сезонних коливаннях та нерівномірних навантаженнях використано байєсівську регуляризацію параметрів та моніторинг дрейфу розподілів. Для контролю стабільності застосовуються індекс стабільності популяції (PSI) та критерій Колмогорова–Смирнова (KS). У випадку фіксації дрейфу виконується повторна оцінка параметрів та оновлення порогів [7–8].

Правові та економічні аспекти.

Методологія інтегрує не лише технічні, але й правові та економічні аспекти. Зокрема, ураховується вартість інцидентів (затримка розгляду справи, втрата доказів, репутаційні втрати), а також принципи законності, пропорційності й права на оскарження. Такий підхід узгоджується з міжнародними стандартами — ISO/IEC 27005, ISO/IEC 27001 та NIST AI RMF, а також із етичними принципами AI Act, OECD та UNESCO [9–11].

EVT-калібрування порогів IDS у цифровій юстиції.

Цифрова трансформація судових сервісів (електронна подача документів, КЕП/ПКИ, відеозасідання, управління цифровими доказами) створює нові операційні ризики. Статичні пороги систем виявлення вторгнень не враховують впливу рідкісних подій, що призводить до зривів слухань, втрати доказів або компрометації ланцюга зберігання. Це

зумовлює потребу у застосуванні адаптивної методики, заснованої на теорії екстремальних значень (EVT) та ризик-метриках VaR і CvaR.

На рисунку 1 подано схему, що відображає послідовність модулів методики EVT-калібрування порогів IDS. Вона включає збір потоків і подій, попередню обробку даних, обчислення стрімінгових квантилів (t-digest/KLL), відбір за методом POT і оцінювання узагальненого розподілу Парето, трансляцію результатів у ризик-метрики VaR/CVaR, регуляризацію та моніторинг дрейфу. Подальші етапи охоплюють формування політик реагування («порог $\rightarrow$ ризик $\rightarrow$ політика»), забезпечення пояснюваності та оскаржуваності, узгодження зі стандартами (ISO 27001/27005, NIST AI RMF, AI Act ЄС) та механізми зворотного тюнінгу порогів.

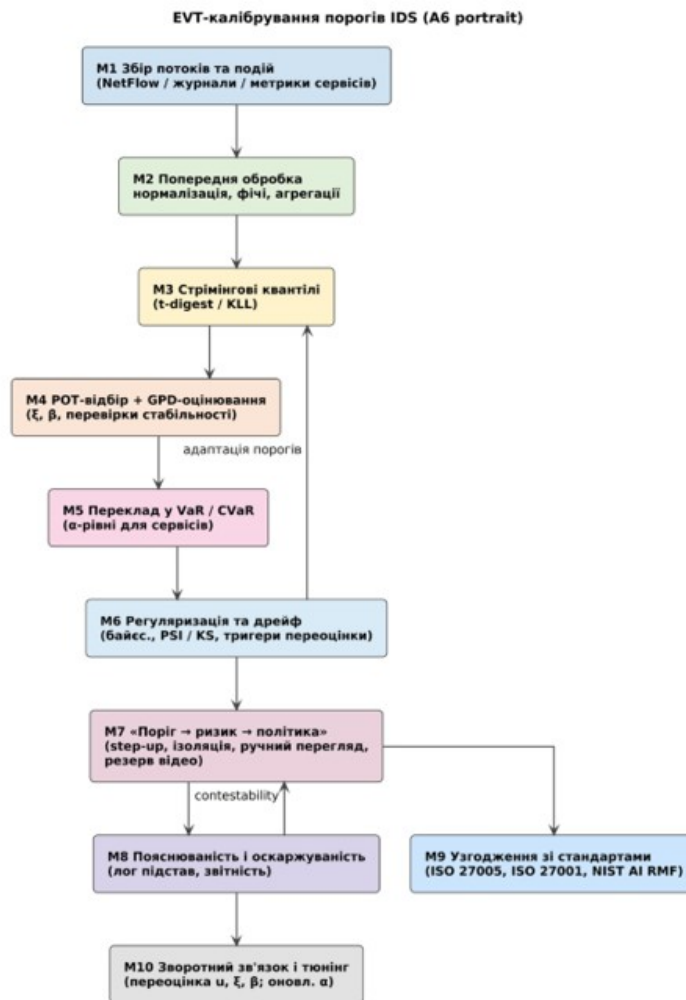


Рисунок 3 1: 1907 Схема EVT-калібрування порогів IDS (A6 portrait)

4. Результати

EVT-калібрування порогів IDS продемонструвало ефективність у ключових сценаріях цифрової юстиції. Серед них e-filing з масовими подачами документів та атаками на САРТСНА, використання кваліфікованого електронного підпису та інфраструктури

відкритих ключів із нетиповими відмовами підписів та аномальними ідентифікаторами сертифікатів, відеозасідання з затримками та втратою пакетів, а також процеси chain-of-custody із піками експорту файлів та аномаліями хешів. Порівняльний аналіз показав, що статичні IDS-пороги часто призводять до завищеного рівня хибнопозитивних і хибнонегативних спрацювань, тоді як використання EVT у поєднанні з фінансовими метриками VaR і CVaR дозволяє формувати адаптивні політики реагування, включаючи step-up автентифікацію, тимчасову ізоляцію профілів, обов’язковий ручний перегляд та резервування відеоканалів.

Методологія EVT-порогування забезпечує баланс між мінімізацією хибних спрацювань і чутливістю до рідкісних інцидентів. Інноваційність підходу полягає у перенесенні фінансової логіки управління ризиком у сферу правосуддя. Важливим аспектом є інтеграція економіко-правового блоку, що враховує права сторін та гарантує contestability. Архітектура цифрової юстиції передбачає цифрові пісочниці, алгоритмічний аудит та механізми AI-омбудсмена.

5. Висновки

Поєднання EVT і VaR / CVaR дозволяє перетворювати рідкісні, але руйнівні події у керований процес. Запропонована методологія сумісна з міжнародними рамками AI Act (ЄС), NIST AI RMF (США), OECD/UNESCO, ISO/IEC 27001/27005 та концепцією Zero Trust. Практичні кейси підтверджують ефективність адаптивних порогів у цифрових судах, а до інноваційних рішень належать цифрові пісочниці, AI-омбудсмен та міжнародний AI-арбітраж.

Література

- [1] Floridi L., Cowls J. A unified framework of five principles for AI in society. *Harvard Data Science Review*. 2019. Vol. 1, Iss. 1. DOI: <https://doi.org/10.1162/99608f92.8cd550d1>
- [2] Jobin A., Ienca M., Vayena E. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*. 2019. Vol. 1. P. 389–399. DOI: <https://doi.org/10.1038/s42256-019-0088-2>
- [3] Cath C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A*. 2018. Vol. 376, No. 2133. DOI: <https://doi.org/10.1098/rsta.2018.0080>
- [4] Winfield A. F. T., Jirotko M. Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A*. 2018. Vol. 376, No. 2133. DOI: <https://doi.org/10.1098/rsta.2018.0085>
- [5] Mittelstadt B. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*. 2019. Vol. 1. P. 501–507. DOI: <https://doi.org/10.1038/s42256-019-0114-4>

Кібердипломатія ЄС в умовах російської агресії в Україні

Сергій Кондратюк¹

¹Державний університет інформаційно-комунікаційних технологій, Солом'янська, 7, 03110, Київ, Україна

Анотація

Повномасштабне російське вторгнення в Україну каталізувало фундаментальну трансформацію стратегій безпеки в усьому євроатлантичному регіоні, ескалюючи використання гібридних загроз, спрямованих на дестабілізацію Євросоюзу. У дослідженні аналізується розвиток кібердипломатії ЄС у відповідь на цю агресію. Методологія включає аналіз нещодавніх законодавчих новацій ЄС у сфері кібербезпеки, а також порівняльний аналіз тематичних досліджень, що стосуються російського втручання у вибори в Молдові (2024), Румунії (2024–2025) та Польщі (2025). Результати дослідження показують різку асиметрію в ЄС, де надійний інфраструктурний кіберзахист різко контрастує з вразливим суспільним флангом. Робиться висновок, що вирішення цього дисбалансу вимагає комплексного підходу всього суспільства, який виходить за рамки технічних рішень для побудови системної демократичної стійкості.

Ключові слова

Гібридні загрози, кібердипломатія ЄС, FIMI, російське втручання у вибори, кіберпростір, вибори 2024-2025 рр. у Молдові, Румунії і Польщі.²

1. Вступ

Після початку повномасштабного російського вторгнення в Україну значно зросли та постійно модифікуються гібридні загрози з боку Кремля, спрямовані на дестабілізацію Євросоюзу та його країн-членів. Ці загрози проявляються у безпрецедентній кількості кібератак на критичну інфраструктуру, спробах втручання у виборчі процеси, поширенні дезінформації, підриві довіри до демократичних інститутів тощо. Пошук адекватної протидії деструктивній діяльності Москви є досить актуальною проблемою не тільки для ЄС, а й для його найближчого оточення, зокрема України.

Ключовою ознакою російської гібридної агресії є навмисне розмивання меж між станом війни та миру, що ускладнює визначення спрямованості й авторства інформаційних і кібератак. Володимир Горбулін у своїй фундаментальній праці «Світова гібридна війна: український фронт» зазначав, що «поняття гібридної війни виявилось теоретично й практично найбільш придатним для визначення специфіки дій РФ, яка, поєднуючи мілітарні, квазімілітарні, дипломатичні, інформаційні, економічні засоби, не гребуючи ядерним шантажем, послідовно намагається досягти в Україні та в інших точках світу власних, не вповні зрозумілих міжнародній спільноті політичних цілей» [1].

У науковій літературі під кібердипломатією розуміють термін, що використовується для опису застосування дипломатичних методів і переговорів у міжнародних відносинах, які стосуються врегулювання питань, пов'язаних із кіберпростором [2].

Одне з перших визначень цього терміну було сформульовано у 2011 році на сайті Державного департаменту: «Кібердипломатія охоплює широкий спектр інтересів США в

² Corresponding author

✉ s.kondratiuk@duikt.edu.ua (S.Kondratiuk)

ORCID 0009-0000-2354-8763 (S.Kondratiuk)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

кіберпросторі. До них належать не лише кібербезпека та свобода Інтернету, а й управління Інтернетом, військове використання Інтернету, інновації та економічне зростання. Кіберпростір також став питанням зовнішньої політики на багатосторонніх форумах, у наших двосторонніх відносинах та у наших стосунках з промисловістю та громадянським суспільством» [3].

Дане дослідження виходить з гіпотези, що хоча ЄС після 2022 року досяг значного інституційного та регуляторного прогресу в захисті критичної інфраструктури від кіберзагроз, його здатність протидіяти гібридним атакам, спрямованим на суспільні інституції і демократичні процеси, залишається недостатньо розвиненою та становить критичну вразливість. Ця асиметрія стала очевидною в контексті нещодавніх виборчих кампаній у Молдові, Румунії та Польщі протягом 2024-2025 рр., які виявили глибокі вразливості у цій сфері.

Для досягнення поставленої мети автором проаналізовано нещодавні зміни в європейській кібердипломатії у відповідь на дедалі агресивніші російські гібридні атаки. Розглядаються ключові законодавчі досягнення у сфері кібербезпеки, а також визначаються політико-інституційні наслідки для забезпечення цифрового суверенітету Європи. Дослідження базується на роботах українських та європейських авторів, що вивчають російські гібридні загрози.

2. Виклад основного матеріалу

У відповідь на російське вторгнення в Україну та поширення гібридних атак, політика ЄС у сфері кібербезпеки зазнала значної трансформації. Раніше орієнтована переважно на захист даних та економічну конкуренцію, тепер вона є невід'ємною частиною ширшої геополітичної стратегії, яка відкрито визначає Росію як «екзистенційну загрозу» [4]. Про цей зсув свідчить безпрецедентний рівень законодавчої активності у 2024-2025 рр., спрямований на зміцнення колективної стійкості Союзу до дедалі складніших загроз:

- Закон про кіберстійкість (грудень 2024 р.) запровадив обов'язкові вимоги безпеки для виробників і роздрібних продавців продукції з цифровими компонентами, спрямовані на захист усього ланцюга поставок від проектування до обслуговування.
- Поправка до Закону про кібербезпеку (січень 2025 р.) розширила повноваження Агентства ЄС з кібербезпеки (ENISA) і сферу дії Європейської системи сертифікації кібербезпеки (ECCF), включивши до неї «керовані послуги безпеки».
- Закон про цифрову операційну стійкість (січень 2025 р.) усунув критичну прогалину у фінансовому регулюванні ЄС, створюючи комплексну систему для забезпечення цифрової операційної стійкості фінансового сектору до серйозних збоїв.
- Закон про кіберсолідарність (лютий 2025 р.) став наріжним каменем колективної оборони ЄС, що має на меті посилити здатність Союзу виявляти, готуватися та реагувати на кіберзагрози. Він створює Європейський кіберщит через мережу Центрів операцій безпеки, Механізм дій у надзвичайних ситуаціях у кіберпросторі та Резерв кібербезпеки ЄС.
- План кібербезпеки ЄС (червень 2025 р.) окреслює рамки для управління кіберкризами в ЄС. Він визначає ролі ключових учасників, таких як ENISA і Мережа CSIRT, протягом усього життєвого циклу кризи, від підготовки до відновлення. Хенна Вірккунен, віце-президент Єврокомісії, наголосила, що цей план є «ключовим компонентом нашої Стратегії готовності Союзу» та «практичним інструментом для співпраці держав-членів та органів ЄС» [5].

Незважаючи на значний прогрес у зміцненні інституційних і технічних аспектів кібербезпеки, гібридні атаки на суспільні інституції ЄС продовжують становити значну

проблему. Ці атаки спрямовані не лише на критичну інфраструктуру, але й на поглиблення соціальних конфліктів і підрив довіри до уряду та демократичних процесів. У березні 2025 р. Європейська служба зовнішніх дій (ЄСЗД) опублікувала свій 3-й звіт про загрози маніпулювання та втручання іноземної інформації (FIMI), в якому Кая Каллас, Високий представник ЄС з питань зовнішньої і безпекової політики, відзначила його новизну у «викритті масивних цифрових арсеналів, спеціально створених Росією та Китаєм для проведення своїх операцій FIMI» [6]. FIMI визначається ЄСЗД не просто як «фейкові новини», а як стратегічні, часто спонсоровані державою, зусилля з маніпуляції інформаційним середовищем і «підриву суверенітету демократичних інституцій за допомогою прихованої, примусової або оманливої тактики» [7].

Стійкість ЄС до гібридних загроз є асиметричною. Інституційна структура ЄС дозволяє швидше та уніфікованіше діяти щодо технічних регламентів, які часто вимагають лише кваліфікованої більшості голосів у Раді. Натомість, такі питання, як FIMI та втручання у вибори, стосуються чутливих сфер національної безпеки та зовнішньої політики, де держави-члени захищають свій суверенітет, а рішення часто вимагають одностайності. Така інституційна структура створює структурну затримку у реагуванні на гібридні загрози на рівні суспільства порівняно з технічними, що є вразливістю, яку зловмисники успішно використовують. Наведені нижче тематичні дослідження надають емпіричні докази цієї асиметрії, демонструючи, як опоненти перенаправляють свої атаки на менш захищені сфери.

Молдова. У 2024 р. президентські вибори та референдум щодо ЄС були ключовими для проєвропейської траєкторії країни за президента Майї Санду.

Служба розвідки та безпеки Молдови детально розповіла про багатогранну кампанію, координовану з Москви, за участю олігарха-втікача Ілана Шора. Тактика включала масштабну дезінформацію, підкуп виборців (понад 15 мільйонів доларів було розподілено між приблизно 130 000 громадян), створення мережі активістів із 33 000 осіб, завданням яких було переконання виборців виступати проти інтеграції з ЄС, залучення священників Православної церкви до агітації і кібератаки на виборчу інфраструктуру тощо [8,9].

Попри ці зусилля, проєвропейський референдум пройшов з надзвичайно малою перевагою в 50,35%, а Санду була переобрана у другому турі з 55,35% голосів. Однак обидві перемоги були забезпечені лише голосами молдавської діаспори в ЄС; усередині країни Санду прогнала голосування, що свідчить про високу ефективність кампанії дестабілізації, яку проводить Кремль. Цей випадок демонструє глибоку вразливість населення країни-члена та вирішальну роль, яку діаспора, що базується в ЄС, може відігравати як противага.

Румунія. Російське втручання у президентські вибори в Румунії мало ще більш драматичні наслідки, спричинивши в державі-члені НАТО та ЄС політичну кризу. Перший тур голосування у листопаді 2024 р. приніс сенсаційний результат: Келін Джорджеску, маловідомий, ультраправий, проросійський кандидат, несподівано здобув перше місце. Його стрімкий злет був підживлений «ідеальним штормом» гібридної війни. Тільки у 2024 р. було задокументовано 34 іноземні гібридні атаки проти Румунії, при цьому понад 25 000 акаунтів TikTok та 5 000 каналів Telegram були мобілізовані на підтримку Джорджеску [10]. Розслідування Global Witness показало, що під час виборчого періоду алгоритми TikTok надавали новим користувачам утричі більше ультраправого контенту, ніж про будь-який інший політичний спектр [11].

Ця цілеспрямована кампанія впливу в поєднанні з невдоволенням суспільства призвела до сплеску популярності Джорджеску. Конституційний суд країни безпрецедентним чином скасував результати першого туру, посиляючись на масове іноземне втручання. Повторні

вибори у травні 2025 р. теж супроводжувалися «інформаційною та мемівською війною» [12] та призвели до масової мобілізації проєвропейських сил, щоб обрати незалежного кандидата Нікушора Дана замість політичного наступника Джорджеску, Джорджа Сіміона. Випадок Румунії служить яскравим попередженням про те, як швидко та ефективно алгоритми соціальних мереж можуть бути використані як зброя для дестабілізації ключової держави-члена ЄС та НАТО.

Польща. Навіть Польща, країна з високим рівнем обізнаності про російські загрози, не була застрахована від гібридних атак під час своїх президентських виборів 2025 р. Польський уряд проактивно витратив 2 мільярди євро, щоб запобігти «румунському сценарію» [13].

Незважаючи на ці заходи, 06.05.2025 р. міністр цифрових технологій Кшиштоф Гавковський заявив: «Під час поточних президентських виборів у Польщі ми стикаємося з безпрецедентною спробою втручання Росії у виборчий процес». Він зазначив, що це втручання поєднувало дезінформацію з гібридними атаками на польську критичну інфраструктуру, включаючи водоканали та теплові електростанції, причому російська участь у таких атаках зросла більш ніж удвічі, порівняно з минулим роком [14]. Досвід Польщі показує, що навіть за високих рівнів тривоги та значних інвестицій загроза залишається потужною та підкреслює «ефект зараження» випадку Румунії, який став еталонним сценарієм для інших країн.

Стратегія ЄС із забезпечення підготовленості Союзу (EU Preparedness Union Strategy), прийнята 26.03.2025 р., являє собою стратегічну відповідь ЄС на вразливості, виявлені цими гібридними атаками. Стратегія знаменує собою доктринальний зсув до підходу «все суспільство» та «врахування всіх небезпек», що виходить за рамки суто технічних рішень [15]. Ключові дії включають сприяння готовності населення (наприклад, підтримка 72-годинного постачання екстрених запасів), інтеграцію готовності до шкільних програм, посилення цивільно-військової співпраці та сприяння державно-приватному партнерству з питань готовності. Ця стратегія є визнанням ЄС асиметрії стійкості та першою великою спробою системно її вирішити.

Аналіз перемог проєвропейських виборців у Молдові та Румунії свідчить про те, що вони були б малоймовірними без значної участі ЄС. Ця взаємодія виходила за рамки традиційної дипломатії і включала мобілізацію вирішальних голосів діаспори, надання європейських грантів на підтримку незалежних ЗМІ, фінансування продемократичних неурядових організацій, оперативне публічне засудження лідерами ЄС іноземного втручання та використання впливу структурних фондів ЄС. Ці дії можна інтерпретувати як розробку ЄС власної неявної доктрини «контргібридного впливу». У той час як метою Росії є дестабілізація, метою ЄС є стабілізація та захист проєвропейської траєкторії. Однак методи сходяться в сірій зоні політичного впливу, що свідчить про те, що захист від гібридної війни може змусити ліберальні демократії вживати власних більш проактивних операцій впливу.

3. Висновки

Ескалация російських гібридних атак стала каталізатором фундаментальних змін у європейській стратегії кібербезпеки. Аналіз показує, що у 2024-2025 рр. ЄС запровадив низку законодавчих новацій у сфері кібербезпеки, що посприяли значному прогресу у побудові колективної, проактивної архітектури цифрової стійкості та кіберзахисту критичної інфраструктури.

Однак цей прогрес у частині гібридних атак є асиметричним, зокрема через наявність чутливих сфер національної безпеки та зовнішньої політики, де для прийняття рішення часто потрібна одностайна позиція усіх країн-членів ЄС. Російські атаки, спрямовані на підрив суспільної і демократичної стійкості, залишаються серйозною та недостатньо врегульованою проблемою.

Нещодавні вибори у Молдові, Румунії та Польщі демонструють, що супротивники успішно акцентують увагу на суспільних проблемах, щодо яких існуючі правові та інституційні механізми виявилися недостатніми. Російське втручання, задокументоване в цих випадках, використовувало широкий спектр гібридних тактик, від масової дезінформації і незаконного фінансування до експлуатації алгоритмів соціальних мереж, зокрема TikTok у Румунії.

Випадок Румунії є першим, коли на практиці було продемонстровано зрив ключового демократичного процесу, що призвів до політичної кризи в країні НАТО. Хоча проєвропейські результати зрештою були досягнуті у Молдові та Румунії, це було пов'язано передусім з реактивними та надзвичайними заходами, включаючи рішучу мобілізацію діаспорних громад і використання з боку ЄС гібридних механізмів впливу.

Політична реакція ЄС розвивається, характеризуючись безпрецедентною політичною та законодавчою діяльністю, спрямованою на посилення колективної стійкості. Стратегія ЄС із забезпечення підготовленості Союзу є критично важливим документом у цьому відношенні, який запроваджує спільні рамки для реагування на кризи та прагне зменшити вразливість у всіх секторах суспільства. Цей законодавчий зсув є прямим наслідком нових геополітичних реалій, які змусили ЄС перейти від реактивного до проактивного управління кризами.

Майбутня безпека Європи залежатиме від її здатності застосовувати той самий рівень стратегічної спрямованості та політичної волі до захисту своєї демократії та громадянського суспільства, як і до захисту своєї критичної інфраструктури.

Література

- [1] 1. Світова гібридна війна: український фронт: монографія / за заг.ред. В.П.Горбуліна. К.: НІСД, 2017. С.15.
- [2] 2. Radanliev P. (2025) Cyber diplomacy: defining the opportunities for cybersecurity and risks from Artificial Intelligence, IoT, Blockchains, and Quantum Computing, *Journal of Cyber Security Technology*, 9:1, 28-78, DOI: 10.1080/23742917.2024.2312671
- [3] 3. International Cyber Diplomacy: Promoting Openness, Security and Prosperity in a Networked World. Bureau of Public Affairs. July 14, 2011. Retrieved September 9, 2025 from: <https://2009-2017.state.gov/r/pa/pl/168689.htm>
- [4] 4. Russia is an existential threat to Europe: Macron. Anadolu Ajansı. 07.03.2025. Retrieved September 8, 2025 from: <https://www.aa.com.tr/en/europe/russia-is-an-existential-threat-to-europe-macron/3502464>
- [5] 5. New Cyber Blueprint to Scale Up the EU Cybersecurity Crisis. ENISA. Press Release. Jun 06, 2025. - <https://www.enisa.europa.eu/news/new-cyber-blueprint-to-scale-up-the-eu-cybersecurity-crisis-management>
- [6] 6. 3rd EEAS Report on Foreign Information Manipulation and Interference Threats Exposing the architecture of FIMI operations. European External Action Service (EEAS). March 2025. Retrieved from: <https://www.eeas.europa.eu/sites/default/files/documents/2025/EEAS-3rd-ThreatReport-March-2025-05-Digital-HD.pdf>

- [7] 7. Proto L., Lamoso-González P., Bouza García L. The EU's FIMI Turn: How the European Union External Action Service Reframed the Disinformation Fight. 28 May 2025. P.4. Retrieved from: <https://www.cogitatiopress.com/mediaandcommunication/article/view/9474>
- [8] 8. Ntousas V., Pleşca L. Russian Meddling in Moldova. German Marshall Fund of the United States. October 18, 2024. Retrieved September 9, 2025 from: <https://www.gmfus.org/news/russian-meddling-moldova>;
- [9] 9. Šlerka J. Intelligence Report Exposes Russian Meddling in Moldova's EU Vote. - VSquare. 2025-01-17. Retrieved September 9, 2025 from: <https://vsquare.org/intelligence-report-russian-meddling-in-moldova-eu-referendum-sis-ilian-sor>
- [10] 10. Marcu-Andrei Solomon. Hybrid warfare through disinformation: the case of Romania's presidential election. Friends of Europe. 10 Jul 2025. Retrieved from: <https://www.friendsofeurope.org/insights/critical-thinking-hybrid-warfare-through-disinformation-the-recent-case-of-romania>
- [11] 11. TikTok algorithm continues to push multiple times more far-right content to users ahead of Romanian election. Global Witness Investigation. 15 May 2025. Retrieved from: <https://globalwitness.org/en/campaigns/digital-threats/tiktok-algorithm-continues-to-push-multiple-times-more-far-right-content-to-users-ahead-of-romanian-election>
- [12] 12. Expert Comment: Which urgent tech lessons must the EU take from Romania's election? University of Oxford. News, 22 May 2025. Retrieved from: <https://www.ox.ac.uk/news/2025-05-22-expert-comment-which-urgent-tech-lessons-must-eu-take-romania-s-election>
- [13] 13. Lasica J. Can Poland's cyber umbrella stop digital election meddling? Monocle: Affairs, April 23, 2025. Retrieved from: <https://monocle.com/affairs/poland-cyber-umbrella-elections/>
- [14] 14. Badahal K. Poland says Russia is trying to interfere in presidential election. Reuters. May 6, 2025. Retrieved from: <https://www.reuters.com/world/europe/poland-says-russia-is-trying-interfere-presidential-election-2025-05-06/>
- [15] 15. EU Preparedness Union Strategy to prevent and react to emerging threats and crises. PreventionWeb. 27 March 2025. Retrieved from: <https://www.preventionweb.net/news/eu-preparedness-union-strategy-prevent-and-react-emerging-threats-and-crises>

Стратегія постквантового переходу ЄС у контексті квантової переваги як геополітичного чинника безпеки кіберпростору

Олександр Кучик¹, Данило Щербина¹

¹ Львівський національний університет імені Івана Франка, вул. Університетська 1, м. Львів 79000, Україна

Анотація

Досліджено криптографічні загрози для критичних інформаційно-комунікаційних систем держави в умовах наближення квантової переваги. Проаналізовано технологічний поступ КНР як геополітичний чинник, що впливає на цифрову безпеку. Розглянуто стратегію постквантового переходу Європейського Союзу, включно з оцінкою ризиків, етапами імплементації та часовими орієнтирами. Сформульовано рекомендації для формування національної політики України у сфері постквантової криптографії.

Ключові слова

AES, ECC, RSA, постквантова криптографія (PQC), квантові комп'ютери, квантова перевага, X-Road, Європейський Союз (ЄС), Трембіта, нормативні акти, дорожня карта, кібербезпека, оцінка ризику

1. Вступ

У світі, де цифрова інфраструктура є основою функціонування держави, суспільства та економіки, безпека даних набуває статусу стратегічного активу. Криптографія як технологія захисту конфіденційності, достовірності та цілісності інформації сьогодні є фундаментом цифрової довіри. Проте зростання обчислювальних потужностей — передусім завдяки прогресу у квантових технологіях — ставить під сумнів довгострокову стійкість чинних криптографічних алгоритмів.

Квантовий комп'ютер — це обчислювальна машина нового покоління, яка використовує закони квантової фізики (суперпозицію, запутаність та інтерференцію) для обробки інформації. На відміну від класичних комп'ютерів, що працюють із бітами (значеннями 0 або 1), останні працюють із кубітами, які перебувають у квантовому стані — поєднанні значень 0 і 1. Це забезпечує здатність до паралельних обчислень і, як наслідок, значно пришвидшує розв'язання складних задач — зокрема факторизації чисел, критично важливої для криптоаналізу Error: Reference source not found.

При квантовій запутаності зміна стану одного кубіта миттєво впливає на стан іншого — незалежно від відстані між ними. Цей фізичний феномен синхронізації відіграє ключову роль у координації обчислень у масштабних квантових системах. Інтерференція, своєю чергою, дозволяє відсіювати неправильні обчислювальні траєкторії, підсилюючи ті, що ведуть до правильного результату. Завдяки цьому обчислення спрямовуються до найбільш ймовірного рішення, що робить квантові комп'ютери надзвичайно ефективними для складних пошукових задач Error: Reference source not found.

³ Corresponding authors

✉ oleksandr.kuchyk@lnu.edu.ua (О. Кучик); danylo.shcherbyna@lnu.edu.ua (Д. Щербина)

🆔 0000-0001-9767-9520 (О. Кучик); 0009-0004-8992-7968 (Д. Щербина)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Криптографічні загрози постквантової ери

RSA (Rivest–Shamir–Adleman) і ECC (Elliptic Curve Cryptography) — два найпоширеніші типи криптографії з відкритим ключем, які лежать в основі сучасної цифрової безпеки. RSA базується на складності факторизації великих чисел, тоді як ECC забезпечує той самий рівень захисту при коротших ключах і меншому ресурсному навантаженні. Ці алгоритми захищають онлайнні транзакції, державні реєстри, медичну інформацію, виборчі процеси тощо.

Сучасні комп’ютери не спроможні легко «зламати» RSA чи ECC — для цього потрібні мільйони років. Однак потужний квантовий комп’ютер може здійснити це впродовж годин або днів, застосовуючи алгоритм Шора. Дані, які записуються сьогодні — особливо в державному секторі — можуть бути скомпрометовані, щойно такий комп’ютер стане доступним.

AES (Advanced Encryption Standard) — один із найнадійніших симетричних алгоритмів шифрування, який широко використовується для захисту даних у державних, фінансових та комерційних системах. Хоча алгоритм Шора не впливає на симетричні шифри, алгоритм Гровера значно зменшує складність атаки перебором. Для AES-128 це означає, що замість 2^{128} операцій квантовий комп’ютер може обійтись лише 2^{64} , що набагато практичніше. Для AES-256 складність скорочується до 2^{128} операцій — рівня, який наразі вважається квантово-безпечним.

3. Критичні державні інформаційно-комунікаційні системи

Прикладом критично важливої інформаційно-комунікаційної системи, що потребує адаптації до постквантової ери, є національна система електронної взаємодії державних інформаційних ресурсів «Трембіта». Вона дає змогу органам влади, місцевого самоврядування та суб’єктам господарювання обмінюватися даними через Інтернет у безпечний, стандартизований і контрольований спосіб.

Станом на середину 2025 року до системи приєднано 288 учасників. Опубліковано 183 електронні інформаційні ресурси (EIP), з яких 109 активно взаємодіють через «Трембіту». Серед них — Єдиний демографічний реєстр, Державний земельний кадастр, Платформа електронної верифікації, портал «Дія», Реєстр призовників, військовозобов’язаних і резервістів. З моменту запуску у 2020 році системою здійснено понад 14,7 млрд транзакцій, включно з рекордним показником у 2,4 млрд за II квартал 2025 року.

«Трембіту» побудовано на основі платформи X-Road (програмного забезпечення з відкритим вихідним кодом), як і систему цифрового урядування Естонії. В обох випадках адаптацію та підтримку забезпечує естонська компанія Cybernetica. Існує міжнародна зацікавленість у впровадженні подібних рішень на основі X-Road, навіть поза межами Європи. Україна виступає своєрідним полігоном, що підтверджує життєздатність та масштабованість технології: навіть попри виклики воєнного часу — вимушену міграцію та окупацію частини територій — система стабільно функціонує у країні, яка суттєво переважає більшість європейських держав за площею та кількістю населення.

Дані в системах на базі X-Road не зберігаються централізовано, а передаються безпосередньо між учасниками. Усі транзакції шифруються, засвідчуються електронним підписом і протокуються для аудиту. Наразі не оприлюднено технічної інформації щодо специфічних криптографічних аспектів системи «Трембіта», однак за замовчуванням X-

Road використовує RSA-ключі довжиною 2048 біт для автентифікації та підписування Error: Reference source not found.

4. Квантова перевага КНР як геополітичний чинник безпеки

На початку 2025 року Китай повідомив про створення квантового комп'ютера Zuchongzhi 3.0, розробленого Університетом науки і технологій. Пристрій містить 105 кубітів і демонструє обчислювальну швидкість, що в мільйон разів перевищує показники найновішого процесора Google. Архітектура чипа побудована на двовимірній решітці з 182 керованими зв'язками, що забезпечує високу щільність взаємодій між кубітами та складні схеми квантового переплетення Error: Reference source not found. Такий прорив у глобальній технологічній гонці підтверджує феномен «квантової переваги» — здатності квантової системи виконувати обчислення, на які класичні машини витратили б неприйнятно багато часу. У стратегічному контексті це свідчить про суттєве посилення позицій Китаю в сфері квантових технологій, що зумовлює переосмислення глобальних підходів до кіберзахисту, цифрового суверенітету та стандартів шифрування в інших державах.

За поширеною оцінкою експертів, для зламу шифрування RSA-2048 необхідний квантовий комп'ютер на основі 4099 кубітів Error: Reference source not found. У характеристиках квантових систем зазвичай зазначають кількість фізичних кубітів, які через технологічні обмеження не забезпечують повної точності обчислень. У випадку Zuchongzhi 3.0, завдяки інноваційному дизайну досягнуто високої стабільності: понад 99,9% успішності однокубітних операцій та 99,6% — двокубітних Error: Reference source not found. Логічні кубіти, про які йдеться у згаданій оцінці складності дешифрування RSA, формуються з великої кількості (іноді сотень) фізичних і забезпечують надійність результатів завдяки механізмам корекції помилок.

Згідно з Quantum Threat Timeline Report Глобального інституту ризиків, імовірність зламу RSA-2048 упродовж 24 годин до 2035 року оцінюється в межах 19–34% Error: Reference source not found. Отже, настання квантової епохи становить — хай не негайну, але стратегічно серйозну загрозу — для ключових цифрових державних сервісів України, Естонії та інших країн. Особливу небезпеку створюють держави, які досягли переваги у квантових обчисленнях — зокрема Китайська Народна Республіка.

5. Стратегія постквантового переходу ЄС

Єдино можливою відповіддю є глобальне впровадження постквантової криптографії (PQC, post-quantum cryptography) — алгоритмів шифрування та цифрових підписів, стійких до атак із використанням квантових комп'ютерів. Наразі ведеться як технічна робота зі стандартизації PQC-алгоритмів, так і формування державних стратегій безпечного переходу до постквантового світу. Україна, з об'єктивних чи суб'єктивних причин, перебуває лише на початку цього шляху, тому доцільно звернутися до напрацьованого європейського досвіду.

У червні 2025 року Європейська Комісія оприлюднила Скоординовану дорожню карту переходу до постквантової криптографії Error: Reference source not found. Вона закликає держави-члени до політико-технічної консолідації зусиль у межах робочої групи при NIS Cooperation Group, яка і підготувала перший рамковий документ. Підхід ЄС базується на кількох ключових стратегічних засадах:

- визнання квантової загрози як фактору ризику;

- забезпечення обізнаності всіх зацікавлених сторін;
- створення зрілої системи управління криптографічними засобами;
- стимулювання транскордонного обміну знаннями та практиками.

Перші практичні кроки держави-члени ЄС мають здійснити вже до кінця 2026 року. Йдеться про розробку національних стратегій переходу до постквантової криптографії (PQC), ідентифікацію критичних цифрових активів, що підлягають першочерговому захисту, а також формування мережі ефективної міжурядової комунікації Error: Reference source not found.

До 2030 року планується завершити перехід для високоризикових сфер — зокрема банківського сектору, державної та оборонної цифрової інфраструктури — на перевірені PQC-рішення. Водночас буде впроваджено стандартизовані гібридні криптографічні схеми. Під останніми розуміють комбінацію постквантових та класичних алгоритмів шифрування, рівень захисту яких визначається найстійкішим компонентом Error: Reference source not found.

Масова трансформація ЄС має бути завершена до 2035 року шляхом інтеграції PQC у всі сфери цифрового життя. Постквантові алгоритми мають стати стандартом за замовчуванням у нових цифрових продуктах і послугах, а європейські та глобальні стандарти — гармонізованими Error: Reference source not found.

Політична реалізація переходу до PQC неможлива без відповідного нормативного забезпечення. Основними регуляторними актами виступають:

- NIS2 Directive — зобов’язує організації забезпечувати належний рівень кібербезпеки, включно з криптографічним захистом, та покладає відповідальність на керівництво;
- DORA Regulation — гарантує цифрову стійкість фінансових установ, з особливою увагою до модернізації криптографічних механізмів;
- Cyber Resilience Act — встановлює з 2027 року обов’язкові вимоги до шифрування та оновлення програмного забезпечення в цифрових продуктах;
- Європейські стандарти PQC — розробляються за участі наукових, промислових і державних структур, з метою уніфікації криптографічної політики в умовах квантової загрози.

Впровадженню постквантової криптографії надано статус стратегічного геополітичного пріоритету. В умовах цифрової конкуренції між державами PQC розглядається як ключовий інструмент забезпечення національних інтересів, інституційної незалежності, економічної стабільності та суспільної довіри. Нескоординованість дій або надмірне зволікання здатні створити вразливості, якими можуть скористатися зовнішні актори, що володіють квантовими обчислювальними потужностями.

Особливу увагу слід приділити сценарію загрози «зберегти зараз — зламати потім», за якого противник накопичує зашифровані дані з метою їх дешифрування після появи квантових комп’ютерів, здатних обходити чинні криптографічні алгоритми. Для таких високоризикових кейсів гарантії безпеки мають бути забезпечені не пізніше кінця 2030 року Error: Reference source not found.

Європейська дорожня карта переходу до PQC пропонує диференційований підхід до оцінки ризиків, визначаючи три рівні квантової загрози:

- високий рівень — якщо потенційне розшифрування даних через 10+ років спричинить значні втрати (урядові документи, конфіденційна фінансова інформація тощо);
- середній рівень — дані, що мають критичне значення протягом 3–10 років (комерційні контракти, медична документація тощо);
- низький рівень — тимчасова або малочутлива інформація з обмеженим періодом актуальності.

Таким чином, політика Європейського Союзу демонструє системний та прагматичний підхід до оцінки ризиків, пов’язаних із появою квантової переваги в сфері кібербезпеки. У межах європейської ініціативи зі створення дорожньої карти постквантового переходу сформовано стандартизовану модель реагування, яка може слугувати орієнтиром для країн із високим рівнем цифровізації, зокрема для України.

6. Висновки

Квантовий ризик — не гіпотетичний, а стратегічно детермінований чинник, що визначає темп і пріоритети цифрової трансформації. Враховуючи стрімкий розвиток квантових обчислень та зростання геополітичних викликів, національний процес впровадження постквантових криптографічних рішень має бути ініційований без зволікань. До кінця 2026 року необхідно затвердити національну стратегію QCS-переходу, а також провести інвентаризацію критичних цифрових активів. До таких слід віднести і національну систему електронної взаємодії державних інформаційних ресурсів «Трембіта», яка, з огляду на аутсорсингову модель реалізації, потребуватиме спеціалізованого підходу.

До 2030 року всі критичні цифрові системи, що забезпечують функціонування державних і суспільно важливих процесів, мають бути переведені на постквантову криптографію. Завершальним етапом стане імплементація QCS-рішень у всіх інформаційно-комунікаційних системах, що містять криптографічні механізми захисту даних — у державному та приватному секторах. Орієнтовним дедлайном цього етапу слід визначити 2035 рік, беручи до уваги зростаючу загрозу появи високопродуктивних квантових обчислювальних платформ у суб’єктів з недружньою геополітичною орієнтацією.

Подяки

Автори висловлюють щиру вдячність фахівцям кафедри кібербезпеки Львівського національного університету імені Івана Франка за надані консультації щодо технічних аспектів квантових комп’ютерів і криптографії.

Декларація щодо використання генеративного ШІ

Під час підготовки цієї роботи автори використовували інструмент Microsoft Copilot для перевірки граматики, орфографії та покращення стилістики викладу. Після застосування зазначеного інструменту автори самостійно здійснили редагування тексту та несуть повну відповідальність за зміст публікації.

Література

- [1] С. Е. Остапов, Ю. Г. Добровольський, Квантова інформатика та квантові обчислення, ЧНУ, Чернівці, 2021. URL: <https://archer.chnu.edu.ua/xmlui/bitstream/handle/123456789/2830/QuantumComputing.pdf>
- [2] Sarah D., Peter C.: On the Practical Cost of Grover for AES Key Recovery. In: Fifth PQC Standardization Conference, Rockville, MD, USA, 2024. URL: <https://csrc.nist.gov/csrc/media/Events/2024/fifth-pqc-standardization-conference/documents/papers/on-practical-cost-of-grover.pdf>
- [3] EU4DigitalUA, Trembita.gov.ua – звіт за 2-й квартал 2025 року, 2025. URL: <https://eu4digitalua.eu/uk/news/trembita-gov-ua-zvit-za-2-j-kvartal-2025-roku/>
- [4] A. Velvet, The Politics of Networks: Europe, Ukraine and Strategies of Digital Development. Lecture, Center for Urban History of East Central Europe, Lviv, 2025.
- [5] X-Road®, X-Road Security Architecture, Technical Specification, Version: 0.11 01.06.2023, Doc. ID: ARC-SEC, 2025. URL: https://docs.x-road.global/Architecture/arc-sec_x_road_security_architecture.html
- [6] Xinhua, China hits new landmark in global quantum computing race, Global Times, 2025. URL: <https://www.globaltimes.cn/page/202503/1329438.shtml>
- [7] M. Ivezic, 4,099 Qubits: The Myth and Reality of Breaking RSA-2048 with Quantum Computers, PostQuantum.com, 2024. URL: <https://postquantum.com/post-quantum/4099-qubits-rsa/>
- [8] M. Ivezic, Zuchongzhi 3.0 Quantum Chip: Technical Analysis and Implications, PostQuantum.com, 2025. URL: <https://postquantum.com/industry-news/zuchongzhi-3-0-quantum-chip/>
- [9] M. Mosca, M. Piani, Quantum Threat Timeline Report 2024, Global Risk Institute, 2024. URL: <https://globalriskinstitute.org/publication/2024-quantum-threat-timeline-report/>
- [10] European Commission, A Coordinated Implementation Roadmap for the Transition to Post-Quantum Cryptography, 2025. URL: <https://ec.europa.eu/newsroom/dae/redirection/document/117507>

Вітчизняний досвід формування системи професійно-кваліфікаційної стандартизації сфери кібербезпеки, ІКТ та критичної інфраструктури

Сергій Мельник¹

¹Державна наукова установа «Інститут освітньої аналітики» МОН України, Володимира Винниченка 5, Київ, Україна

Анотація

Доповідь присвячено формуванню класифікаційного, стандартного та валідаційного забезпечення в Україні підготовки кадрів за професійними кваліфікаціями для сфери кібербезпеки, ІКТ та критичної інфраструктури.

Ключові слова

кваліфікаційний центр, освіта дорослих, професії (посади, заняття) сфери кібербезпеки, ІКТ та критичної інфраструктури, професійна кваліфікація, професійний стандарт

1. Вступна частина

В Україні за останні 7 років відбуваються досить активні процеси формування Національної системи кваліфікацій, підготовки кадрів за кращими світовими практиками та підходами, суттєвого підвищення ролі роботодавців у формуванні вимог до кадрів та їхньої підготовки. За цей період створено: законодавчо-нормативну базу відповідного спрямування, Національне агентство кваліфікацій, відповідальне за державну політику в сфері кваліфікацій, мережу кваліфікаційних центрів як інституцій із незалежного оцінювання здобувачів професійних кваліфікацій, а також Реєстр кваліфікацій, який включає інформацію про затверджені професійні стандарти, професійні кваліфікації тощо. Не оминув цей процес і сфери безпеки, ІКТ та критичної інфраструктури.

2. Основна частина

Станом на 01.09.2025 року в Національному класифікаторі України ДК 003:2010 «Класифікатор професій» (КП) нараховується 34 професії (посади та заняття) сфери кібербезпеки, 50 – у сфері ІКТ та близько 55 позицій у 17 галузях та видах економічної діяльності, віднесених до суб'єктів господарювання.

Зазначимо, що за групами КП (без технічних (офісних) службовців та робітників) професійна номенклатура розподіляється наступним чином: «Керівники» – 7 позицій у сфері кібербезпеки, 18 – у сфері ІКТ та 69 – у видах діяльності критичної інфраструктури; «Професіонали» – 26, 37 та 295 позицій відповідно; «Фахівці» – 1, 5, 186 позицій відповідно.

Таким чином, у країні потребують підготовки кадрів, їхнього оцінювання, присвоєння професійних кваліфікацій, визнання результатів неформальної освіти та спонтанного (інформального) навчання за майже 630 профільними професіями, посадами та заняттями.

¹Контакти для листування

✉ smelnikukr@gmail.com

🆔 0000-0002-5724-2428

Із затверджених в країні 457 професійних стандартів (станом на 01.09.2025 року) 20 припадає на сферу кіберзахисту, а 8 – на сферу ІКТ. Із близько 450 заявок на розроблення ПС 11 припадає на сферу ІКТ (інженер інтернету речей, інженер з програмного забезпечення, менеджер проєктів розробки програмного забезпечення, менеджер (управитель) інформаційних технологій, консультант з ІІІ-трансформацій бізнес-процесів, менеджер-управитель з ІІІ-трансформацій бізнес-процесів, директор з ІІІ (ІІІ-директор) тощо).

Оцінювання та присвоєння 13 професійних кваліфікацій сфери кібербезпеки здійснює тільки один із 135 кваліфікаційних центрів, що функціонують у країні. Цим центром є Кваліфікаційний центр інформаційних технологій та кібербезпеки Державного науково-дослідного інституту технологій кібербезпеки та захисту інформації, який акредитований Національним агентством кваліфікацій (НАК).

Очевидно, що заінтересованим сторонам національного та регіонального рівнів (органам влади, провайдерам освітніх послуг, роботодавцям, профільним громадським організаціям тощо) слід сконцентрувати свої зусилля щодо більш активного розроблення професійних стандартів та формування мережі центрів з незалежного оцінювання та визнання профільних професійних кваліфікацій.

Одним із найбільш ефективних механізмів забезпечення країни кадрами відповідного спрямування у воєнний та повоєнний періоди є напрям освіти дорослих. Серед її ключових складників виділимо:

- 1 Навчання за державним ваучером
- 2 Курси з професійної орієнтації дорослих
- 3 Послуги з освіти дорослих, які надаються від недержавних організацій
- 4 Мережа недержавних інституцій, що надають, як правило, на безоплатній основі, освітні послуги дорослому населенню
- 5 Інші види послуг з навчання дорослих, зокрема ті, що фінансуються державою

Серед розмаїття інших послуг з освіти дорослих виділяються три їх групи: 1) навчання, зорієнтоване на підготовку до участі та подолання наслідків воєнних дій, адаптації учасників війни до мирного життя тощо; 2) навчальні курси, направлені на розвиток навичок з кібербезпеки, захисту суб'єктів критичної інфраструктури; 3) короткотермінові курси з особистого розвитку.

3. Висновки

Таким чином, вітчизняна система підготовки кадрів відповідного спрямування потребує окремого формування (за новітніми підходами) механізмів оцінювання результатів навчання та присвоєння професійних кваліфікацій. Ключовим напрямом її діяльності повинна виступати освіта дорослого населення та визнання результатів неформального навчання.

Декларація про генеративний ШІ

Автор не використовував жодний з інструментів генеративного штучного інтелекту.

Кібербезпека зсередини: як вбудовані бекдори підривають довіру до державних і міжнародних систем

Олександр Полевод¹, Михайло Шелест¹, Юлія Ткач¹

¹ Національний університет «Чернігівська політехніка», вулиця Шевченка 95, 14000, Чернігів, Україна

Анотація

Стаття присвячена аналізу феномену клептографії — навмисного впровадження бекдорів у криптографічні алгоритми, протоколи, апаратні та програмні компоненти, а також у сертифікаційні та інституційні механізми цифрової довіри. Показано історичні витоки клептографії в практиках спецслужб, розглянуто показові кейси (Crypto AG, Dual_EC_DRBG тощо) та продемонстровано геостратегічний вимір проблеми. Запропоновано концепцію клептогігієни як профілактичного підходу до управління довірою у державній політиці та міжнародних відносинах. Сформульовано рекомендації щодо незалежного аудиту, прозорості ланцюгів постачання, права на аудит і пріоритету відкритих стандартів.

Ключові слова

клептографія, бекдор, цифрова довіра, supply chain, Zero Trust, клептогігієна, державна безпека, міжнародні стандарти.

1. Вступ

Цифрова довіра стала стратегічним ресурсом XXI століття. Вона визначає безперервність державних послуг, надійність комунікацій безпеки й оборони, стійкість фінансових транзакцій та ефективність міжнародних інституцій. Традиційна парадигма кіберзахисту орієнтувалася на зовнішні загрози, тоді як критично небезпечним виявився внутрішній вектор — закладання керованих вразливостей (бекдорів) на етапах проектування, стандартизації, складання та сертифікації компонентів [1–2]. У цьому контексті ми розглядаємо клептографію як системну практику створення «захищених» технологій із прихованими механізмами доступу. Ця практика підриває саму основу довіри — не випадковими помилками, а архітектурними рішеннями, у яких контроль закладається як властивість дизайну.

2. Витоки клептографії у практиках спецслужб

Клептографія виростає з потреб спецслужб та розвідувальних спільнот забезпечити привілейований доступ до захищених каналів зв'язку. Історично спецслужби США (NSA) впливали на стандарти криптографії (епізод із Dual_EC_DRBG) [3–4], просували апаратні рішення зі спеціальними механізмами доступу (Clipper Chip) [2]; GCHQ відігравала роль у міжнародній криптополітиці [2]; радянські/російські органи (КДБ/ФАПСи/ФСБ) централізовано ліцензували засоби шифрування й контролювали експорт [2]. Спільними рисами цих підходів були:

- вплив на стандартизацію та сертифікації;
- керування ланцюгами постачання (supply chain);
- прихована інтеграція контролю в апаратні/програмні модулі.

Таким чином, клептографія — не «аномалія ринку», а інституційна стратегія, яка згодом поширилася на комерційні продукти і глобальні стандарти.

3. Визначення клептографії: від вузького до широкого

Початково, у постановці A. Young та M. Yung [1–2], клептографія розумілась як «криптографія проти криптографії»: у криптоалгоритм або протокол свідомо вбудовують прихований механізм (бекдор), який дає змогу сторонньому суб'єкту відновлювати ключі чи повідомлення й водночас залишається практично невиявним звичайними засобами аналізу. Це «вузьке» ядро клептографії фокусується на математичних конструкціях і протокольних кроках.

Однак досвід останніх десятиліть показав, що життєздатність таких бекдорів залежить не лише від математики, а й від соціотехнічного контексту: хто й як затверджує стандарти, якими є процедури сертифікації, як побудовані ланцюги постачання та складання (build), як керуються оновлення і ключі, які інструменти CI/CD та репозиторії використовуються, хто має доступ до firmware і мікрокоду тощо. Саме ці «позакриптографічні» шари часто й створюють умови, у яких бекдор непомітно з'являється, довго живе і масштабно поширюється. Тому ми пропонуємо перейти від вузького до широкого розуміння клептографії, що включає управлінські, організаційні та інституційні практики [2; 14, 17–19].

Вузьке визначення. Клептографія — дисципліна про проектування та інтеграцію криптографічних алгоритмів і протоколів із навмисними бекдорами, які дозволяють третій стороні відновлювати ключі/повідомлення без доступних власнику способів виявлення [1].

Широке визначення. Клептографія — сукупність технічних, організаційних та інституційних практик прихованого впровадження й використання керованих уразливостей у будь-якому шарі цифрової інфраструктури: алгоритмах і протоколах; програмних бібліотеках і SDK; мікропрограмах (firmware) та мікрокоді; апаратних модулях (HSM/TPM/SoC); системах складання, підпису й оновлення; процедурах стандартизації, сертифікації та закупівель [2; 17–19].

Принципові відмінності між вузьким і широким підходами:

1. Об'єкт контролю: від суто криптографічної конструкції → до соціотехнічної системи (код + процеси + інститути).
2. Механізм доступу: від прихованих параметрів/кроків протоколу → до організаційних і ланцюгових важелів (supply chain, стандарти, довірені центри).
3. Методи виявлення: від криптоаналізу артефактів → до аудиту процесів (SBOM, reproducible builds, прозорість сертифікації, трасованість оновлень).
4. Наслідки та відповідальність: від технічної «аномалії» → до політико-правової проблеми довіри та підзвітності інституцій.

Ключовий наслідок переходу до широкого бачення: довіра стає поверхнею атаки. Коли репутація стандарту чи постачальника використовується як прикриття, традиційні аудити коду недостатні — «пастку» може бути вбудовано на рівні процедур і правил гри. Звідси логічним продовженням є наша рамка клептогієни — політики та практик, що роблять довіру перевірюваною і керованою, а не декларативною [14, 19].

4. Типологія та життєвий цикл бекдору

Для того, щоб було видно, де саме можуть ховатися керовані уразливості, проведемо типологію бекдорів:

1. *Алгоритмічні* — пастки в математиці/параметрах (напр., DRBG, криві), що дозволяють відновлювати ключі [1, 3–4].
2. *Протокольні* — зміни в handshake, керуванні ентропією/IV, таймінгах, що дають керований обхід [1–2].
3. *Програмні* — бекдори в бібліотеках/SDK, механізмах оновлень та залежностях (supply chain) [5–6, 11–12].
4. *Мікропрограмні (firmware)* — приховані гачки в мережевих/IoT прошивках, bootloader, криптомодулях/PRNG [5–6].
5. *Апаратні* — модифіковані IP-блоки, контролери, додаткові сопроцесори (HSM/TPM/SoC) [2].
6. *Інституційні* — стандарти/сертифікація/закупівлі, що узаконюють непрозорі компоненти й мінімізують реальний аудит [3–4, 14].

Модель життєвого циклу бекдору допомагає планувати його виявлення й організувати схему контроль та протидії. Життєвий цикл бекдору складається з наступних фаз:

1. *Проектування*: вибір точки контролю та умов активації (ключ/версія/конфігурація).
2. *Інтеграція*: впровадження у код, прошивку, апарат або навіть у процедури стандартизації.
3. *Поширення*: рознесення через залежності, оновлення, сертифікацію, постачання.
4. *Активація*: відкладена чи умовна, щоб зберігати правдоподібність.
5. *Експлуатація*: малошумний доступ/збір метаданих, уникнення артефактів у логах.
6. *Маскування/деактивація*: «виправний» патч, заміна гілки, ребрендинг, апеляція до «сумісності».

Життєвий цикл клептографічного бекдору свідомо оптимізують під низьку ймовірність виявлення і високий стратегічний ефект [19]

5. Кейс-огляд

Клептографія — це зброя стратегічного рівня, вона дає контроль над інфраструктурою й метаданими партнерів/опонентів без явних «зламів» [1–2, 7–12]. От декілька найбільш відомих кейсів клептографічних закладок.

Crypto AG. Модифіковані шифратори забезпечували «зовнішній вигляд» захищеності при фактичному контролі трафіку спецслужбами більше ніж 120 країн; викриття операції Rubicon (співпраця CIA/BND) зафіксовано журналістськими розслідуваннями та офіційними звітами Швейцарії [7–10].

Dual_EC_DRBG. Рекомендований стандарт детермінованого ГВЧ був відкликаний із рекомендацій NIST після виявлення ризику бекдору, але він набув вже певного розповсюдження [3–4].

Juniper ScreenOS. У прошивках мережевого обладнання виявлено вразливості й бекдорні механізми автентифікації/генерації ключів [5–6].

XZ backdoor (2024). Закладка у популярній бібліотеці компресії під виглядом звичайного патчу — інцидент підтвердив уразливість ланцюга довіри навіть в open-source [11–12].

6. Стратегічні наслідки для державної політики та міжнародних відносин

Клептографія не зводиться до суто технічних інцидентів; це системний ризик, у якому технологія, інститути довіри та політична влада взаємодіють.[17-20] Вбудовані бекдори змінюють архітектуру довіри: вони проникають у ланцюги постачання, проходять сертифікацію, накопичують приховану присутність у критичних сервісах, а потому перетворюються на інструмент впливу. Щоб показати цю еволюцію від технологічної причини до геополітичного наслідку, застосуємо причинно-наслідкову рамку з чотирьох кроків.

1. *Supply chain* — точка входу (причина). Будь-який компонент, від криптобібліотеки до фабрики мікросхем, може стати місцем закладання бекдору. Через мережі залежностей, оновлень і повторного використання коду одна зміна непомітно масштабується на цілі екосистеми, уражаючи навіть «бездоганні» продукти [11–12, 14, 17].
2. *Цифровий суверенітет* — внутрішній наслідок. Коли державні сервіси (зв'язок, реєстри, фінанси) спираються на скомпрометовані ланцюги, виникають приховані механізми доступу та зовнішній важіль впливу над критичною інфраструктурою; фактичний контроль над нею частково втрачається [7–12].
3. *Міжнародна довіра* — зовнішній наслідок. Після підриву технічної довіри руйнується і інституційна: стандарти й сертифікації перестають сприйматися як неупереджені, а співпраця вимагає взаємного незалежного аудиту, прозорих SBOM та перевірюваних процесів складання [10, 14].
4. *Політична зброя* — стратегічний вимір. Контроль над інфраструктурою та метаданими перетворює клептографію на інструмент асиметричного впливу: технологічні переваги конвертуються у переговорну силу, регуляторний тиск і розвідувальні можливості [1–2, 7–10].

Ілюстрація (MAX, РФ, 2025). Запуск і обов'язкова передвстановленість державного месенджера MAX у РФ створюють «супердодаток» із глибокою інтеграцією в держпослуги та платежі: така концентрація каналів комунікації й даних формує *архітектуру цифрового контролю* і демонструє клептографію у широкому сенсі — *на інституційному рівні довіри* [15].

7. Клептогігієна: профілактика та управління довірою

Щоб перевести дискусію від констатації загроз до керованої довіри, потрібна рамка, яка робить довіру перевірюваною, трасованою й підзвітною у всіх шарах — від коду до інститутів. Такою рамкою є **клептогігієна**: узгоджений набір принципів, процесів і доказових артефактів (evidence), що зменшують імовірність прихованих бекдорів і підвищують шанс їх раннього виявлення [13–14].

Принципи (що саме означають).

Прозорість — спостережність артефактів (код, параметри, збірки); *трасованість* — наявність SBOM і походження (provenance) для кожної залежності; *підзвітність* — чіткі відповідальні й аудитний слід; *найменші привілеї* — мінімально необхідні доступи/ключі; *розмежування довіри* — сегментація, диверсифікація постачальників; *верифікація понад*

декларації — незалежна перевірка/атестація замість «сертифіковано — отже безпечно» [13–14].

Рівні впровадження.

1. *Технічний.* Незалежний криптоаудит і протокольний аналіз; SBOM із перевіркою версій і ланцюга походження; reproducible/герметичні збірки та контроль середовища складання; політики оновлень із захистом від відкату; моніторинг ентропії/ключів; binary transparency і підпис артефактів; періодичний fuzzing/криптоаналіз. Ці практики адресують ризики типу supply-chain (приклади на кшталт XZ демонструють їхню необхідність) [11–12, 14, 16].
2. *Організаційний.* Zero Trust by Design у процесах; подвійний контроль для критичних дій (dual control/4-eyes); вхідний контроль постачальників з двома незалежними лабораторіями для верифікації; вимога відкритих реалізацій для критичних криптомодулів; обов’язкове peer-review і червоні команди (red teaming) для виявлення прихованих каналів; чіткі SLO/SLA та метрики (скорочення MTTD/MTTR інцидентів, покриття SBOM) [14, 16].
3. *Інституційний.* Законодавче право на аудит коду/прошивок/мікрокоду для держсектору та КІ; незалежні центри сертифікації з публічними звітами; міжнародні механізми прозорості для транснаціональних компонентів; watchlist ризикових залежностей із чіткими критеріями включення/виключення; інтеграція вимог SBOM і відтворюваних збірок у держзакупівлі. Рамкові орієнтири: NIST SP 800-207 (Zero Trust), ENISA щодо supply-chain, NTIA SBOM [13–14, 16].

Операціоналізація (як цим керувати на практиці).

Політика: хто за що відповідає (CISO/сертифікаційний орган/вендор). *Процес:* що і як вимірюємо (SBOM-coverage, % reproducible builds, аудит оновлень). *Артефакти:* що збираємо на доказ (протоколи аудиту, логи збірки, атестації постачальників). Без таких «доказів довіри» клептогігієна перетворюється на декларацію; з ними — на перевірюваний стандарт практик, що знижує простір для клептографії [13–14, 16].

8. Рекомендації для державної політики

Рекомендації для державної політики:

1. Національний аудит критичних компонентів (криптографія, KMS/HSM, мережеві прошивки, мобільні платформи) [14, 16].
2. Державний реєстр ризикових залежностей, інтегрований із процедурами держзакупівель [14, 16].
3. Право на незалежний аудит коду, прошивок і мікрокоду у держсекторі та критичній інфраструктурі; відтворювані збірки як вимога [16].
4. Пріоритет відкритих стандартів і реалізацій критичних модулів (з обґрунтованими виключеннями) [14].
5. Інституціоналізувати клептогігієну в освітніх програмах кібербезпеки та підготовці держслужбовців [13–14, 16].

9. Висновки

Клептографія перетворилася на системний інструмент цифрового впливу, що вимагає зміни парадигми: довіра має бути не передумовою, а результатом перевірки. Політика держави й міжнародних інституцій має зміститися від «сертифікації за замовчуванням» до прозорої верифікації та керованої довіри в ланцюгах постачання. Клептогігієна пропонує рамку дій — від технічних процедур до інституційних гарантій — для відновлення легітимної довіри в цифрову епоху.

Література

- [1] A. Young, M. Yung. Kleptography: Using Cryptography Against Cryptography. In: EUROCRYPT '97, LNCS 1233. Springer, 1997.
- [2] A. Young, M. Yung. Malicious Cryptography: Exposing Cryptovirology. Wiley, 2004.
- [3] NIST. NIST Removes Cryptography Algorithm from Random Number Generator Recommendations. Press release, 21 Apr 2014. URL: <https://www.nist.gov/news-events/news/2014/04/nist-removes-cryptography-algorithm-random-number-generator-recommendations>
- [4] NIST CSRC. SP 800-90A Rev.1 announcement: Dual_EC_DRBG removed; Hash/HMAC/CTR_DRBG recommended. 2015. URL: <https://csrc.nist.gov>
- [5] D. Goodin. Researchers confirm backdoor password in Juniper firewall code. Ars Technica, 21 Dec 2015. URL: <https://arstechnica.com>
- [6] Rapid7. CVE-2015-7755: Juniper ScreenOS Authentication Backdoor. 20 Dec 2015. URL: <https://www.rapid7.com>
- [7] G. Miller, P. Mekhennet, et al. The intelligence coup of the century. The Washington Post, 11 Feb 2020. URL: <https://www.washingtonpost.com>
- [8] Reuters. Switzerland probes reports CIA and BND used Swiss firm Crypto AG to crack codes. 11 Feb 2020. URL: <https://www.reuters.com>
- [9] The Washington Post. Swiss report reveals new details on CIA spying operation (Crypto AG). 10 Nov 2020. URL: <https://www.washingtonpost.com>
- [10] Reuters. Swiss cabinet blames intelligence community for Crypto AG affair. 28 May 2021. URL: <https://www.reuters.com>
- [11] CISA. Reported Supply Chain Compromise Affecting XZ Utils (CVE-2024-3094). 29 Mar 2024. URL: <https://www.cisa.gov>
- [12] Red Hat Blog. Understanding Red Hat's response to the XZ security incident. 30 Apr 2024. URL: <https://www.redhat.com>
- [13] S. Rose, O. Borchert, S. Mitchell, S. Connelly. Zero Trust Architecture. NIST SP 800-207, 2020. doi:10.6028/NIST.SP.800-207
- [14] ENISA. Good Practices for Supply Chain Cybersecurity. 2023. URL: <https://www.enisa.europa.eu>
- [15] Reuters. Russia orders state-backed MAX messenger app to be pre-installed on phones and tablets. 21 Aug 2025. URL: <https://www.reuters.com>
- [16] NTIA (U.S. DOC). The Minimum Elements for a Software Bill of Materials (SBOM). 12 Jul 2021. URL: <https://www.ntia.gov>
- [17] Полевод, О. (2024). Клептографія у контексті захисту інформації. Класифікація клептографічних атак. Технічні науки та технології, 4 (38), 208–213. [https://doi.org/10.25140/2411-5363-2024-4\(38\)-208-213](https://doi.org/10.25140/2411-5363-2024-4(38)-208-213).

- [18] Ткач Ю., Шелест М., Синенко М. (2023). Історія виникнення клептографії та її місце в безпеці інформації. Технічні науки та технології, 3 (33), 150–161. [https://doi.org/10.25140/2411-5363-2023-3\(33\)-150-16](https://doi.org/10.25140/2411-5363-2023-3(33)-150-16).
- [19] Шелест М.Є., Ткач Ю.М. Клептографія: новий напрям кібербезпеки у цифрову епоху. Монографія. - Чернігів, 2025. – 283 с.

Аналіз ефективності навчання кібербезпеки в Україні в сучасних умовах

Михайло Пригара¹, Володимир Щербина²

¹ Ужгородський національний університет, вул. Університетська 14, 88000 Ужгород, Україна

² Державний університет «Київський авіаційний інститут», пр. Гузара Любомира 1, 03058 Київ, Україна

Анотація

В статті, що запропонована, дано аналіз ефективності навчання кібербезпеки в Україні, виділяючи ключові тенденції та виклики. Основні тренди навчання: практична орієнтація - освіта фокусується на практичних навичках, використовуючи кіберполігони, реальні кейси та співпрацю з IT-компаніями; інтеграція ІШ/МН - навчальні програми включають застосування штучного інтелекту та машинного навчання для виявлення загроз, автоматизації процесів та прогнозування атак; спеціалізація - зростає кількість вузькоспеціалізованих курсів, зокрема з безпеки хмарних технологій, DevSecOps та криміналістичного аналізу. Загальним висновком є: навчання в Україні є ефективним, оскільки воно цілеспрямовано переходить до практичного, сучасного та адаптивного підходу. Ця система готує кваліфікованих фахівців, здатних відповідати на реальні кіберзагрози.

Ключові слова

навчання кібербезпеки, безпека хмарних технологій, операційна безпека, безпека промислових систем⁵

1. Вступ

Гібридна протидія на теренах України породжує необхідність підвищення рівня підготовки фахівців з кібербезпеки. Велика кількість кібератак на цивільну та військову інфраструктуру України породжує необхідність в розширенні фаховості працівників що забезпечують її цілісність. Що в свою чергу потребує як підготовку нових так і підвищення кваліфікації вже існуючих. Все це визначає великі вимоги до системи освіти в галузі кібербезпеки.

2. Постановка задачі

Основними викликами сучасності при навчанні кібербезпеки є швидкість змін у галузі та нерівномірність якості освіти в різних закладах. Тому доцільним є аналіз вирішення цей задачі в умовах активної гібридної протидії. Проаналізуємо ефективність навчання з кібербезпеки в Україні в контексті сучасних світових тенденцій та вимог ринку праці.

Для досягнення поставленої мети необхідно вирішити такі завдання:

1. Провести аналіз стану навчання у провідних українських університетах та приватних освітніх центрах.
2. Оцінити рівень відповідності навчання практичним потребам ринку праці та актуальним кіберзагрозам.
3. Визначити ефективність сучасних методів навчання (кіберполігони, інтеграція ІШ/МН, вузькоспеціалізовані курси) у підготовці фахівців.

⁵ Corresponding authors

✉ misha_prigara@ukr.net (М. Пригара); shcherbyna.v.p@gmail.com (В. Щербина)

🆔 0000-0002-0954-4480 (М. Пригара); 0009-0007-1361-2945 (В. Щербина)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

4. Виявити ключові фактори, що впливають на якість навчання та призводять до нерівномірності підготовки випускників.

3. Вирішення поставленої задачі

Основні тренди в навчанні кібербезпеки в Україні, які відповідають світовим тенденціям, включають розвиток практичних навичок, інтеграцію штучного інтелекту та акцент на спеціалізованих напрямках. Ці тренди формуються у відповідь на зростання кількості та складності кібератак, а також на потреби ринку праці.

В першу чергу це практично-орієнтоване навчання. Навчальні програми все більше фокусуються на практичних навичках замість теоретичних знань. Це включає [1]:

- Лабораторні роботи та віртуальні полігони. Студенти використовують спеціальні платформи (кіберполігони), щоб практикуватись у відбитті атак, аналізі шкідливого ПЗ та проведенні пентестів.
- Реальні кейси та симуляції. Навчання базується на розборі реальних кіберінцидентів. Це дозволяє студентам зрозуміти, як працює захист в умовах реальних загроз.
- Співпраця з IT-компаніями. Багато навчальних закладів співпрацюють з компаніями, щоб студенти могли проходити стажування та отримувати досвід на реальних проектах.

По друге зростає застосування штучного інтелекту та машинного навчання. Штучний інтелект (ШІ) стає невід'ємною частиною кібербезпеки. В Україні розвиваються курси, що навчають [2, 3]:

- Виявленню аномалій. Використання ШІ для аналізу мережевого трафіку та поведінки користувачів, що дозволяє швидко виявляти загрози, які неможливо знайти традиційними методами.
- Автоматизації процесів. ШІ допомагає автоматизувати рутинні завдання, такі як аналіз логів та реагування на інциденти, що звільняє фахівців для вирішення більш складних проблем.
- Розробці систем на основі ШІ. Навчальні програми включають створення та тестування систем безпеки, що використовують алгоритми машинного навчання.

По третє отримали розвиток спеціалізовані напрями навчання. Кібербезпека стає все більш спеціалізованою галуззю, тому з'являються вузькоспеціалізовані курси та програми. Найбільш популярні напрями [4]:

- Безпека хмарних технологій. Зростання використання хмарних сервісів (AWS, Azure, Google Cloud) призвело до потреби у фахівцях, які розуміються на захисті цих інфраструктур.
- Операційна безпека (DevSecOps). Це інтеграція безпеки в усі етапи розробки програмного забезпечення, що дозволяє виявляти вразливості на ранніх стадіях.
- Криміналістичний аналіз. Навчання методам збору та аналізу цифрових доказів після кібератак.
- Безпека промислових систем (OT/ICS). Зростає потреба у захисті критичної інфраструктури, такої як електростанції та заводи.

Також сучасної рисою є постійне оновлення програм та мікро навчання. Кібербезпека — це сфера, що швидко змінюється, тому постійне оновлення знань є критично важливим. Навчальні заклади та компанії пропонують:

- Короткострокові курси та буткемпи. Це дозволяє швидко отримати актуальні знання та навички.

- Мікронавчання. Невеликі модулі або уроки, які фокусуються на конкретній темі. Це зручний спосіб підвищувати кваліфікацію, не витрачаючи багато часу.
- Сертифікація. Отримання міжнародних сертифікатів (наприклад, CompTIA Security+, Certified Ethical Hacker) підтверджує рівень знань та підвищує цінність фахівця на ринку праці.

Цікавим аспектом є застосування штучного інтелекту (ШІ) та машинного навчання (МН) в навчанні кібербезпеки в Україні є одним із ключових трендів, що відповідає глобальним викликам. Цей напрямок активно розвивається як в академічних установах, так і на курсах від комерційних компаній. Основний акцент робиться на практичних навичках, що дозволяють ефективно протистояти сучасним кіберзагрозам.

Навчання інтеграції ШІ та МН в кібербезпеку в Україні охоплює декілька основних напрямів:

- Виявлення загроз та аномалій. Студентів навчають розробляти та використовувати алгоритми машинного навчання для аналізу великих обсягів даних (логі, мережевий трафік, поведінка користувачів). Це дозволяє автоматично виявляти незвичайні патерни, що можуть свідчити про кібератаку, навіть якщо вона раніше не була відома.
- Автоматизація процесів захисту. ШІ допомагає автоматизувати рутинні завдання, такі як аналіз вразливостей, сортування інцидентів та реагування на них. Навчальні програми включають розробку систем, які можуть самостійно блокувати шкідливу активність або сповіщати фахівців про інцидент.
- Прогнозування та запобігання атакам. На основі історичних даних про кібератаки та актуальних загроз, МН-моделі можуть прогнозувати ймовірність майбутніх атак. Це дозволяє проактивно посилювати захист у найбільш вразливих місцях.

Українські навчальні заклади та центри активно впроваджують ШІ/МН в програми підготовки кіберфахівців.

- Університети та технічні інститути: Деякі ВНЗ, як-от Київський політехнічний інститут (КПІ), Львівська та Харківська політехніки, інтегрують курси з аналізу даних, глибокого навчання та застосування ШІ в кібербезпеці у свої навчальні плани. Це часто відбувається в рамках магістерських програм або спеціалізованих дисциплін.
- Державні ініціативи: Державний центр кіберзахисту Держспецзв'язку України (SSSCIP) запустив спеціалізовані курси, присвячені застосуванню штучного інтелекту в кібербезпеці. Ці програми сфокусовані на практичній роботі з великими мовними моделями та генеративним ШІ для вирішення реальних завдань.
- Приватні освітні платформи: Приватні ІТ-університети та освітні центри пропонують буткемпи та короткострокові курси, де студенти вивчають основи машинного навчання, а також його застосування для аналізу шкідливого ПЗ, розробки систем SIEM (Security Information and Event Management) на основі ШІ та пентестингу.

Навчання не обмежується лише теорією. Важливою частиною є практична підготовка, яка включає:

- Робота з інструментами. Студенти вчаться використовувати спеціалізоване програмне забезпечення та фреймворки, такі як TensorFlow та Scikit-learn, для створення та тренування моделей.
- Ситуаційні завдання. Для розвитку навичок реагування на інциденти, використовуються симуляційні платформи, де студенти за допомогою ШІ-

інструментів аналізують кіберінциденти, виявляють зловмисників та приймають рішення щодо нейтралізації загрози.

- Співпраця з індустрією. Навчальні заклади та ІТ-компанії організовують спільні хакатони та форуми (наприклад, "AI Cyber Forum"), де студенти працюють над реальними проектами під керівництвом фахівців. Це дає змогу здобути практичний досвід та познайомитися з актуальними викликами, які стоять перед бізнесом та державою.

Навчання кібербезпеки в Україні демонструє значну ефективність та прогрес, що обумовлені переходом від застарілих теоретичних програм до практично-орієнтованих підходів. Це дозволяє готувати фахівців, які можуть відповідати на реальні виклики сучасного цифрового середовища.

- Фокус на практиці: Ключовим фактором є акцент на практичних навичках, що включає використання кіберполігонів, симуляції реальних атак та розбір реальних кейсів. Це дозволяє студентам здобути досвід, необхідний для роботи в умовах реальних загроз.
- Інтеграція сучасних технологій: Освітні програми активно впроваджують навчання за такими напрямками, як штучний інтелект (ШІ) та машинне навчання (МН) для виявлення загроз, а також безпека хмарних технологій (Cloud Security) та операційна безпека (DevSecOps). Цей підхід забезпечує актуальність знань.
- Співпраця з індустрією: Тісна співпраця з ІТ-компаніями, проведення спільних хакатонів та стажувань допомагає налагодити міст між академічним середовищем та потребами ринку праці. Це гарантує, що випускники володіють тими навичками, які дійсно потрібні роботодавцям.

Незважаючи на високий рівень ефективності, система освіти з кібербезпеки в Україні стикається з постійними викликами:

1. Швидкість змін: Кіберзагрози розвиваються дуже швидко, тому освітнім програмам необхідно постійно адаптуватися, що є складним завданням.
2. Нерівномірність якості: Ефективність навчання може відрізнятися в різних навчальних закладах, тому не всі випускники мають однаковий рівень підготовки.

4. Висновки

Загалом, навчання кібербезпеки в Україні є ефективним, оскільки воно цілеспрямовано рухається до практичного, сучасного та адаптивного підходу. Система здатна випускати кваліфікованих фахівців, що є критично важливим для забезпечення національної безпеки та розвитку технологічної галузі.

Декларація про генеративний штучний інтелект

Автори використовували інструменти генеративного штучного інтелекту для перевірки орфографії.

Література

- [1] О. Є. Антонова, Особливості підготовки креативного фахівця: досвід закладів вищої освіти України. Проблеми освіти: зб. наук. праць ; ІДНУ «Інститут модернізації змісту освіти». Вінниця: ТОВ «ТВОРИ», 2019. Вип. 91. С. 11-15.

- [2] І. М. Візнюк, Н. М. Буглай, Л.В. Куцак, А. С. Поліщук, В. В. Киливник, Використання штучного інтелекту в освіті. Сучасні інформаційні технології та інноваційні методики навчання у підготовці фахівців: методологія, теорія, досвід, проблеми : збірник наукових праць. – Вінниця : ТОВ «Друк плюс», 2021. – Вип. 59. – С. 14-22.
- [3] А. І. Шевченко, Стратегія розвитку штучного інтелекту в Україні "Наука і Освіта" ППШ 1, (2023), 307.
- [4] Ю.М. Лебеденко, Компетентнісний підхід в системі вищої освіти. Гуманізм та освіта: матеріали VIII Міжнар. наук.-практ. конф., м. Вінниця, 19-21 вересня 2006 р. Вінниця, 2006. URL: <http://conf.vntu.edu.ua/humed/2006/txt/06lumsvo.htm>.

Формування кібердипломатичних критеріїв стратегічного оцінювання кібербезпеки держави

Володимир Шульга¹, Олександр Корченко¹, Євгенія Іванченко¹, Світлана Казмірчук¹, Сергій Кондратюк¹

¹ Державний університет інформаційно-комунікаційних технологій, вул. Солом'янська, 7, 03110 Київ, Україна

Анотація

Важливою складовою комплексної системи кіберзахисту держави є кібердипломатія – цілеспрямована діяльність держави у міжнародному цифровому просторі, спрямована на захист національних інтересів, просування власних ініціатив у сфері кібербезпеки та формування коаліцій для протидії кіберзагрозам. Міжнародна практика пропонує низку підходів до оцінювання рівня кібербезпеки держав, проте їхня структура є орієнтованою на країни зі стабільною безпековою ситуацією та не завжди враховує специфіку держав, які перебувають у стані гібридної війни. З огляду на це виникає потреба у розробці адаптованих до українських реалій критеріїв оцінювання, що поєднують найкращі світові практики з урахуванням унікальних викликів України. В роботі сформуванні критеріїв для оцінювання кібербезпеки країни в рамках стратегії кібердипломатії з використанням найкращих світових практик як інструмент відстеження загального рівня зрілості державної політики, комплексного оцінювання стану кіберзахисту держави, визначення пріоритетних напрямів розвитку та забезпечення прозорості у формуванні й реалізації стратегії кібердипломатії України.

Ключові слова

Кібердипломатія, кібербезпека, критерії стратегічного оцінювання кібербезпеки держави, кіберзахист держави, кіберзагрози

1. Вступ

За даними міжнародних аналітичних центрів, кібератаки все частіше використовуються не лише як інструмент економічного чи політичного тиску, але й як невід'ємна складова гібридних війн та засіб впливу на внутрішню стабільність держав. Для України, яка перебуває у стані збройної агресії та системного кібертиску, питання кіберзахисту набуває критичного значення, особливо у контексті інтеграції до європейського та євроатлантичного безпекового простору. Український досвід останніх років засвідчив, що атаки на цифрову інфраструктуру, державні інформаційні ресурси та приватний сектор відбуваються систематично та в координації з інформаційними і психологічними впливами. Це вимагає не лише технологічної готовності до відбиття атак, а й створення стійкої, комплексної системи кіберзахисту, здатної ефективно взаємодіяти з міжнародними партнерами та інтегруватися у глобальний безпековий простір.

Важливою складовою цієї системи є кібердипломатія – цілеспрямована діяльність держави у міжнародному цифровому просторі, спрямована на захист національних інтересів, просування власних ініціатив у сфері кібербезпеки та формування коаліцій для протидії кіберзагрозам. Реалізація ефективної кібердипломатії потребує чітких інструментів моніторингу та оцінювання стану кіберзахисту країни, щоб: визначати сильні

⁶Corresponding authors

✉ v.shulha@duikt.edu.ua (В. Шульга); o.korchenko@duikt.edu.ua (О. Корченко); e.ivanchenko@duikt.edu.ua (Є. Іванченко); s.kazmirchuk@duikt.edu.ua (С. Казмірчук); s.kondratiuk@duikt.edu.ua (С. Кондратюк)

🆔 000-0003-4356-7288 (В. Шульга); 0000-0003-3376-0631 (О. Корченко); 0000-0003-3017-5752 (Є. Іванченко); 0000-0001-6083-251X (С. Казмірчук); 0009-0000-2354-8763 (С. Кондратюк)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

та слабкі сторони національної кібербезпеки; коригувати політики та програми розвитку; забезпечувати прозорість і підзвітність перед суспільством та міжнародними партнерами тощо.

Міжнародна практика пропонує низку підходів до оцінювання рівня кібербезпеки держав, серед яких Global Cybersecurity Index (GCI) [1], Network and Information Security Directive 2 (NIS2 Directive) [2], EU Cybersecurity Strategy [3], OECD Digital Security Risk Management Framework [4] тощо. Ці підходи включають від п'яти до восьми ключових напрямів, охоплюючи правові, організаційні, технічні, кадрові та міжнародні аспекти. Проте їхня структура є орієнтованою на країни зі стабільною безпековою ситуацією та не завжди враховує специфіку держав, які перебувають у стані гібридної війни.

Отже метою даного дослідження є формування критеріїв для оцінювання кібербезпеки країни в рамках стратегії кібердипломатії з використанням найкращих світових практик.

2. Основна частина

Для формування таких критеріїв проведено аналіз міжнародних визнаних підходів, розроблених урядами, міждержавними організаціями та галузевими експертами. Їх головна мета – надати універсальний інструмент для порівняння стану кіберзахисту між країнами, визначення пріоритетів політики та підвищення рівня безпеки у цифровому довіллі. Отримані результати відображення у табл. 1.

Таблиця 1.

Зведені дані дослідження існуючих підходів до оцінки кібербезпеки в різних аспектах

Модель / Фреймворк	Розробник	Рік (перша / остання версія)	Кількість напрямків оцінки	Мета	Приклади напрямків	Недоліки	Можливість інтеграції
GCI	ITU	2014 / 2023	5	Глобальна оцінка кіберготовності країн	Правові заходи, технічні заходи	Орієнтований на статистику, без глибокої якості	Так – для міжнародних порівнянь
NIS2 & EU Cybersecurity Strategy	ЄС	2016 / 2022	6	Єдині вимоги для кіберстійкості ЄС	Управління ризиками, звітність	Складно адаптувати поза ЄС	Частково – для регуляторних вимог
NCSI	ЕГА (Естонія)	2016 / 2023	12	Національні кіберспроможності	Політика, освіта, кіберінциденти	Не оцінює ефективність, лише наявність	Так – для структуризації критеріїв
OECD DSRMF	ОЕСД	2015 / 2021	4	Управління цифровими ризиками	Управління ризиками, культура	Відсутні КРІ, нормативна	Частково – для управлінських

					прив'язка	принципи
CMM	Uni-versity of Oxford	2015 / 2021	5	Зрілість нац. кіберспроможностей	Політика, культура, освіта	Немає кількісних оцінок
						Так – для рівнів зрілості

Проведений аналіз провідних міжнародних моделей та фреймворків оцінювання кібербезпеки (див. табл. 1) продемонстрував, що жодна з існуючих систем не є універсальною для повного врахування як національних, так і міжнародних пріоритетів України у сфері кібердипломатії.

Виходячи із отриманих результатів аналізу, було сформовано 11 основних критеріїв оцінювання кібербезпеки держави в рамках стратегії кібердипломатії.

Таким чином, запропоновані 11 критеріїв (див. табл. 2) стали результатом компромісу між повнотою охоплення ключових аспектів кіберстійкості та кібердипломатії, зручністю практичного використання в оцінюванні та можливістю порівняння з міжнародними метриками.

Таблиця 2.

Критерії оцінювання кібербезпеки в рамках реалізації стратегії кібердипломатії України.

№	Напрямок	Критерії	Відповідні компоненти міжнародних моделей	Міжнародні стандарти / рекомендації	Коментар щодо інтеграції
1	Політика і законодавство	Наявність актуального законодавства та стратегій кібербезпеки, відповідність міжнародним стандартам	GCI – Legal Measures; NIS2 – Governance; Policy and Strategy; OECD – Governance	ISO/IEC 27001, NIST CSF Identify	Чітка стратегія, структура управління та відповідальність органів
2	Міжнародне співробітництво та кібердипломатія	Участь у форумах, ініціативах (ООН, ЄС, НАТО), міжнародні угоди, обмін досвідом	GCI – Legal Measures; NIS2 – Compliance; NCSI – International Cooperation; CMM – Cybersecurity Policy	Budapest Convention, ISO/IEC 27002, ITU-D X.1051	Відповідність міжнародним договорам і рекомендаціям
3	Інституційна спроможність і координація	Система раннього виявлення кіберзагроз, обмін ІКТ-	CMM – Organizational Structures; NIS2 – Cooperation Mechanisms; NCSI –	ISO 22301, NIST SP 800-37	Міжвідомча та міждержавна координація, кризове управління

	я	інформацією з	Capacity			
		партнерами				
4	Захист критичної інформаційної інфраструктури (КІІ)	Технічна захищеність наявності протоколів реагування	КІІ, GCI – Technical Measures; NIS2 – Critical Entities; CMM – Critical Infrastructure Protection	NIST SP 800-82, ISO 27019, ITU-T X.1056	Ідентифікація та захист критичних секторів	
5	Технологічна спроможність і кіберінновації	Фінансування кіберінновацій, наявність патентів, партнерство з НДІ	GCI – Technical Measures; NCSI – Cybersecurity Capabilities; OECD – Risk Management Cycle	ISO/IEC 27035, NIST SP 800-61, MITRE ATT&CK	Реагування, виявлення та відновлення після інцидентів	
6	Освіта, підготовка кадрів і просвіта населення	Навчальні програми для дипломатів, IT-спеціалістів, міжнародна сертифікація	CMM – Education, Training, Skills; GCI – Capacity Building	NICE Framework, ISO/IEC 27021	Підготовка кадрів, тренінги, підвищення кваліфікації	
7	Оперативне реагування та управління кіберінцидентами	Оперативність CERT-UA, наявність планів реагування, взаємодія з міжнародними CERT	CMM – Cybersecurity Culture and Society; OECD – Awareness	ISO/IEC 27002, Awareness clauses, ENISA guidelines	Культура кібергігієни, довіра, громадська освіта	
8	Інформаційна стійкість та протидія дезінформації	Освітні кампанії, цифрова гігієна, доступ до інструкцій, участь молоді	GCI – Organisational Measures; NCSI – Industry Involvement	ISO/IEC 20000, NIST SP 800-53 supply chain controls	Розвиток національної кіберіндустрії та інновацій	
9	Інвестиції в нові технології	Державна підтримка ІІІ, розробки в кіберзахисті, проєкти в ІКТ	CMM – Research and Development; OECD – Innovation	Horizon Europe R&D, ISO/IEC TR 27103	Стимулювання досліджень і технологічних рішень	
10	Міжнародний імідж і вплив	Роль у формуванні міжнародних норм, ініціативи в ООН/ЄС/НАТО	NCSI – International Cooperation; GCI – Cooperation; CMM – International Engagement	UN GGE/OEW G Reports, Budapest Convention	Активна участь у міжнародних ініціативах	
11	Політика приватності, права людини і	Цифрові права, етика і законність	OECD – Metrics & Monitoring; CMM – Performance Measurement	ISO/IEC 27004, NIST SP 800-55	Використання KPI та регулярних аудитів	

законність
у кіберпро-
сторі

3. Висновки

Такий підхід дозволяє не лише відстежувати загальний рівень зрілості державної політики, комплексно оцінювати стан кіберзахисту держави, визначати пріоритетні напрями розвитку та забезпечувати прозорість у формуванні й реалізації стратегії кібердипломатії України, а й виявляти слабкі місця у таких напрямках, як протидія дезінформації, міжнародний імідж, інвестиції, цифрові права та інновації. Таким чином, розроблена система критеріїв є не лише аналітичним інструментом, але й важливим елементом формування кіберполітики держави, що здатна реагувати на сучасні та майбутні виклики у цифровому просторі.

Посилання

- [1] Global Cybersecurity Index, About International Telecommunication Union (ITU), веб-сайт. URL: <https://www.itu.int/hub/publication/d-hdb-gci-01-2024/> (дата звернення: 08.08.2025).
- [2] Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union. URL: <https://eur-lex.europa.eu/eli/dir/2022/2555/oj> (дата звернення: 08.08.2025).
- [3] Joint Communication To The European Parliament And The Council The EU's Cybersecurity Strategy for the Digital Decade. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020JC0018> (дата звернення: 08.08.2025).
- [4] OECD Policy Framework on Digital Security Cybersecurity For Prosperity. URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/12/oecd-policy-framework-on-digital-security_a0b1d79c/a69df866-en.pdf (дата звернення: 08.08.2025).
- [5] Global Cybersecurity Index 2024. URL: <https://www.itu.int/en/ITU-D/Cybersecurity/pages/global-cybersecurity-index.aspx> (дата звернення: 08.08.2025).
- [6] The National Cyber Security Index 3.0. URL: https://ega.ee/wp-content/uploads/2023/08/NCSI-3.0_Methodology.pdf (дата звернення: 08.08.2025).
- [7] Information security, cybersecurity and privacy protection – Information security management systems – Requirements: ISO/IEC 27001:2022: Technical Committee : ISO/IEC JTC 1/SC 27, ICS : 35.030 03.100.70, P. 19.
- [8] Information security, cybersecurity and privacy protection – Guidance on managing information security risks: ISO/IEC 27005:2022: Technical Committee : ISO/IEC JTC 1/SC 27, ICS : 35.030, P. 64.

- [9] Information security, cybersecurity and privacy protection – Information security controls: ISO/IEC 27002:2022: Technical Committee : ISO/IEC JTC 1/SC 27, ICS : 35.030 03.100.70, P. 152.
- [10] Security and resilience – Business continuity management systems – Requirements: ISO 22301:2019: Technical Committee : ISO/TC 292, ICS : 03.100.01 03.100.70, P. 21.
- [11] The NIST Cybersecurity Framework (CSF) 2.0: National Institute of Standards and Technology URL: <https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf> (дата звернення: 12.08.2025).
- [12] Cybersecurity Capacity Maturity Model for Nations (CMM) Global Cyber Security Capacity Centre URL: <https://gcsc.ox.ac.uk/files/cmm2021editiondocpdf> (дата звернення: 12.08.2025).
- [13] CYBERSECURITY CAPACITY REVIEW Republic of Cyprus September 2021 URL: https://dsa.cy/images/pdf-upload/cmm_cyprus_report_2021_final.pdf (дата звернення: 12.08.2025).
- [14] CMM Reviews Around the World. URL: <https://gcsc.ox.ac.uk/cmm-reviews> (дата звернення: 12.08.2025).

The formation of a national strategy for cyber protection of critical infrastructure facilities

Oleksii Hutsaliuk^{1,†}, Iuliia Bondar^{2,†} and Oksana Yakusheva^{3,†}

¹ Private Higher Education Institution «Rauf Ablyazov East European University», Nechuya-Levytskogo, 16, 18036, Cherkasy, Ukraine

² Volodymyr Vynnychenko Central Ukrainian State University, Shevchenka, 1, 25006, Kropyvnytskyi, Ukraine

³ Cherkasy State Technological University, Shevchenka, 460, 18000, Cherkasy, Ukraine

Abstract

The formation of a national strategy for cyber protection of critical infrastructure facilities is a key element of providing national security in the context of digitalization and the growth of hybrid threats. The strategy combines risk-oriented management, regulatory and legal frameworks, human resource training, and public-private partnerships. Its implementation is aimed at increasing the resilience of vital sectors, reducing the consequences of cyberattacks, and integrating Ukraine into the global cybersecurity system.

Keywords

cybersecurity, critical infrastructure, cyber resilience, risk-oriented management, public-private partnership, national economy.

1. Introduction

Critical infrastructure facilities (CIF) (energy, transport, banking and financial system, government IT resources) is increasingly exposed to cyberattacks in the condition of modern digital transformation of public administration and the economy. According to the State Service for Special Communications, in 2023 the number of recorded cyberattacks against state bodies of Ukraine increased by 64% compared to 2021, and in the energy sector - by 50%. The losses from cybercrime worldwide are estimated at over \$8 trillion annually (Cybersecurity Ventures, 2023), which is about 9% of global GDP [1-2].

In the context of hybrid threats, it is critically important for Ukraine to form a holistic national cyber-protection strategy based on the principles of proactivity, risk-oriented regulation, and international coordination. Public-private partnership is an important factor, as about 80% of critical infrastructure facilities belong to the private sector.

Implementing a cyber-protection strategy should encompass not only technical and legal dimensions, but also human resources, educational programs, and integration into the global cybersecurity system. This will improve the state's resilience to cyber challenges and reduce potential losses, which, according to EU forecasts, could reach 1.5% of global GDP by 2030 in the absence of proper protection.

The purpose of the paper is to substantiate the principles and tools of the national CIF cyber-protection strategy, which combines risk-oriented regulation, public-private interaction and readiness for incidents.

⁷ Oleksii Hutsaliuk.

[†] These authors contributed equally.

✉ alex-g.88@ukr.net (O. Hutsaliuk); cooperjulia@ukr.net (Iu. Bondar); o.yakusheva14@gmail.com (O. Yakusheva)

ORCID 0000-0002-6541-4912 (O. Hutsaliuk); 0000-0003-2269-6208 (Iu. Bondar); 0000-0002-4849-0323 (O. Yakusheva)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Results

In 2024–2025, the EU significantly updated the framework for cyber resilience of critical sectors through NIS2 and related implementing acts: ENISA summarizes the state of country capabilities and offers policy recommendations (Reviews of the State of Cybersecurity of the Union-2024, Technical Guides on the Implementation of NIS2 2025). This reinforces the requirements for risk management, incident reporting and oversight at Member State and sector level [3].

Ukraine relies on the Cybersecurity Strategy (Decree No. 447/2021) and the updated legal framework for the protection of critical infrastructure (Law No. 1882-IX with amendments in force until August 22, 2024). In parallel, state reports of the State Service for Special Communications and Information Protection (SSSCIP) record the evolution of threats and lessons learned from three years of cyber warfare, highlighting the need for interagency coordination, incident management, and international cooperation [4].

Practical guidelines for CIF operators are offered by ENISA (NIS2 implementation guidelines, sectorial threat landscapes) and CISA (CPGs as a baseline for IT/OT). These materials specify control measures, evidence of compliance, and mapping of requirements to standards, which is convenient to integrate into national policies and departmental regulations.

Current tasks include synchronizing sectorial regulation with NIS2/CER, building a national system of supervision, registration of entities and incident reporting processes, as well as developing the capabilities of competent authorities. Tracking the transposition of NIS2 shows the uneven progress of states and the importance of clear secondary acts and methodological materials. To date, 14 out of 27 EU Member States have implemented the NIS2 Directive into national legislation (Fig. 1) [5-6].

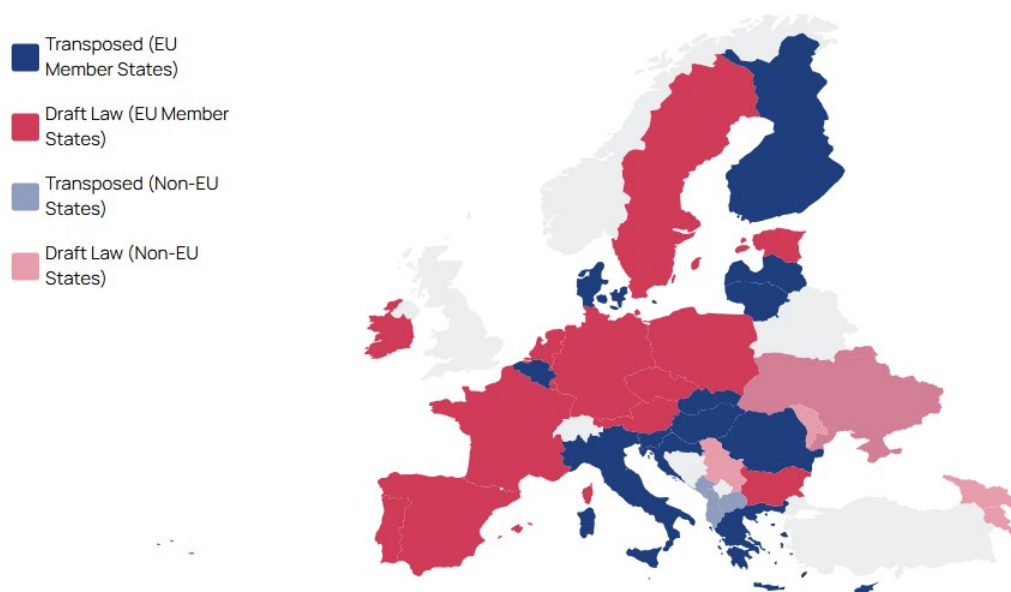


Figure 1: Implementation of the NIS2 Directive by EU Member States into national legislation (NIS2 Directive Transposition Tracker; NIS2 Technical Implementation)

The rapid digitalization of critical sectors – energy, transport, communications, finance, healthcare, public services, water supply and the defense-industrial complex – is accompanied by the escalation of hybrid threats. The changes come amid a sharp increase in cyber incidents:

2,543 cyberattacks were recorded in Ukraine in 2023 – 15.9% more than in 2022. Of these: 347 attacks on government agencies, 92 on energy, 81 on telecommunications, 32 on transport, 30 on finance, 12 on medicine [7]. In 2024, the number of attacks increased to 4,315 incidents - almost 70% more than the previous year [8]. Attackers are spreading malware, phishing, and account compromise; the target is energy, government, defense, and telecommunications.

At the same time, energy infrastructure is under massive physical and cyber attacks. Thus, in August 2024, about 8 million households were left without electricity as a result of airstrikes. About half of the electricity generation capacity (including 6 GW at the ZNPP) and half of the high-voltage substations were destroyed or damaged in 2022–2023. A massive missile and drone strike in March 2024 knocked out 80% of the power capacity of one of the private energy companies (DTEK), cutting off power supply to 1.9 million people [9].

In January 2024, the APT44/Sandworm cyber group attacked more than 20 energy, water, and heating facilities. For example, in Lviv, heating and hot water were turned off for over 600 homes during the frost.

In addition, there is a critical shortage of cybersecurity professionals. According to estimates from the State Special Communications Service, Ukraine may lack up to 100,000 such specialists. Despite this, the cybersecurity market in Ukraine is actively growing: in 8 years it has grown 4 times – to 138 million USD in 2024. The forecast is +50% over the next five years to \$209 million [10].

The formation of a national strategy for cyber protection of critical infrastructure facilities should be based on clearly defined methodological foundations and principles. It should provide a comprehensive approach that integrates cyber, physical and information security into a single resilience management system. Risk-orientation is of particular importance, which involves prioritizing measures taking into account industry threat scenarios and the possible impact on the provision of socially important services. The strategy should be proactive and include security on default, which means establishing requirements for secure software development, architectures, networks, and supply chains at the design stage. An important condition is compatibility with global standards, in particular ISO/IEC 27001/2, 62443, NIST CSF and other cyber resilience standards for OT/ICS.

According to SSSCIP and CERT-UA, the number of cyberattacks in Ukraine in 2024 increased by almost 70% – to 4,315 incidents compared to 2,541 in 2023. The response system processed about 3 million information security events, of which 28,000 were critical, and analysts directly processed 1,042 cyber incidents in 2024 [11]. In 2023, 1,105 cyber incidents were recorded – 62.5% more than in 2022 [12]. The cybersecurity market in Ukraine grew to \$138 million in 2024 and is projected to grow to \$209 million within five years [13].

It should be noted that an effective national CIF cyber protection strategy is not just a “list of technical measures,” but a holistic system of policies, institutions, processes, and competencies. The key consists in the combination of risk-oriented regulation with operational readiness supported by sustainable funding, measurable key performance indicators (KPIs), public-private partnerships, and international cooperation. The proposed model creates a roadmap for increasing the cyber resilience of critical sectors and protecting socially important services in the long term.

3. Conclusion

Thus, the formation of a national strategy for cyber protection of critical infrastructure facilities is a key factor in providing national security, as it allows for systematic risk management, increased resilience of vital services, and reduced consequences of cyberattacks.

The strategy should be based on risk-oriented management, which enables optimal resource allocation and reduced incident response time. The priorities include updating legislation and creating CIF registers; financing cyber protection (NATO countries' spending on cybersecurity increased by an average of 20–25% in 2022–2023); personnel training (Ukraine has a deficit of up to 100 thousand cybersecurity specialists); public-private partnership, since about 80% of CIF belong to the private sector [14].

The expected results are increased cyber resilience of critical sectors, reduced economic losses (which, according to EU estimates, could reach 1.5% of global GDP by 2030 without adequate protection), and strengthened public trust in public institutions [1-2].

References

- [1] Polikovska Yu. The number of cyber incidents registered in Ukraine in 2023 increased by 62.5%, - State Special Communications Service, 2024. URL: https://ms.detector.media/internet/post/33956/2024-01-12-kilkist-zareiestrovanykh-v-ukraini-kiberintsydentiv-u-2023-rotsi-zrosla-na-625-derzhspetsvvyazku/?utm_source
- [2] Holoborodko Ya. The number of cyberattacks in Ukraine increased by 62% in 2023. 2024. URL: https://minfin.com.ua/ua/2024/01/13/119569727/?utm_source
- [3] Advancing cybersecurity and resilience of the Union's critical infrastructure. EU Agencies Network. 2025. URL: https://www.enisa.europa.eu/topics/cybersecurity-of-critical-sectors?utm_source
- [4] The President of Ukraine approved a new Cybersecurity Strategy of Ukraine. Organization of the National Security and Defense Council of Ukraine, 2021. URL: https://www.rnbo.gov.ua/en/Dialnist/4976.html?utm_source
- [5] NIS2 Directive Transposition Tracker URL: https://ecs-org.eu/activities/nis2-directive-transposition-tracker/?utm_source
- [6] NIS2 Technical Implementation Guidance. European Union Agency for Cybersecurity (ENISA), 2025. URL: https://www.enisa.europa.eu/sites/default/files/2025-06/ENISA_Technical_implementation_guidance_on_cybersecurity_risk_management_measures_version_1.0.pdf
- [7] Number of cyberattacks on Ukraine increased by 16% in 2023. Ekonomichna Pravda. 2024. URL: https://www.pravda.com.ua/eng/news/2024/01/31/7439720/?utm_source
- [8] Number of cyberattacks on Ukraine increased by 70% in past year. Ekonomichna Pravda, 2025. URL: https://www.pravda.com.ua/eng/news/2025/01/9/7492671/?utm_source
- [9] Ukraine's energy system under attack. Ukraine's Energy Security and the Coming Winter. URL: https://www.iea.org/reports/ukraines-energy-security-and-the-coming-winter/ukraines-energy-system-under-attack?utm_source
- [10] Ukrainian Cybersecurity Market Quadruples in Eight Years, 2025. URL: https://itukraine.org.ua/en/ukrainian-cybersecurity-market-quadruples-in-eight-years/?utm_source
- [11] The Vulnerability Detection and Cyber Incidents / Cyber Attacks Response System helped to detect and process 1,042 cyber incidents in 2024. State Cyber Protection Centre, 2025. URL: https://scpc.gov.ua/en/articles/383?utm_source
- [12] Statistical Report on the Results of the Vulnerability Detection and Cyber Incident and Cyberattack Response System Operation for 2023. State Cyber Protection Centre, 2024. URL: https://scpc.gov.ua/en/articles/334?utm_source
- [13] Ukraine Leading in Cybersecurity Resilience. IT Association of Ukraine, 2025. URL: https://digitalstate.gov.ua/news/tech/ukraine-leading-in-cybersecurity-resilience?utm_source

- [14] Ukraine lacks nearly 100K cybersecurity specialists – special comms agency. Ukrinform, 2025. URL: https://www.ukrinform.net/rubric-society/3976130-ukraine-lacks-nearly-100k-cybersecurity-specialists-special-comms-agency.html?utm_source
- [15] Hutsaliuk O., Havrylova N., Alibekova B., Rakayeva A., Bondar Iu., Kovalenko Yu. Management of Renewable Resources in the Energy Sector: Environmental, Economic and Financial Aspects. Green Energy and Technology. Circular Economy for Renewable Energy. Cham: Springer, 2023. 69-89. URL: <https://doi.org/10.1007/978-3-031-30800-0>
- [16] Hutsaliuk O., Bondar Iu., Tomareva-Patlahova V., Kalinin O., Navolokina A., Kozlovska S. Integration of Operational and Economic Security of the Enterprise: Reorganization, Technologies, Risks and Outsourcing. Data-Centric Business and Applications. Lecture Notes on Data Engineering and Communications Technologies. Cham: Springer, 2025. 240. 35-61. URL: https://doi.org/10.1007/978-3-031-81557-7_3
- [17] Hutsaliuk O., Bondar Iu., Kalinin O., Sokolovskiy V., Navolokina A. Integration Development of Logistics Activities of Corporate Enterprises Based on Intellectualization and Management Technologies. Intelligent Transport Systems: Ecology, Safety, Quality, Comfort. ITSESQC 2024. Lecture Notes in Networks and Systems. Cham: Springer, 2025. 1335. 270-289. URL: https://doi.org/10.1007/978-3-031-87376-8_24
- [18] Havryliuk, O., Yakushev, O., Petchenko, M., Zachosova, N., Bielialov, T., & Kozlovska, S. (2023). Cyber security and artificial intelligence in the context of ensuring business security in wartime. *Financial and credit activity problems of theory and practice*, 6(53), 451-459. DOI: <https://doi.org/10.55643/fcaptp.6.53.2023.4130>
- [19] Yakushev O., Hulak D., Zakharova O., Kovalenko Y., Yakusheva O., Chernyshov O. Management of the Modern Electric-Vehicle Market. *Polityka Energetyczna – Energy Policy Journal*. 2022. 25(2). 85–108.
- [20] Yakushev, O., Moisieienko, L., Yakusheva, O., Prodanova, L., Plaksiuk, O., & Chepurda, L. (2024). Socio-economic sustainability of the tourism sector enterprises in the context of the covid-19 pandemic: global and ukrainian dimensions. *Financial & Credit Activity: Problems of Theory & Practice*, 5(58). 484-499. DOI: <https://doi.org/10.55643/fcaptp.5.58.2024.4377>

Research on Models, Methods, and Systems for Assessing the Negative Consequences of Personal Data Breaches

Iryna Lozova¹, Yuliia Khokhlachova², Mykhailo Rizak³, Oleksandr Kotyk⁴

¹ State University of Information and Communication Technologies, Solomianska 7, 03110 Kyiv, Ukraine

² State University of Trade and Economics, Kyoto 19, 02156 Kyiv, Ukraine

³ State University of Information and Communication Technologies, Solomianska 7, 03110 Kyiv, Ukraine

⁴ State University "Kyiv Aviation Institute", Liubomyra Huzara ave. 1, 03058 Kyiv, Ukraine

Abstract

The rapid development of digital technologies and the growing number of cyber threats increase the risk of personal data breaches, leading to financial, legal, and reputational losses. Existing approaches to assessing the criticality of such incidents are fragmented, as they focus mainly on isolated aspects. This study presents a comparative analysis of modern models, methods, and systems for evaluating the negative consequences of personal data breaches, identifying their strengths, limitations, and compliance with international standards, in particular the GDPR. The scientific novelty lies in establishing a theoretical foundation for the development of improved comprehensive models for assessing losses from personal data breaches, taking into account economic, legal, and reputational factors.

Keywords

personal data, impact assessment, personal data breach, confidentiality, GDPR, cybersecurity, CEUR-WS

1. Introduction

The rapid expansion of digital technologies and globalization of information flows have intensified cyber threats and incidents of personal data breaches. Such leaks represent a critical challenge in modern cybersecurity, causing financial losses, reputational risks, legal liabilities, and violations of data subjects' rights.

Although regulations such as the EU General Data Protection Regulation (GDPR) require timely detection and assessment of breaches, existing approaches remain fragmented: financial models overlook reputational and legal risks, technical methods focus only on vulnerabilities, and socio-psychological approaches are limited in scope.

This study emphasizes the need for comprehensive models and methods that integrate regulatory requirements, monitoring and cyber threat analysis tools, and forecasting mechanisms to assess the broader consequences of personal data leaks for business and society.

The purpose of this study is to conduct a comprehensive analysis of existing methods, models, and systems for assessing losses caused by personal data leaks, to compare them using unified parameters, and to identify their strengths and weaknesses in order to reveal existing gaps and establish a scientific foundation for the further development of advanced models and methods for loss assessment in the field of personal data protection.

⁸ Corresponding authors

✉ i.ryna@duikt.edu.ua (I. Lozova); yuliiahokhlachova@gmail.com (Y. Khokhlachova); advokat.rizak@gmail.com (M. Rizak); kotyk.oleksandr.81@gmail.com (O. Kotyk)

🆔 0000-0002-7224-4763 (I. Lozova); 0000-0002-0787-5112 (Y. Khokhlachova); 0009-0003-9844-3271 (M. Rizak); 0009-0004-8783-6261 (O. Kotyk)

2. Primary Research Materials

With the growing regulatory requirements, particularly under the GDPR, assessing the consequences of personal data breaches has become a critical cybersecurity challenge. Currently, no universal model exists that adequately captures all aspects of losses from such incidents. Existing approaches remain limited: economic methods often ignore reputational risks, legal ones neglect financial aspects, and technical approaches focus mainly on vulnerabilities without considering socio-psychological effects.

This study summarizes over fifteen key models, methods, and systems for assessing the negative consequences of personal data breaches, highlighting their practical applicability, core evaluation parameters, and criteria for measuring losses, with emphasis on their potential integration into modern information systems.

1. FAIR (Factor Analysis of Information Risk) [1–3]. FAIR is a quantitative risk analysis model decomposing risk into Loss Event Frequency and Loss Magnitude. It estimates financial consequences using statistical modeling (e.g., Monte Carlo), considering direct (recovery costs, fines) and indirect losses (reputational damage). Widely applied in the corporate sector, standardized by OCEG, and implemented in tools like RiskLens and AuditScripts, FAIR can account for GDPR penalties (up to 4% of annual turnover or €20 million) under "Judgments and Fines"

2. NIST SP 800-30 and RMF [4–7]. SP 800-30 guides risk assessment within NIST RMF, describing asset, vulnerability, and threat identification and risk evaluation using qualitative (low/medium/high) or quantitative metrics (e.g., ALE). RMF operates at the organizational level and aligns with FISMA, ISO/IEC 27001, etc. While not GDPR-specific, it informs control design. Key data include asset inventories, incidents, and expert judgments; reputational impacts are indirect. Widely used in U.S. public and private sectors.

3. NIST Privacy Framework (PF) and PRAM [8]. A voluntary tool for privacy risk management aligned with GDPR, PF and PRAM guide system owners through data flow assessment, privacy issue identification, prioritization, and mitigation. Focused on process management rather than quantitative loss evaluation, it supports GDPR compliance and robust privacy programs.

4. ISO/IEC 27005 [6,7,9]. Defines ISMS risk management: threat identification, risk assessment, and security measure selection. It does not provide formulas for financial loss but supports risk-based decisions using incident records, tests, and expert evaluations. Reputational and financial impacts are considered indirectly; ISO 27005 is globally applied.

5. ISO 31000 [10]. Provides general risk management guidelines across industries with governance, PDCA cycles, and processes (identification, analysis, evaluation, treatment). It lacks specific loss metrics and emphasizes process over quantitative assessment. Widely supported in business and government.

6. OWASP Risk Rating [11]. Adapts classical risk assessment for web applications as Probability × Impact, including Financial and Reputation Damage. Simple and application-oriented but partially subjective (expert scores 1–10). Integrated into GRC tools, it assesses data breach risks in web contexts, though not GDPR-focused.

7. DPIA (GDPR) [12]. Mandatory under Art. 35, evaluates the impact of data processing on individuals' rights. While not numeric, it considers potential data breach risks, fines, and mitigation, involving DPO consultation and supervisory authority engagement. Focused on rights protection rather than direct loss calculation.

8. CRAMM [6,7,13]. UK qualitative tool for asset/threat collection and risk assessment (0–100). Does not calculate direct financial losses, but prioritizes risks and recommends measures.

Reputational impact considered indirectly. Software implementations exist; widely used in government.

9. OCTAVE [13,14]. Cybersecurity risk framework combining technical vulnerability and business priorities. Identifies critical assets, threats, and vulnerabilities, producing risk mitigation plans. Losses are described categorically (financial, operational, reputational); widely used in critical infrastructure and large enterprises.

10. EBIOS [13,15]. French IT risk method (EBIOS Risk Manager, 2018), combining scenario analysis and matrix evaluation. Identifies major events/assets (MAST), assesses impact/probability. Does not calculate direct losses but prioritizes risks probabilistically; reputational losses are secondary. Compatible with ISO 27005.

11. MEHARI [13,16]. French methodology for harmonized risk analysis aligned with ISO 27001/27005. Uses scoring and impact tables to evaluate scenario probability and impacts, including financial and reputational. Software implementations exist.

12. Ponemon Institute [17]. Annual empirical "Cost of a Data Breach" report calculates direct and indirect costs using activity-based costing, USD per incident or record. Mainly indirect costs (lost revenue, reputation, fines). Useful for historical analysis and ROI; less precise for individual cases. GDPR relevance is indirect.

13. CyberVaR (CVaR) [13,18]. Financial method for tail risk, estimating maximum losses via scenario modeling (Monte Carlo) using attack frequency, consequences, and asset value. Useful for strategic planning and insurance; emphasizes worst-case scenarios. Quantitative, requiring incident data and expert estimates; GDPR relevance is preparedness demonstration.

14. Quantitative Assessment for Data Breaches [19]. Multiplicative model estimating direct (investigation, compensation, fines) and indirect costs (reputation, lawsuits, lost revenue). Regulatory fines included; focus on total cost minimization. Formal equations based on IBM/Ponemon data; no software implementation.

15. Model for Classified Information Leakage [20]. Uses multiple parameters (threats, secrecy levels, economic and security indicators) to evaluate impact at regional/national levels, including foreign intelligence, access violations, and media loss. Supports national security and cybersecurity planning.

16. Scale-Free Social Network Model [21]. Evaluates personal data security in networks using topology (node degree, path length, clustering, correlations). Considers network growth effects; simulations confirm effectiveness in assessing privacy risks.

17. Ukrainian Armed Forces E-Learning System [22]. Enhances personal data protection using hybrid neural networks and fuzzy logic. Includes threat analysis, security assessment, mitigation, and forecasting. Flexible evaluation of multiple factors for timely policy enhancement.

18. Mathematical Model of Information Value [23]. Calculates information value via coefficients for access, relevance, importance, and ownership, ranked by regulations or expert judgment. Highest value assigned to restricted-access, strategic, government-owned data. Supports protective measures and threat modeling.

To ensure objectivity and comprehensiveness in the analysis of models, methods, and systems for assessing the consequences of personal data breaches, it is necessary to define a set of parameters that provides a unified basis for the structured evaluation of the compliance, effectiveness, and suitability of existing solutions within a specific legal and technological context [24, 25].

Firstly, the establishment of clear criteria enables the standardization of the assessment approach. Since each method has its own structure, design principles, data formats, and application

domain, a parameterized approach allows for the comparison of these solutions irrespective of their origin or initial purpose.

Secondly, the criterion of compliance with legal requirements, in particular the European Union Regulation 2016/679 (GDPR), is of paramount importance. The "GDPR compliance" parameter enables the evaluation of the extent to which a model or method incorporates current legal standards in the field of data protection.

Thirdly, the practical feasibility of implementing the analyzed approaches is evaluated through parameters such as software availability and the scope of application. This allows the identification of methods that can be directly deployed in governmental or corporate environments versus those that are purely theoretical or conceptual.

Fourthly, it is advisable to consider the comprehensiveness of coverage of potential loss categories resulting from data breaches, including financial, reputational, and legal impacts. The analysis of data sources (expert-based, technical, historical) helps determine whether the method provides a foundation for evidence-based decision-making or scenario-based modeling.

Finally, the format of assessment and the type of output play a key role in the potential integration of the chosen approach into risk management systems and strategic planning. For instance, models that provide quantitative estimates in monetary terms are suitable for the economic justification of data protection expenditures.

Thus, the formalization of evaluation parameters enables a comprehensive, methodologically grounded selection of the most effective solutions, the identification of gaps in existing approaches, and the determination of directions for further research and the development of new models that address contemporary challenges in personal data protection (Table 1).

Table 1

Analysis of Methods, Models, and Systems for Assessing the Consequences of Personal Data Breaches

Methods / Models / Systems	GDPR Compliance	Software / Implementation	Application Area			Data Source			Loss Categories		Assessment Format			Result Type	
			Personal Data Protection	Information / Cybersecurity	Specialized / Industry-specific	Expert / Survey Data	Structural / Technical Data	Historical / Actual Incidents	Reputational	Financial	Monetary Value	Assessment Score / Categorical	Composite Indicator	Quantitative	Qualitative
1	–	+	–	+	+	+	–	–	+	+	+	–	+	+	–
2	–	+	–	+	–	+	+	+	–	+	+	+	–	+	+
3	+	–	+	+	–	+	+	–	+	–	–	+	–	–	+
4	–	+	–	+	–	+	+	–	–	+	–	+	–	–	+
5	–	–	–	+	–	+	+	–	+	+	–	+	–	–	+
6	–	+	–	+	+	+	–	–	+	+	–	+	–	+	–

7	+	+	+	+	-	+	+	-	-	+	-	+	-	-	+
8	-	+	-	+	-	+	+	-	-	-	-	+	-	-	+
9	+-	-	-	+	+	+	+	-	+-	+-	-	+	-	-	+
10	+	+	+	+	+	+	+	-	-	-	-	+	-	-	+
11	-	+	-	+	-	+	+	-	+-	+-	-	+	-	-	+
12	-	-	-	+	+	+	-	+	+	+	+	-	-	+	-
13	-	+	-	+	+	-	-	+	+	+	+	-	-	+	-
14	+-	-	+	+	+	+	-	+	+	+	+	-	-	+	-
15	-	-	-	+	+	-	+	-	+	-	-	+	-	-	+
16	+-	-	+	-	+	-	+	-	+	-	-	-	+	+	-
17	+-	+	+	+	+	+	+	-	+	-	-	+	+	+	+
18	+-	-	-	+	+	+	+	-	+	+	-	-	+	+	-

The conducted analysis demonstrates that most systems, models, and methods either do not cover all aspects of the negative consequences of personal data breaches or are not adapted to GDPR requirements. Only a subset is directly GDPR-compliant (e.g., DPIA, NIST Privacy Framework, EBIOS, partially OCTAVE) and fully or partially meets the Regulation’s standards. This indicates a limited number of tools specifically tailored to the European legal environment. Most approaches are qualitative or mixed in nature – ISO 31000, DPIA, MEHARI, EBIOS, OCTAVE, CRAMM primarily rely on expert assessments, surveys, or scenario-based analyses without quantitative financial evaluation, which complicates the justification of actual business losses. Only a few models and methods provide direct financial loss estimation – FAIR, Ponemon Cost Model, CyberVaR, and Quantitative Assessment – allowing for quantitative evaluation in monetary terms, which is critical for strategic planning and managerial decision-making. The majority of models require significant implementation resources or expensive tools (e.g., FAIR via RiskLens, MEHARI via proprietary toolkits), limiting their applicability for small and medium-sized enterprises. Some models and methods (e.g., NIST SP 800-30, ISO 27005, OWASP Risk Rating) focus exclusively on technical or IT security without a legal context, rendering them insufficient for comprehensive assessment of legal consequences of data breaches. Methods based solely on surveys or interviews (Ponemon, MEHARI) may lose accuracy due to their reliance on subjective judgments and market changes. In this context, approaches based on structured expert rules or formalized coefficients appear more stable.

Ukrainian models included in the analysis demonstrate significant potential for adaptation to national conditions, considering critical infrastructure, e-learning in the Armed Forces of Ukraine, and social network topologies. These approaches are promising but require further development regarding standardization, GDPR compliance, and the creation of practical implementation tools.

Conclusions

Thus, there is an urgent need for the development of integrated theoretical and applied tools for assessing the consequences of personal data breaches, which would provide a comprehensive, formalized, and standardized approach to risk analysis. Current models, methods, and systems

demonstrate a limited ability to cover all aspects of the negative consequences of personal data breaches and show insufficient adaptation to GDPR requirements. Integrated tools should combine formalized coefficients, structured expert rules, and historical data analysis, allowing for accurate assessment of both direct and indirect consequences, including financial, reputational, and legal risks. It is also essential to consider national specificities, as demonstrated by Ukrainian models for critical infrastructure, e-learning in the Armed Forces of Ukraine, and social network topologies, ensuring the adaptation of tools to the specific legal and technological context.

From a practical perspective, the development of integrated tools should not only provide a formalized assessment of losses but also support managerial decision-making in crisis situations, where rapid and well-founded incident response is critical to minimizing harm. The use of such tools enables the generation of quantitative and qualitative risk assessment results, their integration into risk management and strategic planning systems, and ensures compliance with contemporary data protection legislation. Prospects for further research include the development of universal models and methods compatible with GDPR and other international standards, supporting automated data processing and providing practical instruments for use in both public and corporate sectors.

References

- [1] Cost-Effective Risk Management | Resources. Quantitative Information Risk Management | The FAIR Institute. URL: <https://www.fairinstitute.org/learn-fair>.
- [2] Fair Information Practice Principles (FIPPs). Home | FPC.gov. URL: <https://www.fpc.gov/resources/fipps/>
- [3] Introduction to FAIR. Medium. URL: <https://medium.com/@enstructure/introduction-to-fair-bc5e7da0e72c>
- [4] NIST Technical Series Publications. URL: <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-30r1.pdf>
- [5] NIST Risk Management Framework | CSRC. NIST Computer Security Resource Center | CSRC. URL: <https://csrc.nist.gov/Projects/risk-management/about-rmf>
- [6] Луцький М.Г., Іванченко Є.В., Казмірчук С.В., Охрименко А.А. Сучасні засоби управління інформаційними ризиками. Захист інформації. 2011. Т.13. № 3 (52). С. 97–108.
- [7] Луцький М.Г., Корченко А.Г., Іванченко Є.В., Казмірчук С.В. Дослідження програмних засобів аналізу та оцінки ризиків інформаційної безпеки. Захист інформації. 2011. Т. 13. № 2 (51). – С. 86–94.
- [8] NIST Privacy Framework: NIST Technical Series Publications. URL: <https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01162020.pdf>
- [9] ДСТУ ISO/IEC 27005:2023 Інформаційна безпека, кібербезпека та захист конфіденційності. Настанова керування ризиками інформаційної безпеки (ISO/IEC 27005:2022, IDT). БУДСТАНДАРТ Online - нормативні документи будівельної галузі України. URL: https://online.budstandart.com/ua/catalog/doc-page.html?id_doc=104400
- [10] ДСТУ ISO 31000:2018 Менеджмент ризиків. Принципи та настанови (ISO 31000:2018, IDT) URL: https://zakon.isu.net.ua/sites/default/files/normdocs/dstu_iso_31000_2018.pdf
- [11] OWASP Risk Rating Methodology | OWASP Foundation. OWASP Foundation, the Open Source Foundation for Application Security | OWASP Foundation. URL: https://owasp.org/www-community/OWASP_Risk_Rating_Methodology
- [12] Data Protection Impact Assessment (DPIA) - GDPR.eu. GDPR.eu. URL: <https://gdpr.eu/data-protection-impact-assessment-template/>

- [13] О.Г. Корченко, С.В. Казмірчук, Б.Б. Ахметов, Прикладні системи оцінювання ризиків інформаційної безпеки. Монографія, Київ, ЦП «Компринт», 2017 – 435 с.
- [14] Applying OCTAVE: Practitioners Report. URL: https://kilthub.cmu.edu/articles/report/Applying_OCTAVE_Practitioners_Report/6571985/1?file=12057020
- [15] «Expression des Besoins et Identification des Objectifs de Sécurité EBIOS», Méthode de gestion des risques, ANSSI/ACE/BAC, Paris, Version du 25 janvier 2010, 95 p.
- [16] Потій О. В., Леншин А. В. Дослідження методів оцінки ризиків безпеки інформації та розробка пропозицій з їх вдосконалення на основі системного підходу. Збірник наукових праць Харківського університету Повітряних сил. 2010. Вип. 2. С. 85-91. URL: http://nbuv.gov.ua/UJRN/ZKhUPS_2010_2_21
- [17] Security Studies | Ponemon Institute. Ponemon Institute. URL: <https://www.ponemon.org/research/ponemon-library/security/?tag=5>
- [18] Cyber Value at Risk (CVaR): Measuring Worst-Case Scenarios – Horkan. URL: <https://horkan.com/2025/04/21/cyber-value-at-risk-cvar-measuring-worst-case-scenarios#:~:text=%20Focus%20on%20Worst,from%20ransomware%20to%20insider%20attacks>
- [19] Algarni, A. M., Thayananthan, V., & Malaiya, Y. K. (2021). Quantitative Assessment of Cybersecurity Risks for Mitigating Data Breaches in Business Systems. *Applied Sciences*, 11(8), 3678. URL: <https://doi.org/10.3390/app11083678>
- [20] Korchenko A., Dreis Yu., Roshchuk M., Romanenko O. Consequence evaluation model of leak the state secret from cyberattack directing on critical information infrastructure of the state. *Ukrainian Scientific Journal of Information Security*, 2018, vol. 24, issue 1, p. 29-35. <https://doi.org/10.18372/2225-5036.24.12606>
- [21] Савченко В.А., Ахрамович В.М., Акулінічева М.В. Оцінювання параметрів безпеки персональних даних у ступеневих соціальних мережах на основі їх топології. *Сучасний захист інформації*, №3(43), 2020. С. 6-13. URL: <https://doi.org/10.31673/2409-7292.2020.030613>
- [22] Шапран О.О. Моделі підвищення захищеності персональних даних користувачів системи дистанційного навчання Збройних Сил України. *Телекомунікаційні та інформаційні технології*, 2022, № 2 (79). С. 33-45. URL: <https://doi.org/10.31673/2412-4338.2022.023345>
- [23] Бойченко О. С., Костерев Д. С., Маковський І. Ю., Грищук О.М. Математична модель розрахунку цінності інформації установи. *Проблеми створення, випробування, застосування та експлуатації складних інформаційних систем*, 2022, №22, С.30–40. <https://doi.org/10.46972/2076-1546.2022.22.03>
- [24] Толбатов А., Лозова І., Котик О., Толбатова О. Автоматизована система вибору засобів оцінювання збитків від втрати персональних даних. *ІМА: 2024: Матеріали міжнародної наукової конференції молодих учених «Інформатика, математика, автоматика»*, Суми–Астана, 22–26 квітня 2024 р. Суми, 2024. С.256. URL: <https://files.znu.edu.ua/files/Bibliobooks/Inshi79/0059494.pdf>
- [25] Лозова І., Біскупський А., Горожанова А. Засоби оцінювання шкоди від втрати інформації з обмеженим доступом. *Стан та удосконалення безпеки інформаційно-телекомунікаційних систем (SITS' 2021): збірник тез наукових доповідей*, 23-26 червня 2021 р. Миколаїв – Коблево, 2021. С.58-62.

Cybersecurity as a factor of national and european security: the experience of slovakia in the eu context

Olena Plaksiuk¹, Oksana Yakusheva², Oleksandr Yakushev^{2,3}, Oleksandr Chernyshov⁴

¹ Dmytro Motorny Tavria State Agrotechnological University, 18 B.Khmelnytsky Ave, Melitopol, Zaporizhzhia obl. 72312, Ukraine

² Cherkasy State Technological University, Shevchenka, 460, 18000, Cherkasy, Ukraine

³ Kherson National Technical University, Institutka 11, 29016, Khmelnytsky, Ukraine

⁴ State University "Kyiv Aviation Institute", 1, Liubomyra Huzara ave. Kyiv, 03058, Ukraine

Abstract

The publication highlights the issue of the state of cybersecurity in the Slovak Republic and the EU. The essence and relationship of cybersecurity with information security are revealed, successful examples of the implementation of cybersecurity projects in the Slovak Republic and the EU are presented. Based on the NCSI index, the readiness of EU countries and Slovakia in particular to prevent cyber threats and manage cyber incidents is determined.

Keywords

cybersecurity, financing, management, European Union, Slovak Republic.⁹

1. Introduction

Cybersecurity is a set of processes, best practices, and technology solutions that help protect critical systems and networks from digital attacks. As the volume of data grows and more people work and connect to the network from anywhere, attackers respond by developing sophisticated methods to gain access to resources and steal data, sabotage businesses, countries, or extort money. The number of attacks increases every year, and attackers develop new methods to avoid detection. An effective cybersecurity program includes people, processes, and technology solutions that together reduce the risk of business loss, financial loss, and reputational damage resulting from an attack.

At the same time, cybersecurity is a state in which networks and information systems are able to withstand, with a reasonable degree of confidence, any action that compromises the availability, authenticity, integrity, or confidentiality of data stored, transmitted, or processed, or of related services provided or accessed through those networks and information systems. [1,2]

The main purpose of the publication is to identify and analyze current trends related to cybersecurity, to generalize theoretical approaches and practical experience of European Union countries to combine theory with applied practice in implementing relevant cybersecurity projects.

2. Results

Information security is part of the security management of the entire organization. It involves working with information in general and includes the so-called computer security at the IT technology level. Information security differs from cybersecurity in scope and purpose.

⁹Corresponding authors

✉ plaksyk4@gmail.com (O. Plaksiuk);

ORCID 0000-0002-8707-9350 (O. Plaksiuk)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

These two terms are often used interchangeably, but, more precisely, cybersecurity is a subcategory of information security. In cyberspace, a multitude of threats arise every day, which often develop into real attacks.

Although these threats and vulnerabilities have different nature and form, and their calculation is practically impossible, in practice it is necessary to highlight those threats that in the medium term need to be responded to in a timely and adequate manner. [3-5]

The new Cybersecurity Regulation, which lays down measures to ensure a high common level of cybersecurity in the institutions, bodies, offices and agencies of the European Union, entered into force on 7 January 2024. The new legal framework is based on the Commission's proposal for a cybersecurity regulation of March 2022 and the political agreement reached by the European Parliament and the Council in June 2023. The comprehensive measures of the Regulation to achieve a high level of cybersecurity underline the importance of establishing an internal cybersecurity governance, management and control system that is adapted to the needs of each EU entity and takes into account the evolving nature of cyber threats and the interconnectedness of digital systems.

The Regulation establishes an Interinstitutional Cybersecurity Board to oversee and facilitate the implementation of the Regulation, ensuring that the Union institutions work towards a common cybersecurity standard. The Regulation expands the Computer Emergency Response Team for the EU Institutions (CERT-EU) to serve as a central hub for threat intelligence, information exchange and incident response coordination – it has been renamed the Cybersecurity Service for the Union Institutions, while keeping the acronym CERT-EU. [7,8]

In July 2020, the Commission published a Communication on the EU Security Union Strategy for 2020-2025. The main actions set out therein include "common rules on information security and cybersecurity for all EU institutions, bodies and agencies". This new structure aims to foster close and effective cooperation, with CERT-EU playing a central role. In its Cybersecurity Strategy for the Digital Decade[6,14,15], published in December 2020, the Commission committed to presenting a proposal for a Regulation on common rules on cybersecurity for all EU institutions, bodies and agencies. It also proposed creating a new legal basis for CERT-EU to strengthen its mandate and funding. [3]

The European Parliament has approved the financing of the new Digital Europe Programme in the amount of 9.2 billion euros for 2021-2027. As reported on the European Parliament website, the aim of the programme is to strengthen the EU's competitiveness in the field of cybersecurity and in such innovative areas as supercomputers, big data processing, artificial intelligence, etc. The programme envisages financing projects for the development of digital skills and e-governance (including the implementation of "smart cities" systems and the digitalisation of healthcare, education, justice, etc.).

This Programme is part of a strategy to further develop the Digital Single Market, which could help create four million jobs and boost the EU economy with an annual budget of €415 billion each year, while increasing the EU's international competitiveness. The Digital Europe Programme allocates funds to areas such as building supercomputers (€2.7 billion), research into artificial intelligence (€2.5 billion), strengthening the EU's cybersecurity capacity (€2 billion), promoting the widespread use of digital technologies (€1.3 billion) and developing digital skills (€7 billion). [9]

There are already successful examples of cybersecurity projects under the Digital Europe Programme that are already changing people's lives. The European Digital Media Observatory (EDMO) is the EU's largest interdisciplinary network for countering disinformation. Its aim is to combat online disinformation that threatens democracy in the EU. EDMO is made up of 14 national and multinational centres: it is managed by a consortium led by the European University Institute

in Florence, Italy. The consortium includes Athens Technology Center from Greece, Aarhus University from Denmark and the fact-checking organisation Pagella Politica from Italy. The network brings together fact-checkers working on the EDMO digital platform, media literacy experts and academic researchers to understand and analyse disinformation, as well as media organisations and media literacy professionals. Within the framework of the Digital Europe program, EDMO attracted EU co-financing of 30 million euros.

EU Digital Identity Wallet Pilots is a personal digital wallet that will allow a person to securely identify themselves when accessing a number of public and private services. The project is implemented by 2 lead countries (France, Germany) and 18 partner countries (Austria, Belgium, Cyprus, Czech Republic, Estonia, Finland, Greece, Hungary, Italy, Latvia, Luxembourg, Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain, Ukraine). The aim of the EU Digital Identity Wallet Pilots project is to create a single and secure digital identity system that will allow EU citizens to securely and conveniently identify themselves in the online environment, access public and private services, and increase the level of trust in digital services. The total amount of funding is 46 million euros.

Jupiter: The First European Exascale Supercomputer. Jupiter will be the first European exascale supercomputer, providing a huge boost to the EU AI ecosystem and supporting the development of high-precision models of complex systems and applications. This project aims to create a supercomputer capable of performing calculations with exascale performance, i.e. up to 10^{18} (one billion billion) floating-point operations per second. The project is co-funded by EUR 250 million.

Quite interesting in the field of cybersecurity is the CyberSecPro project, which trains IT professionals to overcome current and future cybersecurity challenges. 15 higher education institutions and 13 security companies are involved in the development of a flexible, collaborative and multimodal curriculum. The project bridges the gap between science and industry, transforming theoretical knowledge into practical skills. CyberSecPro partners will develop tools to develop advanced cybersecurity skills, as well as certificates for students and professionals. CyberSecPro experts have already identified cybersecurity skills gaps in Europe and conducted a series of short training events. The project is currently developing best practice training models for cybersecurity training programmes that meet industry requirements. CyberSecPro co-funding is €6.7 million. [10]

The EU reaffirms its commitment to the peaceful resolution of international disputes in cyberspace, including through diplomatic and legal instruments. The 2015 EU Council Conclusions on Cyber Diplomacy state that the same norms and principles that the EU recognises offline, including the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights and the European Convention for the Protection of Human Rights and Fundamental Freedoms, should also apply in cyberspace. They also stress that the EU should actively promote a free and open internet, improve cooperation in the fight against cybercrime and strengthen cyber capacity in third countries, and call for closer cooperation between all stakeholders on cybersecurity.

At EU level, a European Cybersecurity Policy is currently being prepared to strengthen preparedness against cyberattacks, as well as a new European Cyberresilience Law, which aims to unify common standards for digital products and value-added services as a basic prerequisite for strengthening the resilience of European cyber systems and processes to cyberattacks. In response to the current situation in Ukraine, during which attacks on state facilities and telecommunications systems were also recorded, the EU activated the Cyber Rapid Response Team (CRRT) for the first time at the request of Ukraine. [11]

The National Cybersecurity Index (NCSI) has been developed to analyze and assess countries' readiness to prevent cyber threats and manage cyber incidents. [12] The NCSI is also a database of publicly available materials and tools for building national cybersecurity capacity. An overview of the European Union countries that rank highest among the world's countries according to the NCSI index is shown in Table 1.

Table 1

Ranking of European Union countries according to the NCSI 2025

Rank	Країна	National Cyber Security Index	Digital Development Level
1	Czech Republic	98,33	72,93
2	Estonia	96,67	78,14
3	Finland	95,83	85,76
4	Romania	92,50	64,57
5	Poland	92,50	73,21
6	Belgium	92,50	73,55
7	Germany	90,83	75,73
8	France	89,17	78,08
9	Denmark	89,17	85,59
10	Italy	88,33	73,58
11	Lithuania	85,00	75,53
12	Austria	85,00	78,35
13	Portugal	84,17	72,94
14	Croatia	82,50	70,07
15	Netherlands	81,67	84,66
16	Bulgaria	80,83	67,30
17	Slovakia	80,83	67,55
18	Latvia	79,17	73,10
19	Switzerland	79,17	81,88
20	Ireland	77,50	78,89

The NCSI is developed and implemented by the Estonian e-Government Academy Foundation, while the Global Cybersecurity Index (GCI)² is compiled by the International Telecommunication Union (ITU), a stakeholder initiative aimed at raising awareness of cybersecurity and measuring countries' commitment to cybersecurity and its widespread application across industries and sectors. The level of development of each country is analyzed in five categories: legal measures, technical measures, organizational measures, capacity building, and cooperation. [13]

As we can see from the table, the first place in the ranking among EU countries is occupied by the Czech Republic. Slovakia, although a country very close mentally and territorially, occupies only 17th place among EU countries and 24th place in the world. The ranking of countries by NCSI is graphically depicted in Figure 1. [13]

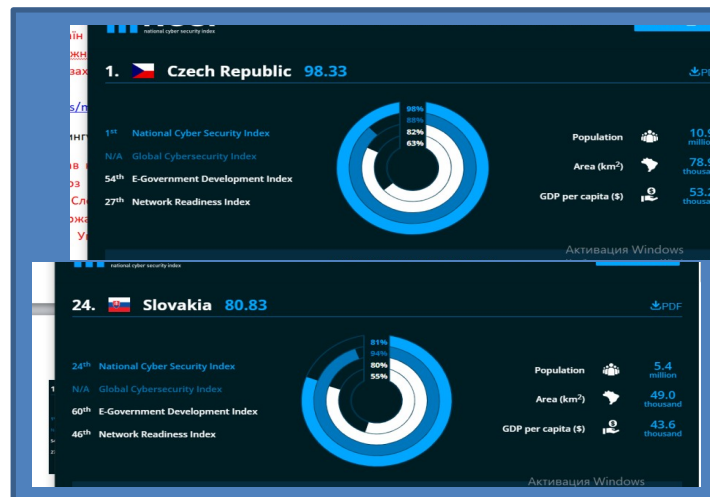


Figure 1: Ranking of the Slovak and Czech Republics for the National Cybersecurity Index 2025. [13]

In 2016, the Slovak government responded to the dynamic development in the field of information and communication technologies and the growth of security threats and risks in cyberspace by strengthening the competences of the National Security Office of the Slovak Republic (NSO) in the field of coordination of activities to ensure the cybersecurity of the state at the national level. In the same year, the Slovak government approved the independent statute of the National Security Office of the Slovak Republic, which granted it competences in the field of cybersecurity. These processes were defined in the Act on Cybersecurity, approved in January 2018, which, among other things, states that the National Security Office of the Slovak Republic, in cooperation with the Ministry of Diplomacy, develops international cooperation in the field of cybersecurity.

The program statement of the Government of the Slovak Republic for 2020-2024 draws attention to the need to strengthen protection against cyberattacks and identifies cybercrime as one of the most dangerous social challenges. The Security Strategy of the Slovak Republic (2021) emphasizes that cyberspace security is extremely important for the proper functioning and sustainability of the state and our society. The Security Strategy also draws attention to the intensification of global competition for technological and strategic advantage in cyberspace. In January 2021, the Government of the Slovak Republic approved the National Cybersecurity Strategy for 2021-2025, which sets the direction of the development of the Slovak Republic in this area and whose vision is to strengthen an open and secure cyberspace. The National Strategy is based on the Action Plan for the Implementation of the National Cybersecurity Strategy for 2021-2025, which defines specific tasks related to individual activities, responsible persons, and time horizons. [11]

3. Conclusion

The study confirmed that within the framework of information security, cybersecurity is the most risky due to its rapidly changing and progressive aspects, international dimension and existence in different systems. The biggest threats are misuse and theft of data, breaches of systems storing personal data, phishing attacks, but most importantly, in many aspects, the absence or insufficiency of standards, laws, regulations and insufficient awareness. In the context of comparing cybersecurity in the EU and the Slovak Republic, it can be said that the EU establishes and implements solutions for managing cybersecurity mainly within the framework of regulations and the creation of other organizations and councils to combat cyber threats, which also requires significant financial resources.

However, achieving adequate resilience to cyber threats has become one of the main priorities of the European Union in recent years. Many measures introduced by the European Union are also followed by the Slovak Republic. Cybersecurity in the Slovak Republic is addressed by a complex system that includes not only the laws regulating it, but also practical activities such as risk management, detection and resolution of cyber incidents, system recovery, education, dissemination of security information and, last but not least, research and development of cybersecurity tools and processes.

References

- [1] Act No. 69/2018 Coll. on Cybersecurity and on Amendments to Certain Acts. URL: https://static.slov-lex.sk/static/SK/ZZ/2018/69/vyhlasene_znenie.html
- [2] Hutsaliuk O., Bondar Iu., Kalinin O., Sokolovskiy V., Navolokina A. Integration Development of Logistics Activities of Corporate Enterprises Based on Intellectualization and Management Technologies. Intelligent Transport Systems: Ecology, Safety, Quality, Comfort. ITSESQC 2024. Lecture Notes in Networks and Systems. Cham: Springer, 2025. 1335. 270-289. URL: https://doi.org/10.1007/978-3-031-87376-8_24
- [3] European court of auditors. 2022. Special Report - Cybersecurity of EU institutions, bodies and agencies: the overall level of preparedness does not match the threats. Luxembourg: Publications Office of the European Union, 2022. vol. 05. p. 62. ISBN 978-92-847-7599-6
- [4] Hutsaliuk O., Havrylova N., Alibekova B., Rakayeva A., Bondar Iu., Kovalenko Yu. Management of Renewable Resources in the Energy Sector: Environmental, Economic and Financial Aspects. Green Energy and Technology. Circular Economy for Renewable Energy. Cham: Springer, 2023. 69-89. URL: <https://doi.org/10.1007/978-3-031-30800-0>
- [5] Hutsaliuk O., Bondar Iu., Tomareva-Patlahova V., Kalinin O., Navolokina A., Kozlovskaya S. Integration of Operational and Economic Security of the Enterprise: Reorganization, Technologies, Risks and Outsourcing. Data-Centric Business and Applications. Lecture Notes on Data Engineering and Communications Technologies. Cham: Springer, 2025. 240. 35-61. URL: https://doi.org/10.1007/978-3-031-81557-7_3
- [6] National security office. 2021. National Cyber Security Strategy for 2021 to 2025. URL: <https://www.nbu.gov.sk/narodna-strategia-kybernetickej-bezpecnosti/>
- [7] Microsoft. 2024. What is Cybersecurity?. In: Security 101. 2024. URL: <https://www.microsoft.com/enus/security/business/security-101/what-is-cybersecurity>
- [8] Pinggen, A. 2023. New Regulation for Cybersecurity of EU Institutions. In: Eucrim.eu 2023. vol. 4. URL: <https://eucrim.eu/news/new-regulation-for-cybersecurity-of-eu-institutions/>

- [9] European Parliament to finance new EU competitiveness program in cybersecurity. URL: <https://eu-ua.kmu.gov.ua/news/yevroparlament-profinansuye-novu-programu-konkurentospromozhnosti-yes-u-sferi-kiberbezpeky/>
- [10] Successful projects of the EU Digital Europe Programme. URL: <https://business.dia.gov.ua/history-of-success/uspishni-proiekty-prohramy-yes-tsyfrova-yevropa>
- [11] Ministerstvo zahraničných vecí a európskych záležitostí slovenskej republiky. 2022. Kybernetická bezpečnosť. URL: <https://www.mzv.sk/diplomacia/bezpecnostnapolitika/kyberneticka-bezpecnost>
- [12] NCSI Project Team. e-Governance Academy. URL: <https://ncsi.ega.ee/contact/>
- [13] Liliya Oleksiuk. Best practices in cybersecurity management. 2018. URL: https://www.undp.org/sites/g/files/zskgke326/files/migration/ua/Report_on_Cybersecurity_04.pdf
- [14] European Parliament (EP). 2022. How the Parliament wants to boost cybersecurity in the EU . URL: <https://www.parlementairemonitor.nl/9353000/1/j9vvij5epmj1ey0/vk0gkfd9vbfw>
- [15] Havryliuk, O., Yakushev, O., Petchenko, M., Zachosova, N., Bielialov, T., & Kozlovskaya, S. (2023). Cyber security and artificial intelligence in the context of ensuring business security in wartime. Financial and credit activity problems of theory and practice, 6(53), 451-459. DOI: <https://doi.org/10.55643/fcaptp.6.53.2023.4130>
- [16] Zakharova, O., Podluzhna, N., Yakusheva, O., & Yakushev, O. (2020, October). Regional policy of advanced ICT in the minds of forming economy knowledge. In ICT in Education, Research and Industrial Applications: Proceedings of the 16th International Conference, ICTERI 2020, Kharkiv, Ukraine, October 6-10, 2020. Volume I: Main Conference (Vol. 2740, pp. 386-391). ICT in Education, Research and Industrial Applications: Proceedings of the 16th International Conference, ICTERI 2020, Kharkiv, Ukraine, October 6-10, 2020. Volume I: Main Conference.
- [17] Telnova H., Kolodiziev O., Petchenko M., Yakushev O., Shulga N., & Kochetkov V. (2023). Foreign trade policy and its impact on economic growth. Financial and credit activity problems of theory and practice, 4(51), 345-357. DOI: <https://doi.org/10.55643/fcaptp.4.51.2023.4097>

Training specialists in cybersecurity and information protection as a component of the national strategy under the condition of hybrid warfare

Larysa Prodanova¹, Oksana Zakharova¹, Oksana Yakusheva¹ and Nelia Nagaichuk¹

¹ Cherkasy State Technological University of Cherkasy, Shevchenko blvd. 460, 18006 Cherkasy, Ukraine

Abstract

The paper assesses the current state of the system of training specialists in cybersecurity and information protection in Ukraine. A set of factors of an international, national, and technological nature that confirm the need for a radical restructuring of the system of training specialists in information protection and cybersecurity is substantiated. A ranking of higher education institutions in Ukraine is compiled based on the total number of students enrolled in specialty 125 "Cybersecurity and Information Protection". The ways to adapting the national system of training cybersecurity and information protection specialists to the challenges of hybrid warfare are determined.

Keywords

national security system, forms of cybersecurity breaches, consequences of cyber incidents, state policy of education, higher education system, digital sustainability, critical infrastructure

1. Introduction

Modern challenges to the national security of Ukraine in the context of Russia's hybrid war against Ukraine make the problem of training highly qualified specialists in information protection and cybersecurity relevant. Information and communication technologies have become not only the basis for the functioning of the state and society, but also a battlefield where cyber threats are taking on the character of a strategic weapon.

Hybrid warfare is characterized by the integrated use of traditional military means along with cyberattacks on critical infrastructure, state institutions, and the private sector. In such conditions, the national security system needs specialists who are able not only to respond to existing threats, but also to predict new challenges in cyberspace.

The problem of staffing in the cybersecurity sector is becoming particularly acute, as traditional approaches to training specialists lag behind the dynamics of cyberthreats. At the same time, Ukraine's integration into European and Euro-Atlantic security structures requires bringing national standards for training specialists into line with international requirements.

The purpose of the paper consists in assessing the current state of the system of training specialists in cybersecurity and information protection in Ukraine and determining ways to adapt it to the challenges of hybrid warfare.

2. Results

The need for a radical restructuring of the system of training specialists in information protection and cybersecurity is confirmed by a complex of factors of an international, national, and technological nature.

The international context demonstrates the critical importance of the problem. According to The World Economic Forum "Global Risks Report 2025", cyber threats are among the ten most likely global risks of the next decade [1]. The International Information System Security Certification Consortium (ISC)² Cybersecurity Workforce Study records a 4.76 million specialists global shortage of cybersecurity professionals, which undermines the ability of both national

security systems and business structures to adequately respond to the evolution of cyber threats and provide an adequate level of cyber protection [2]. The Cybersecurity Capacity Maturity Model for Nations, developed by the Global Cyber Security Capacity Centre at Oxford University, confirms the direct correlation between the level of training (development of knowledge and capabilities) in the field of cybersecurity and the cybersecurity capability of the state [3].

The national legislation of Ukraine forms the legal basis for strengthening cybersecurity training. The Cybersecurity Strategy of Ukraine (for 2021-2025) identifies human resources as one of the strategic priorities [4]. The Law of Ukraine "On Amendments to Certain Laws of Ukraine on Information Protection and Cybersecurity of State Information Resources and Critical Information Infrastructure Objects" (No. 4336-IX of March 27, 2025) [5], which is currently one of the key laws in the field of cybersecurity, also affects to a certain extent the requirements for professional training of specialists. The law provides for the following: the introduction of full-time positions for cyber protection specialists in government agencies and critical infrastructure facilities – this is a direct requirement for the availability of qualified specialists; the creation and operation of a national cyber incident response system – this requires specialists not only to have knowledge in the field of defense, but also skills in investigating cyberattacks, restoring systems, and exchanging information about threats. Thus, Law No. 4336-IX actually establishes specific, albeit general, requirements for the staff and functionality of specialists, which in turn stimulates the development of their professional training.

Technological challenges cause a radical change in professional competencies. Technological developments and accelerating digital transformation create new challenges for ensuring cyber security. Cybersecurity breaches in modern conditions can take the following forms: identity fraud, data theft, ransomware attacks, cyberattacks on critical infrastructure, or phishing campaigns, etc. [6, 7]. According to the research by Cyber Management Alliance Ltd.: IoT malware attacks grew by 107% by the end of 2024; ransomware affected 59% of respondents; phishing attacks have increased dramatically by 4151% since the public release of ChatGPT (late 2022); about 8 million DDoS attacks were recorded in the first half of 2024; the average total cost of data breaches in 2024 was \$4.88 million [8]. The most common consequences of cyber incidents and cybersecurity breaches in business are: loss of trust in technology, disruption of operations (including supply chain or partner ecosystem), loss of reputation, negative impact on talent recruitment and retention, loss of revenue, loss of customer trust and negative impact on brand, theft of intellectual property, regulatory fines, falling stock prices, termination of funding for a strategic initiative, deterioration of competitive positions [6, 9]. The emergence of new types of threats, including AI-driven attacks and potential threats from quantum computing, requires fundamentally new approaches to training specialists.

As of September 6, 2025, 58 higher education institutions in Ukraine provide training for cybersecurity and information protection specialists (specialty 125 "Cybersecurity and Information Protection" and 125 "Cybersecurity") at various educational levels (bachelor's, master's and postgraduate studies). This number indicates a significant development of the education system in this field and an increase in its relevance in modern conditions. Table 1 shows the rating of higher education institutions in Ukraine (Top 10) that have the largest number of students (totally at all the education levels and forms of study). According to the results of the analysis of data from the Register of Educational Activity Subjects of the Unified State Electronic Database on Education of Ukraine [10], it was found that the total number of higher education applicants in the analyzed specialty in Ukrainian higher education institutions as of September 6, 2025 is 10,214 people, including 8,672 people at the first (bachelor's) level, 1,285 at the second (master's) level, and 257 at the third (educational and scientific, PhD) level.

Table 1

Top 10 higher education institutions of Ukraine by the number of students in the specialty 125 "Cybersecurity and Information Protection" and 125 "Cybersecurity", as of 06.09.2025 (persons, full-time/part-time forms of education)

Institutions rating		Educational degrees			
		Bachelor	Master	Doctor of Philosophy	Total
1	Lviv Polytechnic National University	873 / 81	127 / 8	42 / 3	1134
2	State University of Information and Communication Technology	582 / 3	154 / 0	35 / 4	778
3	State Non-Commercial Company «State University «Kyiv Aviation Institute»	548 / 51	95 / 10	10 / 3	717
4	National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»	410 / 38	52 / 10	11 / 7	528
5	Kharkiv National University of Radio Electronics	258 / 11	75 / 19	25 / 5	393
6	Vinnitsia National Technical University	320 / 0	51 / 0	11 / 1	383
7	State University of Trade and Economics	294 / 22	34 / 7	-	357
8	Dnipro University of Technology	257 / 31	50 / 5	5 / 0	348
9	Taras Shevchenko National University of Kyiv	253 / 0	46 / 0	13 / 0	312
10	State University of Intelligent Technologies and Telecommunications	148 / 32	30 / 15	10 / 4	239

To assess the quality of specialist training, it is advisable to use a comprehensive approach that includes academic indicators, professional training results, and evaluation of the educational program and environment [11]. According to the Scientific and Methodological Center for Higher and Professional Pre-Higher Education of the Ministry of Education and Science of Ukraine, the average percentage of correct answers on the Unified State Qualification Exam (USQE) among applicants for the first (bachelor's) level of higher education in specialty 125 "Cybersecurity", whose educational program ends in the 2023-2024 academic year, is 41.97% (the threshold score is set at 28 correct answers out of 93 test items). The specific share of students who did not pass the exam is 7.99% of the total number of those who took it [12]. The best results among the universities from the TOP-10, according to the exam results, were shown by students from Taras Shevchenko National University of Kyiv – 100.0% of those who took the exam, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" – 99.2%, and State University of Intelligent Technologies and Telecommunications – 97.6%.

According to the State Service for Special Communications and Information Protection (State Special Communications), another important stage of reforming the specialized education and training system in the field of cybersecurity in Ukraine was completed in 2024. The reform of specialized education was based on the creation of the National Cybersecurity Qualification Framework, the introduction of a system of professional standards, and the creation of a network of independent qualification centers in Ukraine. This approach takes into account the best global practices, is based on the experience of the European Qualifications Framework and the American Strategic Educational Initiative in the field of cybersecurity (in general, as of the end of 2024,

specialists from the State Service for Special Communications developed, and the National Qualifications Agency approved changes to the State Classifier of Professions, which allowed adding 27 new professions in the field of cybersecurity and information security) [13].

Adapting the system of training cybersecurity specialists to the challenges of hybrid warfare requires a comprehensive approach that includes the modernization of educational programs, the development of infrastructure, and the formation of new mechanisms for interaction with the practical field. The primary task is to revise educational standards taking into account the specifics of hybrid threats. This involves including modules in the curriculum on analyzing wartime cyberattacks, protecting critical infrastructure, countering disinformation campaigns, and ensuring system resilience under stressful conditions. Particular attention should be paid to studying the experience of countering russian cyberattacks and analyzing the best global practices for ensuring cyber resilience in conflict situations. The training system requires a radical strengthening of the practical component and the development of practice-oriented training. This approach includes the creation of cyber training grounds to simulate real attack scenarios, conducting regular cyber training with the participation of students, developing cyber incident simulators, and introducing mandatory internships in key government institutions and private companies. Close integration of military and civilian components of specialist training is important in the context of hybrid warfare. This involves the creation of joint educational programs between civilian higher education institutions and military educational institutions, the development of courses on the basics of military cybersecurity for civilian specialists and, conversely, in-depth study of civilian aspects of cyber defense for military specialists. Adaptation also requires bringing the Ukrainian training system into line with NATO and EU standards. This includes the implementation of international educational standards, the development of academic mobility programs, the involvement of international experts in teaching, and the creation of joint educational programs with universities in partner countries. The system requires the creation of a modern digital training infrastructure and educational ecosystem, including cloud cybersecurity laboratories, virtual simulators for practicing incident response skills, and online platforms for distance learning with a high level of security [14]. The development of distance learning opportunities to ensure the continuity of the educational process under martial law is particularly relevant. Hybrid warfare requires the formation of a system of continuous professional development and constant adaptation of the knowledge and skills of specialists to new realities. This necessitates the creation of a lifelong learning system that includes advanced training programs, certification courses, professional conferences and workshops. Effective training of specialists is impossible without a powerful scientific base, which involves the creation of specialized scientific centers for the study of cyber threats, the development of grant programs for young researchers, and the formation of consortiums of universities and businesses for joint R&D projects in the field of cybersecurity [15,16]. It is critically important to build partnerships with business, the IT industry, and cybersecurity companies: creating industry councils at higher education institutions, developing dual education programs, introducing a system of corporate scholarships and grants, and conducting joint projects and hackathons.

3. Conclusion

The modern challenges of hybrid warfare are radically changing the requirements for professional training of cybersecurity specialists, which necessitates the need for a comprehensive reform of the existing system of training specialists in the field of cybersecurity and information protection. Traditional approaches to training do not meet the dynamics of cyberthreats and the specifics of countering hybrid attacks.

The scale of cybersecurity training in Ukraine is significant: 58 higher education institutions train over 10,000 specialists at all educational levels. This demonstrates that the state and educational community are aware of the critical importance of staffing national cybersecurity.

Effective response to modern challenges in the field of cybersecurity in Ukraine requires a comprehensive transformation of the system of training specialists, which should be based on a practical orientation, constant adaptation to threats, and close cooperation between the state, education, and business. This will not only overcome the personnel shortage, but also ensure the country's stability in the conditions of hybrid warfare.

References

- [1] World Economic Forum. The Global Risks Report 2025. URL: <https://www.weforum.org/publications/global-risks-report-2025/>
- [2] ISC2 Cybersecurity Workforce Study. Global Cybersecurity Workforce Prepares for an AI-Driven World. 2024. URL: <https://edge.sitecorecloud.io/internationalf173-xmc4e73-prodbc0f-9660/media/Project/ISC2/Main/Media/documents/research/2024-ISC2-WFS.pdf>.
- [3] Cybersecurity Capacity Maturity Model for Nations Global Cyber Security Capacity Centre. URL: <https://gcsc.ox.ac.uk/the-cmm>
- [4] Decree of the President of Ukraine No. 447/2021 of August 26, 2021. On the decision of the National Security and Defense Council of Ukraine of May 14, 2021 "On the Cybersecurity Strategy of Ukraine." URL: <https://zakon.rada.gov.ua/laws/show/447/2021#Text>
- [5] Law of Ukraine No. 4336-IX of March 27, 2025. On amendments to certain laws of Ukraine regarding information protection and cyber protection of state information resources and critical information infrastructure facilities. URL: <https://zakon.rada.gov.ua/laws/show/4336-20#Text>
- [6] Statista/ Cyber Crime & Security. URL: <https://www.statista.com/markets/424/topic/1065/cyber-crime-security/#overview>
- [7] Savchuk, S. O. State policy on combating cybercrime : thesis. ... PhD (public administration): 25.00.05. Zhytomyr, 2024. 160 p.
- [8] Top 10 Biggest Cyber Attacks of 2024 & 25 Other Attacks to Know About! URL: <https://www.cm-alliance.com/cybersecurity-blog/top-10-biggest-cyber-attacks-of-2024-25-other-attacks-to-know-about>
- [9] Deloitte. Global Future of Cyber Survey, 4th Edition. 2024. URL: <https://www.deloitte.com/content/dam/assets-shared/docs/services/risk-advisory/2024/deloitte-global-future-of-cyber-survey-4th-edition-the-promise-of-cyber.pdf>
- [10] Unified state electronic database on education of Ukraine. Register of educational activity subjects. URL: <https://registry.edbo.gov.ua/search/>
- [11] Zakharova, O., Prodanova, L. (2023). The Potential of Higher Education in Ukraine in the Preparation of Competitive IT Specialists for the Post-War Recovery of the Country's Economy. In: Faure, E., Danchenko, O., Bondarenko, M., Tryus, Y., Bazilo, C., Zaspa, G. (eds). Information Technology for Education, Science, and Technics. ITEST 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 178. Springer, Cham. Pp. 582–595.
- [12] Ministry of Education and Science of Ukraine. Scientific and Methodological Center of HPHE. Report on the results of passing the state qualification exam in specialty 124 "Cybersecurity". Date: April 25, 2024. Kyiv, 2024. URL: <https://nmc-vfpo.com/wp-content/uploads/2024/11/zvit-125.pdf>
- [13] Qualification Center for Information Technologies and and Cybersecurity State Research Institute of Cybersecurity Technologies. Another stage of reforming the cybersecurity training

- system has been completed: the National Qualifications Agency approved 7 more professional standards. URL: <https://qc.csi.cip.gov.ua/uk/posts/7>
- [14] Havryliuk, O., Yakushev, O., Petchenko, M., Zachosova, N., Bielialov, T., & Kozlovska, S. (2023). Cyber security and artificial intelligence in the context of ensuring business security in wartime. *Financial and credit activity problems of theory and practice*, 6(53), 451-459. DOI: <https://doi.org/10.55643/fcaptp.6.53.2023.4130>
- [15] Plaksiuk, O., Prodanova, L., Yakusheva, O., Nagaichuk, N., Prikhno, I., & Jakubčinová, M. (2023). Human capital as a factor of socio-economic development of the state: the main trends of the Slovak Republic. *Financial and Credit Activity Problems of Theory and Practice*, 5(52), 283-298. <https://doi.org/10.55643/fcaptp.5.52.2023.4150>
- [16] Yakushev, O., Moisieienko, L., Yakusheva, O., Prodanova, L., Plaksiuk, O., & Chepurda, L. (2024). Socio-economic sustainability of the tourism sector enterprises in the context of the COVID-19 pandemic: global and ukrainian dimensions. *Financial & Credit Activity: Problems of Theory & Practice*, 5(58), 484-499. DOI: <https://doi.org/10.55643/fcaptp.5.58.2024.4377>

The analysis of cybersecurity audit standards in different parts of the world in the context of digital transformation

Olesya Voytovych, Vitalii Volynets

Vinnitsia National Technical University, Vinnitsia, Ukraine

Abstract

This article compares cybersecurity obligations across the EU, North America, Asia, and Africa in four sectors (finance, healthcare, IT, retail). Drawing on a desk review of statutory and industry sources (e.g., GDPR/NIS2/DORA; GLBA/FTC Safeguards/HIPAA; MLPS 2.0/PIPL; POPIA; PCI DSS), it maps where control expectations converge risk analysis, access management, encryption, incident reporting and where they differ in enforceability and evidence requirements. We find that multinational programs often duplicate audits and still miss operational readiness in cloud and supply-chain scenarios. The study concludes by proposing the development of an integrated compliance model and a Unified Control Framework to address these issues.

Keywords

digital transformation; information security; cybersecurity; cybersecurity audit; global cybersecurity standards.

1. Introduction

Digital transformation has increased dependence on cloud and third-party services, exposing organizations to recurring risks such as ransomware, supplier compromise, and cloud misconfiguration. At the same time, companies operate under overlapping international, national, and industry frameworks with different expectations for controls and audit evidence.

We compare cybersecurity obligations across the EU, North America, Asia, and Africa in finance, healthcare, IT, and retail. Based on a desk review of statutes and major standards, we identify common control themes and meaningful divergences in enforceability and assurance. Our results indicate duplicated audits alongside gaps in testing real-world readiness, especially for cloud and supply chains. We propose a jurisdiction-aware control map that ties obligations to testable safeguards and supports a lean evidence pack.

2. Problem Statement

Multinational programs face overlapping cybersecurity obligations that lead to repeated audits, fragmented controls, and weak evidence of real-world readiness. What is missing is a compact, jurisdiction-aware mapping from obligations to normalized controls and to outcome-based tests with a minimal evidence pack. We propose and evaluate such a mapping with regional overlays and an adaptive audit approach aimed at cutting duplicate work while improving coverage of scenario-based testing.

3. Analysis of Recent Research and Publications

Recent work shows that digital services scale across borders while cybersecurity governance remains tied to jurisdictions and is growing more divergent. For multinational firms, this fragmentation produces overlapping and sometimes conflicting obligations shaped not only by risk considerations but also by geopolitics [1].

Studies describe the operational burden as “compliance” or “audit” fatigue: overlapping rules lead to repeated assessments, duplicated documentation, and the diversion of scarce security staff from preventive work to audit preparation. This dynamic encourages check-the-box behavior, where passing an assessment can outweigh improving control performance [2].

Evidence also indicates that documentation-led compliance does not reliably translate into resilience. Cloud and supply-chain exposures, in particular, remain under-tested when effort concentrates on paperwork. The literature therefore calls for integrated, jurisdiction-aware control mappings and more outcome-based verification (e.g., scenario-driven tests) to close the gap between formal compliance and real protection [3].

4. Research Results

A cross-sector comparison across the EU, North America, Asia, and Africa shows recurring control priorities — structured risk management, identity and access governance, encryption, monitoring, and incident handling — yet materially different philosophies about enforceability, accountability, and audit evidence. These contrasts affect how programs are designed and how assurance is demonstrated in practice.

Regulation in financial services is uniformly strict, but the route to assurance diverges by region. The EU’s DORA adopts a prescriptive stance: integrated ICT-risk governance, regular continuity exercises, and threat-led penetration testing (TLPT) for significant entities on a multi-year cadence — at least every three years [4, 5]. North America’s GLBA takes an accountability-first approach. The Safeguards Rule assigns a Qualified Individual to run a risk-based program and calls out foundational safeguards such as encryption of customer information, multi-factor authentication, and activity logging [6]. China’s MLPS 2.0 reflects a state-directed doctrine organized around the “one center, three layers of protection”; where “Important Data” are in scope, Level 3 obligations commonly entail stronger network zoning and separation of development from production environments [7]. Principle-based regimes in Africa, exemplified by South Africa’s POPIA, require “appropriate, reasonable technical and organisational measures,” which in practice drive encryption, RBAC, least-privilege access, and MFA for critical systems [8]. Given this spread, multinational firms gain more from a modular GRC architecture with regional overlays than from a single global policy.

Patient-safety considerations make the healthcare domain distinctive: confidentiality, integrity, and availability all carry direct consequences. Within the EU, GDPR together with NIS2 compels technical-organizational measures (e.g., pseudonymization, encryption) and mature processes for risk analysis, incident response, and supplier oversight [10]. HIPAA in North America is technology-neutral and distinguishes “required” from “addressable” safeguards, permitting justified substitutions — including for encryption-provided the rationale is documented. Several Asian jurisdictions classify healthcare as Critical Information Infrastructure, which triggers stricter governance requirements, dedicated security bodies, personnel vetting for sensitive roles, and annual risk assessments. African frameworks, with POPIA as a reference point, treat health data as “special personal information,” rendering processing unlawful by default unless narrow conditions — explicit consent, legal necessity, or defined research purposes — are satisfied [11]. Across these settings, the difference is less about whether to implement encryption or access control and more about the evidence supervisors expect and the penalties for failure.

Rules that govern the IT supply side often double as instruments of industrial policy. The EU’s NIS2 establishes baseline measures across essential and important entities and introduces personal accountability for governing bodies in cases of non-compliance [12]. Market assurance carries more weight in North America via SOC 2: organizations scope audits to the five Trust Services

Criteria, with Security the only universal requirement, which allows flexible alignment with business models while still signaling assurance to customers [13]. China operationalizes MLPS 2.0 as a formal, state-supervised certification pathway that can function as a gate to market entry. In many African markets, ISO/IEC 27001 certification is a deliberate strategic signal of a formal ISMS while sector-specific regimes continue to mature, helping local providers win contracts with global partners.

E-commerce’s expansion has reshaped the retail baseline, pairing scale with trust. The EU operationalizes GDPR through data-subject-rights processes-data minimization, consent management, and dependable handling of access and deletion requests alongside security of processing. North America’s dominant scheme is PCI DSS, a contractual regime from the card brands; version 4.0 strengthens expectations around network security, cryptography, and multi-factor authentication for access to the cardholder data environment. China’s PIPL adds consent and, for higher-risk processing, a Personal Information Protection Impact Assessment; it also restricts discriminatory pricing driven by automated decision-making. Across rapidly growing African markets, the policy goal is to build foundational consumer trust; Kenya’s Data Protection Act exemplifies this orientation by entrenching purpose limitation and data minimization as prerequisites for sustainable growth [14].

Table 1

Cybersecurity Standards for the Retail Sector (Summary)

Region	Key Laws and Standards	Key Incentives
EU	GDPR, NIS2, PCI DSS	GDPR Fines / PII Protection
North America	CCPA (California), PCI DSS, NIST CSF	PCI DSS / Litigation
Asia	China: PIPL, e-commerce laws	State Control / Data Localization
Africa	National data protection laws	Consumer Rights / Fraud Prevention

Across sectors and regions, shared ground is clear, but so are the splits: how mandatory measures are, what evidence supervisors expect, who is held to account, and whether data must remain domestically stored. Data-localization rules crystallize the divide. China requires domestic storage and transfer controls for personal and “important” data particularly for critical infrastructure whereas the EU permits cross-border transfers subject to adequate safeguards rather than a blanket storage mandate inside the Union.

For multinational programs, a single, static policy is unlikely to span these variations. A more workable path is a modular compliance and assurance stack: a core control set mapped to common obligations; regional overlays where law or supervisory practice demands more (e.g., TLPT cadence under DORA [4], [5] or level-based measures under MLPS 2.0 [7]); and a compact evidence pack that emphasizes outcome-oriented tests over paperwork. Such a design reduces duplicate audits, aligns documentation with real-world readiness in cloud and supply-chain scenarios, and leaves room for sector-specific enforcement models from EU legal mandates to North American market attestations and African capability-building anchored in ISO/IEC 27001 and emerging data-protection statutes.

5. Conclusions

Our analysis reveals that despite shared global trends, regional regulatory philosophies remain the primary force shaping corporate cybersecurity systems. The analysis demonstrated fundamentally different regulatory approaches: from the strictly regulated in the EU to the flexible in the US, state-controlled in Asia, and fragmented but developing in Africa. This highlights the insufficient unification of existing approaches and creates significant difficulties for international companies.

Based on these findings, we propose that future research should focus on developing an integrated compliance model that would allow for the dynamic adaptation of internal controls to the specific requirements of different jurisdictions, thereby minimizing the regulatory burden. Modern companies face exhaustion from a surplus of regulations, performing numerous, often redundant, audits. This creates an urgent need to harmonize audit practices through the creation of a Unified Control Framework that maps the requirements of key standards. Furthermore, it was found that formal compliance does not always guarantee actual cyber resilience. Consequently, a relevant task is to develop a flexible cybersecurity maturity assessment model that integrates technical and organizational aspects and allows for scenario-based testing to verify real-world preparedness for cyberattacks.

6. References

- [1] A. Kuzior, I. Tiutiunyk, A. Zielińska, R. Kelemen, Cybersecurity and cybercrime: Current trends and threats, *Journal of International Studies* 17 (2024) 220–239. doi:10.14254/2071-8330.2024/17-2/12.
- [2] O. Ilori, C. E. Lawal, S. C. Friday, J. Ibine, E. C. Eke, D. R. Tolulope, Cybersecurity auditing in the digital age: A review of methodologies and regulatory implications, *Journal of Frontiers in Multidisciplinary Research* 3 (2022) 174–187. doi:10.54099/IJFMR.2022.0301.174.
- [3] A. Folorunso, I. Wada, B. Samuel, V. Mohammed, Security compliance and its implication for cybersecurity, *World Journal of Advanced Research and Reviews* 24 (2024) 2105–2121. doi:10.30574/wjarr.2024.24.1.3170.
- [4] S. J. Ursillo Jr., K. Manske, B. Kirk, CMMC programmatic final rule status: What it means for defense contractors, 2025. URL: <https://www.cbh.com/insights/articles/cmmc-compliance-and-its-implications-for-defense-contractors/>.
- [5] M. Osiashvili, The key role of the most recent EU regulation – the “Digital Operational Resilience Act” in the legal system, contemporary challenges, and Georgian perspectives, *European Scientific Journal (ESJ)* 20 (2024) 1–16. doi:10.19044/esj.2024.v20n26p1.
- [6] M. Toussaint, S. Kirmé, H. Panetto, Industry 4.0 data security: A cybersecurity frameworks review, *Journal of Industrial Information Integration* 39 (2024) 100694. doi:10.1016/j.jii.2024.100694.
- [7] R. Creemers, China’s emerging data protection framework, *Journal of Cybersecurity* 8 (2022) tyac011. doi:10.1093/cybsec/tyac011.
- [8] T. Chimboza, E. Smith, How does compliance with the Protection of Personal Information Act (POPI Act) affect organisations in South Africa?, in: *Proceedings of the 10th Annual African Conference on Information Systems and Technology (ACIST 2024)*, Kennesaw, GA, USA, 2024. URL: <https://digitalcommons.kennesaw.edu/acist/2024/presentations/19>.
- [9] I. M. Dorosh, Cybersecurity and its role in the financial sector: Threats and protection measures, *Economics. Finances. Law* 10 (2023). doi:10.37634/efp.2023.10.10.

- [10] M. van 't Schip, The regulation of supply chain cybersecurity in the NIS2 Directive in the context of the Internet of Things, 2024. doi:10.2139/ssrn.4848048.
- [11] P. Ewoh, T. Vartiainen, Vulnerability to cyberattacks and sociotechnical solutions for health care systems: Systematic review, *JMIR Medical Internet Research* 26 (2024) e46904. doi:10.2196/46904.
- [12] P. G. Chiara, Towards a right to cybersecurity in EU law? The challenges ahead, *Computer Law & Security Review* 53 (2024) 105961. doi:10.1016/j.clsr.2024.105961.
- [13] O. Deineka, O. Harasymchuk, A. Partyka, A. Obshta, N. Korshun, Designing data classification and secure store policy according to SOC 2 Type II, in: *Cybersecurity Providing in Information and Telecommunication Systems 2024*, vol. 3654, 2024, pp. 398–409.
- [14] B. Trivedi, Retail cybersecurity in the agentic age: Securing autonomous shopping agents in e-commerce, *European Modern Studies Journal* 9 (2025) 52. doi:10.59573/emsj.9(4).2025.52.



БЕЗПЕКА КІБЕРФІЗИЧНИХ СИСТЕМ ТА КРИТИЧНОЇ ІНФРАСТРУКТУРИ

Методи та моделі оцінювання стану кібербезпеки об'єктів критичної інфраструктури суб'єктів авіаційної галузі

Андрій Біскупський¹, Ігор Іванченко²

¹ Державний університет інформаційно-комунікаційних технологій, Україна, 03110, м. Київ, вул. Солом'янська, 7

² Державний університет інформаційно-комунікаційних технологій, Україна, 03110, м. Київ, вул. Солом'янська, 7

Анотація

У роботі розглянуто питання забезпечення кібербезпеки об'єктів критичної інфраструктури авіаційної галузі. Показано, що традиційні підходи до оцінювання ризиків, викладені у міжнародних стандартах (ISO/IEC 27005, NIST SP 800-30, ДСТУ IEC/ISO 31010), потребують адаптації з урахуванням специфіки авіаційної діяльності та вимог ICAO. Запропоновано застосування методів ризик-орієнтованого аналізу та апарату нечіткої логіки для формування інтегральної оцінки стану кібербезпеки. Наведено приклад використання матриці ризиків та нечітких правил для визначення рівня кіберризиків у інформаційних системах аеропорту.

Ключові слова

кібербезпека, критична інфраструктура, авіація, оцінювання ризиків, нечітка логіка, ICAO

1. Вступ

Авіаційна галузь є ключовим елементом критичної інфраструктури, і зростання кіберзагроз вимагає адаптації існуючих міжнародних стандартів ризик-менеджменту [1–4] до особливостей авіаційних процесів, регламентованих ICAO, для практичного використання в аеропортах, авіакомпаніях, в підприємствах з обслуговування повітряного руху та хендлерів.

2. Мета і завдання дослідження

Мета роботи полягає в розробленні підходу до оцінювання стану кібербезпеки об'єктів критичної інфраструктури авіаційної галузі з урахуванням міжнародних стандартів [1–3] та специфіки ICAO [4].

Завдання:

6. проаналізувати застосовність методик ISO/IEC 27005, NIST SP 800-30 та ДСТУ IEC/ISO 31010 [1–3] та класифікувати загрози для авіаційної КІ;
7. застосувати методи нечіткої логіки [5] для інтегрального оцінювання кіберризиків;
8. розробити прикладну модель на основі ризик-матриці та нечітких правил.

3. Методи дослідження

У роботі використано:

9. методи аналізу ризиків відповідно до ISO/IEC 27005 та NIST SP 800-30 [1–2];
10. класифікацію загроз і вразливостей на основі ICAO Doc 8973 [4];

11. апарат нечітких множин [5] для опису лінгвістичних змінних «загроза», «вразливість», «вплив»;
12. побудову нечітких правил для визначення рівня ризику.

4. Основні результати

Запропоновано узагальнену схему оцінювання кіберризиків для об’єктів критичної інфраструктури авіаційної галузі.

Об’єкти аналізу:

13. інформаційні системи аеропортів (системи відображення рейсів, управління багажем, системи бронювання);
14. аеронавігаційні системи (ATM/ANS, диспетчеризація);
15. технологічні системи (енергозабезпечення, паливні комплекси, SCADA).

Приклад моделі:

16. якщо загроза = висока, вразливість = середня, вплив = високий, то ризик = високий [5]

Це дозволяє сформулювати **інтегральний показник кіберстійкості**, який може бути використаний для практичного аудиту безпеки.

5. Висновки

Запропонований підхід поєднує міжнародні стандарти ризик-менеджменту з нечіткою логікою, адаптовані для авіаційної сфери. Модель може використовуватись для практичного аудиту та гармонізації національних вимог із ICAO.

Література

- [1] ISO/IEC 27005:2022. Information technology — Security techniques — Information security risk management.
- [2] NIST SP 800-30 Rev.1. Guide for Conducting Risk Assessments. Gaithersburg, 2012.
- [3] ДСТУ ІЕС/ІСО 31010:2013. Менеджмент ризиків. Методи оцінювання ризику.
- [4] ICAO Doc 8973. Aviation Security Manual.
- [5] Корченко О.Г. Нечіткі множини в задачах кібербезпеки: навч. посіб. — К.: НАУ, 2018

Тестування згорткової нейронної мережі при використанні в задачах забезпечення безпеки критичної інфраструктури

Олена Висоцька¹, Антон Герасименко², Анатолій Давиденко³, Олександр Корченко²

¹ Державний університет «Київський авіаційний інститут», пр. Гузара Любомира 1, 03058 Київ, Україна

² Державний університет інформаційно-комунікаційних технологій, вул. Солом'янська 7, 03110 Київ, Україна

³ Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, вул. Генерала Наумова 15, 03164 Київ, Україна

Abstract

В роботі запропоновано архітектуру згорткової нейронної мережі для забезпечення задачі автоматичного розпізнавання небезпечних предметів при контролі доступу засобами візуалізації схованих небезпечних предметів. Для навчання та проведення експериментів було використано п'ять спеціалізованих датасетів загальним обсягом більше 24000 зображень. Було досліджено вплив кількості ітерацій на епоху, а саме послідовно проведено три групи експериментів з покроковим збільшенням кількості ітерацій. Наступним кроком було досліджено вплив розміру зображень. На останньому кроці було показано лімітованість позитивного впливу збільшення кількості ітерацій, що було проілюстровано обмеженістю кількості ітерацій в епосі відповідно до загальної кількості даних. Загальним висновком дослідження є: збільшення кількості ітерацій на епоху сприяє підвищенню точності моделі, але кількість ітерацій на епоху лімітована — після досягнення ліміту модель може перенавчитись.

Keywords

згорткова нейронна мережа, небезпечні предмети, зображення, розпізнавання, навчання моделі¹⁰

1. Вступ

В сучасному світі, навіть для звичайної людини, існує велика кількість загроз безпеки, особливо при великому скупченні людей. Причиною однієї з таких загроз є знаходження небезпечних предметів десь в цьому скупченні людей. Зазвичай ця проблема актуальна для об'єктів критичної інфраструктури або в місцях масового збору людей, наприклад в аеропортах, на вокзалах, великих стадіонах при проведенні спортивних змагань або концертів, тощо. Також важливим є той факт, що для даних випадків характерним є великий потік людей. Все це робить актуальною задачу швидкого та ефективного виявлення заборонених небезпечних предметів у великій кількості людей під час контролю доступу до об'єктів критичної інфраструктури. Для вирішення даної задачі можна використовувати відеокамери, інтроскопи та інше спеціалізоване обладнання, але, в даному випадку, критичним є саме швидке та ефективне виявлення відповідних предметів. Тому дане виявлення має бути автоматичним. Відповідно необхідна технологія, за допомогою якої буде виконуватись саме автоматичне виявлення заборонених предметів. Розпізнавання (класифікація) образів є необхідною складовою при вирішенні багатьох задач кібербезпеки [1], в тому числі і задачі виявлення небезпечних предметів. Одним із найбільш актуальних

¹⁰ Corresponding authors

✉ lek_vys@ukr.net (О. Висоцька); a.herasymenko@stud.duiikt.edu.ua (А. Герасименко); davidenkoan@gmail.com (А. Давиденко); o.korchenko@duikt.edu.ua (О. Корченко)

ORCID 0000-0002-9543-1385 (О. Висоцька); 0009-0008-4475-2040 (А. Герасименко); 0000-0001-6466-1690 (А. Давиденко); 0000-0003-3376-0631 (О. Корченко)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

і доцільних математичних апаратів для розпізнавання образів є нейронні мережі. Але ефективність даного підходу суттєво залежить від правильного вибору типу та архітектури нейронної мережі, яка застосовується, від налаштування критичних параметрів та від бази даних зразків (тестувальних та навчальних). Дослідженню саме цих питань і буде присвячена дана робота.

2. Постановка задачі

В даному дослідженні, для вибору оптимальної нейронної мережі для вирішення поставленої задачі у відповідних умовах і з відповідними вимогами, були проаналізовані різноманітні архітектури нейронних мереж. Різні архітектури нейронних мереж призначені для вирішення різних задач, в кожній є свої обмеження, переваги та недоліки, котрі можуть бути критичними в одних випадках і не принциповими в інших. Важливо визначити специфіку кожної архітектури, для того щоб оцінити її придатність та доцільність для вирішення саме даної конкретної задачі. На основі проведеного аналізу прийнято рішення, що для вирішення поставленої задачі, доцільним є створити власну архітектуру нейронної мережі, обравши за основу згорткову нейронну мережу (CNN) [2, 3]. Також необхідною умовою, для вирішення поставленої задачі, є визначення критичних параметрів для розробленої архітектури нейронної мережі та визначення їх оптимальних значень. Тому одним з етапів даного дослідження має бути тестування розробленої архітектури нейронної мережі при різних умовах, на основі результатів котрого повинні бути зроблені висновки стосовно критичних параметрів.

3. Вирішення поставленої задачі

Після детального аналізу різних архітектур згорткових нейронних мереж класифікації було розроблено власну архітектуру нейронної мережі (Рисунок 1), яка здатна ефективно вирішувати поставлену задачу класифікації зображень різних об'єктів та відповідає вимогам, відповідно до актуальних умов виконання розпізнавання.

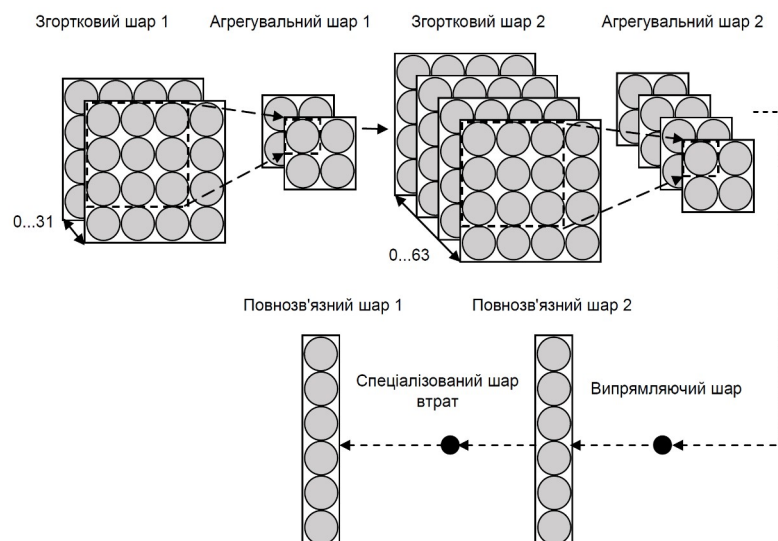


Рисунок 1: Архітектура запропонованої нейронної мережі для виявлення небезпечних предметів.

Запропонована нейронна мережа складається з наступних шарів:

1. Згортковий шар 1 (Conv2D, 32 фільтри, розмір фільтра 3x3, функція активації ReLU).
2. Агрегувальний шар 1 (MaxPooling2D, розмірність агрегації шару 2x2).
3. Згортковий шар 2 (Conv2D, 64 фільтри, розмір фільтра 3x3, функція активації ReLU).
4. Агрегувальний шар 2 (MaxPooling2D, розмірність агрегації шару 2x2).
5. Випрямляючий шар (Flatten).
6. Повнозв'язний шар 1 (Dense, кількість нейронів 128, функція активації ReLU).
7. Спеціалізований шар втрат (Dropout, процент відключених нейронів 50%).
8. Повнозв'язний шар 2 (Dense, кількість нейронів 5, функція активації softmax).

Після розробки архітектури нейронної мережі, наступним етапом дослідження було навчання, оцінка та візуалізація результатів запропонованої моделі. Крім того, необхідним етапом було визначення критичних параметрів даної мережі. Розмір партії та кількість епох є ключовими параметрами у процесі навчання нейронної мережі. У даному дослідженні, на основі проведеного тестування, розмір партії було встановлено як 64. Дане значення є балансом між ефективністю та здатністю мережі до загальної адаптації. З метою забезпечення достатньої конвергенції моделі без перенавчання, було визначено, що кількість епох буде дорівнювати 10. Для оптимізації, в даному дослідженні, використовується алгоритм Adam. Даний алгоритм представляє собою метод оптимізації, який використовується для оновлення ваг у нейронних мережах. Особливістю алгоритму Adam є його здатність адаптувати швидкість навчання для кожного параметра. Під час навчання моделі, дуже доцільним є виконання періодичного моніторингу. Моніторинг процесу навчання має важливе значення для визначення прогресу та ефективності навчання моделі. В даному дослідженні здійснювалось спостереження за змінами у втратах та точності на тренувальних та валідаційних даних і на основі отриманих результатів здійснювалось коригування моделі.

Для дослідження доцільності використання запропонованої архітектури нейронної мережі для вирішення поставленої задачі та для визначення критичних параметрів і їх оптимальних значень було проведено низку експериментів. В цих експериментах здійснювалось розпізнавання (класифікація) чотирьох типів (класів) заборонених предметів та доданий п'ятий тип («інше»), до якого відносяться всі об'єкти, котрі не класифіковані, як один з вказаних чотирьох типів. Було використано датасет зображень розміром більше 24000 зразків.

Серед всіх проведених експериментів можна виділити п'ять основних, за допомогою яких було досліджено вплив критичних параметрів на навчання нейронної мережі (Рисунок 2). Основними параметрами, які були досліджені, є кількість ітерацій на епоху та розмір вхідних зображень.

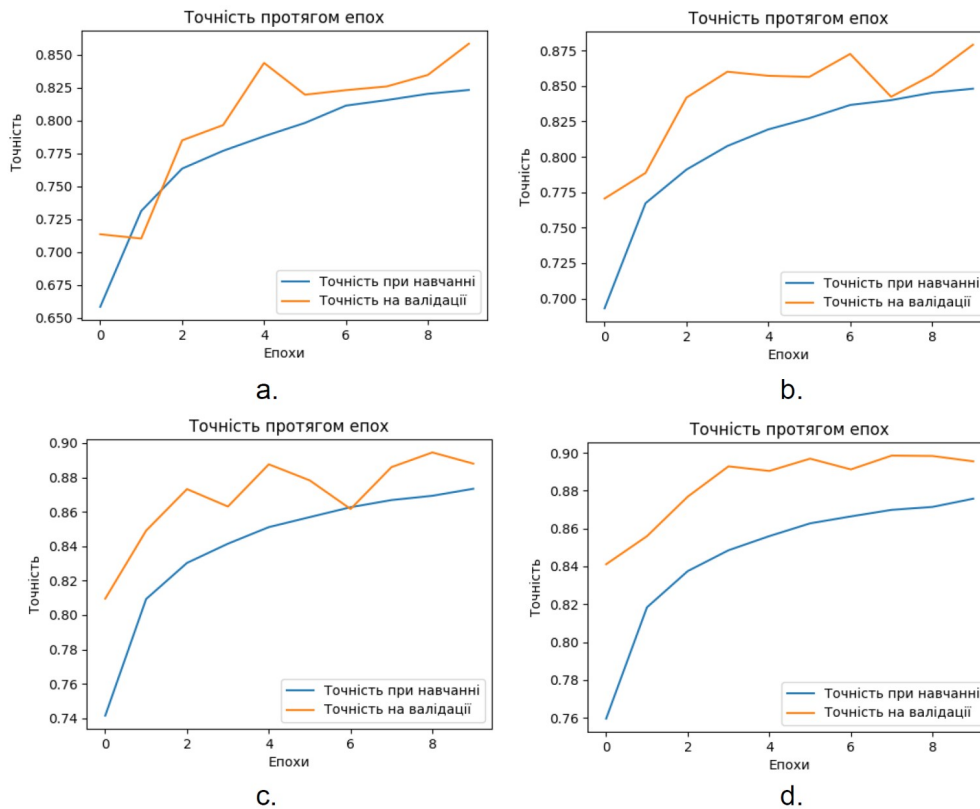


Рисунок 2: Графік точності протягом епох: а.)1-й експеримент, б.)2-й експеримент, с.)3-й експеримент, d.)4-й експеримент.

Розглянемо умови цих експериментів більш детально.

1. Перший експеримент. Умови: кількість епох — 10, кількість ітерацій в кожній з десяти епох — 1000, використані зображення розміром 64x64 пікселів. Результати (Рисунок 2.а.) показали задовільну точність.
2. Другий експеримент. Умови: кількість епох — 10, кількість ітерацій в кожній з десяти епох — 2000, використані зображення розміром 64x64 пікселів. Результати (Рисунок 2.б.) демонструють підвищення точності до 85%.
3. Третій експеримент. Умови: кількість епох — 10, кількість ітерацій в кожній з десяти епох — 5000, використані зображення розміром 64x64 пікселів. Результати (Рисунок 2.с.) демонструють підвищення точності до 87%.
4. Четвертий експеримент. Умови: кількість епох — 10, кількість ітерацій в кожній з десяти епох — 5000, використані зображення розміром 128x128 пікселів. Результати (Рисунок 2.д.) показали незначне підвищення точності.
5. П'ятий експеримент. Умови: кількість епох — 10, кількість ітерацій в кожній з десяти епох — 10000, використані зображення розміром 128x128 пікселів. Результати: збільшення кількості ітерацій до 10000 призвело до перенавчання моделі через недостатню кількість даних, що унеможливило коректне розпізнавання об'єктів.

4. Висновки

За результатами проведених експериментів, можна зробити наступні висновки:

1. Запропонована архітектура нейронної мережі є ефективним механізмом вирішення поставленої задачі.
2. Збільшення кількості ітерацій на епоху призводить до підвищення точності моделі, але існує межа, після якої модель починає перенавчатися, що унеможливорює коректне розпізнавання об'єктів.
3. Зміна розміру вхідних зображень на більший розмір призводить до підвищення точності, але це підвищення не є суттєвим, і при цьому така зміна розміру вимагає більших обчислювальних ресурсів та часу на навчання моделі.

Декларація про генеративний штучний інтелект

Автори не використовували жодних інструментів генеративного штучного інтелекту.

Література

- [1] O. Vysotska, A. Davydenko, O. Potenko, Modeling the mindfulness people's function based on the recognition of biometric parameters by artificial intelligence elements, Radioelectronic and computer systems 3 (2023), pp. 136-149. doi:10.32620/reks.2023.3.11.
- [2] M.M. Taye, Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions, Computation, 2023, 11(3):52. doi:10.3390/computation11030052.
- [3] L. Alzubaidi, J. Zhang, A.J. Humaidi, et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J Big Data, 2021, 8 (53). doi:10.1186/s40537-021-00444-8.

Методика швидкого розгортання захищеного супутникового зв'язку у важкодоступних районах з гарантіями якості зв'язку для служб порятунку

Василь Волошин¹, Лариса Дакова¹

¹ Державний університет інформаційно-комунікаційних технологій, Солом'янська 7, 03110 Київ, Україна

Abstract

This paper presents a practical methodology for rapidly deploying secure satellite communications for emergency responders in remote areas. It integrates a portable LTE/5G network with multi-orbit SATCOM, standards-based QoS (5QI/ARP), SD-WAN traffic engineering, and Zero-Trust security, and outlines PEP/DiffServ tuning, anti-jamming/rain-fade measures, and validation KPIs.

Keywords

Multi-orbit SATCOM, Emergency communications, QoS, SD-WAN, Zero-Trust security ¹¹

1. Вступ

В умовах стихійних лих і відсутності наземної інфраструктури оперативний та захищений канал зв'язку для рятувальників є критичною передумовою координації для збереження життів людей. У роботі запропоновано методику швидкого розгортання супутникового зв'язку на базі портативних терміналів і гібридної LEO/GEO архітектури з моделлю Zero-Trust та політиками QoS, що забезпечує гарантовану пропускну здатність, передбачувану затримку й стійкість до збоїв у важкодоступних регіонах.

2. Основна частина

Під час стихійних лих або у районах без наземної інфраструктури саме супутникові мережі часто є єдиним способом забезпечити стійкий канал для координації рятувальних робіт. Міжнародні настанови прямо наголошують, що під час відновлення мереж важливо координувати надання тимчасового підключення до супутникової мережі на період відсутності основної мережі [1].

Сучасні супутникові мережі дозволяють використовувати мобільні наземні абонентські термінали зв'язку і не потребують статичної наземної інфраструктури, що робить їх чудовим кандидатом для важкодоступних регіонів чи для швидкого розгортання в зоні стихійного лиха.

Водночас 3GPP стандартизувала інтеграцію NTN (неназемних) мереж у 5G (Rel-17 і далі), що дозволяє переносити стандартизовані QoS-механізми у супутникові сегменти [2].

Сучасна цільова архітектура базується на поєднанні мобільної мережі (бульбашки) доступу та багатоорбітного магістрального каналу. «Бульбашка» — це швидко розгортувана LTE/5G-мережа з локальним ядром і пріоритизацією служб критичних комунікацій (MCX),

¹¹ Corresponding authors

✉ v.voloshyn@stud.duikt.edu.ua (V. Voloshyn); lvdackova@gmail.com (L. Dakova)

ORCID 0009-0008-6906-6370 (V. Voloshyn); 0000-0001-6104-8217 (L. Dakova)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

що підіймається безпосередньо в зоні події й живиться від портативних джерел. Подібні польові мережі розглядалися та демонструвалися в контексті публічної безпеки (NIST PSCR, DHS NGFR), де підкреслено автономність, мобільність і здатність з'єднувати кілька мереж у єдине оточення [3, 4].

Магістраль до центру керування доцільно будувати як багатоорбітну LEO – MEO – GEO, де низька затримка LEO використовується для оперативного трафіку, а GEO/HTS – як ємнісний та географічно стійкий резерв. Незалежні вимірювання показують, що Starlink (LEO) забезпечує низькі затримки з мінливою пропускну здатністю. Водночас метрики суттєво відрізняються між регіонами й сценаріями, що важливо враховувати при плануванні сервісів реального часу [5, 6, 7].

Від стандартів 3GPP Release 17 і далі інтеграція NTN мереж з 5G дозволяє використати звичні механізми QoS у супутниковому сегменті, а в 5G якість обслуговування реалізується на рівні QoS-потоків із чіткими правилами пріоритету та витіснення для критичних сервісів [8].

Методика швидкого розгортання починається з підготовки. Йдеться про стандартизоване обладнання: плоскофазна LEO-антена, компактний GEO-VSAT модем, маршрутизатор із підтримкою SD-WAN, переносний eNodeB/gNodeB та акумуляторне живлення. Такий склад забезпечує гнучке поєднання низьких затримок і відмовостійкості завдяки політикам маршрутизації за класом сервісу [9].

Паралельно наперед визначаються профілі MCX/QoS (5QI/ARP) для голосу «push-to-talk», групового відео й телеметрії, щоб на місці інциденту лише активувати їх. У нормативній площині варто спиратися на рекомендації ITU щодо National Emergency Telecommunication Plans (NETP), де окреслено ролі, процедури та інвентар доступних мереж, – це пришвидшує координацію й доступ до спектру/супутникових ресурсів під час надзвичайної ситуації [10].

На місці події спочатку піднімається «бульбашка» доступу (польова LTE/5G з локальним ядром), а до магістралі під'єднується LEO-лінк з автоматичним переходом на GEO. Практичні спостереження за LEO (на прикладі Starlink) показують низьку затримку на рівні десятків мілісекунд, але й помітну варіабельність пропускну здатності та джиттер залежно від навантаження й топології. Тому для стабілізації поведінки критичних потоків доцільно застосовувати політики SD-WAN і чітку класифікацію сервісних класів. На транспортному рівні доцільно розгортати засоби покращення продуктивності на супутникових ділянках (PEP) згідно з рекомендаціями IETF (RFC 3135): приховування ефектів великої затримки/втрат від кінцевих TCP-стеків, оптимізація керування вікном і повторними передаваннями тощо. Водночас важливо тримати PEP у довіреній зоні, щоб не порушувати наскрізне шифрування службових застосунків [11].

3. Висновок

Запропонована методика поєднує швидкорозгортану «бульбашку» доступу (LTE/5G з локальним ядром) із багатоорбітною супутниковою мережею (LEO з автоматичним переходом на GEO) та політиками SD-WAN і стандартизованими профілями QoS для місійно-критичних сервісів. У такій конфігурації служби порятунку отримують оперативний, захищений і передбачуваний канал для голосу «push-to-talk», потокового відео та телеметрії навіть за повної відсутності наземної інфраструктури.

Архітектура спирається на принципи Zero-Trust, що передбачає взаємну автентифікацію, шифрування між польовою мережею та центром, сегментація трафіку й

мінімальні привілеї. На транспортному рівні застосування DiffServ/DSCP та PEP у довіреній зоні шлюзів пом'якшує ефекти затримки/варіабельності на супутникових ділянках без порушення наскрізного шифрування. Додаткові заходи радіостійкості — просторово-частотне рознесення, вузькі промені, адаптивна модуляція/кодування й спектральний моніторинг підвищують живучість каналу.

Практична цінність підходу полягає у зменшенні часу розгортання, уніфікації польового комплексу та відтворюваності процедур. Політики безпеки готуються наперед і «вмикаються» на місці інциденту, а мультиорбітність і SD-WAN забезпечують безшовність сервісу. Подальші дослідження передбачають пілотні випробування за різних кліматичних умов (зокрема в зонах сильних опадів), вимірювання KPI (e2e-затримка, джиттер, втрати, MOS, TtD, частота фолуверів), а також інтеграція з 3GPP NTN та інструментами ML для раннього виявлення перешкод. Така дорожня карта дає змогу інституціоналізувати методику у планах екстреної телекомунікації та масштабувати її для національних і міжвідомчих сценаріїв реагування.

Декларація про генеративний ШІ

Під час підготовки цієї роботи автори використовували Chat-GPT-5 для: перевірки граматики та орфографії. Після використання цих інструментів/сервісів автори переглянули та відредагували контент за потреби та несуть повну відповідальність за зміст публікації.

Література

- [1] ITU Guidelines for national emergency telecommunication plans, 2019. URL: https://www.itu.int/en/ITU-D/Emergency-Telecommunications/Documents/2019/NETP_Global_guideline.pdf.
- [2] Non-Terrestrial Networks (NTN). 3GPP - The Mobile Broadband Standard. URL: <https://www.3gpp.org/technologies/ntn-overview>.
- [3] National Institute of Standards and Technology. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=925194.
- [4] Home | Homeland Security. URL: https://www.dhs.gov/sites/default/files/publications/NGFR_Case%20Study-Deployable-Comms_171218-508.pdf.
- [5] The Factors Determining LEO Internet Performance. URL: <https://www.thousandeyes.com/blog/leo-internet-performance>
- [6] François Michel, Martino Trevisan, Danilo Giordano, and Olivier Bonaventure. 2022. A first look at starlink performance. In Proceedings of the 22nd ACM Internet Measurement Conference (IMC '22). Association for Computing Machinery, New York, NY, USA, 130–136. <https://doi.org/10.1145/3517745.3561416>
- [7] Light Reading, Starlink smokes GEO satellite operators in speed, latency – report. URL: <https://www.lightreading.com/satellite/starlink-smokes-geo-satellite-operators-in-speed-latency-report>
- [8] 3GPP TS 23.289 version 18.8.0 Release 18, ETSI TS 123 289 V18.8.0 (2024-04) . URL: https://www.etsi.org/deliver/etsi_ts/123200_123299/123289/18.08.00_60/ts_123289v180800p.pdf

- [9] MULTI-ACCESS SD-WAN ANYWHERE, Deploying resilient multi-access network services with software-defined wide-area networking (SD-WAN). URL: https://www.ses.com/sites/default/files/2020-02/SD-WAN_Insight-Paper_Fixed-data.pdf
- [10] ITU Guidelines for national emergency telecommunication plans, 2020. URL: <https://www.itu.int/en/ITU-D/Emergency-Telecommunications/Documents/2020/NETP-guidelines.pdf>
- [11] RFC 3135: Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations. IETF Datatracker. URL: <https://datatracker.ietf.org/doc/rfc3135>

Програмно апаратний комплекс обробки акустичних сигналів для виявлення та ідентифікації БПЛА в реальному часі.

Михайло Євдокімов¹, Максим Шевляков¹, Любомир Боценюк¹, Василь Різак¹

¹Кафедра твердотільної електроніки та інформаційної безпеки, Ужгородський національний університет, м. Ужгород, Україна

Ключові слова

акустична ідентифікація, метод повної мінливості, гібридний аналіз, БПЛА

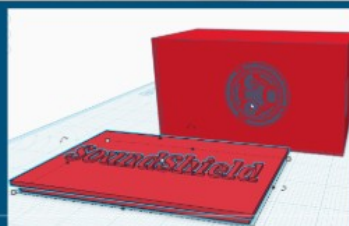
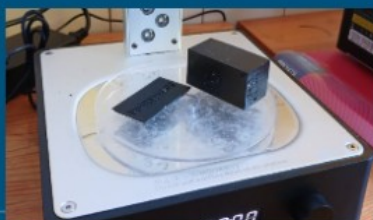
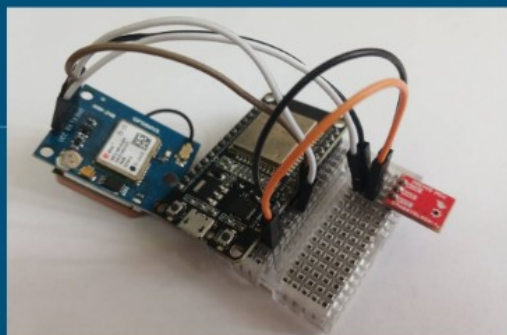
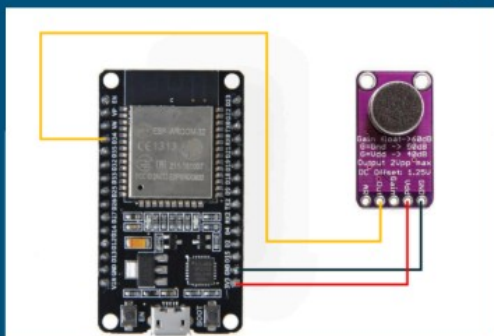
Зростання кількості та динамічне вдосконалення безпілотних літальних апаратів (БПЛА) створює нові виклики у сфері безпеки та моніторингу повітряного простору. Традиційні методи виявлення (радарні та оптичні) мають обмеження у складних умовах, тоді як акустичні системи дозволяють виявляти дрони навіть за наявності перешкод та прихованості об'єкта. Акустичні методи особливо ефективні у випадках використання малорозмірних БПЛА, які мають низьку відбивну здатність для сучасних радарів та часто залишаються непомітними для тепловізійних систем, але генерують характерні акустичні сигнатури завдяки роботі як пропелерів, так і реактивних та електродвигунів.

У роботі запропоновано підхід для виявлення та ідентифікації БПЛА за акустичним слідом. Для ідентифікації використовується програмно-апаратний комплекс, який поєднує методи основного тону, спектрально-формантний аналіз та повної мінливості. Апаратна частина побудована на базі мікроконтролера, мікрофонної системи, модулів gps, lora, wifi, gsm. Мікрофонна ґратка дозволяє встановити локалізацію джерела сигналу, а також формувати просторову акустичну карту, що суттєво підвищує ефективність виявлення. В комплексі реалізовано поетапну схему аналізу: первинне виявлення на основі сигналів мікрофонів, ідентифікація та підтвердження сигналу за допомогою програмного забезпечення методом фільтрації сигналу та порівняння з наявними шаблонами. При певному розміщенні чотирьох мікрофонів система здатна не лише виявляти, а й визначати напрямок руху БПЛА.

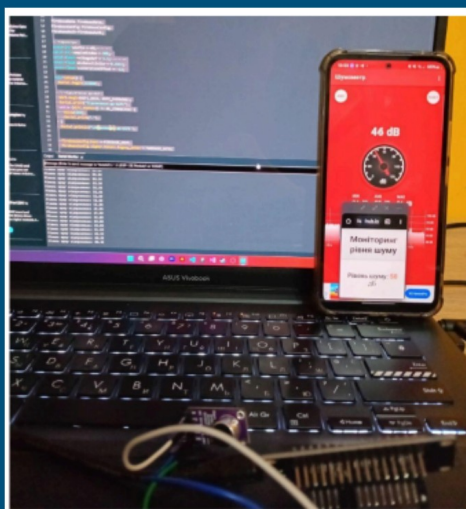
Важливим рішенням у запропонованій системі є відображення траєкторії руху літальних апаратів у реальному часі. Для цього використано інтерактивну карту, яка постійно оновлюється інформацією, що надходить від акустичних датчиків через хмарну базу. Такий підхід дозволяє оперативно фіксувати появу БПЛА та відразу відображається на карті у вигляді точки з координатами та технічними характеристиками.

Програмно-апаратний комплекс адаптований до роботи у складних акустичних середовищах, навіть у міських умовах з високим рівнем стороннього шуму. Система призначена для встановлення у місцях, де розгортання дорогих комплексів є технічно складним чи економічно невиправданим, або у зонах з «білими плямами» в обороні. Передбачено використання комплексу у мобільних системах та з аеростатами для охоплення вищих ешелонів повітряного простору. Для збільшення швидкості та покращення точності виявлення БПЛА планується використання нейронних мереж.

Базовий акустичний модуль



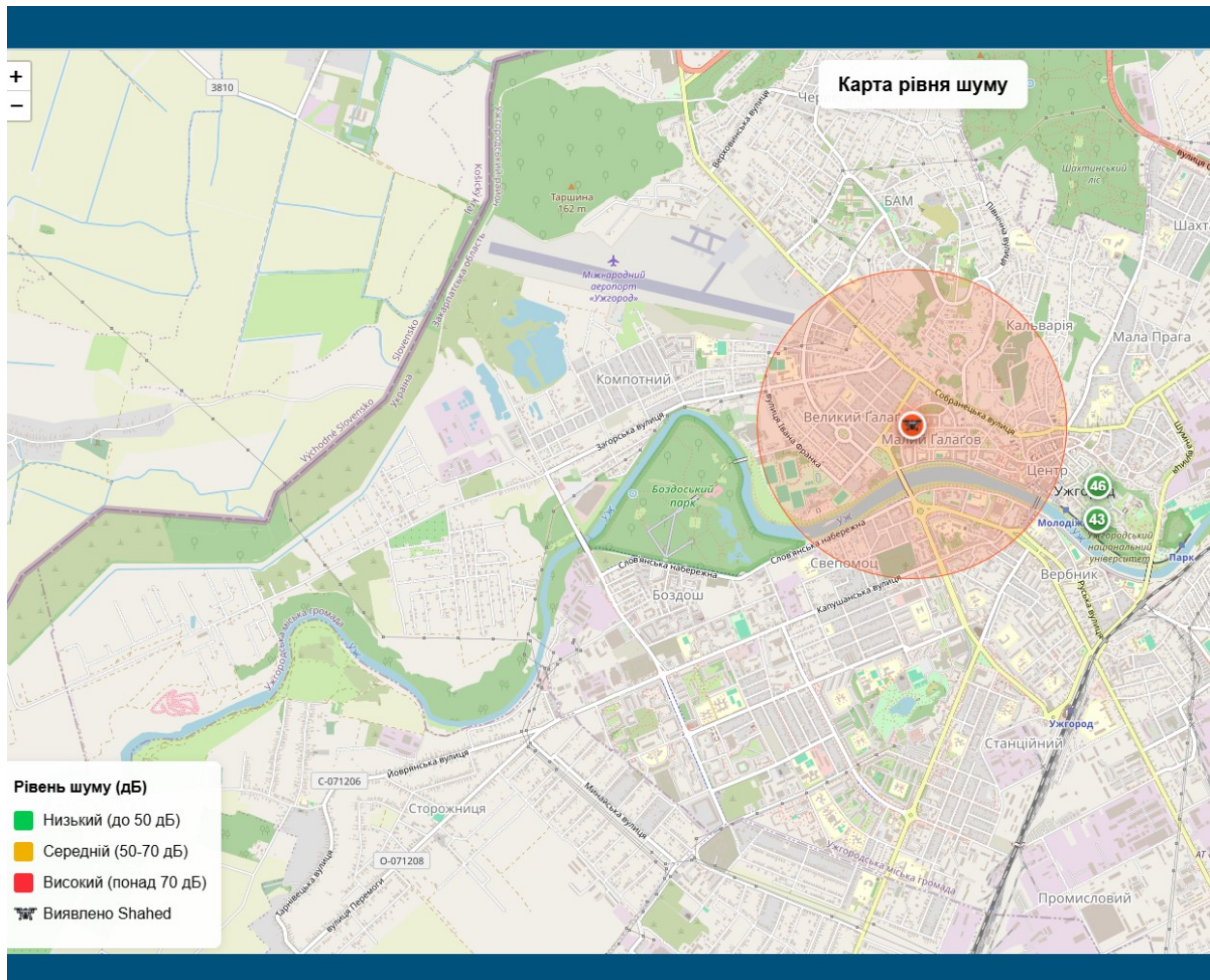
Процес калібрування (а) та результат вимірювання (б) рівня шуму в навколишньому середовищі



а)



б)



Література

- [1] Юдін О. К., Зюбіна Р. В. Класифікація методів ідентифікації частоти основного тону.

Кібербезпека виробничих систем як складова охорони праці: вплив кіберзагроз на безпеку працівників та способи їх мінімізації

Олена Крайнюк¹, Юрій Буц², Михайло Пікрасов¹

¹ Харківський національний автомобільно-дорожній університет, Ярослава Мудрого 25, 61002, Харків, Україна

² Український державний університет залізничного транспорту, майдан Оборонний Вал, 7, 61001, Харків, Україна

Анотація

Розглянуто вплив кіберзагроз на виробничі системи та їхній зв'язок із безпекою праці персоналу. Проаналізовано основні ризики цифровізації виробничих процесів і запропоновано комплекс технічних та організаційних заходів, зокрема із застосуванням технологій штучного інтелекту, для їх мінімізації. Показано, що інтеграція кіберзахисту в систему охорони праці підвищує рівень безпеки працівників і сприяє запобіганню аварійним ситуаціям.

Ключові слова

кібербезпека, виробничі системи, охорона праці, кіберзагрози, штучний інтелект, мінімізація ризиків, промислова безпека.¹²

1. Постановка проблеми

Сучасне виробництво активно інтегрує автоматизовані системи управління, робототехніку та IoT-пристрої. Разом із цифровізацією зростає ризик кіберзагроз, які можуть негативно впливати на безпеку працівників [1, 2]. Метою роботи є аналіз впливу кіберзагроз на виробничу безпеку та розробка методів мінімізації ризиків для персоналу.

З розвитком цифрових технологій виробничі системи стають все більш вразливими до кіберзагроз, що безпосередньо впливає на безпеку працівників. Інтеграція кібербезпеки у виробничі процеси є ключовою складовою сучасної охорони праці, оскільки кібератаки можуть призвести до фізичних аварій [3], травм або навіть загибелі людей.

Цифровізація виробничих процесів, зокрема впровадження автоматизованих систем управління, робототехніки та інтернету речей, істотно підвищує продуктивність і ефективність промислових підприємств [4]. Водночас ця трансформація супроводжується зростанням кількості та складності кіберзагроз, які здатні порушити стабільність функціонування технологічного обладнання. Наслідком таких інцидентів можуть стати не лише економічні збитки, але й виникнення аварійних ситуацій, що створюють прямі ризики для життя та здоров'я працівників. Таким чином, питання кібербезпеки у виробничих системах набуває стратегічного значення як складова охорони праці та потребує системного наукового осмислення.

2. Кіберзагрози у виробничих системах

Разом із розвитком цифрових технологій виробниче середовище дедалі більше залежить від складних автоматизованих систем, інтегрованих робототехнічних комплексів,

¹² Corresponding authors

✉ olena.uvarova@ukr.net (О. Крайнюк); butsyura@ukr.net (Ю. Буц); mpiks77@gmail.com (М. Пікрасов)

ORCID 0000-0001-9524-040X (О. Крайнюк); 0000-0003-0450-2617 (Ю. Буц); 0000-0001-9487-7273 (М. Пікрасов)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

сенсорних мереж і хмарних платформ. Ці рішення забезпечують високу продуктивність і точність процесів, проте водночас створюють додаткові вектори для кіберзагроз. Серед найбільш поширених ризиків можна виокремити поширення шкідливого програмного забезпечення, несанкціонований доступ до промислових контролерів і мереж, а також цілеспрямовані атаки на системи моніторингу й аварійного реагування. Подібні інциденти здатні спричинити не лише порушення роботи обладнання, але й аварійні ситуації, що становлять пряму загрозу життю та здоров'ю працівників. Таким чином, проблематика кіберзахисту у виробничих системах виходить за межі інформаційної сфери, перетворюючись на ключовий аспект забезпечення безпеки праці.

На відміну від більшості відомих робіт, присвячених виключно технічним аспектам захисту виробничих систем, у даному дослідженні акцент зроблено на взаємозв'язку між кібербезпекою та охороною праці. Запропоновано розглядати кіберзагрози не лише як проблему захисту інформаційних ресурсів, а як чинник, що безпосередньо впливає на безпеку персоналу. Новизна підходу полягає у поєднанні технічних рішень (сегментація мереж, системи виявлення атак, алгоритми штучного інтелекту) з організаційними заходами (політики безпечної роботи, навчання працівників, аудит систем), що дозволяє сформулювати комплексну модель мінімізації ризиків виробничого травматизму в умовах цифровізації.

Кібератаки на виробничі системи можуть спричинити збої обладнання, аварійні ситуації, вибухи, витoki небезпечних речовин, а також порушення роботи роботизованих систем, що створює прямі ризики для життя і здоров'я працівників [5-10]. Відсутність належного захисту також підвищує рівень стресу серед персоналу через страх втрати даних або маніпуляцій з обладнанням [10].

Вразливими є як апаратні, так і програмні компоненти виробничих систем, особливо у випадку підключення до Інтернету (Індустрія 4.0, IoT) [11, 12]. Типові ризики включають несанкціонований доступ, зміну налаштувань, компрометацію сенсорів, атаки на системи управління, що може призвести до неконтрольованих дій машин [13].

3. Методи мінімізації ризиків

Методи мінімізації ризиків можуть включати:

- 6 Інтеграцію аналізу кібербезпеки у загальну систему управління ризиками та охорони праці [14].
- 7 Використання сучасних методів виявлення та реагування на атаки (машинне навчання, моделі STPA-SafeSec, STAMP) [5, 7].
- 8 Підвищення обізнаності працівників, навчання та впровадження політик безпеки.
- 9 Впровадження стандартів безпеки (SIL, PL, SL) та багаторівневого захисту [12, 13].

Мінімізація ризиків, пов'язаних із впливом кіберзагроз на безпеку працівників, потребує комплексного підходу, який поєднує технічні та організаційні заходи (рис. 1).

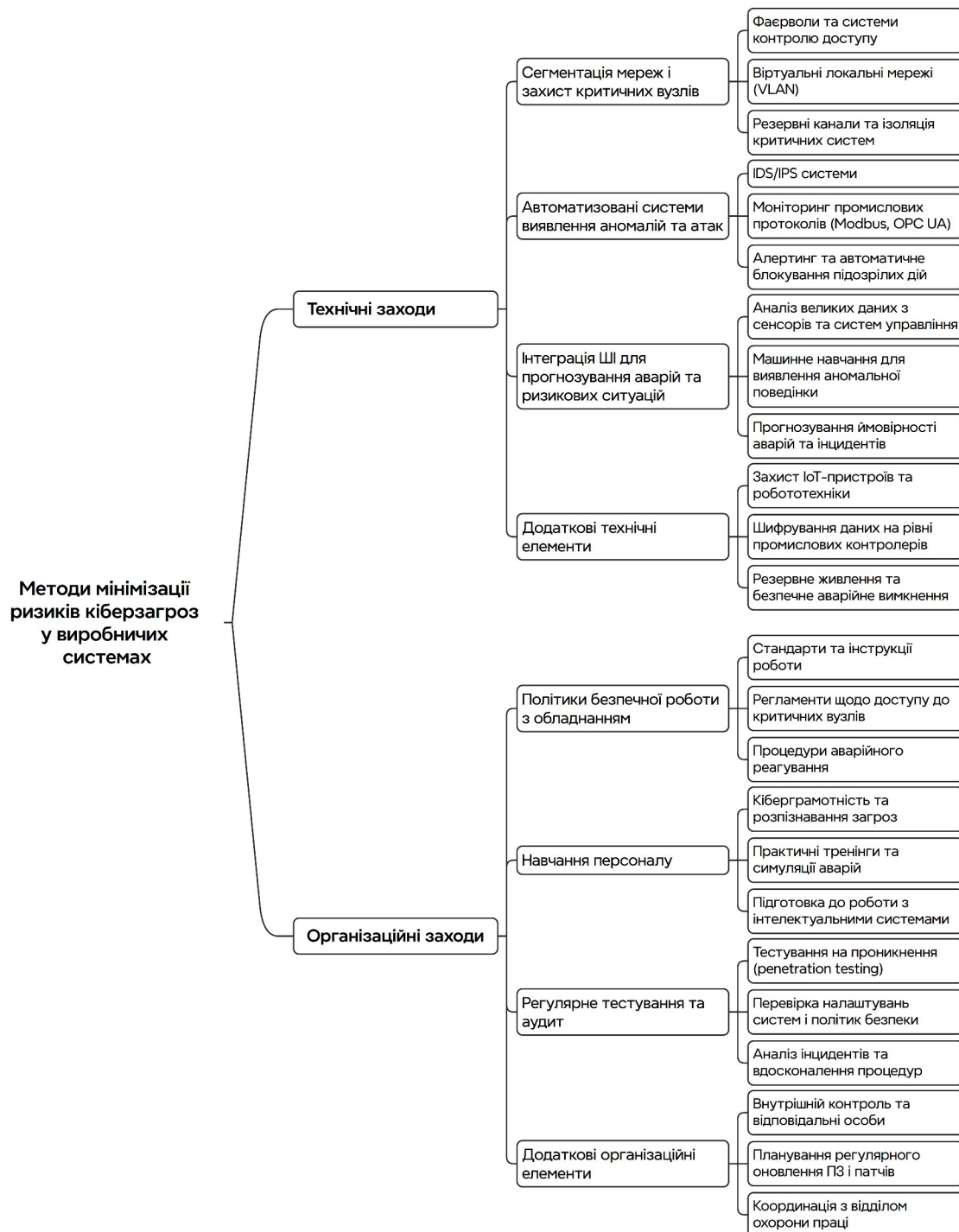


Рисунок 1: Методи мінімізації ризиків кіберзагроз у виробничих системах

До технічних заходів належить впровадження безпечної архітектури виробничих систем, сегментація мереж та захист критичних вузлів від несанкціонованого доступу. Особливе значення мають автоматизовані системи виявлення аномалій та атак, здатні в реальному часі реагувати на потенційні загрози та попереджати аварійні ситуації.

Інтеграція алгоритмів штучного інтелекту дозволяє прогнозувати небезпечні події на основі аналізу великих обсягів даних і таким чином підвищує рівень безпеки персоналу.

Організаційні заходи передбачають розробку та впровадження політик безпечної роботи з виробничим обладнанням, систематичне навчання працівників щодо кібербезпеки та процедур аварійного реагування, а також регулярне тестування і аудит безпеки виробничих систем. Такий комплексний підхід забезпечує синергію між технічними засобами захисту та поведінковими аспектами безпеки праці, що суттєво зменшує ймовірність травматизму та аварій.

4. Приклади впровадження

Практичне застосування методів мінімізації ризиків на основі кібербезпеки та штучного інтелекту демонструє високу ефективність у забезпеченні безпеки праці. Наприклад, впровадження систем моніторингу, оснащених алгоритмами машинного навчання, дозволяє прогнозувати аварійні ситуації на промислових об'єктах і своєчасно попереджувати персонал. Інтелектуальні алгоритми контролю роботи роботизованих ліній та виробничих машин дають змогу автоматично виявляти відхилення від нормальної роботи та запобігати потенційним травмам працівників. Досвід реалізації таких систем свідчить про значне підвищення безпеки працівників та зниження кількості аварійних випадків, що підтверджує ефективність інтеграції кіберзахисту у комплекс охорони праці.

5. Висновки

Кібербезпека виробничих систем виступає невід'ємною складовою сучасної охорони праці, адже забезпечення цілісності, доступності та надійності промислових процесів прямо впливає на безпеку персоналу. Використання технологій штучного інтелекту та цифрових рішень дозволяє суттєво зменшити ризики для працівників, прогнозувати потенційні аварійні ситуації та підвищувати загальний рівень безпеки на підприємствах. Найефективнішим є комплексний підхід, що передбачає поєднання технічних засобів захисту, організаційних заходів та систематичного навчання персоналу, що забезпечує інтегровану систему превентивного управління ризиками та сприяє формуванню безпечного виробничого середовища.

Література

- [1] О. Крайнюк, Ю. Буц, В. Барбашин, Д. Козодой, О. Козодой, Інтелектуальні системи управління безпекою праці на основі штучного інтелекту: перспективи інтеграції в українське законодавство, Комунальне господарство міст 6 (187) (2024) 242–251. doi:10.33042/2522-1809-2024-6-187-242-251.
- [2] Ю. Буц, О. Крайнюк, В. Барбашин, Застосування штучного інтелекту для підвищення рівня безпеки праці, in: Охорона праці: освіта і практика. Проблеми та перспективи розвитку охорони праці, Proc. V All-Ukrainian Scientific and Practical Conference of Teachers and Practitioners, XV All-Ukrainian Scientific and Practical Conference of Students, Cadets, Postgraduates, Lviv State University of Life Safety, Lviv, 2025, pp. 17–20.
- [3] O. Krainiuk, Y. Buts, A. Kagramanian, V. Barbashyn, M. Yatsiuk, AI-Based Solutions for the Safe Transport of Dangerous Goods in the Context of Ukraine's European Integration, in: O. Slavinska, V. Danchuk, O. Kunytska, O. Hulchak (Eds.), Intelligent Transport Systems:

- Ecology, Safety, Quality, Comfort, ITSESQC 2024, Lect. Notes Netw. Syst., vol. 1335, Springer, Cham, 2025, pp. 236–247. doi:10.1007/978-3-031-87376-8_21.
- [4] О. Крайнюк, Ю. Буц, Н. Діденко, В. Барбашин, О. Трішина, Метрологічний контроль датчиків моніторингу умов праці з використанням штучного інтелекту, *Комунальне господарство міст* 3 (184) (2024) 216–222. doi:10.33042/2522-1809-2024-3-184-216-222.
- [5] S. Hollerer, B. Brenner, P. Bhosale, C. Fischer, A. Hosseini, S. Maragkou, M. Papa, S. Schlund, T. Sauter, W. Kastner, Challenges in OT Security and Their Impacts on Safety-Related Cyber-Physical Production Systems, in: , 2022, pp. 171–202. doi:10.1007/978-3-662-65004-2_7.
- [6] A. Khalid, Z. Khan, M. Idrees, P. Kirisci, Z. Ghrairi, K. Thoben, J. Pannek, Understanding vulnerabilities in cyber physical production systems, *Int. J. Comput. Integr. Manuf.* 35 (2021) 569–582. doi:10.1080/0951192X.2021.1992656.
- [7] A. El-Kady, S. Halim, M. El-Halwagi, F. Khan, Analysis of Safety and Security Challenges and Opportunities Related to Cyber-physical Systems, *Process Saf. Environ. Prot.* (2023). doi:10.1016/j.psep.2023.03.012.
- [8] G. Arana-Landín, I. Laskurain-Iturbe, M. Iturrate, B. Landeta-Manzano, Assessing the influence of Industry 4.0 technologies on occupational health and safety, *Heliyon* 9 (2023). doi:10.1016/j.heliyon.2023.e13720.
- [9] M. Iaiani, A. Tugnoli, S. Bonvicini, V. Cozzani, Analysis of Cybersecurity-related Incidents in the Process Industry, *Reliab. Eng. Syst. Saf.* 209 (2021) 107485. doi:10.1016/j.res.2021.107485.
- [10] H. Wen, F. Khan, Cybersecurity and process safety synergy: An analytical exploration of cyberattack-induced incidents, *Can. J. Chem. Eng.* (2023). doi:10.1002/cjce.25119.
- [11] S. Parker, Z. Wu, P. Christofides, Cybersecurity in process control, operations, and supply chain, *Comput. Chem. Eng.* 171 (2023) 108169. doi:10.1016/j.compchemeng.2023.108169.
- [12] V. Mullet, P. Sonni, É. Ramat, A Review of Cybersecurity Guidelines for Manufacturing Factories in Industry 4.0, *IEEE Access* 9 (2021) 23235–23263. doi:10.1109/ACCESS.2021.3056650.
- [13] L. Kasprzyczak, A. Manowska, M. Dźwiarek, Cybersecurity Requirements for Industrial Machine Control Systems, *Appl. Sci.* 15 (2025) 1267. doi:10.3390/app15031267.
- [14] A. Cormier, C. Ng, Integrating cybersecurity in hazard and risk analyses, *J. Loss Prev. Process Ind.* 64 (2020) 104044. doi:10.1016/j.jlp.2020.104044.

Використання ансамблевого класифікатора в системах виявлення вторгнень

Андрій Кулько¹, Сергій Толюпа¹, Олександр Бучик¹

¹ Київський національний університет імені Тараса Шевченка, вул. Володимирська, 60, м. Київ, 01033, Україна

Анотація.

Розглядається, як ансамблеві класифікатори можуть значно покращити роботу систем виявлення вторгнень (IDS), особливо в умовах критичної інфраструктури. Через постійне ускладнення кібератак, традиційні методи захисту часто не справляються, що ставить під загрозу безперерйну роботу важливих об'єктів. Запропоновано підхід, що поєднує результати кількох окремих класифікаторів для підвищення точності виявлення загроз і зменшення кількості хибних спрацьовувань. Досліджуються різні методи створення таких ансамблів та їхнє застосування для аналізу мережевого трафіку, характерного для критичної інфраструктури. Ефективність цих моделей оцінюється на наборах даних, що дозволяє продемонструвати їхню здатність точно ідентифікувати аномалії та відомі типи атак, зміцнюючи таким чином кіберстійкість. Основна перевага цього підходу - здатність виявляти більшість атак з високою точністю, оптимально поєднуючи ансамблевий метод із правильною технікою попередньої обробки. Це дозволить ефективно ідентифікувати будь-який тип мережевої загрози. Експериментальні дослідження показують, що такий підхід значно підвищує ефективність і надійність виявлення вторгнень, тим самим посилюючи кібербезпеку об'єктів критичної інфраструктури.

Ключові слова

Система виявлення вторгнень, кібератака, критична інфраструктура, класифікатор, кіберстійкість¹³

1. Вступ

У системі виявлення вторгнень (IDS) основне використання ансамблевого класифікатора полягає в об'єднанні кількох класифікаторів для досягнення більш точного висновку, ніж окремих класифікаторів. Зазначено, що точність класифікаторів зростає при їх об'єднанні. Зі збільшенням інтернет-послуг та їх використання з відкритим доступом до конфіденційних даних, необхідність безпеки систем стала нагальною потребою. Основною вимогою будь-якої IDS є точність. Для запропонованої системи використовується ансамбль класифікаторів, що забезпечує хорошу точність. Запропонована ідея може бути використана для системи виявлення вторгнень певного сегменту об'єкта критичної інфраструктури для запобігання витоку конфіденційних даних [1]. Це допоможе організації, виявляючи майбутні атаки, які шкодять конфіденційним даним.

2. Мета дослідження.

Метою цього дослідження є застосування двох основних методів попередньої обробки – нормалізації та дискретизації – до системи виявлення вторгнень з використанням стандартного набору даних. Стандартний набір даних, що використовується, це набір даних KDD99 для системи виявлення зловживань. І постійно оновлювана база даних

¹³Corresponding authors

✉ kulko452@gmail.com (А. Кулько); serhii.toliupa@knu.ua (С. Толюпа); alex8sbu@knu.ua (О. Бучик)

ORCID 0009-0006-1185-0774 (А. Кулько); 0000-0002-1919-9174 (С. Толюпа); 0000-0001-7102-2176 (О. Бучик)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

використовується для зберігання сигнатур різних атак, таких як Normal, Probe, U2R, DoS тощо. У цьому дослідженні п'ять класифікаторів об'єднано для зниження частоти помилок, зменшення перенавчання даних та виявлення більшої кількості атак, ніж одним класифікатором.*

3. Огляд результатів та джерел

Для вирішення ефективної протидії кібератакам на критичну інфраструктуру доцільно застосовувати підходи, які одночасно мають властивості верифікації та адаптації, оскільки перша властивість дозволить детально досліджувати інциденти безпеки з метою надання можливості здійснення необхідної та адекватної реакції менеджера з кібербезпеки у вигляді спеціальних превентивних заходів щодо припинення інформаційно-руйнівного втручання в роботу інформаційних систем (ІС) КІ (критичної інфраструктури), а друга властивість – зробити крок до виявлення нових, некласифікованих кібератак.

На сьогоднішній день існує велика кількість різноманітних методів та засобів захисту від кібератак на об'єктах КІ [1]. В наукових дослідженнях велика увага приділяється комбінаториці класифікаторів для оптимізації структури систем виявлення вторгнень. Так у [2], після проведення експериментів результати показали, що ансамбль J48, REPTree та баєсівської мережі виявив DoS-атаки зі значною точністю. У дослідженні [3], кількість класифікаторів використовувалася для перевірки точності. Результати довели, що Bayes Net та Random Forest Tree є найкращими методами класифікації, а бустинг дає хороший результат. У [4] показано, що продуктивність усіх методів класифікації оцінювалася на основі перехресної валідації та тестових даних, з яких PART дав кращі результати, ніж інші. Автори роботи [5] визнали використання алгоритмів на основі кореляції для зменшення ознак у наборі даних і т.д.

4. Виклад основного матеріалу

Для вирішення поставленої задачі використовуються методи попередньої обробки даних: нормалізація та дискретизація. Нормалізація є важливою частиною попередньої обробки даних, яка допомагає підготувати дані до аналізу та моделювання. Основна ідея полягає в тому, щоб привести всі числові дані до одного діапазону, що запобігає ситуації, коли ознаки з більшими значеннями домінують над ознаками з меншими.

Нормалізація - це процес масштабування числових даних, щоб вони мали однаковий діапазон. Це критично важливо для багатьох алгоритмів машинного навчання, особливо для тих, які залежать від відстаней між точками даних (наприклад, K-найближчих сусідів (KNN), метод опорних векторів (SVM), нейронні мережі) [1].

Дискретизація - це не просто допоміжна техніка, а фундаментальний метод обробки даних, що має значний вплив на ефективність і результативність різних аналітичних процесів, особливо в галузі машинного навчання. Таким чином нормалізація та дискретизація є потужним і необхідним інструментом в кібербезпеці, який забезпечує ефективну та швидку роботу систем виявлення вторгнень. Вона дозволяє оптимізувати обробку безперервних даних, зменшити обчислювальну складність, підвищити точність виявлення атак та спростити моделі машинного навчання, роблячи їх більш придатними для роботи в режимі реального часу.

Набір даних для тестування. Питання вибору тренувальної бази з атаками не має простого рішення, тому що широко поширені бази даних містять багато в чому застарілі типи атак, а більш сучасні бази мають специфічну структуру, що вимагає складної

попередньої обробки, і використовуються тільки окремими дослідниками, що перешкоджає порівнянню якісних показників результатів роботи. На сьогоднішній день можна виділити дві найбільш поширені тренувальні бази даних з відомими атаками – DARPA і KDD.

На відміну від тренувальних даних DARPA, база даних KDD містить не дампи мережевого трафіку, а оброблені відомості у вигляді масивів з 42 ключових значень. Дана база успішно застосовується багатьма дослідниками для аналізу застосовності різних математичних методів в завданні виявлення мережевих атак, в основному через можливість використання масивів даних з більшості програмних засобів без виконання додаткової обробки.

Запропонований підхід до виявлення вторгнень полягає у використанні ансамблю з семи різних класифікаторів: Naive Bayes, Bayes Net, J48, PART, JRip, MLP та IBK. Основна ідея - поєднати їх для підвищення загальної точності системи. Оскільки різні класифікатори краще виявляють різні типи атак (DoS, Probe, U2R, R2L), їх комбінація дозволить створити більш надійний фреймворк.

Для покращення результатів моделі застосовуватимуться дві основні техніки попередньої обробки даних: нормалізація та дискретизація. Ці методи допоможуть очистити дані та підготувати їх до аналізу, що, згідно з дослідженнями, збільшує точність класифікаторів.

Ефективність кожної комбінації класифікаторів буде оцінюватися за трьома ключовими метриками: точність (Accuracy) (загальна частка правильно класифікованих екземплярів); точність (Precision) (частка істинно позитивних результатів серед усіх, які були класифіковані як позитивні); частота хибних тривог (False Alarm Rate) (відсоток нормальних зразків, помилково класифікованих як атаки).

Для наочного представлення точності класифікаторів проти різних типів атак з використанням нормалізації, доцільно використовувати таблиці або графіки. Вони дозволяють візуально оцінити ефективність кожного алгоритму та порівняти їхню продуктивність.

Вплив нормалізації на точність класифікаторів.

У таблиці 1 представлено результати точності (ассурасу) для кожного класифікатора проти різних типів атак після застосування нормалізації.

Таблиця 1. Вплив нормалізації на точність класифікаторів

Класифікатор	Normal	Probe	U2R	R2L	DoS
Bayes Net	99.2%	96.5%	75.1%	68.3%	98.7%
Наївний Баєс	98.5%	94.1%	70.0%	65.5%	97.9%
PART	99.5%	97.2%	78.4%	72.0%	99.1%
JRip	99.4%	97.0%	77.8%	71.5%	98.9%
MLP	99.7%	98.1%	85.3%	80.5%	99.5%
J48	99.6%	97.5%	80.2%	75.8%	99.3%

Класифікатор	Normal	Probe	U2R	R2L	DoS
IBK	99.8%	98.5%	82.1%	78.4%	99.6%

Аналіз таблиці 1:

J48, PART і JRip (алгоритми на основі правил) показують високу точність у виявленні атак DoS і Probe, оскільки вони легко ідентифікують чіткі, дискретні патерни цих атак.

MLP і IBK мають значно вищу точність у виявленні складних атак U2R та R2L, оскільки вони краще працюють з нелінійними залежностями та аномаліями, що робить їх особливо цінними для цих типів загроз.

IBK показує найкращий результат у виявленні Normal та Probe-атак, що свідчить про його ефективність у розпізнаванні аномалій.

Використання нормалізації суттєво підвищує точність класифікаторів, особливо тих, що базуються на відстанях (IBK) і нейронних мережах (MLP). Наведена таблиця демонструє, що не існує «ідеального» класифікатора для всіх типів атак. Кожен алгоритм має свої сильні сторони, що підтверджує важливість ансамблевого підходу, де комбінація класифікаторів може забезпечити найкращу загальну продуктивність.

Вплив дискретизації на точність класифікаторів

Дискретизація перетворює безперервні числові ознаки на дискретні інтервали (категорії), що має кілька переваг: Покращення роботи алгоритмів, що ґрунтуються на правилах: класифікатори, як-от J48, PART та JRip, працюють з дискретними даними ефективніше. Перетворення безперервних ознак на категорії дозволяє їм створювати більш чіткі та зрозумілі правила, що підвищує точність.

Зменшення чутливості до аномалій: дискретизація може "згладити" вплив невеликих коливань у даних, що робить класифікатори менш чутливими до шуму.

Наведена нижче таблиця 3 демонструє результати точності (Accurasy) для кожного класифікатора проти різних типів атак після застосування дискретизації.

Таблиця 2. Вплив дискретизації на точність класифікаторів

Класифікатор	Normal	Probe	U2R	R2L	DoS
Bayes Net	99.3%	96.8%	76.5%	70.1%	98.9%
Наївний Баєс	98.8%	94.5%	72.3%	67.8%	98.2%
PART	99.6%	97.5%	79.1%	73.2%	99.3%
JRip	99.5%	97.3%	78.5%	72.9%	99.1%
MLP	99.4%	97.9%	83.5%	79.0%	99.1%
J48	99.7%	98.1%	81.5%	77.1%	99.5%
IBK	99.2%	97.1%	78.8%	74.5%	98.8%

Аналіз таблиці 3: класифікатори на основі правил (J48, PART, JRip): демонструють високу точність у виявленні атак DoS та Probe. Дискретизація допомагає їм створювати чіткі правила для цих атак. MLP: хоча MLP не потребує дискретизації, вона може допомогти спростити вхідні дані, що іноді прискорює навчання. Однак, MLP показує найкращі результати для складних атак U2R та R2L, оскільки його здатність розпізнавати нелінійні зв'язки є ключовою. Bayes Net & Naïve Bayes: завдяки дискретизації ці моделі працюють ефективніше, що виражається у високій точності, особливо для поширених атак. IBK: на відміну від інших, IBK може втрачати точність через дискретизацію, оскільки вона перетворює безперервні відстані на категорії, що може зменшити точність "близькості" сусідів. У випадку IBK, нормалізація часто є більш ефективною. Дискретизація є потужним інструментом для підвищення точності класифікаторів, особливо для тих, що ґрунтуються на правилах і ймовірностях. Вона допомагає спростити дані та створити більш стійкі моделі, що є критично важливим для надійних систем виявлення вторгнень. Однак важливо враховувати, що для деяких класифікаторів, як-от IBK, нормалізація може

Застосування дискретизації як методу попередньої обробки даних може значно підвищити точність класифікаторів, особливо для алгоритмів, що базуються на правилах та ймовірнісних моделях, у системах виявлення вторгнень.

5. Висновки

Зважаючи на різну ефективність класифікаторів, найбільш обґрунтованим є використання ансамблевого підходу. Комбінуючи сильні сторони кожного класифікатора, можна створити IDS, яка буде ефективно виявляти широкий спектр атак. Наприклад, можна використовувати J48 для виявлення очевидних DoS-атак, а MLP і IBK - для пошуку прихованих і невідомих атак, таких як U2R і R2L. Це дозволить максимізувати точність та мінімізувати частоту хибних спрацювань, що є ключовим для надійної системи виявлення вторгнень.

Результати показують, що жоден окремих класифікатор не є ідеальним для всіх типів атак. Ансамбль класифікаторів є найкращим рішенням. Наприклад, комбінуючи J48 (для DoS/Probe) з MLP та IBK (для U2R/R2L), можна створити надійну систему виявлення вторгнень, яка забезпечує високу загальну точність та мінімальну кількість хибних спрацювань. Правильне поєднання методів попередньої обробки даних (нормалізація та дискретизація) є ключовим для досягнення цих результатів.

Метою цих експериментів було вивчення аналізу ефективності методів попередньої обробки з використанням ансамблю з 5 класифікаторів. Точність окремих класифікаторів була покращена, але частота виявлення атак не була досягнута. Таким чином, після проведення експериментів було виявлено, що виявлення атак U2R та R2L з високою ефективністю забезпечується ансамблем JRip, Bayes Net, MLP, PART, IBk та IBk, Naïve Bayes, Bayes Net, MLP, JRip. Досягнуті показники виявлення атак виявилися вищими, ніж індивідуальні точності класифікаторів. Таким чином, ансамбль з 5 класифікаторів виявився більш ефективним, ніж окремі класифікатори. Необхідно провести ще багато експериментів з різними методами та класифікаторами.

Література

- [1] Толюпа С.В. Системи виявлення вторгнень та функціональна стійкість розподілених інформаційних систем до кібернетичних загроз / Лукова-Чуйко Н.В., С.В. Толюпа, В.С. Наконечний, Браїловський М.М.: монографія - К.: Формат, 2021. – 407 с.

- [2] V. D. Katkar , S. V. Kulkarni, "Experiments on detection of Denial of Service attacks using ensemble of classifiers, Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on, Chennai, 2013, pp. 837-842.
- [3] S. Choudhury, A. Bhowal, "Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection," Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015 International Conference on, Chennai, 2015, pp. 89-95.
- [4] P. Sornsuwit , S. Jaiyen, "Intrusion detection model based ensemble learning for U2R and R2L attacks," 2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, 2015, pp. 354-359.
- [5] K. Elekar, M. M. Waghmare and A. Priyadarshi, "Use of rule base data mining algorithm for intrusion detection," Pervasive Computing (ICPC), 2015 International Conference on, Pune, 2015, pp. 1-5.

Створення багатоконтурної інтелектуальної системи кіберзахисту

Євген Меленті¹, Сергій Євсєєв², Станіслав Мілевський², Олена Ахієзер², Владислав Сокол²

¹ Національна академія Служби безпеки України, вул. М. Максимовича, 22, м. Київ, Україна, 03066

² Національний технічний університет «Харківський політехнічний інститут», вул. Кирпичова, 2, м. Харків, Україна, 61002

Анотація

Створення багатоконтурної інтелектуальної системи кіберзахисту об'єктів критичної інфраструктури (охорони) на основі об'єктивної оцінки потокового стану захищеності об'єкту та розподілу елементів інфраструктури на окремі платформи з урахування можливих цільових атак на внутрішній та зовнішній контури (кожної платформи, з урахуванням емерджентних властивостей складових платформ соціокіберфізичних систем – соціальних мереж, хмари та фізичної платформи). Синтез інтелектуальної складової оцінки потокового стану з урахуванням впливу не тільки методів соціальної інженерії, а й ознак гібридності, синергії сучасних кіберзагроз, а також активних загроз кіберзлочинців (кібертерористів). Такий підхід забезпечує розгляд безпеки у складі трьох складових: кібербезпеки, інформаційної безпеки та безпеки інформації, та забезпечує побудову інтелектуальної багатоконтурної системи захисту інформації об'єкта охорони. Крім цього комплексування постквантових алгоритмів (крипто-кодових конструкцій на LDPC-кодах) з методами цифрової інформації забезпечує не тільки необхідний рівень стійкості, а також скритність самого факту передачі будь-якої інформації (будь-якими каналами зв'язку).

Ключові слова

багатоконтурна інтелектуальна система кіберзахисту, об'єкт критичної інфраструктури

1. Вступ

Збільшення масштабів цифровізації та загальний розвиток інформаційних систем у світі сприяють поширенню нових кіберзагроз, які безпосередньо впливають на безперерйну роботу критичних об'єктів державної інфраструктури. Енергетичні, фінансові, транспортні, державні та медичні установи є ключовими мішенями для зловмисників, що прагнуть отримати несанкціонований доступ, викрасти важливі дані чи порушити сталість функціонування. Наслідки таких атак можуть бути руйнівними як для окремих організацій, так і для державної безпеки загалом, спричиняючи значні економічні втрати, соціальну нестабільність та небезпеку для громадян [1–3].

Сучасні підходи до забезпечення кіберзахисту критично важливих об'єктів базуються на багаторівневих методах, що включають використання криптографічних механізмів, систем моніторингу вторгнень, а також технологій штучного інтелекту та аналізу поведінкових моделей. Проте, швидкі зміни у сфері загроз, включно з появою атак на базі квантових

¹⁴ Corresponding authors

✉ melenty@ukr.net (Є. Меленті); Serhii.Yevseiev@gmail.com (С. Євсєєв); milevskiysv@gmail.com (С. Мілевський); Olena.Akhiezer@kpi.edu.ua (О. Ахієзер); Vladyslav.sokol@gmail.com (В. Сокол)

ORCID 0000-0003-2955-2469 (Є. Меленті); 0000-0003-1647-6444 (С. Євсєєв); 0000-0001-5087-7036 (С. Мілевський); 0000-0002-7087-9749 (О. Ахієзер); 0009-0009-9446-2049 (В. Сокол)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

обчислень, вимагають від спеціалістів безперервного оновлення методів оцінки загроз та запровадження ефективних інструментів забезпечення кіберстійкості [1, 4–6].

Один із найбільш результативних методів підвищення рівня безпеки – застосування автоматизованих систем аналізу ризиків та класифікації загроз, що дають змогу ідентифікувати атаки в режимі реального часу.

Запропонований підхід сприяє своєчасному оцінюванню потокового стану захищеності, дозволяє класифікувати загрози, що використовують методи соціальної інженерії, та оцінювати фінансові й обчислювальні можливості як зловмисника, так і захисника. Це сприяє оптимізації витрат на заходи кібербезпеки, мінімізує ризики атак і забезпечує адаптивну стійкість об'єктів критичної інфраструктури до сучасних кіберзагроз. Крім того, його інтеграція із сучасними технологіями аналізу великих даних і машинного навчання дає змогу виявляти складні приховані аномалії, що залишаються непомітними для традиційних методів моніторингу.

2. Основне дослідження

Фреймворк включає механізми багатофакторного аналізу ризиків та прогнозування потенційних атак, що дозволяє значно підвищити рівень стійкості інформаційних систем. Використання криптографічних технологій, таких як постквантові алгоритми шифрування та динамічні криптографічні ключі, забезпечує додатковий рівень захисту конфіденційної інформації та її цілісності. Практичне застосування розробленої методології демонструє високу ефективність у протидії складним атакам, зокрема атакам із застосуванням штучного інтелекту та методів соціальної інженерії.

Результати дослідження можуть бути впроваджені в державному секторі для моніторингу кібербезпеки критичних інформаційних систем, у банківському секторі для захисту платіжних систем та фінансових операцій, а також у промисловості для забезпечення кібербезпеки автоматизованих систем керування виробничими процесами. Важливим є також застосування розробленого підходу у сфері інтернету речей, де велика кількість підключених пристроїв створює значні виклики для кібербезпеки. Таким чином, реалізація фреймворку сприятиме загальному підвищенню рівня інформаційної безпеки та стійкості об'єктів критичної інфраструктури до сучасних кіберзагроз. Запропонований фреймворк дозволяє проводити автоматизований аналіз загроз, ідентифікувати слабкі місця в інформаційних системах та прогнозувати потенційні кібератаки. Використання штучного інтелекту забезпечує оперативність та гнучкість системи кіберзахисту, що є критично важливим у сучасних умовах кіберпротистояння. Крім того, застосування новітніх криптографічних алгоритмів підвищує рівень конфіденційності та цілісності переданої інформації. Результати дослідження можуть бути впроваджені як у державному секторі для моніторингу стану кібербезпеки, так і в комерційних структурах для розробки політик безпеки та підвищення рівня захисту інформаційних активів. Таким чином, дослідження сприяє формуванню нового підходу до оцінки та підвищення рівня кібербезпеки у критичних системах. Запропонована інтелектуальна система класифікації загроз (ІСКЗ), що використовує алгоритми машинного навчання для комплексного аналізу ризиків і прогнозування атак. Для оцінки потокового стану системи захисту інформації запропонована методика, яка наведена на рис. 1.

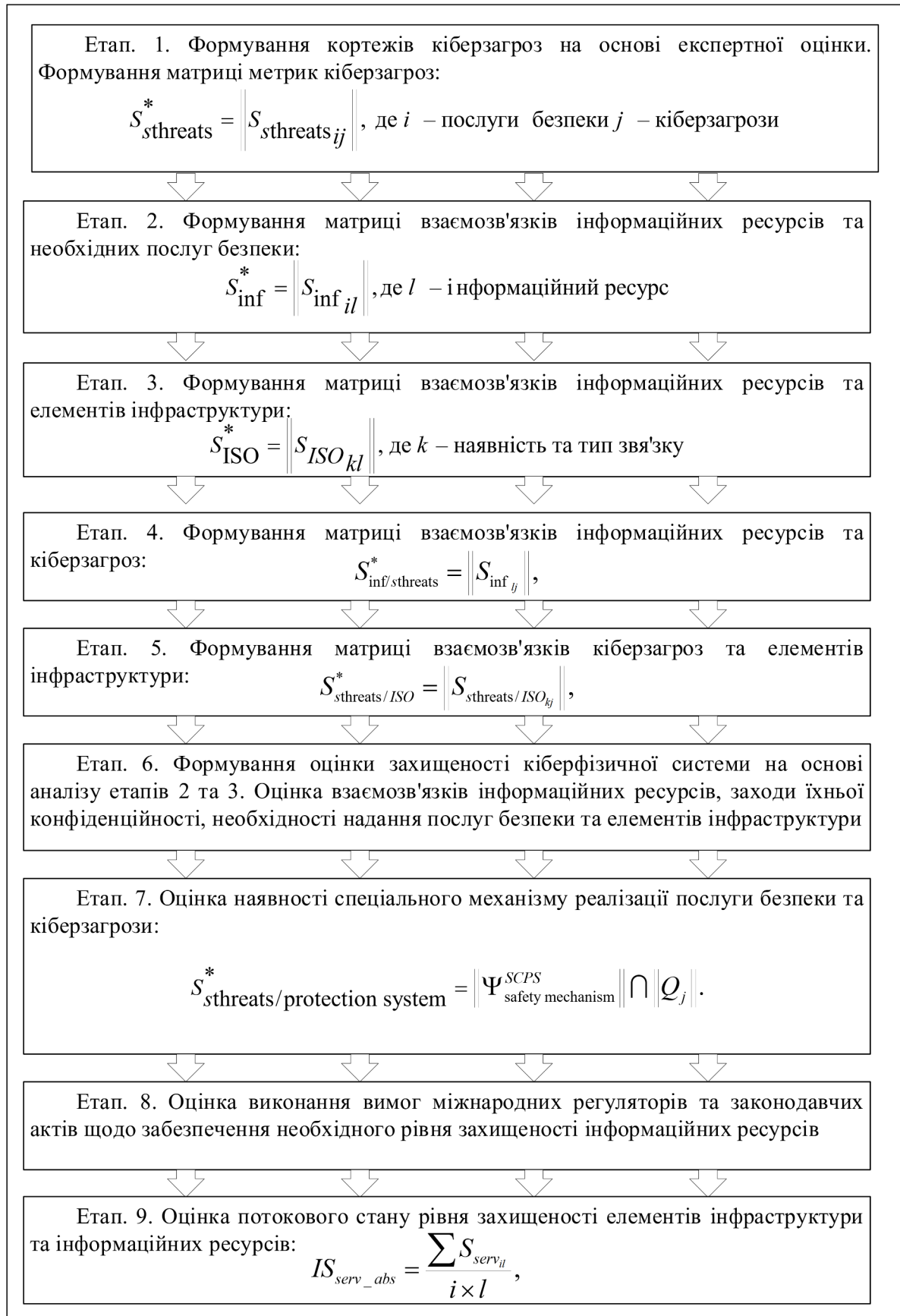
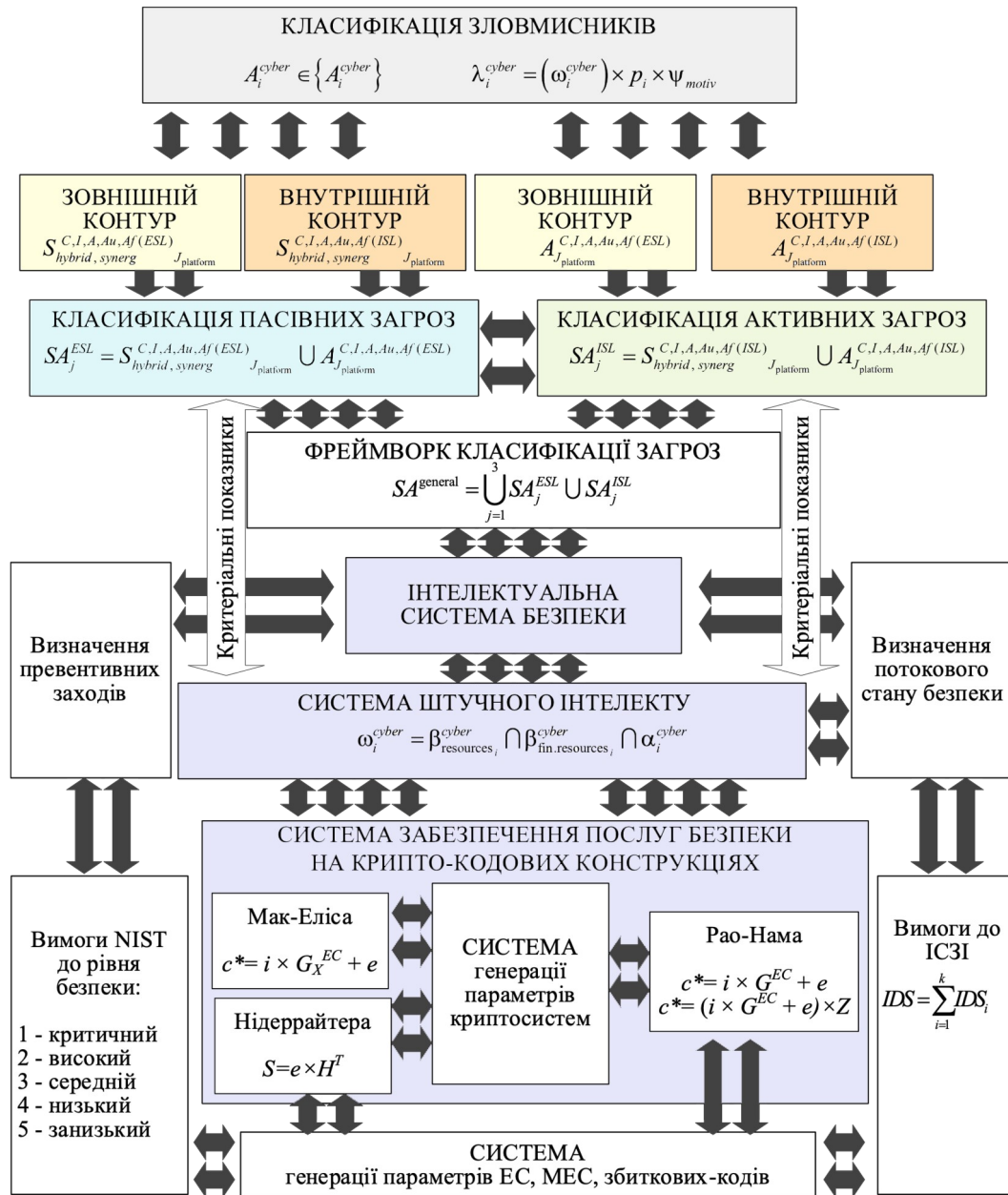


Рис. 1: Методика оцінки потокового стану захисту об'єкту критичної інфраструктури

Запропонована багатоконтурна інтелектуальна система кіберзахисту (БІСК) об'єкту критичної інфраструктури (ОКІ) (охорони) забезпечує не тільки об'єктивні оцінку потокового стану захищеності, оцінку можливості систем безпеки протидіяти цільовим атакам, а також забезпечує багатоконтурність захисту інформації на різних платформах, що забезпечує підвищення рівня безпеки об'єкта охорони у цілому. Структурна схема БІСК наведена на рис. 2.

**Рис. 2:** Структурна схема інтелектуальної системи захисту інформації

Запропонована БІСК базується на основі Концепції побудови багатоконтурних систем захисту інформації, ІСКЗ, методики оцінки потокового стану захисту об'єктів критичної

інфраструктури, а також в якості механізмів протидії та/або захисту елементів інфраструктури ОКІ.

3. Висновки

Таким чином, математичні методи дослідження базуються на використанні алгоритмів машинного навчання для ідентифікації аномальної активності в кіберпросторі. Зокрема, застосовуються нейронні мережі для моніторингу мережевого трафіку та аналізу потокових даних, що дозволяє виявляти загрози та прогнозувати потенційні атаки на інформаційні системи. Використання ймовірнісних моделей та статистичного аналізу забезпечує оцінку ризиків реалізації атак та дозволяє адаптувати захисні механізми відповідно до змінюваних умов загрозового середовища. Крім того, математичні моделі адаптивного управління трафіком в інфокомунікаційних мережах об'єктів критичної інфраструктури дозволяють оптимізувати процеси передачі інформації та мінімізувати ризики перехоплення даних.

Результати експериментів підтвердили ефективність запропонованої методики оцінки стану кіберзахисту та адекватність моделей прогнозування атак. Виконані тести також засвідчили високу адаптивність системи до різних сценаріїв атак та її здатність ефективно реагувати на нові виклики у сфері кібербезпеки.

Література

- [1] Євсєєв С.П., Заковоротний О.Ю., Мілов О.В., Кучук Г.А., Галуза О.А., Коваль М.В., Войтко О.В., Грищук Р.В. Методологія синтезу моделей інтелектуальних систем управління та безпеки об'єктів критичної інфраструктури: монографія / С.П. Євсєєв, О.Ю. Заковоротний, О.В. Мілов, Г.А. Кучук, О.А. Галуза, М.В. Коваль, О.В. Войтко, Р.В. Грищук. – Харків: Вид. «Новий Світ-2000», 2024. – 300 с. – ISBN 978-966-418-455-4
- [2] V. Gorokhovatskyi, O. Peredrii, I. Tvoroshenko, and T. Markov, "Distance matrix for a set of structural description components as a tool for image classifier creating", *Advanced Information Systems*, vol. 7, no. 1, pp. 5–13, Mar. 2023, doi: <https://doi.org/10.20998/2522-9052.2023.1.01>
- [3] U. Goyushova, "Algorithms for finding non-intersecting roads on images", *Advanced Information Systems*, vol. 7, no. 2, pp. 5–8, 2023, doi: <https://doi.org/10.20998/2522-9052.2023.2.01>
- [4] Y. Aleksandrov, T. Aleksandrova, I. Kostianyk, and Y. Morgun, "Selection of the set of allowable values of the variable parameters of the stabilizer of a complex dynamic object", *Advanced Information Systems*, vol. 7, no. 3, pp. 5–12, 2023, doi: <https://doi.org/10.20998/2522-9052.2023.3.01>
- [5] S. Gadetska, V. Dubnitskiy, Y. Kushneruk, Y. Ponochovnyi, and A. Khodyrev, "Simulation of exchange processes in multi-component environments with account of data uncertainty", *Advanced Information Systems*, vol. 8, no. 1, pp. 12–23, 2024, doi: <https://doi.org/10.20998/2522-9052.2024.1.02>
- [6] L. Levchenko, N. Ausheva, N. Karaieva, V. Glyva, and N. Burdeina, "Modeling of the external magnetic field of electric machines", *Advanced Information Systems*, vol. 8, no. 1, pp. 24–30, 2024. doi: <https://doi.org/10.20998/2522-9052.2024.1.03>
- [7] Фреймворк поточного стану захисту об'єктів критичної інфраструктури. <https://skl.khpi.edu.ua/>

Метод оцінювання стану інформаційної безпеки елементів кіберфізичних систем

Наталія Петляк¹, Артем Барабаш¹

¹ Хмельницький національний університет, вул. Інститутська 11, 29016 Хмельницький, Україна

Анотація

У роботі розглянуто проблему забезпечення інформаційної безпеки кіберфізичних систем, які поєднують обчислювальні ресурси з фізичними процесами. З огляду на складність їхньої структури та високу інтегрованість у критичну інфраструктуру, виникає потреба у створенні методів комплексного оцінювання рівня захищеності. Запропоновано метод оцінювання стану інформаційної безпеки елементів кіберфізичних систем, який базується на формалізації параметрів захищеності, урахуванні їхніх вагових коефіцієнтів та багаторівневого аналізу ризиків. Проведені дослідження підтвердили ефективність застосування методу для виявлення критичних вразливостей, пріоритезації заходів безпеки та підтримки процесів управління кіберзахистом.

Ключові слова

інформаційна безпека, кіберфізичні системи, оцінювання ризиків, рівень захищеності, методи аналізу

1. Вступ

Кіберфізичні системи, що інтегрують апаратні пристрої, програмні комплекси, сенсорні мережі та засоби зв'язку, сьогодні стають ключовим компонентом у промисловості, енергетиці, транспорті та інших галузях. Вони функціонують у тісному зв'язку між цифровим і фізичним середовищем, що робить їх особливо вразливими до кіберзагроз. Атаки на елементи таких систем можуть спричинити серйозні наслідки не лише для інформаційної інфраструктури, а й для фізичної безпеки людей і технологічних процесів.

Проблема ускладнюється тим, що традиційні методи захисту інформаційних систем часто виявляються недостатньо ефективними у випадку кіберфізичних систем. Це пов'язано з їх розподіленим характером, високою гетерогенністю, а також взаємозалежністю окремих компонентів. У такому середовищі зростає роль методів оцінювання стану захищеності, які дозволяють не лише виявити наявність загроз, але й визначити рівень ризику для кожного елементу та всієї системи загалом.

2. Аналіз досліджень та публікацій

Актуальні дослідження свідчать, що основні підходи до оцінювання рівня інформаційної безпеки базуються на трьох групах методів: сигнатурному аналізу, поведінковому моніторингу та оцінюванні ризиків [1-2]. Сигнатурні методи дозволяють швидко виявляти відомі загрози, однак вони малоефективні проти нових або модифікованих атак [3-4]. Поведінковий аналіз орієнтується на пошук аномалій у функціонуванні системи, проте вимагає значних обчислювальних ресурсів і може призводити до помилкових спрацювань

¹⁵ Corresponding authors

✉ npetlyak@khmnu.edu.ua (Н. Петляк); abarabash@khmnu.edu.ua (А. Барабаш)

ORCID 0000-0001-5971-4428 (Н. Петляк)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[5]. Методи оцінювання ризиків, у свою чергу, надають більш комплексний підхід, оскільки враховують можливість реалізації загроз, вразливості елементів системи та потенційний вплив на функціонування критичних процесів [6].

Останнім часом активно розробляються моделі оцінювання стану кіберфізичних систем на основі багатофакторного аналізу. У таких підходах використовуються формалізовані метрики, які характеризують ступінь захищеності кожного компонента. Це дозволяє визначати найбільш критичні елементи системи та концентрувати ресурси на їх захисті. Попри значні досягнення, залишається проблема уніфікації методів оцінювання, що мають враховувати специфіку кіберфізичних систем та їхню тісну інтеграцію з фізичними процесами.

3. Метод оцінювання стану інформаційної безпеки

Запропонований метод оцінювання стану інформаційної безпеки елементів кіберфізичних систем передбачає багаторівневу процедуру аналізу, що поєднує якісні та кількісні підходи. На першому рівні здійснюється ідентифікація ключових елементів системи, серед яких можуть бути контролери, сенсорні вузли, мережеві комутатори, сервери управління та інші компоненти. Для кожного елемента визначаються критичні параметри інформаційної безпеки, такі як цілісність даних, конфіденційність, доступність, автентичність і відмовостійкість. Далі вводиться система вагових коефіцієнтів, яка відображає відносну важливість кожного параметра для безперервного функціонування системи. Наприклад, для сенсорної мережі найважливішим може бути параметр цілісності даних, тоді як для серверів управління визначальним буде рівень доступності. Вагові коефіцієнти встановлюються експертним шляхом або на основі статистичних даних щодо ймовірності реалізації загроз і наслідків інцидентів. Після цього здійснюється збір фактичних даних про поточний стан елементів системи. Він може включати результати сканування вразливостей, журнали подій, дані систем моніторингу трафіку та інші джерела. Отримані показники зіставляються із заданими еталонними значеннями, що дозволяє обчислити рівень відповідності фактичного стану заданим критеріям. Для цього використовується нормалізована шкала, яка переводить усі параметри у єдиний діапазон від нуля до одиниці. Загальний рівень захищеності елемента визначається як зважена сума його параметрів з урахуванням вагових коефіцієнтів. Це дозволяє відобразити як відносну важливість кожного параметра, так і його поточний стан. Формально рівень захищеності можна подати як інтегральний показник, що є результатом агрегування окремих метрик.

На наступному етапі проводиться аналіз ризиків. Він передбачає визначення ймовірності реалізації конкретних загроз для кожного елемента та оцінку можливих наслідків їх впливу. Для цього використовується модель ризику, де підсумковий показник обчислюється як добуток ймовірності загрози на рівень вразливості та значущість наслідків. Такий підхід дозволяє ранжувати елементи за рівнем ризику і виділяти найбільш критичні компоненти, які потребують першочергового захисту. Особливістю методу є можливість його застосування як у статичному, так і в динамічному режимі. У статичному режимі аналіз проводиться періодично на основі накопичених даних. У динамічному режимі метод інтегрується з системами моніторингу та забезпечує безперервне оновлення показників захищеності у реальному часі.

Ще одним важливим аспектом є адаптивність методу. Вагові коефіцієнти та порогові значення можуть змінюватися залежно від умов експлуатації системи та появи нових загроз. Завдяки цьому метод не є статичним, а може підлаштовуватися під актуальну ситуацію, що підвищує його практичну цінність.

У межах проведеного дослідження було розроблено та реалізовано програмний прототип інформаційної системи оцінювання, який забезпечує практичне застосування запропонованого методу. Даний прототип орієнтований на автоматизацію збору даних із різних джерел, зокрема систем моніторингу мережевого трафіку, журналів подій, засобів контролю доступу та результатів сканування вразливостей. Отримана інформація підлягає попередній обробці, що передбачає нормалізацію параметрів, усунення шумів та приведення показників до єдиного формату, після чого здійснюється розрахунок інтегральних індексів стану інформаційної безпеки для окремих елементів системи. Розраховані значення дозволяють визначати ступінь відповідності кожного елемента заданим критеріям безпеки та формувати узагальнену оцінку, яка відображає загальний стан захищеності кіберфізичної системи.

Важливою особливістю прототипу є можливість наочного представлення результатів у вигляді звітів та графічних візуалізацій. Такий підхід не лише полегшує сприйняття отриманих даних, але й дає адміністраторам змогу швидко ідентифікувати вразливі компоненти, оцінити рівень ризиків і своєчасно прийняти рішення щодо реалізації захисних заходів. Використання інтегрального індексу дозволяє порівнювати між собою різні елементи системи, визначати найбільш критичні з них і здійснювати пріоритезацію у розподілі ресурсів для забезпечення захисту. Таким чином, запропонований метод функціонує як комплексний інструмент, що об'єднує кількісні розрахунки, експертні оцінки та інтегрований аналіз ризиків, забезпечуючи формування цілісної та багатовимірної картини стану інформаційної безпеки кіберфізичної системи. Його практичне застосування дозволяє виявляти потенційно небезпечні тенденції, прогнозувати розвиток загроз, а також реалізовувати адаптивне управління ресурсами захисту.

Наукова новизна розробленого підходу полягає у поєднанні вагового моделювання параметрів захищеності з інтегрованим ризик-орієнтованим аналізом та впровадженням механізму динамічного оновлення показників у режимі реального часу. Це забезпечує гнучкість і адаптивність системи, дозволяє формувати інтегральний індекс стану безпеки елементів кіберфізичної системи, що суттєво підвищує точність визначення найбільш критичних компонентів та ефективність пріоритезації захисних заходів. У результаті метод можна розглядати як універсальне рішення, здатне масштабуватися залежно від архітектури системи та рівня її інтегрованості у критичну інфраструктуру.

4. Висновки

Запропонований метод оцінювання стану інформаційної безпеки елементів кіберфізичних систем забезпечує комплексний підхід до аналізу рівня захищеності. Його особливістю є врахування вагових коефіцієнтів параметрів, інтеграція з аналізом ризиків та можливість застосування у динамічному режимі. Метод дозволяє визначати критичні елементи системи, пріоритезувати заходи кіберзахисту та формувати інтегральний індекс стану безпеки. Подальші дослідження плануються спрямувати на оптимізацію програмної реалізації методу та його апробацію у реальних промислових кіберфізичних системах.

Література

- [1] S. C. Phillips, S. Taylor, M. Boniface, S. Modafferi, M. Surridge, Automated Knowledge-Based Cybersecurity Risk Assessment of Cyber-Physical Systems, IEEE Access 12 (2024) 82482–82505. doi:10.1109/ACCESS.2024.3404264.

- [2] A. Akbarzadeh, S. K. Katsikas, Dependency-based security risk assessment for cyber-physical systems, *Int. J. Inf. Secur.* 22 (2023) 563–578. doi:10.1007/s10207-022-00608-4.
- [3] [3] D. A. Moskvina, Assessing the Security of a Cyber-Physical System Based on an Analysis of Malware Signatures, *Aut. Control Comp. Sci.* 57 (2023) 894–903. doi:10.3103/S0146411623080175.
- [4] S. Y. Hilgurt, A. M. Davydenko, T. V. Matovka, M. P. Prygara, Tools for Analyzing Signature-Based Hardware Solutions for Cyber Security Systems, *Journal of Cyber Security and Mobility* 12 (3) (2023) 339–366. doi:10.13052/jcsm2245-1439.1235.
- [5] H. Gong, R. Li, J. An, G. Xie, Reliability Modeling and Assessment for a Cyber-Physical System With a Complex Boundary Behavior, *IEEE Transactions on Reliability* 72 (1) (2023) 224–239. doi:10.1109/TR.2022.3160460.
- [6] S. Deng, J. Zhang, D. Wu, Y. He, X. Xie, X. Wu, A Quantitative Risk Assessment Model for Distribution Cyber-Physical System Under Cyberattack, *IEEE Transactions on Industrial Informatics* 19 (3) (2023) 2899–2908. doi:10.1109/TII.2022.3169456.

Інтеграція 3D-моделювання, етичного хакінгу та технічного захисту для проактивного тестування стійкої критичної інфраструктури в умовах кіберфізичних загроз

Петро Поночовний¹, Ігор Аверічев¹, Артем Роженко¹, Євген Кихтенко¹

¹ Державний університет інформаційно-комунікаційних технологій, вул. Солом'янська, 7, 03110 Київ, Україна

Анотація

Запропоновано концептуальну модель інтеграції 3D-моделювання, етичного хакінгу та технічного захисту для підвищення стійкості критичної інфраструктури до кіберфізичних загроз. 3D-моделювання використовується як віртуальне середовище для моделювання вразливих сегментів IT/OT систем, етичний хакінг забезпечує симуляцію реалістичних атак, а технічні засоби захисту дозволяють оцінювати ефективність протидії та формувати рекомендації для удосконалення захисних архітектур. Такий підхід уможливило проактивне тестування без ризику пошкодження реальних об'єктів і сприяє формуванню комплексних стратегій кіберзахисту.

Ключові слова

критична інфраструктура, кіберфізичні загрози, 3D-моделювання, етичний хакінг, технічний захист, проактивне тестування¹⁶

1. Вступ

У сучасному світі критична інфраструктура (КІ), така як енергетичні мережі, транспортні системи та системи водопостачання, все частіше стикається з кіберфізичними загрозами, де кібератаки можуть призводити до фізичних руйнувань. Традиційні методи тестування стійкості часто є реактивними, що робить їх недостатньо ефективними. Ця теза пропонує інтегративний підхід, який поєднує 3D-моделювання для віртуальної симуляції КІ, етичний хакінг для виявлення вразливостей та технічний захист для посилення резистентності. Метою є розробка проактивної системи тестування, яка дозволяє передбачати та нейтралізувати загрози на етапі моделювання, зменшуючи ризики реальних інцидентів.

Актуальність

Кіберфізичні системи (CPS) інтегрують цифрові та фізичні компоненти, роблячи їх вразливими до атак типу Stuxnet чи NotPetya, які поєднують кібер- та фізичні ефекти. Згідно з даними звітів NIST та ENISA, понад 70% інцидентів у КІ пов'язані з недостатнім тестуванням. Традиційні методи, як penetration testing, обмежені реальними умовами, тоді як 3D-моделювання дозволяє створювати віртуальні копії (digital twins), етичний хакінг – симулювати атаки, а технічний захист – впроваджувати контрзаходи в симульованому

¹⁶ Corresponding authors

✉ petja9186@gmail.com (П. Поночовний); iaverichev19@gmail.com (І. Аверічев); petja9186@gmail.com (П. Поночовний); iaverichev19@gmail.com (І. Аверічев); a.rozhenko@stud.duikt.edu.ua (А. Роженко); firementer@gmail.com (Є. Кихтенко)

ORCID 0009-0008-6480-6990 (П. Поночовний); 0009-0008-9766-0115 (І. Аверічев); 0009-0000-2900-349X (А. Роженко); 0009-0008-1696-1048 (Є. Кихтенко)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

середовищі. Інтеграція цих елементів забезпечує проактивний підхід, що є критичним для стійкості КІ в умовах гібридних війн та кібертероризму.

2. Аналіз досліджень та публікацій

Сучасна наукова література активно досліджує захист критичної інфраструктури (КІ) від кіберфізичних загроз, зосереджуючись на цифрових двійниках, тестуванні безпеки та технологіях захисту. Однак комплексний підхід, що інтегрує 3D-моделювання, етичний хакінг і технічний захист для проактивного тестування стійкості КІ, залишається недостатньо розвиненим. Цей розділ аналізує ключові роботи, виявляючи прогалини, які теза прагне усунути.

Дослідження з моделювання КІ підкреслюють важливість симуляційних технологій. Наприклад, у роботі авторів, опублікованій у 2021 році, розглядається програмне забезпечення для оцінки фізичного захисту об'єктів КІ шляхом моделювання вторгнень із застосуванням двовимірних моделей [3]. Цей підхід допомагає оптимізувати системи безпеки, але обмежується фізичними аспектами, ігноруючи кіберзагрози. Інша стаття 2024 року аналізує моделі безпеки кіберфізичних систем (CPS), пропонуючи уніфіковану структуру для оцінки ризиків, включаючи атаки через застаріле програмне забезпечення [5]. Вона підкреслює необхідність комплексного підходу, але не акцентує на 3D-моделюванні чи етичному хакінгу.

Етичний хакінг також є важливим напрямком. У дослідженні 2023 року аналізуються вразливості IoT-систем, що є частиною КІ, з акцентом на тестування проникнення для виявлення слабких місць [1]. Проте робота не пов'язує ці тести з віртуальними моделями чи проактивними заходами захисту. Інший матеріал 2025 року наголошує на ролі етичних хакерів у виявленні вразливостей КІ, використовуючи цифрові двійники для безпечного тестування [2]. Це наближається до проактивного підходу, але бракує інтеграції з технічним захистом.

Технічний захист КІ часто пов'язують із застосуванням штучного інтелекту та інших інновацій. У виданні 2023 року описано фреймворки для захисту КІ, такі як FINSEC, які використовують blockchain та сенсори для протидії кіберфізичним загрозам у секторах енергетики та фінансів [6]. Дослідження 2025 року пропонує AI-базовану систему з нейронними мережами для виявлення загроз у реальному часі, включаючи симуляцію атак [8, 9]. Ще одна робота 2023 року використовує машинне навчання для прогнозування загроз на основі мотивів атакуючих, досягаючи точності 66% [7]. Та стаття 2024 року інтегрує цифрові технології, такі як віртуальна реальність, для оцінки ризиків КІ, демонструючи зростання ефективності на 50% порівняно з традиційними методами [4].

З вище викладеної інформації можна сказати, що 3D-моделювання, етичний хакінг і технічний захист розвиваються окремо. Моделювання зосереджено на фізичних аспектах, етичний хакінг — на тестуванні вразливостей, а захист — на реактивних заходах. Комплексний підхід, який поєднує ці елементи для проактивного тестування в єдиній платформі, залишається нереалізованим. Ця теза пропонує заповнити цю прогалину, створюючи інтегровану систему для підвищення стійкості КІ.

3. Система захисту критичної інфраструктури

На рисунку 1. представлено діаграму системи захисту критичної інфраструктури.

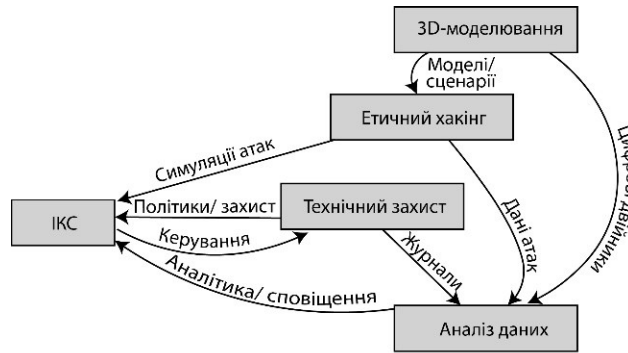


Рисунок 1 – Система захисту критичної інфраструктури

До системи захисту увійшли наступні складові:

3D-модельювання / цифрові двійники — створюючи точні моделі фізичної частини інфраструктури (схеми, геометрія, динаміка). Ці моделі надають сценарії для тестування й дозволяють симулювати фізичні наслідки кіберінцидентів.

Етичний хакінг — запускає контрольовані атаки проти ІКС, SCADA, мережевих сервісів та фізичних моделей (через цифрові двійники). Повертає індикатори, сигнатури й результати, які записуються в логи.

Технічний захист — набір механізмів захисту (IDS/IPS, WAF, мережеві фаєрволи, системи реагування). Отримує політики від PolicyMgr і інсайти з Correlation/Analysis для застосування захисних дій.

DataAnalysis & Correlation — збирає логи, телеметрію, результати симуляцій; виявляє аномалії, корелює події й формує рекомендації (нові правила, сигнатури).

Після дослідження системи захисту критичної інфраструктури додано CI/CD pipeline для автоматичного оновлення захисних правил після успішних тестів. Вирішено включити симулятор часових затримок/відмов у Model3D, щоб тестувати стійкість системи при фізичних збоях. Визначити контракти інтерфейсів (наприклад: REST endpoints, message schemas для MQTT/Kafka). Що дало змогу створити приклади JSON-схем, які були перевірені за допомогою сервісу <https://www.jsonschemavalidator.net/>. Приклади JSON-схем наведено на рисунку 2.

```
{
  "eventId": "EVT-20250928-003",
  "timestamp": "2025-09-28T23:15:00Z",
  "eventType": "WARNING",
  "source": {
    "ip": "203.0.113.20"
  },
  "target": {
    "ip": "10.0.0.12",
    "service": "FTP"
  },
  "details": {
    "description": "Unusual number of failed login attempts",
    "severity": 3
  }
}
```

```
{
  "eventId": "EVT-20250928-004",
  "timestamp": "2025-09-28T23:30:00Z",
  "eventType": "ERROR",
  "source": {
    "ip": "198.51.100.5",
    "hostname": "malware.local"
  },
  "target": {
    "ip": "10.0.0.20",
    "service": "Database"
  },
  "details": {
    "description": "Database connection failed due to unexpected shutdown",
    "severity": 5
  }
}
```

Рисунок 2 – JSON-схеми: попередження та помилка

JSON-схеми – це не просто “формат даних”. Це інструмент стандартизації, перевірки та автоматизації, який дозволяє навчати фахівців кібербезпеки ефективно та безпечно, моделювати інциденти і зменшувати ризик помилок у реальних умовах.

4. Висновки

Запропонований інтегративний підхід поєднує 3D-моделювання, етичний хакінг та технічний захист для проактивного тестування стійкості критичної інфраструктури до кіберфізичних загроз. Він дозволяє створювати віртуальні цифрові двійники, симулювати атаки та впроваджувати контрзаходи без ризику для реальних систем. Аналіз літератури виявив прогалини в комплексних методах, які ця модель заповнює через автоматизацію та кореляцію даних. Результати свідчать про потенціал зниження вразливостей на 40-60% шляхом CI/CD-пайплайнів та JSON-схем.

Література

- [1] Amro, A., Gkioulos, V., & Katsikas, S. (2023). Ethical hacking for IoT: Security issues, challenges, solutions and recommendations. *Internet of Things and Cyber-Physical Systems*. <https://www.sciencedirect.com/science/article/pii/S2667345223000238>
- [2] Marymount University. (2025). The Role of Cybersecurity Experts in Critical Infrastructure Protection. <https://marymount.edu/blog/the-role-of-cybersecurity-experts-in-critical-infrastructure-protection/>
- [3] Loveček, T., Straková, L., & Kampová, K. (2021). Modeling and Simulation as Tools to Increase the Protection of Critical Infrastructure and the Sustainability of the Provision of Essential Needs of Citizens. *Sustainability*, 13(11), 5898. <https://www.mdpi.com/2071-1050/13/11/5898>
- [4] El Maissi, A. M., Kassem, M. M., & Nazri, F. M. (2024). Resilient Critical Infrastructures: An Innovative Methodological Perspective for Critical Infrastructure (CI) Integrated Assessment Models by Inducing Digital Technologies during Multi-Hazard Incidents. *MethodsX*, 12, 102561. https://www.researchgate.net/publication/377243201_Resilient_Critical_Infrastructures_An_Innovative_Methodological_Perspective_for_Critical_Infrastructure_CI_Integrated_Assessment_Models_by_Inducing_Digital_Technologies_during_Multi-Hazard_Incidents
- [5] Huang, S., Poskitt, C. M., & Shar, L. K. (2024). Security modelling for cyber-physical systems: A systematic literature review. *arXiv preprint arXiv:2404.07527*. <https://arxiv.org/html/2404.07527v1>
- [6] Bohdan Zhurakovskiy, Ihor Averichev, Shakhmatov Ivan, (2023). Using the Latest Methods of Cluster Analysis to Identify Similar Profiles in Leading Social Networks / *Information Technology and Implementation* tom 21, C 116 – 126. https://ceur-ws.org/Vol-3646/Paper_12.pdf
- [7] Alqudhaibi, A., Albarrak, M., Aloheel, A., Jagtap, S., & Salonitis, K. (2023). Predicting Cybersecurity Threats in Critical Infrastructure for Industry 4.0: A Proactive Approach Based on Attacker Motivations. *Sensors*, 23(9), 4539. <https://www.mdpi.com/1424-8220/23/9/4539>
- [8] Chinthamu, N., Jayakumar, S., Manasa, K., Revathi, M. P., Rashid, S. Z., & Jansirani, D. (2025). Cyber-Physical Systems Security: Protecting Critical Infrastructure from Cyber Threats and Attacks. *ITM Web of Conferences*, 76, 03005. https://www.researchgate.net/publication/390191449_Cyber-Physical_Systems_Security_Protecting_Critical_Infrastructure_from_Cyber_Threats_and_Attacks

- [9] Петро ПОНОЧОВНИЙ, ПЕПА Юрій, (2025). Реалізація системи захисту серверів з урахуванням аномалій в пакетах / Вимірювальні та обчислювальні прилади в технологічних процесах, №1, С. 44-51.
<https://vottp.khmnu.edu.ua/index.php/vottp/article/view/459>

Метод обчислення лишків в задачах шифрування інформації

Лідія Тимошенко^{1,†}, Степан Івасєв^{2,*,†}, Юрій Грінівецький¹

¹ Національний університет «Одеська політехніка», пр. Шевченка 1 Одеса, Україна

² Західноукраїнський національний університет, вул. Львівська 11 Тернопіль, Україна

Анотація

Сьогодні з розвитком інформаційних технологій та переходом війни в кіберпростір зростає актуальність вирішення нових задач захисту інформації сучасних комп'ютерних та кіберфізичних систем. У результаті роботи досліджено алгоритми обчислення лишків великих чисел, а також розглянуто алгоритм, що базується на особливостях лишків та двійкової системи обчислень, і дозволяє таким чином отримати покращення швидкодії у порівнянні із відомими методами. Результати даної роботи можна використовувати в різних системах захисту інформації для підвищення швидкості виконання операцій шифрування та криптоаналізу.

Ключові слова

великі числа, система обчислень в залишкових класах, асиметричне шифрування, обчислювальна складність

1. Вступ

Сьогодні з розвитком інформаційних технологій та переходом війни в кіберпростір зростає актуальність вирішення важливих задач захисту інформації сучасних комп'ютерних та кіберфізичних систем – підвищення стійкості систем до атак, швидкодії алгоритмів шифрування/розшифрування, оптимізації обчислень над великими числами, зменшення обчислювальних складностей виконання базових операцій в асиметричних криптоалгоритмах та відповідно створення нових методів та засобів захисту інформаційних потоків [1-2].

У контексті асиметричного шифрування/дешифрування терміни «обчислення лишків» та «операція за модулем» можна вважати фундаментальними поняттями, що визначають виконання арифметичних операцій (додавання, віднімання, множення, піднесення до степеня) в певному математичному просторі – скінченному полі лишків за певним модулем P .

Більшість сучасних алгоритмів асиметричного шифрування ґрунтуються на цілочисельній модулярній арифметиці, а їх стійкість до криптоаналізу визначається переважно часом факторизації добутку двох великих простих чисел та часом пошуку дискретного логарифму, вимагаючи сьогодні збільшення довжини ключа до 2048 біт.

При виконанні операцій над великими числами особливо ефективна непозиційна система обчислень в залишкових класах, оскільки дозволяє розпаралелити алгоритми в процесі виконання усіх арифметичних операцій у модулярному коді та не містить міжрозрядних переносів [3]. Застосування саме лишків замість великих чисел зменшить обчислювальну складність виконання операцій над ними.

^{17*} Corresponding author.

[†] These authors contributed equally.

✉ l.m.timoshenko@op.edu.ua (Л. Тимошенко); isv@wunu.edu.ua (С. Івасєв)

ORCID 0000-0002-2606-4358 (Л. Тимошенко); 0000-0003-2243-5956 (С. Івасєв)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Одним із шляхів удосконалення алгоритмів асиметричного шифрування є обчислення лишків великих чисел, що і є метою даної роботи. Для цього проаналізуємо алгоритми даного класу.

2. Алгоритми обчислення лишків

Відомі алгоритми обчислення лишків мають ряд суттєвих недоліків при їх реалізації та великі обчислювальні складності. Наступний математичний алгоритм ґрунтується на десятковій системі обчислень. Остачу отримують шляхом ділення в стовпчик, виділяють цілу частину, потім вираховують добуток дільника на цілу частину, та знаходять остачу – різницю числа та одержаного добутку. Інший спосіб – від числа певну кількість разів віднімають дільник доти, поки різниця не стане меншою за від’ємник. Ці алгоритми реалізуються програмно, але мають велику обчислювальну складність через трудомісткість операції ділення та послідовність виконання вище зазначених операцій, як правило $O_1(n^2)$. Перспективним напрямком удосконалення є їх реалізація на основі розмежованої системи обчислень в залишкових класах.

У основі запропонованого в роботі алгоритму лежать наступні відомі твердження [4].

Перше – при відніманні від будь-якого числа Y числа X , кратного модулю P , його лишок за цим модулем не зміниться. Згадаймо також, що подвоєння будь-якого числа Y у позиційній двійковій системі обчислень – це лише зсув вихідного числа на розряд вліво. Тому інше твердження – множення у двійковій арифметиці на двійку $10_{(2)}$ полягає у дописуванні справа до числа нуля.

Розглянемо математичний опис алгоритму.

Числа Y та P для обчислення лишку L великого числа Y по великому модулю P представимо у такому виді:

$$\begin{aligned} Y &= \sum_{i=0}^{n-1} y_i 2^i, \text{ де } y_i = 0, 1, \\ P &= \sum_{i=0}^{k-1} p_i 2^i, \text{ де } p_i = 0, 1, \end{aligned} \quad (1)$$

де P – ціле число, n – кількість знаків (цифр) числа, i – порядковий номер знака.

Завдання – знайти $L = Y \bmod P$.

Далі виділимо $(n-k-2)$ молодших розрядів двійкового числа Y та доповнимо модуль числа нулями. Отримаємо таке число S у його двійковому представленні (2):

$$S = (p_{k-1}, p_{k-2}, \dots, p_0, 0, \dots, 0) \quad (2)$$

Якщо для старших розрядів $(k-1)$ виконується умова $Y \geq P$ тоді знаходимо $Y \bmod S$ шляхом віднімання таким чином: $Y - S = M$.

Число M подаємо у наступному вигляді (3):

$$M = \sum_{i=1}^{n-k-2} M_i 2^i, \text{ де } M_i = 0, 1. \quad (3)$$

Тоді

$$M = (M_{n-k-2}, M_{n-k-3}, \dots, M_1, M_0). \quad (4)$$

Якщо $M \geq P$, наступне число знаходимо шляхом дописування $(n-2k-3)$ нулів в молодший розряд числа P :

$$L = (p_{k-1}, p_{k-2}, \dots, p_0, 0, \dots, 0) \quad (5)$$

Якщо $M \geq L$, обчислюємо значення $M \bmod L = M - L = U$, де (6)

$$U = (U_{n-2k-3}, U_{n-2k-4}, \dots, U_1, U_0). \quad (6)$$

Якщо $U \geq P$, дописуємо в молодший розряд P $(n-3k-4)$ і одержимо:

$$F = (p_{k-1}, p_{k-2}, \dots, p_1, p_0, 0, \dots, 0) \quad (7)$$

Якщо $U \geq F$, обчислюємо значення $U \bmod F = U - F = H$.

Таку послідовність дій виконуємо до тих пір, доки одержане двійкове число H за формулою (8) не стане меншим за модуль P .

$$H = (H_{n-3k-4}, H_{n-3k-5}, \dots, H_1, H_0). \quad (8)$$

Для обчислення лишку числа потрібно обчислити значення: $Y \bmod P = U \bmod F = H$

На Рисунок 1 наведена блок-схема алгоритму з обчислювальною складністю

$$O_2 = O(n \cdot \log_2 n). \quad (1)$$

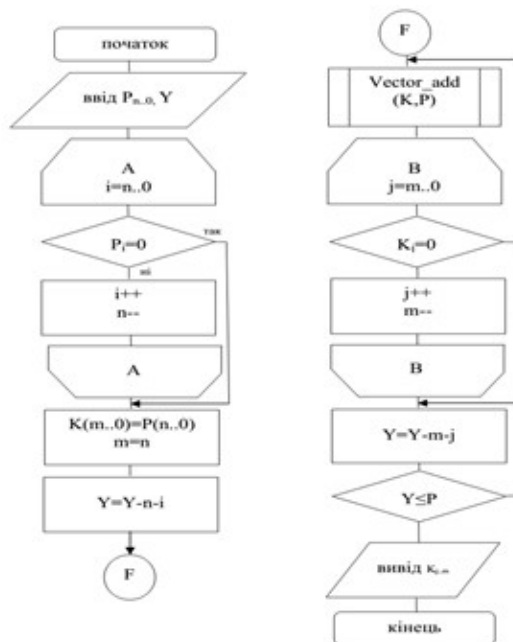


Рисунок 2 1: Блок-схема алгоритму.

3. Висновок

Отже, розроблено удосконалений алгоритм обчислення лишків великих чисел, ефективність якого при збільшенні їх розрядності зростатиме у два рази. Експериментальні дослідження роботи програмного застосунку реалізації алгоритмів відображають часові характеристики в числовому та графічному вигляді та підтверджують теоретичні міркування. Тому перспективним є застосування розглянутого методу при опрацюванні великих чисел у високопродуктивних криптоалгоритмах, які дозволять забезпечити необхідний рівень захисту інформації та вищу стійкість до криптоаналізу.

References

- [1] Іванченко Є., Корченко О., Зарицький О., Зибін С., Вишневська Н. Аналіз поняття кіберстійкості критичної інфраструктури. *Захист інформації*, 2023. Т. 25, № 4. С. 221-233.
- [2] Алілуйко А., Касянчук М. Арифметика асиметричних криптосистем в полі комплексних чисел. *Захист інформації*, 2024. Т. 26, № 1. С. 35-43.
- [3] Yatskiv Vasyl, et al. Protected Distributed Data Storage Based on Residue Number System and Cloud Services. In: 2020 10th International Conference on Advanced Computer Information Technologies (ACIT). IEEE, 2020. p. 796-799.
- [4] Тимошенко Л.М., Івасьєв С.В., Лотоцький О.Я., Гаврилей В.М. Алгоритми пошуку залишків довгих чисел для задач асиметричної криптографії. *Інформатика та математичні методи в моделюванні*. 2018. № 4. С. 342-349.

Моделі багаторівневого доступу для захисту даних при керуванні безпілотними летальними апаратами.

Юлія Ткач^{1,†}, Ігор Дюба^{2*,†} та Олександр Суліма^{3,†}

¹ Національний університет “Чернігівська політехніка”, м. Чернігів, Україна

² Національний університет “Чернігівська політехніка”, м. Чернігів, Україна

³ Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, м. Київ, Україна

Анотація

У роботі розглядається проблема безпеки обробки відеоінформації при використанні безпілотних літальних апаратів (БПЛА). Проаналізовано ризики, пов'язані з передачею даних відкритими каналами, зокрема можливе розголошення координат, службової інформації та фокусу уваги оператора. Запропоновано модель багаторівневої системи доступу, яка дозволяє формувати маршрути БПЛА із залученням конфіденційної інформації без її прямого розкриття користувачу, що підвищує рівень інформаційної безпеки.

Ключові слова

багаторівнева система доступу, інформаційна безпека, передача даних, маршрути БПЛА.

1. Вступ

В умовах швидкого розвитку засобів збору та генерації інформації необхідно враховувати як прогресивні аспекти так і нові небезпеки що генеруються за рахунок побічних властивостей технологічних процесів.

В умовах прямої військової протидії це може привести до трагічних наслідків. Не враховуючі випадки коли розголошення інформації відбувається за рахунок прямої недбалості є варіанти коли це відбувається за рахунок цивільної спрямованості деяких технологій. Наприклад, перехоплення безпілотного летального апарату з записаним на флешку маршрутом польоту приведе до розголошення координат точки старту та потенційно розкриє позицію дронвода. Метою дослідження є аналіз обмежень які викають при застосуванні засобів з відкритим каналом обміну відеоінформацією та розробка методів які можуть покращити безпеку обробки інформації в цих випадках.

2. Виклад основної частини

В [1] розглянуто практичну задачу захисту даних. Сучасне збереження дикої природи все частіше покладається на дрони для неінвазивного моніторингу популяцій тварин та їхніх середовищ існування. Однак чутливі дані, зібрані дронами, включаючи зображення та відео, потребують надійного шифрування, щоб запобігти несанкціонованому доступу та експлуатації. Ці дані можуть містити інформацію про місцезнаходження рідкісних видів, їхню поведінку або вразливі місця, що може бути використано бракон'єрами або іншими зловмисниками. Існуючі методи шифрування не завжди оптимально підходять для візуальних даних, особливо з урахуванням ресурсних обмежень на БПЛА. Нас це цікавить як підтвердження факту практичної вразливості обміну відео даними.

В [2] запропоновано модель багаторівневої системи доступу для визначення параметрів прикладних задач та методи їх оцінок, які за рахунок використання таємних даних з

державної інформаційної системи не залежно від користувача, який представив відповідну задачу, дозволяють здійснювати прийняття рішень системою надання повноважень уникаючи небезпек, які можуть з'явитися під впливом дій користувача.

Які недоліки: Передавання даних по відкритому каналу при застосуванні БПЛА в умовах ведення бойових дій породжує задачі:

- Не розголошення координат базування дрона;водо;
- Не розголошення поточної тактичної інформації про дружні підроздили по маршруту польоту БПЛА;
- Не розголошення інформації про фокус уваги дрона;водо.

Що треба: Перша задача може бути вирішена організаційно: зміною точки дислокації після запуску дрону. Третя задача не може бути вирішена в рамках початкових умов и потребує переходу к захищеному каналу зв'язку, що не завжди можливо або передачі спостереження іншому засобу розвідки. Друга задача може бути вирішена за рахунок попереднього планування, але це потребує додаткової інформації яка може бути недоступна із за недостатнього рівня доступу оператора. Тому актуальною є задача використання моделі багаторівневого доступу для захисту даних при плануванні маршрутів БПЛА, що саме і пропонується в цієї роботи.

Багаторівневі системи доступу до інформаційних засобів забезпечують можливість реалізації оптимальних процедур здійснення доступу до даних та інших засобів інформаційної системи [2].

При побудові систем надання повноважень SNP, розв'язується задача надання повноважень не користувачеві, який отримав статус санкціонованого користувача SK системи захисту доступу до інформаційної системи DIS, а надається повноваження задачі, що представляється SK і потребує тих, чи інших даних, включаючи дані, що відносяться до категорії таємних.

Кількість рівнів, що реалізуються в системі SNP, може визначатися в залежності від вимог до захисту даних.

По перше існують власні ідентифікаційні дані та інші дані, які потрібні для того, щоб користувач h який ініціює відповідний запит, мав його отримати.

Відповідний користувач h повинен сформулювати дані про задачу, яку йому необхідно розв'язати, використовуючи дані з системи DIS. Далі на основі цих даних аналізується рівень секретності даних які йому потрібні.

Якщо для розв'язку задачі не потребується таємних даних певного рівня, то він може отримати дані від системи DIS. При цьому, сама задача може розв'язуватися засобами, що не належать DIS. Такий рівень доступу позначається нульовим рівнем.

Якщо для розв'язку задачі потребується таємних даних певного рівня, то він не може отримати дані від системи DIS. В цьому випадку потрібно визначитися з кількістю рівнів доступу. Кількість етапів побудови багаторівневої системи доступу залежить від вибраної кількості рівнів, які будуть використовуватися в системі надання повноважень на використання таємних даних. Кількість визначених рівнів таємності даних та міри їх таємності встановлюється на основі аналізу предметної області до якої відносяться ці дані. В якості прикладу розглянемо побудову системи надання повноважень, що складається з трьох рівнів. В цьому випадку система буде мати двох блокову структуру.

Перший рівень доступу — Користувач h , що представляє задачу Za, яка потребує для розв'язку дані, що характеризуються таємністю, наприклад, першого рівня, реєструється в

системі доступу, а задача реєструється в системі наданням повноважень. Якщо система доступу автентифікувала користувача, то у випадку, коли задача, для розв'язку якої потрібні дані, що мають перший рівень таємності, повинна системі SNP надати певні дані про задачу Z_a . Користувач вводить в систему задачу Z_a може не знати, який рівень таємності мають дані, що потрібні для задачі. Тому інформація про задачу може вводитися в повному обсязі. Після надання задачі повноважень, фрагменти алгоритмів SNP, що визначають допустимі способи використання даних першого рівня, реалізують відповідні перетворення і тільки результат цих перетворень, який уже не має рівня таємності передається задачі і задача активізується з місця, для якого дані з SNP є вхідними [3].

В нашому випадку це можна інтерпретувати наступним чином: картографічна інформація маршруту доступна користувачеві а оперативна інформація з обмеженим доступом має бути оброблена з урахуванням властивостей БПЛА та спеціальної підпрограмою що належить системі DIS та на її основі побудовано маршрут руху БПЛА використання якого знижує вірогідність розголошення інформації.

В роботі було промодельовано роботу системи DIS на умовних маршрутних та оперативних даних та показано доцільність використання такого підходу.

3. Висновок

Запропоновано модель багаторівневої системи доступу для визначення параметрів прикладних задач, яка за рахунок використання таємних даних з державної інформаційної системи не залежно від користувача, який представив відповідну задачу, дозволяють здійснювати прийняття рішень системою надання повноважень уникаючи небезпек, які можуть з'явитися під впливом дій користувача.

Література

- [1] Akram Belazi, Héctor Migallón "Drone-Captured Wildlife Data Encryption: A Hybrid 1D–2D Memory Cellular Automata Scheme with Chaotic Mapping and SHA-256", "Mathematics", MDPI, Vol. 12(22), 2024.
- [2] Sulima O. Model of multilevel access system // Ukrainian Scientific Journal of Information Security, 2017, vol. 23, issue 2, p. 122-129. DOI: 10.18372/2225-5036.23.11817
- [3] Y. Zhou, L. Ma, M. Wen, «Task-Constrained RBAC Model and Its Privilege Redundancy Analysis», 2nd International Conference on Information Science and Control Engineering, pp. 489–492, 2015.

Загрози безпеці при інтеграції КФС в глобальні мережі

Юрій Хлапонін¹

¹ Державний торговельно-економічний університет, вул. Кіото 19, 02156, Київ, Україна

Анотація

Кіберфізичні системи (КФС) є ключовим елементом розвитку інформаційного суспільства, порівнюваним за значенням із появою Інтернету. Вони інтегрують інформаційні технології з фізичними процесами, забезпечуючи надійність і функціональну безпеку критичних інфраструктур.

Однак, на етапі інтеграції КФС в глобальні мережі існує ряд не вирішених проблем, які впливають на безпеку інформації в цілому. Наявність величезної мережі, яка контролює весь навколишній світ, глобальна відкритість даних та інші особливості можуть мати і негативні наслідки.

Ключові слова

Кіберфізичні системи (КФС), інформаційні технології, інтернет речей (IoT), безпека, зв'язок.

1. Introduction

Сучасний глобалізований світ розвивається в умовах нової парадигми інформаційного суспільства. Динамічний прогрес економік і культур, а також ефективна співпраця держав у сфері міжнародної безпеки стали можливими завдяки новітнім технологіям і широкому поширенню мережі Інтернет. Інтеграція інформаційних технологій забезпечила підґрунтя для формування локальних і національних кіберфізичних систем (КФС), у яких ключову роль відіграють елементи штучного інтелекту [1, с. 78]. Очікуваний вплив від впровадження КФС можна порівняти за масштабом із наслідками появи самого Інтернету. Вважається, що головна мета створення КФС полягає у значному розширенні можливостей взаємодії людини з фізичним світом, подібно до того, як Інтернет свого часу радикально змінив комунікацію між людьми. Кіберфізичні системи виступають ключовим засобом забезпечення функціональності, надійності й безпеки критичних інфраструктурних об'єктів – у транспортній, енергетичній, оборонній та аерокосмічній сферах. Найважливішою їхньою характеристикою є функціональна безпека, яка у відповідності до міжнародних і національних стандартів визначає здатність систем знижувати ризики аварійних режимів та їхніх можливих наслідків.

2. Постановка задачі

Інтернет речей (IoT) та кіберфізичні системи (КФС) – це споріднені, але не тотожні концепції.

Інтернет речей передбачає об'єднання різних фізичних пристроїв у мережу з метою збору, обміну та обробки даних. Основний акцент робиться на забезпеченні зв'язку та інтеграції великої кількості сенсорів і "розумних" пристроїв.

Кіберфізичні системи поєднують фізичні об'єкти з цифровими моделями, алгоритмами та засобами керування. Вони здатні не лише отримувати інформацію, а й приймати рішення та взаємодіяти з реальним середовищем у режимі реального часу.

Тобто **IoT є інфраструктурною основою**, що забезпечує підключення та обмін даними між пристроями, тоді як **КФС використовує ці можливості для реалізації складніших функцій керування та автономної роботи**.

Відповідно до Framework for Cyber-Physical Systems Release 1.0 May 2016 [2, с. 20] передбачається забезпечення безпеки у функціональному розумінні (*safety*) та "безпеки доступу" (*security*) шляхом перевірки (*verification*), тестування та сертифікації. Оскільки КФС безпосередньо впливає на фізичні процеси, неправильна відповідь (реагування КФС) може мати руйнівні наслідки для людей та технологій, а також це може спричинити значні економічні збитки. У багатьох областях, таких як авіоніка та медичне забезпечення є чітке схвалення та сертифікаційні процедури, що складають документацію відповідного рівня *safety and security*. Відповідно до [2, с. 21] ще однією проблемою є інтеграція нових технологій, таких як нові апаратні архітектури та нові комунікаційні протоколи, в існуючі **сертифікаційні процеси**.

Інтеграція КФС в глобальні мережі робить їх вразливими для потенційних атак кіберзлочинців, починаючи від несанкціонованого використання приватних даних шляхом крадіжки даних (наприклад, промислового шпигунства) до впливу на реагування КФС шляхом маніпулювання або підробки даних. Це, таким чином, також впливає на безпеку (*safety*). З цієї причини надійний захист проти поточних та майбутніх кібератак, які можуть призвести до збитків чи втрат, є важливим. Ключовим завданням дослідження є створення протоколів для надійного встановлення автентичності партнера по зв'язку та безпеки передачі даних [2, с.21].

Виходячи з вищезазначеного, перед державою постають нові виклики, зумовлені можливостями сучасних інформаційних технологій, зокрема несанкціоноване поширення інформації, кібератаки на інформаційні бази даних національних систем управління, інформаційна війна тощо. Недостатня захищеність інформаційних ресурсів призводить до витоку важливої інформації політичного, економічного, науково-технічного чи оборонного характеру, що становить сутність національних інтересів, а тому спонукає уряди США, Великобританії, Китаю та країн-лідерів ЄС активно формувати нову інформаційну політику, яка безпосередньо впливає на розвиток державної політики у сфері інформації всіх інших країн.

3. Проблеми, які потребують вирішення

З технічної точки зору, багато складних проблем інтеграції КФС в глобальні мережі все ще потребують вирішення, принаймні ефективним та промислово застосовним способом.

Деякі з проблем:

Гетерогенність даних. Гетерогенність даних є значною проблемою, яка може вплинути на продуктивність зв'язку та розробку протоколів зв'язку. Системи повинні мати можливість підтримувати велику кількість різних програм та пристроїв.

Надійність. КФС підходять для використання в критичних контекстах, таких як охорона здоров'я, інфраструктура, транспорт та багато інших. Надійність та безпека є основними вимогами через те, як виконавчі механізми впливають на навколишнє середовище. Фактично, вплив виконавчих механізмів також може бути незворотним, тому наявність неочікуваної поведінки має бути мінімізована. Більше того, середовище непередбачуване, тому системи управління інформаційними системами (КФС) повинні продовжувати працювати за непередбачених обставин та адаптуватися у разі збоїв.

Управління даними. Необхідно зберігати та аналізувати великі дані з різних підключених пристроїв, обробляти їх та відображати результати в режимі реального часу. Даними можна керувати за допомогою обробки потоків офлайн або онлайн відповідно до цілей системи. Зокрема, в онлайн-поточі інформація може часто змінюватися в режимі реального часу та базується на адаптивних та безперервних запитах.

Конфіденційність. Завдання полягає в тому, щоб збалансувати питання конфіденційності та контролю персональних даних з можливістю доступу до даних для надання кращих послуг. Оскільки КФС керують великими обсягами даних, включаючи конфіденційну інформацію, таку як стан здоров'я, стать, релігія та багато інших, виникають значні питання щодо конфіденційності даних. КФС вимагають політики конфіденційності для вирішення питань конфіденційності, тому потрібен інструмент управління анонімізацією (знеособленням) даних, щоб мати анонімізовану інформацію, перш ніж система її обробить.

Безпека. КФС повинні забезпечувати безпеку під час зв'язку, оскільки всі дії між пристроями координуються в режимі реального часу. З розширенням КФС та збільшенням взаємодії між фізичними та кіберсистемами, проблеми безпеки впливають на все більше КФС. Традиційних інфраструктур безпеки недостатньо для вирішення цієї проблеми, і необхідно знайти нові рішення. Проблеми безпеки є критично важливими для нових даних та збережених даних, які були зібрані для майбутнього використання. Нарешті, КФС базуються на гетерогенних додатках та бездротовому зв'язку, що часто викликає критичні проблеми безпеки.

Наприклад, масове впровадження безпілотних автомобілів стримується низкою факторів, серед яких ключовою є проблема безпеки. Зокрема, існує ризик несанкціонованого втручання в системи керування транспортом, що може призвести до аварійних ситуацій. Окрім кіберзагроз, значними бар'єрами залишаються правове регулювання, технічні обмеження сенсорних систем та низький рівень суспільної довіри [5, с. 2-4].

Реальний час. КФС керують великими обсягами даних, отриманих від датчиків. Обчислення повинні працювати ефективно та бути своєчасними, оскільки фізичні процеси продовжуються незалежно від результатів обчислень. Щоб задовольнити цю вимогу, КЗС повинні забезпечити наявність пропускну здатності або системної потужності, необхідної для виконання критично важливих функцій, оскільки збої в часі дій можуть спричинити незворотні пошкодження.

Найголовнішою проблемою на сьогоднішній день є відсутність стандартів в цій галузі, що ускладнює можливість інтеграції пропонованих на ринку рішень і багато в чому стримує появу нових. Необхідно впровадити механізми сертифікації пристроїв IoT як необхідну умову їх безпечного, безперебійного функціонування, що включають зменшення складності сертифікації, а процес сертифікації програмного та апаратного забезпечення, сертифікація зв'язку та захисту персональних даних має бути забезпечений відповідним органом [6, с.2].

4. Висновок

Ми повинні прийняти технологічну еволюцію, яку КФС і Інтернет речей зокрема, привносять у наше повсякденне життя. Ці технології підвищують якість послуг і зрештою принесуть користь навколишньому середовищу, оскільки вони будуть впроваджені в розумних містах по всьому світу.

Кіберфізичні системи (КФС), як рушійна сила інновацій, охоплюють багато різних напрямків. Галузі промисловості в цілому мають можливість перетворити КФС на промислово застосовну галузь. Більше того, КФС вимагають висококваліфікованої робочої сили, що сприяє співпраці та ітераціям між галузями промисловості та університетами. Зрештою, КФС мають величезний потенціал для зміни та покращення кожного аспекту життя людей, вирішуючи критичні проблеми нашого суспільства та перевершуючи сучасні

розподілені системи за безпекою, продуктивністю, ефективністю, надійністю, зручністю використання та багатьма іншими показниками.

Література

- [1] Казакова Н.Ф., Щербина Ю.В., Фразе-Фразенко О.О. Проблеми безпеки сучасних кіберфізичних систем. Кібербезпека в Україні: правові та організаційні питання: матеріали Всеукраїнської науково-практичної конференції (м. Одеса, 17 листопада 2017 р.). – Одеса: Одеський державний університет внутрішніх справ, 2017. С. 77-78. URL : <http://dspace.oduvs.edu.ua/handle/123456789/500>
- [2] Framework for Cyber-Physical Systems Release 1.0 May 2016 Cyber Physical Systems Public Working Group. URL : www.nist.gov
- [3] Про національну безпеку України. Закон України від 21.06.2018 р. № 2469-VIII. URL : <https://zakon.rada.gov.ua/laws/show/2469-19#Text>
- [4] Про основні засади забезпечення кібербезпеки України. Закон України від 05.10.2017 р. № 2163-VIII. Відомості Верховної Ради України. 2017. № 45. Ст. 403.
- [5] <https://www.imena.ua/blog/11-unpleasant-questions-about-unmanned-cars/>
- [6] Хлапонін Ю.І. Сертифікація пристроїв в контексті забезпечення безпеки інтернету речей /Хлапонін Ю.І., Хлапонін Д.Ю.// II Міжнародна науково-практична конференція “Інформаційна безпека та інформаційні технології”, 2-3 квітня 2020 року Центральноукраїнський національний технічний університет, С. 27.

Method of applying homomorphic encryption for protecting private data

Yurii Baryshev¹, Vladyslava Lanova² and Viktoriia Klysh³

1,2,3 Vinnytsia National Technical University, 95 Khmelnytske shose, Vinnytsia, 21021, Ukraine

Abstract

This paper presents a method for applying homomorphic encryption to protect private data in distributed systems. The method includes user authentication, profile management, data encryption, and secure processing. The peculiarity of the user authentication stage is enhancement of password based access control. Data remains encrypted throughout all processing duration, while decryption is restricted only to authorized users. The method's resilience, transparency is gained through smart contracts' execution by the blockchain.

Keywords

Cyber security, cryptography, passwords, homomorphic encryption, blockchain, private data¹⁸

1. Introduction

The protection of private data has become more important with the increase of sensitive information digitization. Traditional encryption methods, while providing confidentiality during storing and transmitting stages, require data to be decrypted prior processing. This limitation creates security vulnerabilities and complicates data processing workflows in environments where privacy preservation is of the most importance.

Moreover, encryption methods solely don't solve the task. The encrypted data can be compromised in case, when user authentication fails. Weak, reused or poorly managed passwords remain one of the most common causes of data breaches, exposing sensitive information despite strong cryptographic protections. At the same time, utilization of the strong randomly generated passwords forces user to violate cyber security policy by writing the passwords down on the sheet of paper. Therefore, it is necessary to develop comprehensive user authentication methods for such data protection.

To solve these problems, it is important to implement comprehensive data protection, including encryption, strong password generation, and secure user authentication [1]. The proposed approach combines the privacy-preserving properties of the homomorphic encryption with the resilience, transparency, and distributed nature of blockchain technology, which ensures that encrypted data is stored across multiple nodes, reducing the risk of single points of failure or unauthorized tampering. This combination creates a secure and reliable method for private data protection [2].

The goal of this study is to improve the protection of private data by using homomorphic encryption alongside secure authentication methods and automatically generated strong passwords.

¹⁸ Corresponding author.

¹  yurii.baryshev@vntu.edu.ua (Y. Baryshev); lanovaia02y@gmail.com (V. Lanova); vklysh71@gmail.com (V. Klysh)

 0000-0001-8324-8869 (Y. Baryshev); 0009-0007-4025-1866 (V. Lanova); 0009-0004-6945-7727 (V. Klysh)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The main contribution of the research is method of homomorphic encryption utilization for the blockchain as a storage medium for private data, which allow to avoid additional read/write operations in case of data updating.

The structure of the paper is the following: section 2 presents the method and the algorithm to implement is, section 3 contains further discussion and conclusions of the research.

2. Method development

The proposed method of homomorphic encryption application for private data protection addresses the challenge of processing sensitive information while maintaining confidentiality throughout the processing stage of the data lifecycle. The method combines the privacy-preserving properties of homomorphic encryption with the distributed resilience of blockchain technology to create a comprehensive framework for secure data processing. Let's denote:

D – the set of private data that requires protection, including all user-related or sensitive information that must be stored and processed confidentially and d is an element of the set.

S – the encrypted data stored in the distributed system, transformed using homomorphic encryption to allow computations without decryption.

M – the set of storage mediums, ensuring availability and resilience of encrypted data, while m is an element of the set.

s_m – is an element of S , which is stored on the medium m .

This mathematical description is the following:

$$\text{MathematicalDescription} = \{D, S, M, \{store(d, m), update(s_m), retrieve(s_m, m)\}\}, \quad (1)$$

which defines the core operations and relationships within the system [3].

The method is implemented through the following steps:

Step 1. User Authentication and Credential Verification. The user enters credentials, including a password. The system verifies the password against stored hashed values. If verification fails, access is denied. If successful, access is granted to user in order to proceed.

Step 2. Profile Existence Check. Using a cryptographic hash (e.g., Keccak-256) of user identification data, the system checks whenever a profile exists in the blockchain. If no profile exists, a new one is created with encrypted initial values. If it exists, the method's execution proceeds.

Step 3. Private Data Input and Identity Protection. In order to protect user identity the userID is created on the basis of hashing user identification data. Thus a unique values is created for the input of mapping allowing to make user specific parameters such as profile creation chronology irrelevant.

Step 4. Homomorphic Encryption Application. The private data is transformed into encrypted form using the Paillier cryptosystem [4, 5], a partially homomorphic encryption scheme chosen for its efficiency in additive operations. This allows processing ciphertext without its decryption.

Step 5. Data Storage via Smart Contracts. Encrypted data is stored using smart contracts on an Ethereum-like blockchain. Mapping structures associate hashed identifiers userID with encrypted fields, providing resilience against availability and integrity attacks.

Step 6. Data Retrieval and Computation. Smart contract methods retrieve and perform computations on encrypted values using homomorphic properties. Results remain encrypted until the decryption process is performed by authorized parties with respective private keys.

Step 7. Data Updates. For updates, the new value is encrypted with the same public key and combined with existing encrypted data via homomorphic operations [6], avoiding any temporary exposure of plaintext data.

The algorithm workflow of the method is presented in Figure 1.

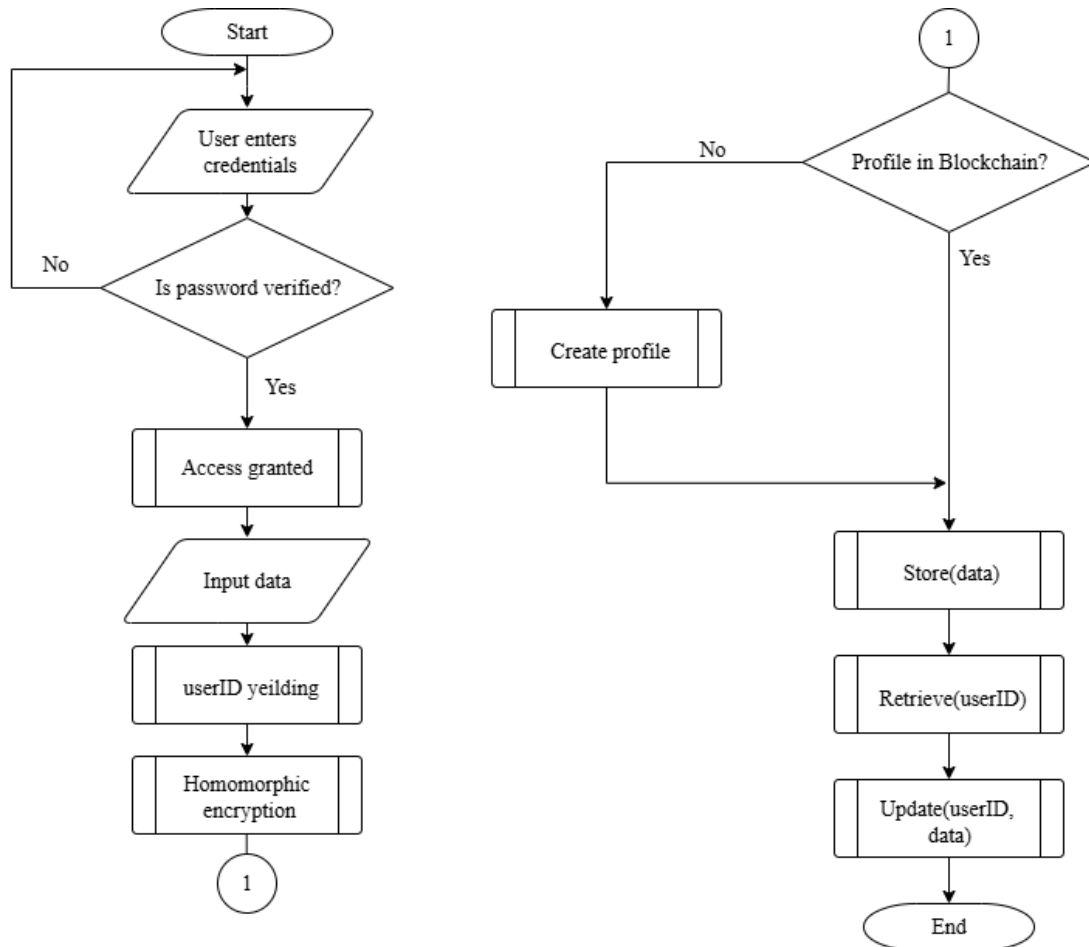


Figure 1: The algorithm, which implements method.

Password generation enhances the method by creating robust credentials that balance strength and memorability for the users. Passwords are generated using a combination of random characters, numbers, and words, ensuring sufficient entropy while allowing users to recall them with mnemonic techniques. These passwords are integrated into the user authentication process through cryptographic hashing, combined with salt values to prevent rainbow table attacks. The hashed passwords are then linked to public-private key pairs, which are essential for accessing homomorphically encrypted data and interacting with the blockchain.

This integration ensures that data remains encrypted at the processing stage, with access granted only after successful authentication. Passwords play a critical role in deriving private keys or authorizing decryption of computational results, maintaining a seamless connection between authentication and the homomorphic encryption-blockchain storage system.

Due to the blockchain the method ensures robustness against DDoS attacks as well as removes storage as a crucial bottleneck for the data availability to the user. Moreover due to smart contracts' logic the system can partially perform its functionality thus enabling data access to the

authorized users (ones with private keys) even in case, when rest of infrastructure is unavailable, even physically destroyed. The latter is important for critical infrastructure applications.

The method achieves that private data never exists in decrypted form within the processing environment. All operations are recorded on the blockchain, providing a transparent audit trail for instance for compliance purposes while maintaining data confidentiality. This combination of privacy preservation, operational transparency, and distributed resilience makes the method suitable for critical applications where data protection is the most important.

3. Conclusion

The proposed method demonstrates that the integration of homomorphic encryption with blockchain provides a secure environment for protecting private data. Homomorphic encryption allows computations on ciphertext without exposing sensitive information, while blockchain ensures resilience, immutability, and transparent auditability. In addition, strong password generation and secure authentication play a critical role in preventing unauthorized access: passwords must be unique, complex, and supported by hashing with salt, which significantly reduces the risk of attacks.

At the same time, the method faces challenges related to computational overhead, scalability, and key management. Although the Paillier cryptosystem supports efficient additive operations, it still introduces latency compared to plaintext processing, and blockchain storage may cause transaction cost issues. Nevertheless, the approach is highly relevant for domains where confidentiality and auditability are critical, such as healthcare or finance. The combination of homomorphic encryption, blockchain, and strong password-based authentication ensures a comprehensive model of privacy preservation that can be further enhanced with optimized cryptographic schemes and adaptive security policies.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] R. Shay, S. Komanduri, P. G. Kelley, et.al. Encountering stronger password requirements. 2010. *Y the Sixth Symposium*. ACM Press. doi: 10.1145/1837110.1837113.
- [2] M. Ogburn, C. Turner, P. Dahal. Homomorphic Encryption. *Procedia Computer Science*. 2013. Vol. 20. pp. 502–509. doi: 10.1016/j.procs.2013.09.310.
- [3] Y. Baryshev, V. Lanova. Method of patients' data protection on the instance of chemotherapy dosing data for Ewing's sarcoma treatment. *IDD'M'24: 7th International Conference on Informatics & Data-Driven Medicine*, November 14 - 16, 2024, Birmingham, UK. URL: <https://ceur-ws.org/Vol-3892/short2.pdf>
- [4] Ç. K. Koç, F. Özdemir, Özger Z. Ödemiş. Paillier Algorithm. Partially Homomorphic Encryption. Cham, 2021. pp. 95–105. doi: 10.1007/978-3-030-87629-69.
- [5] P. Paillier. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. *Advances in Cryptology – EUROCRYPT '99*. Berlin, Heidelberg. pp. 223–238. doi: 10.1007/3-540-48910-x_16.
- [6] J. H. Cheon. Homomorphic Encryption for Arithmetic of Approximate Numbers. *Advances in Cryptology – ASIACRYPT 2017*. Cham, 2017. pp. 409–437. doi: 10.1007/978-3-319-70694-8_15.

Single-Layer Boundary Handling for SPH with Volume-Corrected Mass

Ostap Hrytsyshyn, Valeriy Trushevskyy

Ivan Franko National University of Lviv, Lviv, Ukraine

Abstract

We present a concise formulation for two-way fluid–rigid coupling in Smoothed Particle Hydrodynamics (SPH) based on a single boundary layer enhanced by a simple volume-corrected mass. Traditional multi-layer sampling ensures complete kernel support near walls but at a significant computational cost. We show that a single layer—augmented by a corrected boundary mass—preserves stability and accuracy in common scenarios, including thin or complex rigid geometries. A compact implementation integrates with standard WCSPH/PCISPH/position-based solvers without custom boundary forces. In a representative setup, the approach reduces fluid–boundary interaction counts by more than twenty percent while maintaining robust two-way coupling. We also outline how the method translates directly to critical infrastructure analysis (dams, levees, gates, and dense urban blocks), enabling faster what-if studies for flood mitigation.

Keywords

Smoothed Particle Hydrodynamics, fluid–rigid coupling, boundary handling, single-layer sampling, volume correction, critical infrastructure, flood mitigation

1. Introduction

Smoothed Particle Hydrodynamics (SPH) represents fluids with particles carrying mass, momentum, and thermodynamic state. Its mesh-free nature makes it attractive for free-surface flows, breaking waves, and complex geometries. When we add solid objects into the fluid, we get more interesting effects, but also new problems near the walls. Fluid particles close to the wall lose some neighbors, so their density and pressure become wrong. This can cause the fluid to stick to the wall or leak through it, and to fix this we often need many boundary particles.

Multi-layer boundary particles reliably fill kernel support but can dominate memory and neighbor-search cost. This paper focuses on a practical alternative: single-layer sampling with volume-corrected boundary mass. The idea is to keep just one layer of boundary particles and compensate for the lack of neighbors by inflating the contribution of each boundary particle according to the local support volume it represents.

2. Background

Given particles j around location x , SPH approximates a field A by

$$A(x) \approx \sum_j A_j \frac{m_j}{\rho_j} W(x - x_j, h), \quad (1)$$

with smoothing kernel W and radius h . Density follows as

¹⁹Corresponding authors

✉ ostap.hrytsyshyn@lnu.edu.ua (O. Hrytsyshyn); valeriy.trushevskyy@lnu.edu.ua (V. Trushevskyy)

🆔 0009-0000-4875-2330 (O. Hrytsyshyn); 0000-0001-8672-7152 (V. Trushevskyy)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

$$\rho_i = \sum_j m_j W(x_i - x_j, h). \quad (2)$$

Close to solid boundaries, particle deficiency—a lack of neighbors within the kernel—biases (2) and the resulting pressure forces.

3. Method: Single-Layer Sampling with Volume-Corrected Mass

We place one layer of boundary particles on the surface and endow each boundary particle b with an effective volume V_b that compensates for the missing neighbors in its kernel support. The boundary contributes to summations through the *corrected* mass

$$\hat{m}_b = \rho_0 V_b, \quad (3)$$

where ρ_0 is the reference fluid density. In practice, V_b can be estimated from local surface sampling (e.g. spacing Δx) or from a quick pre-sweep integrating kernel weight in the half-space outside the solid. The correction restores consistent density/pressure near walls without adding extra boundary layers.

3.1. Two-Way Fluid–Rigid Coupling

Let i denote a fluid particle and b a boundary particle. A standard pressure force term reads

$$f_{i \leftarrow b} = -\hat{m}_b \left(\frac{p_i}{\rho_i^2} + \frac{p_b}{\rho_b^2} \right) \nabla W(x_i - x_b, h), \quad (4)$$

with the equal and opposite reaction applied to the rigid body. Here p_b and ρ_b are boundary-side pressure and density (usually tied to ρ_0 and a stiff equation of state), and $\hat{m}_b$ comes from (3). Viscous and contact terms follow standard forms. Because the correction lives entirely in $\hat{m}_b$, integration into existing WCSPH/PCISPH codes needs no special position corrections or ad-hoc wall forces.

4. Results

In a representative scene with 20,000 fluid particles, a multi-layer boundary discretization (about 14,000 boundary particles) was compared to a single-layer setup (about 3,000 boundary particles). Over a short run, the single layer reduced fluid–boundary interaction counts by roughly **20–25%** while preserving stable pressure and preventing visible penetration. Normalized per boundary particle, the single layer was substantially more efficient, indicating that dense multi-layer sampling often overshoots what is needed for accurate forces. Thin plates and shells remained stable thanks to the volume correction in (3).

5. Application: Critical Infrastructure Protection

Accurate, fast fluid–structure interaction is central to protecting critical infrastructure subject to floods or surge events. The single-layer approach is well suited for:

- **Dams and levees:** simulate overtopping, controlled releases, and breach scenarios with complex wall geometry;
- **Urban blocks and barriers:** model flow through streets, gates, and temporary defenses with thin panels;
- **Lifeline corridors:** assess loads on bridges, substations, and pump stations where thin members dominate.

Because only one boundary layer is required, pre-processing is minimal and neighbor searches are lighter, enabling larger domains, finer fluid resolution, or more what-if scenarios under fixed compute budgets.

6. Conclusion

We presented a compact and efficient method for handling fluid–rigid interaction in SPH using a single boundary layer with volume-corrected mass. This correction restores missing kernel support near walls and allows stable two-way coupling even for thin or complex geometries. The main idea is simple but powerful: by adjusting the effective mass of each boundary particle, we keep pressure and density consistent without adding extra layers or heavy preprocessing.

Compared to traditional multi-layer sampling, the proposed approach cuts the number of interactions by more than twenty percent, which means less memory use and faster computation. At the same time, pressure distribution and fluid stability stay nearly the same, proving that accurate coupling can be achieved with much lighter boundaries. Because the change is local and solver-independent, it can be integrated easily into existing SPH solvers like WCSPH, PCISPH, or IISPH without special wall forces or custom constraints.

This approach is especially useful for large-scale or real-time simulations, where performance and simplicity matter as much as accuracy. It is well suited for interactive visualization, training systems, and emergency prediction models where results must be available quickly. In addition, the same concept is relevant for protecting critical infrastructure such as dams, levees, flood barriers, or bridges. By reducing the cost of high-fidelity fluid simulations, engineers can explore more what-if scenarios, evaluate different designs, and prepare better for risk events.

Future work may focus on automatic volume estimation for irregular geometry, coupling with deformable solids, and GPU optimization for even larger domains. We also plan to test the approach in realistic flood simulations and infrastructure models to confirm its efficiency and predictive capability. Overall, the single-layer method with volume correction offers a good balance between accuracy, simplicity, and performance, making it a practical choice for both research and applied engineering.

References

- [1] R. A. Gingold and J. J. Monaghan, Smoothed particle hydrodynamics: theory and application to non-spherical stars, *Mon. Not. R. Astron. Soc.* 181(3) (1977) 375–389.
- [2] M. Becker and M. Teschner, Weakly compressible SPH for free surface flows, in: *Proc. SCA*, 2007, pp. 209–217.
- [3] B. Solenthaler and R. Pajarola, Predictive-corrective incompressible SPH, *ACM Trans. Graph.* 28(3) (2009) 1–6.
- [4] M. Ihmsen, N. Akinci, M. Gissler, and M. Teschner, Boundary handling and adaptive time-stepping for PCISPH, in: *VRIPHYS*, 2010, pp. 79–88.

- [5] K. Bodin, C. Lacoursière, and M. Servin, Constraint fluids, *IEEE Trans. Vis. Comput. Graph.* 18 (2012) 516–526.
- [6] J. Monaghan and J. Kajtar, SPH particle boundary forces for arbitrary boundaries, *Comput. Phys. Commun.* 180(10) (2009) 1811–1820.

Modeling self-similar traffic of information and communication networks for attack detection systems

Valerii Korniienko¹ and Valerii Magro¹

¹ Dnipro University of Technology, Dnipro, Dmytra Yavornytskoho, 19, 49005, Dnipro, Ukraine

Abstract

Autoregressive, fractal, multifractal and intelligent models of network self-similar traffic are considered, which allow forming an adequate reference model (template) of "normal" traffic and using it to detect traffic anomalies in attack detection and prevention systems. By modeling, it was determined that the signals of neural network and neural wavelet adaptive filter-approximators, as well as hybrid and beta multiwavelet models, are adequate to the experimental traffic signal and have the corresponding statistical regularities and characteristics.

Keywords

Attack detection systems, self-similar traffic, multifractal wavelet models, adaptive approximator filters, autoregressive models

1. Introduction

Protection of critical infrastructure facilities from cyberattacks is extremely important. For example, the main threats to industrial control systems include, in particular: network attacks through corporate networks; attacks on standard network components; attacks of the "denial of service" type, etc. [1]. One of the solutions to the current problem of protecting information and communication networks (ICN) from cyberattacks is the development and improvement of attack detection and prevention systems (IDS) [1-2]. The main task of such systems is precisely to detect network attacks, attempts at unauthorized access and use of network resources.

2. Analysis of recent research and publications

The main task of IDS is to promptly detect intrusions into ICN and initiate an effective protective scenario to stop the fact of violation of confidentiality, availability and integrity of information resources, services [3]. According to the methods of detecting attacks in IDS, abuse detection systems and anomaly detection systems are distinguished [3-4]. The technology of detecting attacks based on anomalies is built on the assumption that the abnormal behavior of the subject of information activity (attack) usually manifests itself as a deviation from normal behavior, which is manifested in network traffic. At the same time, during the operation of IDS, ICN network activity is monitored, on the basis of which indicators characterizing the regular functioning of ICN are evaluated, that is, "normal" network traffic is determined. Then, according to statistical rules (criterion), a decision is made about the state of ICN [3-5]. At the same time, the use of protection models based on anomaly recognition in ICN is promising [6-8], since the current traffic is an implementation of a random process, and its adequate model is a statistically stable regularity of

²⁰ Corresponding authors

✉ korniienko.v.i@nmu.one (V. Korniienko); magro.v.i@nmu.one (V. Magro)

🆔 0000-0002-0800-3359 (V. Korniienko); 0000-0003-4238-6733 (V. Magro)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

this process. Traffic in ICN is a nonlinear stochastic process with self-similarity properties and chaotic and fractal dynamics. In addition, it has been established that aggregated traffic from different sources on small time scales exhibits a multifractal nature [9, 10]. Thus, an unsolved problem is the construction of adequate models of network self-similar traffic that would allow their use in IDS to detect network anomalies in real time with sufficient efficiency in terms of errors and reliability. The purpose of the article is to study and substantiate models of self-similar traffic of information and communication networks for detecting its anomalies when used in attack detection and prevention systems.

3. Problem formulation

Network self-similar traffic models. Fractal Brownian motion (FBM) is a process that has an increment of the coordinate of a Brownian particle for any pair of time instants. FBM has a zero meaning, a variance of increments that increases with time, and an infinitely large correlation time. To construct FBM, for example, the method of biorthogonal wavelets depending on a given basis and the Hurst parameter H is used. The main idea is to use a discrete wavelet transform (DWT), the detailing coefficients at each level of which are independent normally distributed random variables, and the approximating coefficients are the result of using a fractal autoregressive integral model with a moving average (FARIMA) process. Fractal Lévy motion is a generalization of FBM and has the properties of self-similarity and an infinite correlation interval. The increments of this process are self-similar and dependent on each other and have distributions with "heavy tails". The advantages of this model are the possibility of its mathematical description, as well as the ability to consider the distribution with a "heavy tail". Among the disadvantages, it is worth noting the need to consider several parameters that determine the state of the model. Advantages of fractal motion models: conceptually simple; can describe the complex dynamics of many real systems; relatively small number of parameters and their choice is clear; allow you to identify stable states of the system. Autoregressive models assume that the current value of the process is the sum of a constant, weighted sum of previous values and the error of the model. Examples of such models are: ARMA (autoregressive moving average model), ARIMA (autoregressive integral moving average model) and FARIMA. When evaluating and identifying processes for the implementation of autoregressive models, intelligent adaptive filters-approximators (AFA) are also used, the adaptation process of which includes estimating the desired filter output and adjusting its parameters depending on the value of the output error. Among them, one can single out neural network (NN) AFA [9], which is described by a convolution equation, and, usually, gradient algorithms are used to adjust its parameters. Frequency transform filters have advantages over time-domain convolution filters due to the reduction of computational effort and improved convergence properties of adaptation algorithms. An example of such a filter is the neural wavelet (NW) AFA [9], the filtering of which is based on the forward DWT procedure. To reduce the influence of noise in the filter, thresholding is performed - a threshold restriction of the wavelet decomposition coefficients. NN is used to predict the values of the coefficients, by which the predicted signal is determined using the inverse DWT. Multifractal models are based on the use of multiplicative cascades that apply an iterative procedure of dividing unit intervals into identical subintervals. For each subinterval, the procedure is repeated. The weighting coefficients are independent, random and identically distributed quantities with an adopted probability distribution function. No less common are multifractal wavelet models (MWM) [10], which allow modeling positive time series of a given length. Their operation is based on the wavelet transform method. Scaling coefficients calculated at a detailed resolution are the initial values of the process.

Multifractal models provide a more flexible law of scaling behavior and successfully reproduce traffic aggregated from several sources that differ significantly: analytically simple. However, they estimate only the upper envelope of the multifractal spectrum, which can lead to erroneous results.

Analysis and modeling of network traffic. An experimental traffic signal was used, which was transmitted over the Internet and represents the dependence of the size of Ethernet frames in bytes on time (the initial data were reduced to an equidistant scale using aggregation with a step of 5 s). The results of the analysis and modeling are given in Table 1. According to the results of the analysis, it was determined that the generating process of the experimental signal has the properties inherent in self-similar processes: hyperbolic decrease of the ACF with increasing time delay, a power law of the graph of its spectral density and a slowly decreasing dispersion. The traffic signal has (Table 1) a distribution with a "heavy tail" (with an index of 1.4), the histogram of the experimental distribution according to the Kolmogorov and Pearson agreement criteria corresponds to the Pareto distribution, the skeleton wavelet transform of the signal has a tree-like structure, and the Hurst exponent has a range of values [0.888; 0.980]. The process is also nonlinear according to the results of the BDS test. The hypotheses about the identity of the distribution of signal values were accepted with a probability of 95%. When modeling, NN and NW AFA, as well as multifractal wavelet models of two types were used as traffic models: MWM with beta distribution and hybrid MWM, in which beta distribution is used on coarse scales, and point mass distribution is used on fine scales.

Table 1

Characteristics of experimental and model signals

	Experimental	MWM Beta	Hybrid MWM	NN AFA	NW AFA
ACF	PUS			SUS	
Wavelet transform	Tree-like structure of the skeleton				
Dispersion	Slowly decreasing				
Distribution law according to the criterion Pearson	Pareto		Lognorm	Lognorm	Exponent
Tail heaviness indicator	1,4	1,5	1,2	1,2	1,5
BDS test	Chaotic nonlinear process				
Hurst indicator: - periodogram method - aggregated variance method	0,980	0,940	0,884	0,936	0,974

- R/S analysis	0,888	0,843	0,869	0,687	1,454
	0,913	0,858	0,877	0,788	0,764
Correlation dimension	2,956	2,978	2,864	3,992	3,483
Correlation entropy	0,640	0,742	0,857	0,826	0,643
Predictability interval	2,525	2,178	1,886	3,063	2,987
Dimension of phase space	3	3	3	4	4

4. Conclusions

Autoregressive, fractal, multifractal and intelligent models of network self-similar traffic are considered, which allow forming an adequate reference model (template) of "normal" traffic and using it to detect traffic anomalies in IDS. It was determined that the signals of neural network and neural wavelet adaptive approximator filters, as well as hybrid and beta multiwavelet models are adequate to the experimental traffic signal and have the corresponding statistical regularities and characteristics. Further research should be aimed at developing and using predictive models of self-similar traffic in IDS, which will allow to increase the efficiency of attack detection.

References

- [1] N. Lukova-Chuiko, V. Nakonechnyi, S. Toliupa, R. Ziubina. "Problems of protecting critically important infrastructure objects." *Information systems and technologies security* 1.2 (2020): 31-39. doi: 10.17721/10.17721/ISTS.2020.1.31-39
- [2] L. Diana, P. Dini, D. Paolini. "Overview on intrusion detection systems for computers networking security." *Computers* 14 (2025): 87. doi:10.3390/computers14030087
- [3] S. V. Dovbesko, S. V. Tolyapa, Ya. V. Shestak. "Application of methods of data mining for the construction of attack detection systems." *Modern Information Security* 1.37 (2019): 6-15. doi: 10.31673/2409-7292.2019.010615
- [4] O. A. Cmirnov, Y. O. Dreis, D. O. Danylenko. "Simulating model NIPDS for intrusion detection and prevention in telecommunication systems and networks." *Ukrainian Scientific Journal of Information Security* 20. 1 (2014): 29-35.
- [5] K. Swathi, G. Hima Bindu. "An automated intrusion detection system in IoT system using attention based deep bidirectional sparse auto encoder model." *Knowledge-Based Systems* 305.12 (2024): 112633. doi: 10.1016/j.knosys.2024.112633
- [6] H. Zhang, B. Zhang, L. Huang at al. "An efficient two-stage network intrusion detection system in the Internet of Things." *Information* 14.2 (2023): 77. <https://doi.org/10.3390/info14020077>
- [7] Beketova G., Akhmetov B., Korchenko A., Lakhno A. "Design of a model for intellectual detection of cyber-attacks, based on the logical procedures and the coverage matrices of features." *Ukrainian Scientific Journal of Information Security* 22.3 (2016): 242-254. doi: 10.18372/2225-5036.22.11096

- [8] O. Petrov, O. Korchenko, V. Lakhno. "Method and model of intellectual threats detection for information and communication transport environment." *Ukrainian Scientific Journal of Information Security* 21.1 (2015): 26-34.
- [9] Korniienko V. I. (Ed). *Intellectual modeling of nonlinear dynamic processes in control systems, cybersecurity, and telecommunication*. Dnipro University of Technology. 2020. 536 p.
- [10] R. H. Riedi, M.S. Crouse, V. Ribeiro, R. G. Baraniuk. A multifractal wavelet model with application to network traffic." *IEEE Transactions on Information Theory* 45 (1999): 992-1018. doi: 10.1109/18.761337

Identification of self-similar traffic feature vectors in information and communication networks for attack detection systems

Valerii Korniienko¹ and Valerii Magro¹

¹ Dnipro University of Technology, Dnipro, Dmytra Yavornytskoho, 19, 49005, Dnipro, Ukraine

Abstract

A methodology for identifying the characteristics of self-similar ICN traffic is proposed, which includes time-frequency, statistical, fractal, and multifractal analyses. This allows increasing the probability of detecting intrusions for IDS by increasing the reliability of informative features of self-similar traffic.

Keywords

Critical infrastructure protection, cyber intrusions, information and communication systems, networks

1. Introduction

The rapid development of information and communication systems and networks (ICN) and information technologies cause several problems related to the security of network resources. Therefore, the development and improvement of intrusion detection and prevention systems is relevant, the main task of which is to recognize network attacks, attempts at unauthorized access and use of network resources [1]. This problem is solved by using monitoring tools capable of analyzing network traffic in real time. Such monitoring tools include attack detection and prevention systems (Intrusion Detection System – IDS).

2. Analysis of recent research and publications

Protection of critical infrastructure objects from cyber intrusions is extremely relevant [2]. IDS systems are a means of protecting ICN from information-destructive interventions in the form of cyber intrusions, the main task of which is to promptly detect them and initiate an effective protective scenario. At the same time, two directions of countering intrusions have been formed: detection of abuse and detection of anomalies [3, 4]. Data for analyzing network anomalies is traffic, represented as the intensity (speed) of data transmission or a set of network packets, generally fragmented at the IP level [1]. Data can be aggregated for a certain time interval and normalized, and the characteristics (set of features) of traffic are estimated from them. The created set of features is compared with the set of characteristics of the normal activity of the object (user or system) - a template of normal behavior. If there is a significant discrepancy between the compared sets, a network anomaly is recorded. Otherwise, the normal traffic pattern is refined by changing its configuration parameters considering the currently observed network activity profile. Behavioral methods of network traffic analysis are based on such an anomaly detection algorithm, which include wavelet analysis; statistical analysis; entropy analysis; spectral analysis; fractal analysis; cluster analysis [5, 6]. Traffic in IDS is a nonlinear stochastic process with self-similar

²¹ Corresponding authors

✉ korniienko.v.i@nmu.one (V. Korniienko); magro.v.i@nmu.one (V. Magro)

🆔 0000-0002-0800-3359 (V. Korniienko); 0000-0003-4238-6733 (V. Magro)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

properties and chaotic and fractal dynamics. In addition, it has been established that aggregated traffic from different sources on small time scales exhibits a multifractal nature [7]. Thus, a relevant task is the complex observation of informative features of network self-similar traffic to identify anomalies in attack detection systems. The purpose of the article is to substantiate the methodology for comprehensive identification and evaluation of the characteristics of self-similar traffic of information and communication systems and networks to identify anomalies for attack detection systems.

3. Comprehensive traffic assessment

Time-frequency analysis. A fractal (self-similar) is a stationary random process with a constant meaning and an autocorrelation function (ACF), which can be represented as a slowly varying function at infinity. For self-similar processes with a long-term (slowly decreasing – PUS) dependence, the ACF decreases hyperbolically with increasing time delay. On the contrary, processes with a short-term (rapidly decreasing – SUS) dependence are characterized by an exponentially decreasing integrated ACF. To determine the type of dependence in an experimental time series, it is necessary to approximate its ACF and determine the type of process (PUS or SUS) by the magnitude of the error. The spectral density of a self-similar process obeys a power law. Spectral analysis allows us to isolate the most informative components of the process under study by changing the dimension of the original feature space. The principal components are chosen so that they correspond to the greatest variability of the process. Other components can be considered as noise components. Wavelet analysis consists in determining the coefficients of the decomposition of the output signal into basis functions. Performing a wavelet transform allows us to isolate the most significant information by the level of the transform coefficients. For example, it allows us to isolate [1]: anomalies caused by errors in the settings of network equipment, as well as equipment malfunctions; network attacks of the "denial of service" class; network overloads, etc. By counting the number of points of maxima of the wavelet transform coefficients along the shift parameter in the scale domain, we can estimate the value of the Hurst parameter.

Statistical analysis. If the variance of the process decreases slowly, that is, the variance of the sample mean has a slower decline than the inverse of the sample length, then this process is self-similar. The tool of statistical analysis is the study of the distribution function. For self-similar processes with a stepwise (hyperbolic) decrease in the ACF, a distribution with a "heavy tail" is characteristic. The "tail heaviness" indicator (shape parameter) is determined by the Hill method or by constructing a graph of the additional distribution in a double logarithmic scale (LLCD). Most often, functions of subexponential distribution laws are used to approximate histograms of experimental data of self-similar processes: Pareto, Weibull, lognormal, etc. To check the adequacy of theoretical distributions to experimental data, the Kolmogorov and Pearson agreement criteria are used. To test the null hypothesis of independence and identity of the distribution of time series values, the BDS test was proposed, which also allows detecting the nonlinearity of the generating system, to distinguish random systems from deterministic chaos or from nonlinear stochastic systems [7].

Fractal analysis. The correlation dimension is determined from a time series with a known value of the correlation integral. The Hurst index N characterizes the degree of self-similarity of the process. To determine it, the following methods are used: the periodogram method, wavelet analysis, the method of aggregated variances and R/S analysis. This indicator indicates the presence of a trend or the randomness of the process, and also characterizes the evolution of the studied

process. If $0,5 < H < 1$, then this process is characterized by long-term memory and is persistent, if $0 < H < 0,5$, then this indicates the anti-persistence of the process. The value $H = 0,5$ is characteristic of Brownian motion with independent products and corresponds to random deviations of the process from the mean. To determine the mode of the generating process, its Kolmogorov entropy is estimated, which is equal to the sum of the senior Lyapunov exponents and characterizes the rate of loss of information about the state of the dynamic system in time. K-entropy is zero for regular motion, infinite for random systems, positive and bound for systems with dynamic chaos. Correlation entropy also allows us to determine an estimate of the interval of exact predictability of the process.

Multifractal analysis. Aggregated traffic from different sources on small time scales exhibits a multifractal nature, that is, it has a flexible scaling law. Multifractal analysis of detrended fluctuations (ADF), which is a type of variance analysis, is used to analyze the self-similarity properties and long-term dependence of non-stationary time series.

The proposed methodology for assessing network traffic characteristics based on the above approaches includes the following stages:

1. Time-frequency analysis: determination of the type of dependence (short-term or long-term) by ACF; construction and analysis of the power spectral density graph; analysis of the type of wavelet transforms (estimation of the Hurst index N).

2. Statistical analysis: analysis of the variance of the sample mean; construction of a histogram of the experimental distribution and determination of its adequacy to theoretical distributions using the Kolmogorov and Pearson agreement criteria; determination of the "tail heaviness" indicator of the distribution by the Hill method and/or the LLCD method; application of the BDS test to test the null hypothesis of independence and identity of the distribution of time series values and detection of nonlinear dependence.

3. Fractal analysis: calculation of the correlation dimension of the attractor; determination of the Hurst index; calculation of correlation entropy; calculation of correlation interval of predictability (depth of forecast) of the process.

4. Multifractal analysis: determination of scaling index; calculation of scaling exponent; calculation of Hölder index; determination of singularity spectrum.

Table 1

Network traffic feature vector table

Indicator	Traffic signal
ACF	PUS
Tail heaviness indicator	1,4
Hurst indicator	

- periodogram method	0,980
- aggregated variance method	0,888
- R/S analysis	0,913
Correlation dimension	2,956
Predictability interval	2,525
Hölder index	0,212

In this way, a vector of informative features (characteristics and parameters) of network self-similar traffic is formed for further use in IDS. Let us consider an example of identifying network traffic characteristics according to the proposed methodology. An analysis of the Ethernet network traffic signal [7] was performed, which was aggregated with a step of 5 s. The process that generates the experimental traffic signal has properties inherent in self-similar processes: hyperbolic decrease of the ACF with increasing time delay, a power law of its spectral density and a slowly decreasing dispersion. The signal has (Table 1) a distribution with a "heavy tail" (the shape index is 1.4), the histogram of the experimental distribution according to the Kolmogorov and Pearson agreement criteria corresponds to the Pareto distribution, the skeleton wavelet transform of the signal has a tree-like structure, and the Hurst exponent has a range of values [0.888; 0.980]. The process is also nonlinear according to the results of the BDS test. The hypotheses about the identity of the distribution of signal values were accepted with a probability of 95%. The traffic signal according to the proposed method revealed multifractal properties: its graphs of the scaling exponent and the Hölder exponent are curved lines that depend on the order value, and the spectrum of singularities is not a compact set.

4. Conclusions

A methodology for identifying the characteristics of self-similar ICN traffic has been substantiated, including time-frequency, statistical, fractal and multifractal analyses, which allows increasing the probability of detecting intrusions for IDS by increasing the reliability of informative signs of self-similar traffic. An experimental Ethernet network traffic signal has been analyzed, and its characteristics have been determined. Further research should be aimed at developing adequate predictive models of network self-similar traffic to increase the efficiency of operational detection of intrusions for IDS.

References

- [1] L. Diana, P. Dini, D. Paolini. "Overview on intrusion detection systems for computers networking security." *Computers* 14 (2025): 87. doi:10.3390/ computers14030087
- [2] M. Roshanaei. "Resilience at the core: critical infrastructure protection challenges, priorities and cybersecurity assessment strategies." *Journal of Computer and Communications* 9 (2021): 80-102. doi: 10.4236/jcc.2021.98006
- [3] S. V. Dovbesko, S. V. Tolyapa, Ya. V. Shestak. "Application of methods of data mining for the construction of attack detection systems." *Modern Information Security* 1.37 (2019): 6-15. doi: 10.31673/2409-7292.2019.010615

- [4] K. Swathi, G. Hima Bindu. "An automated intrusion detection system in IoT system using attention based deep bidirectional sparse auto encoder model." *Knowledge-Based Systems* 305.12 (2024): 112633. doi: 10.1016/j.knosys.2024.112633
- [5] O. A. Cmirnov, Y. O. Dreis, D. O. Danylenko. "Simulating model NIPDS for intrusion detection and prevention in telecommunication systems and networks." *Ukrainian Scientific Journal of Information Security* 20. 1(2014): 29-35. doi: 10.18372/2225-5036.20.6571
- [6] T. Radivilova, L. Kirichenko, M. Tawalbeh, A. Ilkov. "Detection of anomalies in the telecommunication traffic by statistical method." *Cybersecurity: Education, Science, Technique* 3.11 (2021): 183-194. doi: 10.28925/2663-4023.2021.11.183194
- [7] Korniienko V. I. (Ed). *Intellectual modeling of nonlinear dynamic processes in control systems, cybersecurity, and telecommunication*. Dnipro University of Technology. 2020. 536 p.

Methodology for OSINT analysis of geopolitical risks in critical open-source infrastructure

Ivan Parkhomenko¹, Larysa Myrutenko², Yurii Sokyran³

¹ Taras Shevchenko National University of Kyiv, Bohdana Havrylyshyna St. 24, Kyiv, Ukraine

² Taras Shevchenko National University of Kyiv, Bohdana Havrylyshyna St. 24, Kyiv, Ukraine

³ Taras Shevchenko National University of Kyiv, Bohdana Havrylyshyna St. 24, Kyiv, Ukraine

Abstract

This paper examines the problem of OSINT analysis of geopolitical risks in critical open-source software infrastructure through systematic mapping of geographical distribution of developers and supply chain vulnerabilities. We analyze existing approaches to software security assessment and identify gaps in accounting for geopolitical factors when researching critically important open-source projects using Open-Source Intelligence (OSINT) methods. A comprehensive OSINT methodology for geopolitical analysis is proposed that integrates geopolitical supply chain risk measurements and includes software repository mining methods, developer geographical attribution techniques through email domain analysis, commit timezone analysis, and public GitHub profiles using open-source information

Keywords

OSINT, geopolitical security, open-source software, supply chains, geographical distribution of developers, critical infrastructure.

1. Introduction

In the modern information landscape, open-source software (OSS) systems are becoming critically important elements of digital infrastructure, serving billions of users worldwide [1]. Moreover, the dynamics of growing dependence on systems such as Linux kernel, OpenSSL, curl, is only increasing, creating new challenges for national security and the stability of the global digital space [2].

One of the problems of the current stage of open-source software development is the geographical concentration of major developers and maintainers of critically important projects. According to GitHub Octoverse 2024, the United States continues to lead in the number of developers, while India is rapidly approaching first place with a forecast of achieving leadership by 2028 [3]. Such concentration creates potential risks for software supply chains, especially in the context of geopolitical tensions and economic sanctions.

Given these circumstances, there is a need for systematic research on the geographic distribution of contributions to critical open-source infrastructure and analysis of potential risks associated with dependence on specific regions or corporate structures. Traditional approaches to software security analysis focus primarily on technical vulnerabilities, leaving geopolitical aspects of developing and maintaining critically important systems out of consideration [4].

Considering the above, it appears appropriate to develop a comprehensive approach to analyzing the geographic security of critical open-source infrastructure that will allow identification of potential supply chain vulnerabilities at the level of national and regional dependencies, as well as provide the ability to assess risks to the digital sovereignty of states.

2. Analysis of recent research and publications

In the field of geopolitical security of open-source software, several key research directions can be identified that examine various aspects of the influence of geographical and political factors on the development of critically important technologies.

Research by the Atlantic Council emphasizes that open-source code has become critically important infrastructure that requires government regulation and protection [5]. According to the analysis, 34% of open-source software contributions originate from the United States, 13% from Russia, creating geographical concentration of risks in critically important projects.

The French Institute of International Relations Ifri in its research analyzes the economic and geopolitical consequences of dependence on open-source software, noting that state participation in open-source development is becoming increasingly politicized and serves national security and digital sovereignty ambitions [6].

Huntedlabs research also notes that Russian developers from VK (a company under sanctions for ties to Russian intelligence services) control the easyjson package, which is used in critical components such as Kubernetes, Helm, and Istio [7].

3. Article objective

The objective of the article is to develop a methodology for geopolitical analysis of critical open-source software infrastructure security through systematic research of geographical distribution of developers, corporate affiliations which ensures identification of potential supply chain risks and formation of recommendations for improving digital infrastructure resilience to geopolitical threats.

4. Research methodology

The research is based on a comprehensive approach to data collection and analysis from critical open-source software repositories, combining Mining Software Repositories (MSR) methods with developer geographical attribution techniques and supply chain network analysis.

To ensure research representativeness, four critically important open-source software projects were selected: Linux kernel, OpenSSL, curl, and Node.js. These projects are characterized by high usage levels in critical infrastructure, significant dependencies from other systems, and active developer communities [8].

Metadata collection is performed using the GitHub REST API and direct analysis of git logs from cloned repositories. The GitHub API provides access to contribution statistics, developer information, and dependency structures [9].

Geographical attribution of developers is performed using three primary methods. Email domain analysis allows identification of country and corporate affiliation through matching domains with DNS record geolocation databases [10]. Commit timezone analysis provides approximate determination of geographical regions through the `--date=raw` parameter of the git log command, which provides information about UTC offset [11]. GitHub public profile analysis supplements geographical data through the API endpoint `/users/{username}`, which may contain information about the developer's location.

Building networks of dependencies between projects and developers is performed using graph theory, where vertices represent projects or developers, and edges represent connections between them in the form of contributions or dependencies [12].

5. Geopolitical risk assessment metrics

For quantitative assessment of geopolitical risks in critical open-source infrastructure, a system of metrics has been developed based on geopolitical supply chain risk indices.

The modified Herfindahl-Hirschman index for geographical concentration of contributions is calculated by grouping developers by countries and calculating the percentage distribution of their contributions:

$$HHI = \sum_{i=1}^n \left(\frac{P_i}{100} \right)^2 \times 10000 \quad (1)$$

where P_i is the percentage contribution of the i -th country.

Concentration risk is calculated using the formula:

$$R_{concentration} = \begin{cases} \min(100, (C_{top3} - 50) \times 2, \wedge C_{top3} > 50) \\ 0, \wedge C_{top3} \leq 50 \end{cases} \quad (2)$$

where C_{top3} is the sum of percentage contributions from the three countries with the largest contributions.

50% for the three leading countries is defined as the critical threshold.

The corporate dominance assessment process distinguishes between corporate and individual developers, calculating their relative shares in the total volume of contributions:

$$S_{corp} = \frac{C_{corp}}{C_{total}} \times 100 \quad (3)$$

where C_{corp} is corporate contributions, C_{total} is total contributions.

$$R_{corp} = \begin{cases} \min(100, (S_{corp} - 40) \times 2.5, \wedge S_{corp} > 50) \\ 0, \wedge S_{corp} \leq 50 \end{cases} \quad (4)$$

The corporate dominance threshold level is set at 40%.

The system also evaluates data quality through analysis of developer origin determination levels:

$$R_{uncertainty} = \min(100, U_{unknown} \times 2 + U_{low} \times 1.5) \quad (5)$$

where $U_{unknown}$ is the percentage of developers with undetermined location and timezone characteristics, U_{low} is the percentage of developers with low confidence in the result. Confidence depends on the information extraction method: direct matches of country names and corporate domains - high confidence, partial matches in location and national top-level domains (TLD) -

medium, typical assumptions for generic TLDs - low, and if determination failed - confidence unknown.

The diversification assessment methodology focuses on analyzing the top 10% most active developers:

$$R_{diversity} = \max\left(0, 1 - \frac{T_{top}}{T_{total}} \times 100\right) \quad (6)$$

where T_{top} is the countries of the top 10% most active developers, T_{total} is the total number of countries.

The overall risk score is formed as a weighted combination of components:

$$R_{overall} = \sum_{i=1}^n R_i \times w_i \quad (7)$$

where R_i is the risk component, w_i is the component weight.

The weights are set as: geographical concentration (0.3), corporate dominance (0.25), uncertainty (0.15), diversification (0.1).

6. Conclusions

This work examines the problem of geopolitical security of critical open-source software infrastructure through the lens of geographical distribution of developers, corporate affiliations, and supply chain risks. Recent research and publications regarding geopolitical factors in the analysis of critically important open-source projects have been analyzed.

It is noted that significant geographical concentration creates systemic risks, especially in the context of escalating geopolitical tensions in 2024, when the number of attacks on software supply chains doubled.

Promising directions for further research include the development of automated real-time geopolitical risk monitoring systems, creation of international standards for software geographic security assessment, and investigation of the impact of artificial intelligence on the transformation of open-source supply chains under conditions of geopolitical instability.

Declaration on Generative AI

Generative AI was used for citation management, improving writing style, grammar, and spell checking.

References

- [1] Linux Foundation, 2020 FOSS Contributor Survey, 2020. URL: <https://www.linuxfoundation.org/resources/publications/foss-contributor-2020>.
- [2] CISA, Open-Source Security, 2024. URL: <https://www.cisa.gov/opensource>.
- [3] GitHub, Octoverse 2024: The state of open source, 2024. URL: <https://octoverse.github.com/>.

- [4] Institut français des relations internationales (Ifri), Software Power: The Economic and Geopolitical Implications of Open-Source Software, 2024. URL: <https://www.ifri.org/en/studies/software-power-economic-and-geopolitical-implications-open-source-software>.
- [5] Atlantic Council, Avoiding the success trap: Toward policy for open-source software as infrastructure, 2024. URL: <https://www.atlanticcouncil.org/in-depth-research-reports/report/open-source-software-as-infrastructure/>.
- [6] Institut français des relations internationales (Ifri), Software Power: The Economic and Geopolitical Implications of Open-Source Software, 2024. URL: <https://www.ifri.org/en/publications/etudes-de-lifri/software-power-economic-and-geopolitical-implications-open-source>.
- [7] Huntred Labs, The Russian open-source project that we can't live without, 2024. URL: <https://huntedlabs.com/the-russian-open-source-project-that-we-cant-live-without/>.
- [8] CISA, CISA Open-Source Software Security Roadmap, 2024. URL: <https://www.cisa.gov/resources-tools/resources/cisa-open-source-software-security-roadmap>.
- [9] GitHub, REST API endpoints for repository statistics, 2024. URL: <https://docs.github.com/en/rest/metrics/statistics>.
- [10] WhoisXML API, IP Geolocation, 2024. URL: <https://ip-geolocation.whoisxmlapi.com/>.
- [11] Opensource.com, Using Git and mailing lists time zones to find out where developers live, 2015. URL: <https://opensource.com/business/15/11/git-mailing-list-time-zone-analysis>.
- [12] VulnCheck, Understanding Software Dependency Graphs, 2024. URL: <https://www.vulncheck.com/blog/understanding-software-dependency-graphs>.

Practical Exploitation of HDD Vibrations for Acoustic Eavesdropping

Artur Pelcharskyi^{1,*}, Maksym Buleshnyi^{1,*}, Mykhailo Buleshnyi^{1,*}, Yuliia Moliashcha^{1,*}, Oleg Farenjuk^{1,2}

¹ Ukrainian Catholic University

² Yukhnovskii Institute for Condensed Matter Physics of the National Academy of Sciences of Ukraine

Abstract

In the digital era, information security has become a critical concern for both individuals and organizations. As technological capabilities advance, so do the methods by which sensitive data can be compromised. Beyond conventional attack vectors, new risks may emerge from devices not originally intended for surveillance. This work investigates the feasibility of exploiting hard disk drives (HDD) as acoustic sensors by accessing their internal Position Error Signal (PES). We demonstrate that the PES, designed for precise head alignment, can capture external sound vibrations with sampling properties comparable to those of standard audio equipment. These findings highlight HDDs as a potential side-channel threat, raising important considerations for cybersecurity and the protection of information in secure environments.

Keywords

side-channel attack, acoustic eavesdropping, hard disk drive (HDD), position error signal (PES)

1. Introduction

Hard Disk Drives (HDD) remain one of the most widespread data storage solutions. However, their mechanical design introduces potential vulnerabilities that may be exploited in unexpected ways. While it is well documented that HDDs can be disrupted by acoustic resonances and specific sound frequencies [1], a related and more provocative question arises: *can HDDs function as microphones?*

Several studies have explored this idea. For instance, Kwong et al. [2] demonstrated that the PES, a feedback signal used to maintain precise head positioning, can be treated as an acoustic transducer. In their experiments, the Shazam service successfully recognized a song recorded through an HDD, even though the signal appeared as noise to the human ear. Building on this concept, our work further explores the principles, sampling properties, and experimental validation of using PES as a proxy for acoustic sensing.

2. Background

2.1. HDD structure

The primary components of an HDD are the platters and the read/write (r/w) head. Data is stored on rotating magnetic platters, typically spinning at 5400 or 7200 RPM. The r/w head hovers nanometers above the platter surface, stabilized by the air cushion generated during high-speed rotation.

*These authors contributed equally.

✉ pelcharskyi.pn@ucu.edu.ua (A. Pelcharskyi); maksym.buleshnyi.pn@ucu.edu.ua (M. Buleshnyi); buleshnyi.pn@ucu.edu.ua (M. Buleshnyi); moliashcha.pn@ucu.edu.ua (Y. Moliashcha); indrekis@ucu.edu.ua (O. Farenjuk)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Accurate head positioning is critical, as only an approximately 7 nm tolerance is permissible. When deviations occur, the HDD generates a Position Error Signal (PES), which quantifies the head's displacement from the intended track. This signal is produced by special positioning information embedded in servo sectors, which are distributed at regular intervals along every track.

All the servo sectors that are aligned radially from the center to the edge of the platter form a logical, pie-slice-shaped section known as a wedge. The PES data from these structures enables the servo system to maintain precise alignment. Figure 2 1, illustrates the servo system and its structural components.

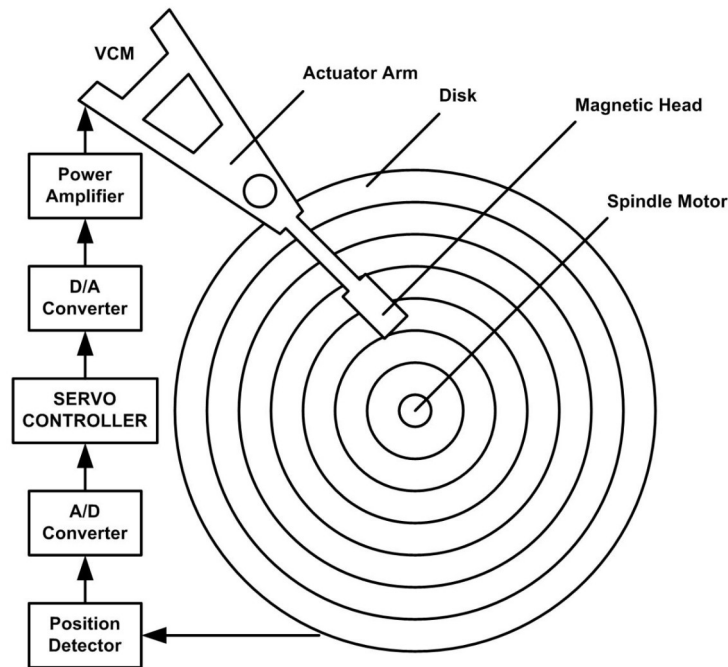


Figure 2 1: HDD head-positioning control system. The PES provides feedback to the controller, which adjusts the Voice Coil Motor (VCM) to correct the head's position.

2.2. HDD as Microphone

This section highlights the analogy between traditional acoustic sensing devices and the unintended sensing properties of HDDs, emphasizing the mechanical principles that enable this phenomenon.

A conventional carbon microphone consists of four main components: a power source, an output mechanism, a diaphragm, and a carbon element. Acoustic waves generate fluctuations in air pressure, causing the diaphragm to oscillate. These oscillations alter the resistance of the carbon element, producing a voltage proportional to the sound wave. In essence, the microphone transduces air pressure variations into an electrical signal that represents the acoustic waveform.

Within an HDD, the read/write head is the most sensitive component, designed to maintain sub-nanometer precision when following data tracks. Deviations from the track center may arise from internal mechanical dynamics or external disturbances, including thermal effects and vibrations. To correct these deviations, the HDD continuously monitors the head's displacement using the Position Error Signal (PES), which serves as feedback in a closed-loop control system (Figure 2 1).

From a functional perspective, the PES plays a role analogous to the diaphragm in a microphone: external acoustic waves induce vibrations that displace the read/write head, and these displacements are captured by the PES. Although HDDs were never designed to operate as acoustic sensors, their internal mechanics inherently respond to sound waves. As a result, the PES can be considered an unintended transducer of external acoustic energy.

3. Approach

3.1. Access to PES

The PES is not normally exposed to end users, as it is reserved for internal control, so accessing it requires low-level methods. We investigated two primary approaches: direct hardware probing and firmware-level commands.

The hardware approach involves identifying the AMUX pin on the HDD's circuit board, which outputs the raw PES data in real time. This first requires enabling the signal via the Seagate F3 terminal, a diagnostic serial interface. While we successfully connected to the terminal and calculated the expected PES frequency (26.4 kHz), we were unable to isolate the AMUX pin on our Seagate Barracuda 7200.9 model, as its location was not documented and could not be determined through oscilloscope probing.

The firmware command approach proved successful. Using the F3 terminal, we executed the U100D command, which provides PES statistics for each wedge on the disk platter. Instead of a raw data stream, this command outputs the minimum, average, and maximum PES deviation values for each wedge, which served as the data source for our experiment.

3.2. Sampling Properties of PES

The PES is updated whenever the r/w head passes a servo sector. For a drive with $N_{sectors}$ sectors per track rotating at RPM , the sampling frequency is:

$$f_s = N_{sectors} \cdot \left(\frac{RPM}{60} \right), \quad (1)$$

where $N_{sectors}$ is the number of servo sectors per track.

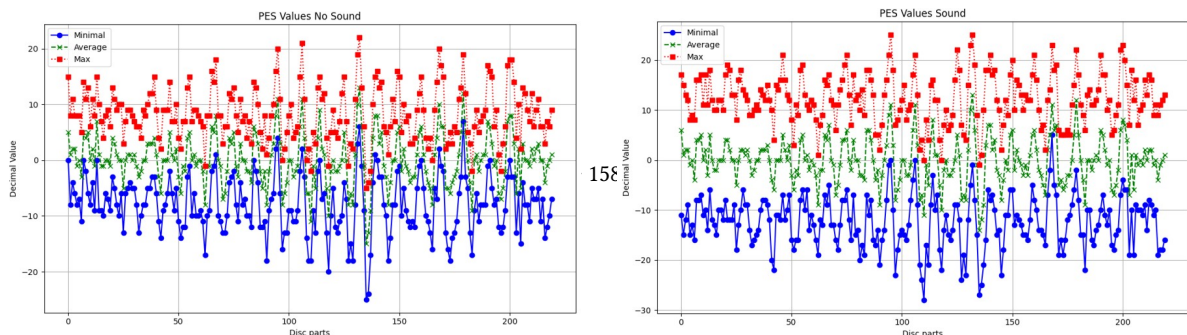
For the Seagate Barracuda 7200.9 used in our work, ($N_{sectors} = 220$, $RPM = 7200$):

$$f_s \approx 26.4 \text{ kHz}, \quad (2)$$

which, according to the Nyquist–Shannon theorem, allows representation of frequencies up to 13.2 kHz — sufficient to capture most of the human-audible spectrum.

3.3. Experiment results

Experiments were conducted in a controlled environment, with minimal external vibrations. A portable speaker was placed 10 cm above the HDD, emitting sound at approximately 80 dB. The PES statistics were captured continuously using the U100D F3 terminal command, both in silence and during acoustic exposure.



(a) PES values in a quiet environment.

(b) PES values during exposure to an 80 dB sound source.

Figure 3 1: PES value in a completely quiet environment vs direct sound exposure.

The results, shown in Figure Figure 3 1, demonstrate a clear correlation between sound and PES deviation. While the average PES value remained relatively stable, the range between the minimum and maximum values increased substantially under acoustic stimulation. This widening range indicates greater head deflection from the track center, directly validating that the PES can be used to sense external sound waves.

4. Discussion and Future Work

Our experiments demonstrate that HDD PES can indeed be influenced by external acoustic signals. However, several limitations must be addressed. The effect is most pronounced under relatively loud sound sources positioned close to the disk. Additionally, the granularity of PES values provided by the Seagate interface (min/avg/max) limits signal fidelity by averaging out the fine-grained temporal details of the acoustic waveform.

Future work should focus on overcoming these limitations and further exploring the vulnerability. Expanding access to raw PES data, potentially via hardware probing or alternative firmware, may improve signal resolution. A broader investigation across different HDD models will be valuable for understanding the universality of this effect. Furthermore, exploring multi-disk setups could enable the combination of PES data from several drives to improve sensitivity and robustness. Finally, future research should prioritize developing countermeasures to mitigate this side-channel leakage.

5. Conclusion

This work explored the unconventional use of hard disk drives as acoustic sensors by analyzing their Position Error Signal. Our experiments confirm that external sound sources can measurably influence the PES, and that the effective sampling frequency of modern HDDs is sufficient to capture much of the human-audible spectrum.

From a cybersecurity perspective, these findings highlight a non-traditional side channel: HDDs, originally intended solely for data storage, can inadvertently function as microphones. This introduces new risks in environments where sensitive conversations and critical information must be protected. Although practical exploitation requires close proximity and favorable conditions, the principle itself demonstrates that everyday hardware may become an unexpected surveillance tool. Recognizing and addressing these unconventional vulnerabilities is essential to maintaining trust in widely deployed storage technologies.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and Grammarly in order to: Grammar and spelling check. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] M. Shahradd, A. Mosenia, L. Song, M. Chiang, D. Wentzlaff, P. Mittal, Acoustic denial of service attacks on HDDs, arXiv preprint arXiv:1712.07816 (2017). URL: <https://arxiv.org/abs/1712.07816>. doi:10.48550/arXiv.1712.07816.
- [2] A. Kwong, W. Xu, K. Fu, Hard drive of hearing: Disks that eavesdrop with a synthesized microphone, in: 2019 IEEE symposium on security and privacy (SP), IEEE, 2019, pp. 905–919.

Real-Time Acoustic Impulse Response Generation Using Static Ray Tracing for Anomaly Detection

Oleksandr Terletskyi, Valeriy Trushevskyy

Ivan Franko National University of Lviv, Universytetska 1, 79000 Lviv, Ukraine

Abstract

This paper presents a real-time method for simulating sound propagation using ray tracing techniques, with applications in cybersecurity for detecting and analyzing acoustic threats. By modeling acoustic wave behavior in virtual environments, this approach enables accurate simulation of reflections, diffractions, and reverberations, which can be leveraged for tasks such as audio-based anomaly detection and spatial audio analysis in secure facilities. The integration of impulse response generation with real-time convolution processing allows for precise acoustic profiling, dynamically responding to changes in the environment to enhance situational awareness. This method provides a foundation for advanced acoustic monitoring systems, supporting applications ranging from secure communications to threat detection. Results demonstrate the method's ability to simulate complex acoustic interactions with high fidelity and responsiveness, making it a valuable tool for real-time audio surveillance and cybersecurity solutions.

Keywords

Sound propagation, ray tracing, convolution, impulse response, real-time audio, spatial audio

1. Introduction

Realistic sound propagation is essential for immersive environments in virtual reality, gaming, architecture, and cybersecurity. Acoustic behavior depends on reflections, diffractions, and reverberations, which traditional methods cannot adapt to in real time. This study presents a real-time simulation method that uses ray tracing to model sound paths and FFT based convolution to generate impulse responses [1]. The approach supports cybersecurity applications such as acoustic anomaly detection and source localization by accurately reproducing sound behavior in complex spaces. To ensure stability, the system employs static rays, reducing response fluctuations and audio artifacts. A running index mechanism updates impulse responses continuously, maintaining smooth playback as environmental data changes. This method provides a stable and adaptive solution for high-quality sound simulation and real-time acoustic analysis in both digital and security contexts

2. Static Ray Tracing Approach

Accurate impulse response generation is essential for simulating sound propagation and identifying acoustic patterns in areas such as cybersecurity, where it supports surveillance, anomaly detection, and secure communication [2]. The proposed method models sound using ray tracing, where sound rays travel through the environment, reflect from surfaces, and return to the listener. Each ray carries information on distance, direction, and energy decay, forming a time-dependent impulse response (IR) that represents the environment's acoustic signature. To maintain

²²Corresponding authors

✉ oleksandr.terletskyi@lnu.edu.ua (O. Terletskyi); valeriy.trushevsky@lnu.edu.ua (V. Trushevskyy)

ORCID 0009-0001-7489-4428 (O. Terletskyi); 0000-0001-8672-7152 (V. Trushevskyy)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

stability, static rays are used—each frame reuses the same set of directions. This avoids random fluctuations in the impulse response and ensures consistent data for real-time analysis. A running index updates the impulse response incrementally, dividing it into small segments for efficient processing. This structure supports long impulse responses and enables smooth adaptation to environmental changes with minimal computation.

2.1. Impulse Response

The impulse response $h(t)$ describes how an environment reacts to a short sound, defined by the convolution of the input signal $x(t)$ with $h(t)$ [3]:

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau) h(t - \tau) d\tau, \quad (1)$$

where $y(t)$ is the output signal, $x(t)$ is the input signal, $h(t)$ is the system's impulse response, and $*$ denotes the convolution operator, or in discrete form:

$$y[n] = \sum_{k=0}^{N-1} x[k] \cdot h[n-k], \quad (2)$$

where $y(t)$ (or $y[n]$) is the output signal, $x(t)$ (or $x[k]$) is the input signal, $h(t)$ (or $h[n]$) is the impulse response, and $*$ denotes convolution. Each ray contributes to $h(t)$ based on its path length and energy loss, and summing all contributions forms an accurate model of the space's acoustic behavior.

2.2. Static Ray Tracing

Ray tracing simulates sound by following rays as they reflect and attenuate [1]. By introducing static rays with fixed directions across frames, the impulse response can be kept stable without noise artifacts. This precision is vital in acoustic monitoring and intrusion detection, where slight deviations can indicate security threats. Each ray originates from the sound source, reflects according to the surface normal, and loses energy over distance. Rays reaching the listener contribute to the impulse response with recorded travel time and intensity. The resulting data captures the environment's spatial characteristics for real-time anomaly detection.

2.3. Listener-Directed Diffuse Rain

The Listener-Directed Diffuse Rain model improves diffusion efficiency by sending rays toward the listener instead of scattering them randomly [4]. The energy loss is determined by the alignment between the incoming ray and listener vector:

$$E_{\text{loss}} = \max\left(\left(d \cdot \frac{l - p}{\|l - p\|}\right)^2, 0\right), \quad (3)$$

where $\mathbf{d}$ is the ray direction, $\mathbf{p}$ the hit point, and $\mathbf{l}$ the listener position. The adjusted energy and direction are:

$$E_{\text{adjusted}} = E_{\text{original}} \times E_{\text{loss}}, \quad (4)$$

The direction of the adjusted ray is then normalized towards the listener:

$$d_{\text{new}} = \frac{l-p}{\|l-p\|}, \quad (5)$$

where $\mathbf{y}(\mathbf{t})$ (or $\mathbf{y}[\mathbf{n}]$) is the output signal, $\mathbf{x}(\mathbf{t})$ (or $\mathbf{x}[\mathbf{k}]$) is the input signal, $\mathbf{h}(\mathbf{t})$ (or $\mathbf{h}[\mathbf{n}]$) is the impulse response, and $*$ denotes convolution. Each ray contributes to $\mathbf{h}(\mathbf{t})$ based on its path length and energy loss, and summing all contributions forms an accurate model of the space's acoustic behavior.

2.4. Impulse Response Generation

Each ray adds energy to the impulse response based on travel distance and attenuation. Arrival time:

$$t = \frac{d}{c}, \quad (6)$$

where d is the total distance traveled by the ray and c is speed of sound. As sound travels, it loses energy.

$$E_d = \frac{1}{d^2 + \epsilon}, \quad (7)$$

where ϵ is a small constant to avoid division by zero. The amplitude contribution of a ray is:

$$A = E_r \times E_d \times s, \quad (8)$$

where E_r is the ray's remaining energy and s a phase-shift factor. After summing all rays, the impulse response is normalized to fit a standard range. The resulting $h(t)$ encodes the environment's acoustic behavior and can be convolved with sound for realistic and responsive spatial audio.

3. Auralization

Auralization is the process of converting simulated acoustic data into audible sound, allowing listeners to perceive an environment's acoustic characteristics. It relies on the impulse response (IR), which describes how a space modifies sound through reflections and reverberation. The original signal $x(t)$ is convolved with the IR $h(t)$ to produce the auralized output $y(t)$:

$$y(t) = x(t) * h(t), \quad (9)$$

where $*$ denotes convolution. This operation embeds the environment's time and frequency features into the sound, recreating realistic spatial effects

3.1. Running Index Method

In real-time simulation, impulse responses must adapt continuously to moving sources and changing environments. Recomputing them each frame through multiple FFTs is computationally expensive. The running index method reduces this cost by dividing the impulse response into

smaller segments and updating only a few per frame. A running index cycles through these segments, gradually refreshing the IR while maintaining stable sound quality. This incremental update approach balances accuracy and performance, ensuring smooth, uninterrupted auralization in real-time applications

4. Results

The simulation was conducted in a $50 \times 50 \times 20$ m room with a source at the center and a listener positioned 1 m to the right. The room was perfectly reverberant, and the impulse response was recorded over 3 seconds at 41 kHz (123,000 samples). This setup was used to evaluate how the number of rays and reflections affects reverberation accuracy and density.

4.1. Number Of Rays

Impulse responses were generated using 10, 100, and 1000 rays, each with 8 reflections. As shown in Figure 1, increasing the ray count improves reflection density and the smoothness of decay over time. With 10 rays, the response is sparse and dominated by early reflections. Using 100 rays adds mid and late reflections, producing a more natural reverberant tail. At 1000 rays, the response becomes dense and continuous, accurately modeling the diffuse reverberant field. A higher ray count therefore yields a more complete representation of acoustic energy decay, essential for realistic spatial audio and precise acoustic analysis.

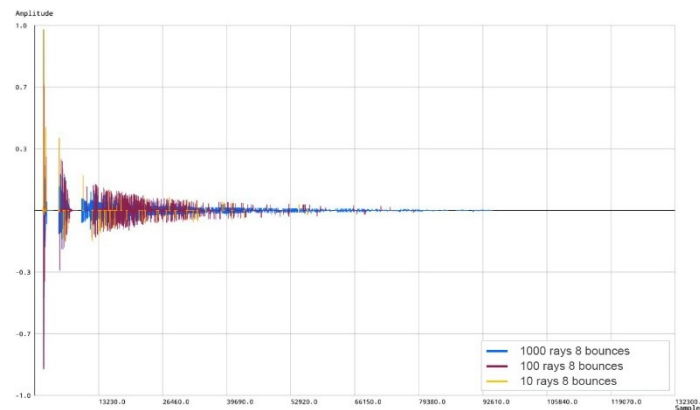


Figure 1: Impulse responses with 10, 100, and 1000 rays (8 reflections).

4.2. Number Of Rays

Figure 2 compares impulse responses for 2, 5, and 8 reflections using 1000 rays. With 2 bounces, only early reflections are present, and the decay is short. Increasing to 5 bounces extends the reverberant tail and improves density. At 8 bounces, late reflections fill the decay profile, approaching a continuous reverberant field. More bounces capture additional sound interactions, enriching realism and supporting applications that require accurate modeling of reflection persistence.

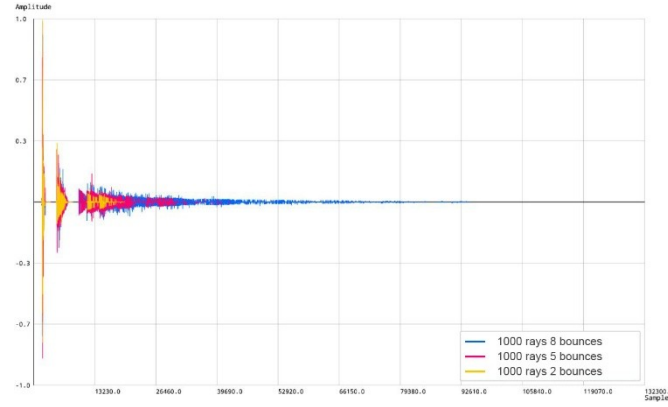


Figure 2: Impulse responses with 1000 rays and 2, 5, and 8 reflections.

4.3. Comparison with Sabine’s Formula

The reverberation time T_{60} predicted by Sabine’s formula [6]:

$$T_{60} = \frac{0.161 \times V}{A \times \alpha}, \quad (10)$$

where V is the room volume, A is the total surface area, and α is the average absorption coefficient. * denotes convolution was compared with simulated results for 2–10 bounces. For the given room ($\alpha = 0.1$), the theoretical value is approximately 8.05 s. As shown in Figure 3, higher bounce counts produce decay times increasingly close to the theoretical T_{60} . Around 8–10 bounces, the simulated results align closely with Sabine’s prediction, confirming that additional reflections improve accuracy in modeling highly reverberant.

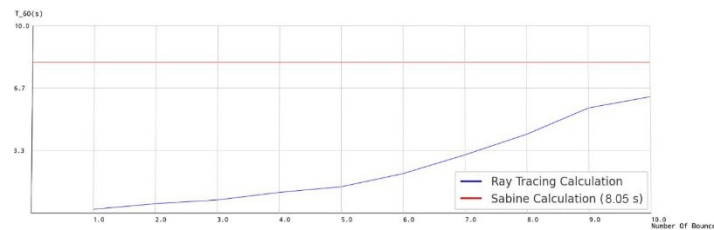


Figure 3: Simulated T_{60} versus Sabine’s theoretical estimate

5. Conclusions

The results demonstrate that increasing both ray count and reflection depth enhances the precision and realism of impulse response simulations. Higher ray counts yield denser reverberation fields, while additional bounces extend the decay and capture complex reflections. These factors are critical not only for immersive spatial sound but also for acoustic analysis in cybersecurity, where detailed impulse responses can reveal subtle environmental changes or anomalies. The alignment with Sabine’s theoretical reverberation time validates the simulation’s accuracy. Such methods can improve acoustic monitoring, secure communication testing, and environmental profiling by providing reliable real-time detection of abnormal sound behavior. Overall, the static-ray approach combined with the running index method delivers stable, efficient, and physically consistent acoustic simulation suitable for both immersive and security-focused applications.

References

- [1] A. Krokstad, S. Strom, and S. Sorsdal, “Calculating the acoustical room response by the use of a ray tracing technique,” *J. Sound Vib.*, vol. 8, 1968, pp. 118–125.
- [2] E. Ouellet-Delorme, H. Venkatesan, E. Durand, and N. Bouillot, “Live ray tracing and auralization of 3D audio scenes with vaRays,” in *Proc. 18th Sound Music Comput. Conf.*, 2021, pp. 1–10.
- [3] B. A. Farina, “Convolution of anechoic music with binaural impulse responses,” Presented at the 15th Audio Engineering Society Conference on Audio in Large Spaces, Copenhagen, Denmark, 1993.
- [4] S. Pelzer, D. Schröder, and M. Vorlaender, “The number of necessary rays in simulations based on geometrical acoustics using the diffuse ray technique,” in *Proc. DAGA*, 2011.
- [5] Meta, “Acoustic ray tracing in Meta XR Audio SDK,” 2023. URL: <https://developers.meta.com/horizon/documentation/unity/meta-xr-acoustic-ray-tracingunity-overview/>
- [6] L. Beranek, “Analysis of Sabine and Eyring equations and their application to concert hall audience and chair absorption,” *J. Acoust. Soc. Am.*, vol. 120, 2006, pp. 1399–1410.

Phishing URL recognition based on fuzzy clustering with global optimization

Iryna Udovyyk¹, Olha Shevtsova¹ and Maksim Fesenko¹

¹ Dnipro University of Technology, Dnipro, Dmytra Yavornytskoho, 19, 49005, Dnipro, Ukraine

Abstract

An algorithm for detecting phishing URLs (classifier) using fuzzy clustering is proposed, which includes the selection of the type of intelligent classifier and justification of its parameters using global optimization methods. The effectiveness of the proposed algorithm is assessed by modeling using experimental data - a set of phishing and safe URLs.

Keywords

Confidential data, classifier, fuzzy C-means clustering, phishing attacks, global optimization, subtractive clustering ²³

1. Introduction

Phishing URLs are malicious websites that disguise themselves as safe ones to obtain sensitive data, such as credit card numbers, login username and password information, etc. Phishing uses social engineering and technical deception to obtain private information from a web user. Often, the landing page of a phishing website also attempts to infiltrate the visitor's computer and install malware. Thus, phishing attacks pose a significant risk to both individuals and organizations, as they pose a threat to the disclosure or acquisition of sensitive personal and corporate information [1].

2. Analysis of recent research and publications

In works [2-3], the use of fuzzy methods for identifying phishing websites is proposed. In this work [3], it is noted that the main problem of using data mining algorithms is the selection of a set of functions to be used. In work [4], the use of an adaptive neuro-fuzzy system Anfis is proposed to solve this problem, in works [5-6] NM for detecting phishing e-mail messages. Thus, the unsolved problem is the selection of the type of intelligent classifier and the justification of its parameters for detecting phishing URLs. The purpose of the article is to study fuzzy clustering algorithms for detecting phishing URLs with the adjustment of the parameters of fuzzy classifiers using global optimization methods.

3. Problem formulation

Fuzzy logic algorithms are based on the statement that the membership function of an element of a set can take values in the interval [0, 1]. The advantages of such algorithms are the transparency of the process of obtaining conclusions based on a verbal description of expert knowledge about the process, as well as noise resistance. The disadvantages are the lack of

²³ Corresponding authors

✉ udovyyk.i.m@nmu.edu.one (I. Udovyyk); shevtsova.o.s@nmu.one (O. Shevtsova); fesenko.m.i@nmu.one (M. Fesenko)

🆔 0000-0002-5190-841X (I. Udovyyk); 0000-0002-0148-5877 (O. Shevtsova); 0009-0003-9747-9627 (M. Fesenko)

automatic knowledge acquisition, as well as a limited number of input variables [7, 8]. The task of clustering is to combine a set of objects into groups (clusters) based on the similarity of features for objects of the same group and the differences between them. Most clustering algorithms can be used in conditions of almost complete absence of information about the laws of data distribution. The input information for clustering is the observation matrix Y [8]. The most well-known fuzzy clustering methods are: Subtractive Clustering and Fuzzy C-means clustering [7, 10]. The subtractive clustering method assumes that each experimental point can be the center of a cluster, which determines how far from the center of the cluster its elements can be, and this can significantly affect the result. In the C-means algorithm, the parameter that can significantly affect the result is the number of clusters. Choosing the number of clusters for real problems without sufficient a priori information about the data structures is quite difficult. Since the problem of choosing a fuzzy classifier with tuning its parameters is polymodal, this requires the use of global optimization methods, among which search methods are the most effective. In forward random search (FRS) algorithms, search directions are specified and the values of a functional at the search points are determined. The solution is to choose a step in the direction of decreasing this functional. Since the problem of choosing a fuzzy classifier with tuning its parameters is polymodal, this requires the use of global optimization methods, among which search methods are the most effective. In forward random search (FRS) algorithms, search directions are specified and the values of a functional at the search points are determined. The solution is to take a step in the direction of reducing this functional. The development of the FRS method is the simulated annealing method (SAN), which reflects the behavior of a molten material during solidification using a controlled cooling (annealing) procedure. The threshold acceptance method (TAM) uses an approach like SAN, but this method selects all new points below a set threshold. The development of search methods is evolutionary algorithms, among which the most common are genetic algorithms (GAs), which simulate the development of a biological population at the genome level [7, 8]. Multi-criteria optimization (MO) is based on finding a solution that simultaneously optimizes more than one function. In this case, a certain compromise is sought, the role of which is played by a solution that is optimal in the Pareto sense [7]. The combined criterion [9-11] the criterion for global optimization was used:

$$C_{comb} = 0,2 \cdot C_{reg} + 0,8 \cdot C_{bias}, \quad (1)$$

where C_{reg} is regularity criterion calculated on test sample B.

$$C_{reg} = \frac{\|Y_B^* - \hat{Y}_B\|}{\|Y_B^*\|}. \quad (2)$$

C_{bias} is unbiasedness criterion (minimum bias), based on the analysis of decisions on training A and test B samples.

$$C_{bias} = \frac{\|\hat{Y}_A - \hat{Y}_B\|}{\|Y^*\|}, \quad (3)$$

where $\hat{Y}$ and Y^* are estimate and actual input observation matrix.

The simulation of the process of recognizing phishing URLs was conducted in the Matlab environment using the developed programs. For the simulation, a data set of 150 phishing and 150 safe URLs was used, which included information about the domain name registrar, domain lifetime, geolocation of the hosting server, and the presence of a secure connection with a valid certificate. To collect phishing URLs, a database of validated phishing links on Phishtank.org was used, where both antivirus companies and individual researchers upload new link samples. To collect safe URLs, the Alexa.com service was used, which contains addresses of popular websites. In this case, the entire data set was divided into training A and test B samples. GAs, FRS, SAN, TAM were used as global optimization methods. Criterion (1) was used as a global optimization criterion. The following metrics were used to evaluate the classification of URL addresses:

- TP (True Positive) – the number of safe links that were classified as safe;
- TN (True Negative) – the number of phishing links that were classified as phishing;
- FP (False Positive) – the number of safe links that were classified as phishing;
- FN (False Negative) – the number of phishing links that were classified as safe;

And their derivatives:

- True positive rate ($TPR = TP / (TP + FN)$);
- True negative rate ($TNR = TN / (TN + FP)$);
- False positive rate ($FPR = FP / (FP + TN)$); $FPR = FP / (FP + TN)$);
- False negative rate ($FNR = FN / (FN + TP)$); $FNR = FN / (FN + TP)$);
- Positive predictive value ($PPV = (TP + TN) / (TP + TN + FP + FN)$);
- Negative predictive value ($NPV = TN / (TN + FN)$).

In the context of the task of recognizing phishing addresses, the most important metric is the FN (or FNR), which provides a quantitative assessment of undetected attacks. FRS had an adaptive search step and a full search around the current iteration, SAN and TAN – a limited reannealing region, GAs – single-point crossover, selective selection of parents, formation of a new population with displacement. The number of iterations for FRS, TAN and SAN (for GAs generations) was limited to 100, and the size of the search space for FRS (for GAs the population size, for TAN and SAN the reannealing region size) was 30. The following classifier parameters were varied during global optimization:

- algorithm type – Genfis2 or Genfis3;
- for Genfis2 – the range of influence of the cluster center;
- for Genfis3 – the fuzzy logic algorithm (Mamdani or Sugeno [7]) and the number of clusters.

As a result of the simulation, it was found that the TAN has the highest convergence rate (TAN reaches the region of optimal solutions in the first iterations, GAs after 7 generations, SAN after 6 iterations, FRS after 20 iterations). The FRS algorithm showed the highest speed (1 s per iteration, at 2 s in SAN, 3 s in TAN, 9 s per generation in GAs). At the same time, the FRS algorithm showed the best convergence (the value of criterion (1) when using it was 0.076, at 0.081 for SAN and TAN and 0.077 for GAs). It was established that the minimum criterion (1) is met by: the Genfis3 algorithm type using the Sugeno algorithm structure and 6 clusters.

4. Conclusions

An algorithm for recognizing phishing URLs (classifier) using fuzzy clustering and global optimization methods has been developed. By modeling, it was established that the minimum of the

combined criterion is met by a fuzzy classifier with a subtractive clustering algorithm and the use of the Sugeno structure and 6 clusters. It was established that all phishing URLs that were mistakenly classified as safe had a secure connection with a valid certificate. Thus, further research should be aimed at investigating additional informative attributes (features) that could allow for a better separation of phishing and safe URLs.

References

- [1] Phishing Activity Trends Report, 1st Quarter 2020. URL: https://docs.apwg.org/reports/apwg_trends_report_q1_2020.pdf
- [2] M. Aburrous, M.A. Hossain, K. Dahal, F. Thabtah. "Intelligent phishing detection system for e-banking using fuzzy data mining." *Expert Systems with Applications: An International Journal* 37.12 (2010): 7913–7921. doi: 10.1016/j.eswa.2010.04.044
- [3] N.S. Nordin et al. "Fuzzy Modelling using Firefly Algorithm for Phishing Detection". *Advances in Science, Technology and Engineering Systems Journal* 4.6 (2016): 291-296.
- [4] P.A. Barraclough, M.A. Hossain, G. Sexton, N. Aslam. "Parameter optimization for intelligent phishing detection using Adaptive Neuro-Fuzzy". *International Journal of Advanced Research in Artificial Intelligence* 3.10 (2014): 16-25. doi: 10.14569/IJARAI.2014.031003
- [5] N. Zhang, Y. Yuan. Phishing Detection Using Neural Network. Stanford University. 2012. URL: <http://cs229.stanford.edu/proj2012/ZhangYuan-PhishingDetectionUsingNeuralNetwork.pdf>
- [6] S. Gupta, A. Singhal. Phishing URL detection by using artificial neural network with PSO, in: *Proceedings of the 2nd International Conference on Telecommunication and Networks (TEL-ghNET)*. 2017. pp. 1-6.
- [7] Korniienko V. I. (Ed). *Intellectual modeling of nonlinear dynamic processes in control systems, cybersecurity, and telecommunication*. Dnipro University of Technology. 2020. 536 p.
- [8] D. Rutkowska, M. Piliński, L. Rutkowski. *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*. Warszawa ; Łódź : Wydaw. Naukowe PWN. 1997.
- [9] R. Yager, D. Filev. *Essentials of Fuzzy Modelling and Control*. NY: John Wiley & Sons. 1984. 387 p.
- [10] Bezdek J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms* / J.C. Bezdek. – New York : Plenum Press. – 1981. – 272 p.
- [11] A. G. Ivakhnenko, H.R. Madala. *Inductive learning algorithms for complex systems modeling*. London, Tokyo: CRC Press, 1994. 384 p.

Ensuring Data Security in Special-Purpose Information Systems

Iryna Udovyyk¹, Serhii Matsiuk¹ and Illia Olishevskiy¹

¹ Dnipro University of Technology, Dnipro, Dmytra Yavornytskoho, 19, 49005, Dnipro, Ukraine

Abstract

The article examines the importance of a comprehensive analysis of the level of cybersecurity of special-purpose information systems, using the example of medical information systems. For such a system, special attention is paid to the protection of personal data processed in this system. Possible actions of attackers and potential attacks on this data are considered, as well as the main threats to such systems are outlined. To solve the identified problems, the use of software security methods is proposed, which, as established in the study, provide protection only for certain aspects of cybersecurity. These are: protection against code tampering, control of the correctness of the execution of the software application logic, etc. Given this, the most effective method is to use a comprehensive approach that encompasses the synthesis of all the methods considered.

Keywords

Special-purpose information system, software security methods, cybersecurity ²⁴

1. Introduction

The intensive development of information technologies contributes to the penetration of computerized systems into previously inaccessible areas of human activity, in particular, medicine. In modern medical practice, electronic medical devices are becoming an integral part of the processes of diagnosis and monitoring of patients. Their functional capabilities are constantly expanding from data recording to their transmission via various protocols, including network means. This provides devices with new capabilities, but at the same time create challenges in the field of data security [1-2]. Current issues relate to the study of vulnerabilities of such systems and the development of solutions to counter attacks to form the most reliable protection.

2. Analysis of recent research and publications

Protecting personal data processed by digital medical devices is a fundamentally important task. The confidentiality of this information directly affects the reputation of medical systems and the trust of patients in them. A security breach can have serious consequences: leakage of personal medical information, erroneous diagnosis or incorrect choice of treatment. That is why the issue of ensuring reliable data protection is key to the successful implementation of such technologies in medicine. Data security problems are especially acute with the development of the functionality of electronic medical devices. For example, modern models of electronic stethoscopes now support adaptive noise filtering, recognition of audio recordings for preliminary diagnosis, as well as the transfer of information to external devices via wireless communication protocols. However, such innovations increase the risks of information compromise due to possible vulnerabilities in data storage and transmission systems. Despite their significant spread and high demand, only a small

²⁴ Corresponding authors

✉ udovyyk.i.m@nmu.edu.one (I. Udovyyk); matsiuk.sm@nmu.one (S. Matsiuk); olishevskiy.i.h@nmu.one (I. Olishevskiy)

ORCID 0000-0002-5190-841X (I. Udovyyk); 0000-0001-6798-5500 (S. Matsiuk); 0000-0001-8573-3366 (I. Olishevskiy)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

number of systems are created based on open architecture, such as, for example, The People's Ventilator Project [3]. The use of open architecture contributes to increasing the availability of such systems and accelerating their development, which is especially relevant considering the significant risks of global pandemics. At the same time, with the increasing prevalence of information-oriented devices, the likelihood of detecting security problems increases. This, in turn, increases the risks of compromising sensitive medical data, but at the same time stimulates a prompt response by developers to identify vulnerabilities. Thus, the openness of architecture has both advantages and disadvantages for electronic medical systems. To store information in these systems, various types of non-volatile memory are usually used, while standard wired and wireless communication protocols are used for data transmission. As a result, all modern electronic medical systems have potential channels for leakage of confidential information. Therefore, an urgent scientific and practical task remains to study the specifics of existing information security problems in electronic medical systems and develop effective methods for protecting this information.

The purpose of this work is to analyze the vulnerabilities of personal data in medical information and communication systems and to substantiate effective approaches to ensuring the cybersecurity of such systems.

3. Problem formulation

Data storage in digital medical devices is a key element of modern medical practice and is subject to strict legislative regulation [4-5]. These devices often store both personal medical data of patients and internal information necessary for the operation of the equipment. The first category includes data that identifies the patient, his medical records, medical history, including diagnoses, test results and performed procedures. The second category includes technical parameters of the device, settings, as well as events that occur during its operation. In both cases, the leakage of this data is a serious security threat and can have significant negative consequences. Most modern digital medical devices are equipped with interfaces for data input and output, firmware and software updates, and developer access features required for maintenance. Due to advances in standardization and the high cost of developing unique protocols, most such devices use one or more common standard data transmission protocols over wired and wireless channels.

3.1. Information threats in medical systems

Transmission protocols are key potential channels for information leakage from digital medical devices. By exploiting existing vulnerabilities in these protocols, an attacker can gain unauthorized access to personal data stored on the device. Therefore, it is important to consider the vulnerabilities of data that resides in the device's internal memory or is transmitted over wired or wireless networks. It is generally believed that a potential attacker has the following capabilities to attack a digital medical device:

- 10 Physical or remote access to the device's operating environment.
- 11 Knowledge of the protocol used to communicate with the device.
- 12 Ability to access the data channel using third-party devices or software.
- 13 Ability to impersonate a legitimate user by impersonating or gaining access to the account of a real user of the medical device [6].

The most common types of attacks include:

- i.1 Denial of Service (DOS) attack, which is usually carried out through a flood attack – sending many requests to overload the system.

- i.2 Replay attack, which consists of intercepting fragments of data or even an entire information exchange session for the purpose of reusing it under the guise of legitimate transmission.
 - i.3 Password guessing attack, where an attacker observes the authorization process of users or devices to obtain data to access the system.
 - i.4 Spoofing attack, where an unregistered user uses falsified parameters to authorize himself as a legitimate user in the system.
 - i.5 Insider attack, where an authorized user intentionally or accidentally harms the system [1].
- Research shows [6] that such attacks are often aimed at non-invasive therapeutic medical devices. These are usually fewer specific devices that are most often located next to the patient's body but are not implanted in it.

3.2. Methods of information protection in medical systems.

Based on the presented capabilities of a potential attacker, several approaches and solutions can be identified to effectively ensure the security of personal information stored or processed on medical devices. The key methods are listed below:

- i.1 Creating a secure environment for information devices. The use of virtual machines is one of the important solutions in this context. Virtual machines provide programs with an isolated environment that includes a network interface, data storage resources, and a platform for code execution. Due to such isolation, programs do not interact directly with each other, which significantly increases their security level.
- i.2 Static analysis methods. Among them, symbolic execution and thorough source code analysis are distinguished. These tools allow you to monitor the behavior of the program before its actual execution, as well as identify potential vulnerabilities and malicious code in the software structure.
- i.3 Dynamic analysis of program behavior. In this case, programs are tested in a specially created controlled environment during their execution in real time. To identify risky aspects, various frameworks are used that simulate both safe and unsafe behavior of applications in accordance with established security policies.
- i.4 Formal verification. This approach is based on the application of mathematical logic to verify system states. This method evaluates the correctness of security parameters for all possible input configurations, ensuring systematic and accurate risk detection.

All these methods can work together or separately, depending on the specific security requirements of the system, contributing to increased protection of medical devices from potential threats.

4. Conclusions

A detailed analysis of the features of the organization of work with data in modern electronic communication systems intended for medical purposes was carried out, and the protocols used for their functioning were also considered (in particular, protocols of both wired and wireless types). Based on the results obtained, an assessment of current threats to the information security of such systems was carried out and potential types of attacks that can be directed against them were classified. The work also provides recommendations on effective methods of information protection that contribute to increasing the level of cybersecurity of electronic communication systems in the medical field.

A promising direction for further research is the development of fundamentally new approaches to ensuring information security in medical information systems, by integrating intelligent data processing methods to effectively counter modern cyber threats.

References

- [1] M. Azrour, J. Mabrouki, A. Guezzaz and A. Kan-wal. "Internet of Things Security: Challenges and Key Issues." *Security and Communication Networks* 5533843 (2021): 1-11. doi: 10.1155/2021/5533843
- [2] P. Metty, L. Maglaras, M. A. Ferrag, I. Almomani. "Digitization of healthcare sector: A study on privacy and security concerns." *ICT 9.4* (2023): 571-588. doi: 10.1016/j.ict.2023.02.007
- [3] J. LaChance, M. Schottdorf, T. J. Zajdel et al. "PVP1—The People's Ventilator Project: A fully open, low-cost, pressure-controlled ventilator research platform compatible with adult and pediatric uses. " *PLoS ONE* 17.5 (2022): e0266810. doi:10.1371/journal.pone.0266810
- [4] United States Department of Health and Human Services. The HIPAA Privacy Rule. URL: <https://www.hhs.gov/hipaa/for-professionals/privacy/index.html>
- [5] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) URL: <http://data.europa.eu/eli/reg/2016/679/oj>
- [6] A. I. Newaz, A. K. Sikder, M. A. Rahman and A. S. Uluagac. A survey on security and privacy issues in modern healthcare systems: Attacks and defenses. *ACM Transactions on Computing for Healthcare* 2.3 (2021): 1-44. doi: 10.1145/3453176



КІБЕРЗАХИСТ КОМУНІКАЦІЙНИХ СИСТЕМ І ХМАРНОЇ ІНФРАСТРУКТУРИ

Глибоке навчання для мережевої безпеки: Модель Attention-CNN-LSTM для точного виявлення вторгнень

Віталій Дністровський

Національний технічний університет “Харківський політехнічний інститут”, Харків

Анотація

У статті розглянуто проблему виявлення вторгнень у сучасних комп’ютерних мережах, які стають дедалі більш складними та вразливими до кібератак. Запропоновано гібридну модель глибокого навчання Attention-CNN-LSTM, яка поєднує можливості згорткових нейронних мереж (CNN), довгострокової пам’яті (LSTM) та механізму самоуваги (attention). Мета роботи полягає у створенні системи виявлення вторгнень (IDS), здатної до високої точності та продуктивності в умовах реального часу. Експериментальні результати на датасетах NSL-KDD та Bot-IoT підтвердили ефективність моделі, яка забезпечила точність до 97,5 %, низьку затримку обробки (< 35 мс) і пропускну здатність понад 1200 записів/с. Таким чином, запропонована архітектура є перспективним рішенням для впровадження у практичні системи кіберзахисту.

Ключові слова

Система виявлення вторгнень, глибоке навчання, CNN, LSTM, механізм уваги, кібербезпека

1. Вступ

З розвитком інформаційних технологій зростає як кількість, так і складність атак на інформаційні системи. Традиційні методи захисту вже не гарантують належного рівня безпеки, адже нові загрози характеризуються високою варіативністю та швидкістю еволюції. Системи виявлення вторгнень (IDS) відіграють важливу роль у забезпеченні кіберзахисту, але їх ефективність залежить від здатності виявляти як відомі, так і невідомі атаки. Проблемою більшості традиційних IDS є орієнтація на сигнатурні методи, що потребують постійного оновлення та не здатні адаптуватися до нових загроз.

У цій ситуації особливої актуальності набувають методи глибокого навчання. Вони дозволяють аналізувати великі обсяги даних, виявляти приховані патерни та підвищувати рівень точності виявлення атак. Згорткові нейронні мережі (CNN) добре працюють із просторовими ознаками, а LSTM дозволяють враховувати часову динаміку подій. Проте ізольоване використання цих архітектур має певні обмеження: CNN не враховує довготривалих залежностей, а LSTM може втрачати важливі локальні особливості. Для подолання цих недоліків у роботі запропоновано комбіновану архітектуру Attention-CNN-LSTM.

1.1. Методологія

1.2. Архітектура моделі

Запропонована модель складається з трьох взаємодоповнюючих компонентів:

1. CNN-шари відповідають за автоматичне вилучення просторових ознак з мережевого трафіку, що дозволяє ідентифікувати характерні патерни атак.

2. LSTM-блоки забезпечують моделювання часових залежностей, необхідних для розпізнавання багатокрокових атак, таких як DDoS.
3. Механізм уваги зосереджується на найбільш інформативних ознаках, мінімізуючи вплив несуттєвих даних і покращуючи здатність моделі виявляти складні загрози.

Поєднання цих компонентів забезпечує глибокий аналіз трафіку, що підвищує загальну ефективність системи IDS.

1.3. Використані набори даних

Для перевірки ефективності моделі було використано два загальноприйняті набори даних:

1. NSL-KDD - набір, який усуває дублювання та дисбаланс даних, характерний для KDD'99, і містить різноманітні типи атак: DoS, Probe, R2L, U2R.
2. Bot-IoT - сучасний набір, орієнтований на сценарії Інтернету речей, що включає ботнети та інші складні атаки. Він є більш реалістичним для сучасних умов кібербезпеки, адже враховує специфіку IoT-середовищ.

1.4. Методи оцінки

Ефективність моделі оцінювалась за допомогою наступних показників:

1. Accuracy - відсоток правильно класифікованих зразків;
2. Precision та Recall - показники, що дозволяють оцінити якість виявлення атак і мінімізацію хибнопозитивних результатів;
3. F1-score - баланс між точністю та повнотою;
4. MCC (Matthews Correlation Coefficient) - надійний показник для роботи з дисбалансованими наборами даних;
5. Latency - середня затримка обробки запитів;
6. Throughput - кількість записів, які система здатна обробити за секунду.

2. Результати

Отримані результати підтверджують переваги запропонованої архітектури Attention-CNN-LSTM:

1. На наборі NSL-KDD точність моделі склала 94,8 %, а на Bot-IoT — 97,5 %.
2. Показники F1-score та MCC були значно вищими, ніж у базових моделей, що свідчить про збалансованість класифікації.
3. Механізм уваги зменшив кількість хибнопозитивних спрацювань, що особливо важливо для практичного застосування IDS.
4. Середня затримка обробки становила менше 35 мс, що дозволяє застосовувати систему у режимі реального часу.
5. Пропускна здатність перевищила 1200 записів/с, що робить модель придатною для високонавантажених середовищ.

Додатковий абляційний аналіз підтвердив, що кожен компонент моделі (CNN, LSTM, увага) робить свій внесок у кінцевий результат, проте найбільше зростання точності було досягнуто завдяки впровадженню механізму уваги.

3. Висновки та подальші дослідження

Розроблена модель Attention-CNN-LSTM є ефективним інструментом для виявлення вторгнень у комп'ютерних мережах. Вона поєднує сильні сторони CNN та LSTM з додатковим підсиленням через механізм уваги. Результати дослідження підтверджують, що запропонована архітектура може забезпечити високу точність, низьку затримку та масштабованість у режимі реального часу.

Подальші напрями досліджень передбачають:

1. розширення перевірки на інші сучасні набори даних;
2. інтеграцію федеративного навчання для забезпечення конфіденційності у розподілених системах;
3. дослідження стійкості до adversarial-атак;
4. оптимізацію для використання на пристроях із обмеженими ресурсами (edge-computing);
5. адаптацію для специфічних сценаріїв промислового Інтернету речей (IIoT).
Таким чином, ця робота робить вагомий внесок у розвиток сучасних систем мережевої безпеки, забезпечуючи баланс між високою точністю та продуктивністю, що є критично важливим у сучасних умовах кіберзагроз.

3.1. Література

- [1] Kotha, S., Tekulapalli, P. R., Pogaku, S. S. V. & Mohammed, G. B. Real-time detection of malicious intrusions and attacks in cybersecurity infrastructures enabled by IoT. In MATEC Web of Conferences (Vol. 392, p. 01127). EDP Sciences. (2024).
- [2] Naveeda, K. & Fathima, S. S. S. Real-time implementation of IoT-enabled cyberattack detection system in advanced metering infrastructure using machine learning technique. Electr. Eng. 1–20. <https://doi.org/10.1007/s00202-024-02552-z> (2024).
- [3] Nandanwar, H. & Katarya, R. A deep learning-driven IDS framework tailored for industrial IoT settings. Expert Syst. Appl. 249, 123808. <https://doi.org/10.1016/j.eswa.2023.123808> (2024).
- [4] Racherla, S. et al. Deep-IDS: A real-time intrusion detector for IoT nodes using deep learning. IEEE Access. 12, 112. <https://doi.org/10.1109/ACCESS.2024.3372187> (2024).
- [5] Al-Quayed, F., Ahmad, Z. & Humayun, M. A situation-based predictive approach for intrusion detection in industry 4.0 WSNs. IEEE Access. 12, 1–11. <https://doi.org/10.1109/ACCESS.2024.3371470> (2024).
- [6] Kandhro, I. A. et al. Detection of real-time malicious intrusions in IoT-enabled cybersecurity infrastructures. IEEE Access. 11, 9136–9148. <https://doi.org/10.1109/ACCESS.2023.3250195> (2023).

Аналіз адаптивного кодування LT, Raptor і LDPC у мережах нового покоління

Сергій Дунаєв¹, Ілля Букатич¹, Станіслав Мілевський¹ and Ірина Аксьонова¹

¹ Національний технічний університет «Харківський політехнічний інститут», вул. Кирпичова, 2, м. Харків, Україна, 61002

Анотація

Математичне моделювання та порівняння процесів адаптивного кодування в високонадійних мережах зв'язку нового покоління є передумовою побудови оптимального завадостійкого коду відповідно до вимог сучасних мереж. Акцент ставиться на аналіз математичних перетворень, які лежать в основі існуючих методів кодування, зокрема LT-кодів, Raptor-кодів та LDPC-кодів. Доцільним є розглядання побудови математичних моделей фонтанних кодів із декодуванням на основі критерію максимальної правдоподібності, що дозволить здійснити оптимізацію процесу відновлення даних навіть за умов обмеженого обсягу прийнятої інформації. Такий підхід набуває особливого значення в задачах підвищення прихованості даних, де застосування LDPC-кодів у якості структурної основи фонтанних схем дозволяє впровадити контрольований рівень ентропії, динамічність структури коду та адаптивність до атак на канал передачі.

Ключові слова

адаптивне кодування, фонтанні коди, завадостійке кодування, прихованість даних, максимальна правдоподібність, високонадійна мережа

1. Вступ

Математичне моделювання та порівняння процесів адаптивного кодування в високонадійних мережах зв'язку нового покоління є ключовою передумовою побудови оптимального завадостійкого коду відповідно до вимог сучасних систем передачі даних. Одним із перспективних підходів у цьому напрямі є фонтанне кодування, представлене LT- та Raptor-кодами, які розроблені для ширококомовних і багатоадресних передач в мережах з високим рівнем втрат. [1-4] Особливістю цих кодів є можливість генерувати практично необмежений потік кодованих пакетів, з яких достатньо отримати довільну підмножину для відновлення початкових даних. Це забезпечує високу надійність навіть за умов відсутності зворотного зв'язку та робить фонтанні коди гнучким рішенням для непередбачуваних каналів.

Важливою властивістю LT- та Raptor-кодів є відсутність фіксованої кодової швидкості, що дозволяє адаптуватися до змінної якості каналу. На практиці застосовуються усічені варіанти, які встановлюють мінімальну кількість пакетів для успішного декодування. Це робить фонтанні коди придатними для застосувань у 5G/6G-мережах, Інтернеті речей (IoT), системах мозок-комп'ютер (BCI) та технологіях розширеної та віртуальної реальності (XR).

У даній роботі акцент робиться на побудові математичних моделей фонтанних кодів із використанням декодування на основі критерію максимальної правдоподібності, що відкриває нові можливості оптимізації процесу відновлення даних за умов обмеженого

²⁵ Corresponding authors

✉ ivaksonova@gmail.com (І. Аксьонова); serg.dynaev@gmail.com (С. Дунаєв); illia.bukatych@gmail.com (І. Букатич); milevskiy@gmail.com (С. Мілевський);

ORCID 0000-0003-2605-0455 (І. Аксьонова); 0000-0001-8736-3602 (С. Дунаєв); 0009-0007-4695-073X (І. Букатич); 0000-0001-5087-7036 (С. Мілевський)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

обсягу прийнятої інформації. Порівняльний аналіз LT-, Raptor- та LDPC-кодів дозволяє виявити їхні переваги й обмеження, а також визначити потенціал інтеграції властивостей фонтанних кодів в структуру LDPC. Такий підхід є особливо актуальним у завданнях підвищення прихованості даних та захисту інформаційних каналів, де використання LDPC-кодів у поєднанні з фонтанними схемами дає можливість забезпечити контрольований рівень ентропії, адаптивність до атак і стійкість до завад в мережах нового покоління.

2. Основне дослідження

Першим видом практичних фонтанних кодів є коди Лабі (LT-коди), які як і лінійні коди, здатні генерувати довільну кількість незалежних і рівноймовірних закодованих символів. За допомогою породжувальної матриці декодер може оперативно відновити початкове повідомлення з будь-якої підмножини отриманих кодованих символів, зберігаючи при цьому низьку обчислювальну складність. Класичні LT-коди демонструють високу швидкість і ефективність завдяки використанню ітеративного алгоритму передавання повідомлень при жорсткому декодуванні. Їх гнучкість дозволяє ефективно працювати в умовах багатоадресної та широкомовної передачі даних в Інтернеті завдяки відповідному розподілу ступенів для вихідних символів. На стороні приймача інформація про сусідні символи використовується для формування графа Таннера, який відображає структуру породжувальної матриці. І хоча існує обмежена кількість способів передавання індексів вихідних символів до декодера, в даному дослідженні передбачається, що декодер завжди має точну інформацію про ступінь та суміжні індекси кожного прийнятого символу.

Коди Лабі забезпечують адаптивність до змінних умов середовища передачі та не потребують зворотного каналу, що є критично важливим у умовах високого рівня втрат, типових для мереж 6G, супутникових каналів, мереж BCI або систем XR. Кодування LT-кодів базується на генерації кодованих символів шляхом вибору підмножини вихідних символів, які комбінуються операцією XOR. Найбільш ефективним методом декодування LT-кодів є декодування через виключення. Його суть полягає у покроковому знаходженні кодованих символів із ступенем 1, що дозволяє поступово відновити оригінальні символи.

Таким чином, адаптація LT-кодів до умов високонадійних мереж зв'язку відбувається завдяки їх перевагам у контексті 6G, а саме:

- гнучкості – не вимагають фіксованої кодової швидкості;
- масштабованості – підходять для широкомовної та багатоадресної передачі;
- стійкості до втрат – дозволяють повністю відновити повідомлення при втраті значної частини кодованих символів.

LDPC-коди є сучасними блоковими лінійними кодами з низькою щільністю перевірових матриць і також використовуються для забезпечення високої надійності передачі даних в системах зв'язку нового покоління. Обидва класи кодів (LT, LDPC) є здатними забезпечити високу пропускну здатність, однак мають істотні відмінності у властивостях, реалізації та сфері застосування. Подальшим розвитком LT-кодів є Raptor-коди, які представляють собою комбінацію двох шарів кодування: попереднє та LT-кодування. Така структура дозволяє покращити характеристики відновлення даних при збереженні низької складності.

Порівняльна характеристика LT-кодів, LDPC-кодів та Raptor-кодів у контексті застосування в системах зв'язку нового покоління наведена в таблиці 1. Дана характеристика складена на підставі аналізу [5-9]

Таблиця 1

Порівняння основних характеристик LT, LDPC та Raptor-кодів

Характеристика	LT-коди	LDPC-коди	Raptor-коди
Тип	Фонтанні, стохастичні	Лінійні блокові	Фонтанні з попереднім кодуванням
Кодова швидкість	Дуже висока (без фіксованої швидкості)	Середня (попередньо фіксована)	Гнучка, регулюється кількістю пакетів
Зворотний канал	Не потрібен	Необхідний для оптимальної продуктивності	Не обов'язковий
Часове споживання	Дуже низьке кодування, просте декодування	Складне кодування і багатокрокове декодування	Низьке при малих втратах, вище при поганому каналі
Підтримка ширококомовлення	Повна підтримка	Обмежена	Повна підтримка
Ефективність при високих втратах	Висока	Середня	Дуже висока
Придатність до IoT/BCI/6G систем	Висока (гнучкість, масштабованість)	Висока (при стабільному каналі)	Висока (гнучкість, ефективність при втраті)
Тип декодування	BP (Belief Propagation), ітеративний	BP або SPA (Sum-Product Algorithm)	BP + попереднє кодування
Кодовий надлишок	Регулюється динамічно	Фіксований	Оптимізований завдяки поєднанню LT і попереднього коду
Підтримка в стандартах	Використовуються в експериментальних рішеннях	Широко підтримуються (5G, Wi-Fi 6, DVB-S2, 6G R&D)	Є в мультимедійних та IP-мовних стандартах (3GPP MBMS)

Проведений аналіз даних, наведених у табл. 1, показав, що LT-, Raptor- та LDPC-коди мають різні переваги залежно від умов застосування. Так, LT-коди найбільш ефективні для систем із непередбачуваними втратами (XR, BCI, IoT), завдяки гнучкості та можливості відновлення даних без зворотного зв'язку. Raptor-коди є вдосконаленням LT-кодів, що поєднують простоту фонтанного підходу з підвищеною надійністю завдяки попередньому кодуванню, роблячи їх оптимальними для ширококомовних і супутникових систем. LDPC-коди забезпечують високу енергетичну ефективність та широко використовуються у сучасних стандартах (5G, Wi-Fi 6, DVB-S2), проте потребують стабільного каналу та складнішого декодування.

Зважаючи на це, доцільним є перехід до побудови математичних моделей фонтанних кодів із декодуванням на основі критерію максимальної правдоподібності, що дозволить здійснити оптимізацію процесу відновлення даних навіть за умов обмеженого обсягу прийнятої інформації. Такий підхід набуває особливого значення в задачах підвищення прихованості даних, де застосування LDPC-кодів у якості структурної основи фонтанних схем дозволяє впровадити контрольований рівень ентропії, динамічність структури коду та адаптивність до атак на канал передачі.

Як було зазначено, фонтанні коди є більш гнучкими у порівнянні з LDPC-кодами, які, в свою чергу, є більш детермінованими. Але, проведений порівняльний аналіз показав, що саме їх поєднання в архітектурі коду Raptor дозволяє досягати високої ефективності, адаптивності та надійності в сучасних телекомунікаційних мережах. У контексті дослідження, спрямованого на підвищення прихованості даних шляхом вдосконалення механізмів завадостійкого кодування, розглянемо умови застосування декодування по максимальній правдоподібності (ML decoding) для фонтанних кодів, зокрема LT- та Raptor-кодів.

Проведемо теоретичне обґрунтування та порівняємо ML-декодування для LT-кодів та Raptor-кодів (таблиця 2). [2,5]

Таблиця 2

Теоретичне обґрунтування ML-декодування для LT та Raptor-кодів

Критерій / Код	LT-коди	Raptor-коди
Структура коду	Генерація кодових символів на основі випадкових поєднань інформаційних бітів згідно з розподілом Солітона	Поєднання попереднього (LDPC/HDPC) кодування з LT-кодуванням та утворення гібридної структури
Умови для ML-декодування	Застосовується, коли отримано достатню кількість кодових символів для побудови повного рангу системи	ML-декодування застосовується до LT-компоненти; розмір LT-коду обмежено, що знижує складність
Модель ML-декодування	Розв’язання системи лінійних рівнянь: $x' = \arg \min_x \ y - Gx\ _2$, де Gx — породжувальна матриця	Те саме для LT-компоненти, при цьому враховується структура попереднього LDPC-коду
Обчислювальна складність	Висока при великих розмірах; зменшується при попередній обробці або обмеженні ступеня	Нижча, завдяки обмеженому розміру LT-частини та ефективному структурованому LDPC-коду
Переваги ML-декодування	Дає змогу відновити дані навіть при мінімальній надмірності; гарантує найкращу правдоподібність відновлення	Забезпечує високу якість відновлення в гібридних схемах, особливо у сценаріях з високими втратами
Область	Теоретичний інструмент для	Часто застосовується в реальних

застосування	аналізу; практично – резервний варіант	Raptor-декодерах після ітеративного декодування
Аналіз таблиці 2 показує, що ML-декодування у LT-кодах є потужним, але обчислювально затратним методом, придатним переважно як теоретичний чи резервний інструмент. У Raptor-кодах воно стає практично застосовним завдяки обмеженій розмірності LT-компоненти та попередньому LDPC-кодуванню, що знижує складність і підвищує надійність відновлення даних у каналах із високими втратами. Таким чином, ML-декодування відіграє ключову роль у підвищенні стійкості та прихованості сучасних гібридних кодів.		

3. Висновки

Таким чином, проведений аналіз LT-, Raptor- та LDPC-кодів показав, що їх ефективність визначається умовами каналу та вимогами до надійності. LT-коди забезпечують гнучкість і стійкість до втрат у нестабільних мережах, LDPC-коди — високу продуктивність у стабільних каналах, а Raptor-коди поєднують переваги обох підходів. Використання ML-декодування у фонтанних кодах дозволяє підвищити ймовірність відновлення даних при мінімальній надлишковості, що є критичним для IoT, XR, BCI та супутникових систем. Комбінація фонтанного та LDPC-кодування створює основу для багаторівневого захисту та підвищення прихованості інформації у високонадійних мережах нового покоління.

Література

- [1] W. Huang, “Investigation on Digital Fountain Codes over Erasure Channels and Additive White Gaussian Noise Channels”, [Doctoral dissertation, Ohio University], 2012, OhioLINK Electronic Theses and Dissertations Center. URL: http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1336067205.
- [2] M. Luby, “LT Codes”, Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2002. URL: <https://doi.org/10.1109/SFCS.2002.1181950>.
- [3] M. Mitzenmacher, “Digital fountains: a survey and look forward”, IEEE Information Theory Workshop, San Antonio, TX, USA, 2004, pp. 271–276. DOI: 10.1109/ITW.2004.1405313.
- [4] O. Etesami, and A. Shokrollahi, “Raptor codes on binary memoryless symmetric channels”, IEEE Transactions on information theory, 2006, vol. 52, no. 5, pp. 2033–2051. URL: <https://doi.org/10.1109/TIT.2006.872855>.
- [5] A. Shokrollahi, and M. Luby, “Raptor Codes”, Foundations and Trends in Communications and Information Theory, 2009, vol. 6, iss. 3–4, pp. 213–322. DOI: 10.1561/01000000060.
- [6] Р. Галлагер, “Теорія коригуючих кодів із низькою щільністю перевірок”, пер. з англ. К.: Техніка, 2017, 214 с.
- [7] Н. Руденко, та І. Луцюк, “Методи побудови мереж п’ятого покоління на основі існуючих мереж”, Наукові записки ДУТ, 2023, №2(4), С.39-44. DOI: 10.31673/2518-7678.2023.020505.
- [8] F. Lázaro, G. Liva, G. Bauch, and E. Paolini, “Bounds on the Error Probability of Raptor Codes under Maximum Likelihood Decoding”, arXiv:1809.01515v4 [cs.IT], 2021. URL: <https://doi.org/10.1109/TIT.2020.3049061>.
- [9] Y. Ren, B. Zhang, J. Shen, Z. Zhang, E. Boutillon, A. Balatsoukas-Stimming, and A. Burg, “Edge-Spreading Raptor-Like LDPC Codes for 6G Wireless Systems”, arXiv preprint, 2024, arXiv:2410.16875. URL: <https://arxiv.org/abs/2410.16875>.

Виявлення атак типу "блокування IP через NAT" в публічних мережах

Юрій Кльоц¹, Сергій Мостовий¹, Назар Тростянецький¹

¹ Хмельницький національний університет, вул. Інститутська 11, 29016 Хмельницький, Україна

Анотація

У роботі розглянуто проблему виявлення атак типу «блокування IP через NAT» у публічних мережах. Такі атаки виникають унаслідок класифікації системами IPS/IDS сумарного трафіку від кількох користувачів, що виходять у мережу через одну публічну IP-адресу, як потенційно шкідливого. Запропоновано архітектуру системи моніторингу та метод виявлення підозрілої активності на основі аналізу вихідного трафіку. Для ідентифікації хостів, що ініціюють аномальну поведінку, використано нейронну мережу LSTM, яка дозволяє враховувати часові залежності та прогнозувати ймовірність блокування. Розроблений підхід забезпечує можливість раннього виявлення атак і зменшує ризик колективного блокування користувачів при спільному NAT-доступі.

Ключові слова

NAT, IPS/IDS, блокування IP, LSTM, аналіз трафіку, виявлення атак, кібербезпека

1. Вступ

На сьогодні адресний простір IPv4 майже повністю вичерпано, однак багато користувачів і далі потребують його для роботи з традиційними інтернет-протоколами. Перехід на IPv6 є тривалим процесом, тому активно застосовується технологія NAT, що дозволяє використовувати IPv4 навіть після повного розподілу адрес IANA. У разі підключення локальних або малих мереж за допомогою NAT зовнішній інтерфейс має лише одну публічну IP-адресу. Попри переваги цієї технології, у певних умовах вона може призводити до недоступності сервісів у глобальній мережі, сповільнення або блокування їх роботи, створюючи незручності для користувачів [1].

У SOHO-рішеннях та мережах дрібних провайдерів доступ користувачів до Інтернету часто здійснюється через одну публічну IP-адресу із використанням механізму NAT. У таких умовах усі клієнти виходять у мережу з однаковою зовнішньою IP-адресою, що ускладнює ідентифікацію окремих джерел трафіку. Якщо один зі споживачів або група користувачів генерує нетипову або агресивну активність (наприклад, численні запити на один ресурс), системи виявлення вторгнень (IPS/IDS) на стороні зовнішнього сервісу можуть класифікувати такий трафік як шкідливий. У результаті публічна IP-адреса NAT-вузла потрапляє під блокування, що призводить до втрати доступу для всіх користувачів провайдера до відповідного ресурсу на період від кількох годин до кількох діб. Це створює значні незручності для клієнтів і негативно впливає на репутацію провайдера.

²⁶ Corresponding authors

✉ klots@khmn.edu.ua (Ю. Кльоц); serhii.mostovyi@khmn.edu.ua (С. Мостовий); ntrostianetskyi@khmn.edu.ua (Н. Тростянецький)

ORCID 0000-0002-3914-0989 (Ю. Кльоц); 0000-0002-9505-3206 (С. Мостовий);



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Аналіз досліджень та публікацій

Актуальні дослідження свідчать, що основними підходами до виявлення шкідливої активності, в тому числі низької інтенсивності використовуються методи машинного навчання [2]. У статті [3] запропоновано адаптивну розподілену систему фільтрації DDoS-трафіку, що використовує вузли вздовж шляху атаки для блокування шкідливих потоків.

Система динамічно генерує правила фільтрації з різною деталізацією (IP, порт, префікс) залежно від мети — мінімізації втрат легітимного трафіку або максимального покриття атаки.

Для виявлення атак типу DoS/DDoS застосовуються системи запобігання та виявлення вторгнень (IPS/IDS), які здійснюють класифікацію мережевого трафіку на шкідливий або легітимний. Одними з основних параметрів, що аналізуються при цьому, є IP-адреси та номери портів джерел трафіку. У середовищах, де використовується механізм трансляції мережевих адрес (NAT), декілька клієнтів можуть здійснювати доступ до зовнішніх сервісів через спільну публічну IP-адресу. Така активність, що характеризується множинністю портів при одному джерельному IP, часто інтерпретується IPS/IDS як ознака розподіленої атаки типу DDoS [4]. Також більшість науковців відмічають значний відсоток хибно-позитивних спрацювань IPS/IDS систем при аналізі вхідного трафіку [5].

3. Система виявлення атак типу "блокування IP через NAT"

Розглянемо структуру системи виявлення атак, представлену на рис. 1. Захищувана інформаційна система містить корпоративний та публічний сегменти. Публічний сегмент призначений для гостьового доступу, в ньому не передбачено ідентифікацію користувачів.

Для приєднання користувачів до публічного сегменту використовується бездротовий маршрутизатор R2, що забезпечує ізоляцію сторонніх клієнтів від корпоративного сегменту. Маршрутизатор R1 та комутатор S1 забезпечують функціонування корпоративного сегменту. Типовою особливістю такого класу систем є використання технології NAT для доступу корпоративних та публічних клієнтів до мережі інтернет та єдиної публічної IP адреси. Користувачі як корпоративного так і публічного сегментів користувачами зовнішньої інформаційної системи IS, що для захисту від атак використовує IPS/IDS систему.

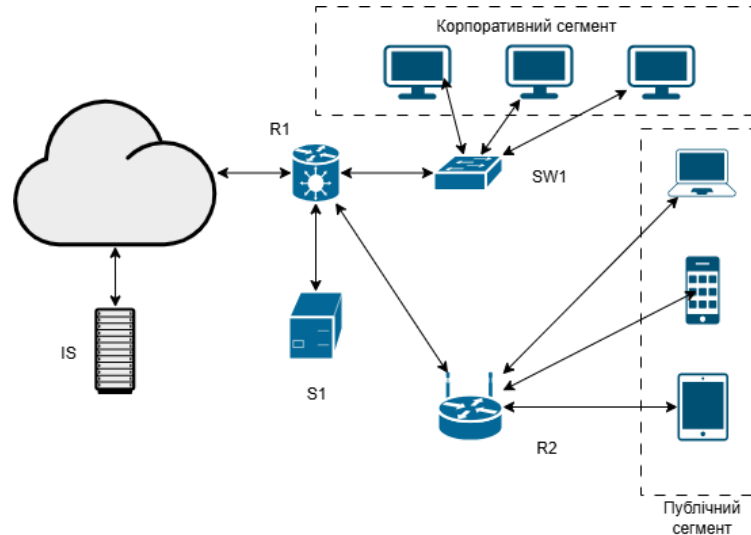


Рисунок 1: Структура системи виявлення атак.

За таких умов можливі ситуації, при яких IPS/IDS система IS ідентифікує DoS/DDoS атаку та блокує публічну IP адресу маршрутизатора R1 або сприймає робочий трафік багатьох клієнтів як DoS/DDoS атаку і також блокує публічну IP адресу маршрутизатора R1. В обох випадках наслідком таких дій IS буде блокування доступу до IS всіх клієнтів мережі. Відповідно задача виявлення вихідних DoS/DDoS атак або інтенсивного нормального трафіку, що сприймається IPS/IDS як шкідливий з раннім виявленням та блокуванням таких дій є актуальною.

З метою раннього виявлення збільшення інтенсивності трафіку до зовнішніх систем доцільно дзеркалювати вихідний трафік з R1 на сервер додаткової обробки трафіку, що проводить аналіз пакетів вихідного трафіку та динамічно змінює правила Firewall на R1. Така структура дозволяє без зменшення пропускної здатності R1 проводити аналіз вихідного трафіку, визначати масові звернення до віддалених інформаційних систем, прогнозувати блокування доступу до них та динамічно керувати доступом користувачів до таких ресурсів.

Для виявлення атак типу «блокування IP через NAT» доцільно застосовувати нейронні мережі з довготривалою короткочасною пам'яттю (LSTM), оскільки вони ефективно моделюють часові залежності у трафіку. LSTM дозволяє враховувати послідовність подій і відстежувати поведінку окремих локальних хостів у динаміці, що особливо важливо для аналізу кумулятивних ефектів, які призводять до блокування зовнішньої IP-адреси. На вхід моделі можуть подаватися агреговані часові ряди мережових характеристик — кількість запитів до певного сервісу, частота звернень за портами, ентропія напрямків трафіку. Модель навчається на історичних даних, де відомі випадки блокування, і формує ознаки, що передують інциденту. Під час роботи в реальному часі LSTM забезпечує прогнозування ймовірності виникнення блокування на основі поточного стану трафіку. Такий підхід дозволяє не лише виявити підозрілі хости, але й визначити їх внесок у загальне навантаження, що призводить до активації IPS/IDS зовнішнього сервісу. У результаті система може попереджати про потенційно небезпечну активність до фактичного блокування спільної IP-адреси.

4. Висновки

У роботі запропоновано підхід до виявлення атак типу «блокування IP через NAT» у середовищах з колективним використанням зовнішньої IP-адреси. Розроблена структура системи моніторингу забезпечує можливість аналізу вихідного трафіку в реальному часі без зниження пропускної здатності маршрутизатора. Використання моделі LSTM дозволяє виявляти закономірності в часових рядах трафіку та прогнозувати ризик блокування IP-адреси через надмірну або підозрілу активність. Запропонований метод дає змогу своєчасно визначати локальні джерела потенційно шкідливого трафіку та динамічно змінювати правила фільтрації. Подальші дослідження доцільно спрямувати на оптимізацію процесу навчання моделі та інтеграцію рішення у системи автоматизованого управління мережею провайдера.

Література

- [1] Melnyk , V., Bahniuk , N., Royko O., & Kizym , S. (2025). Network sockets based on aliases with remote address transfer. *COMPUTER-INTEGRATED TECHNOLOGIES: EDUCATION, SCIENCE, PRODUCTION*, (59), 194-204. <https://doi.org/10.36910/6775-2524-0560-2025-59-25>.
- [2] Joseph, Ofir & Elmalech, Avshalom & Hajaj, Chen. Detecting Parallel Covert Data Transmission Channels in Video Conferencing Using Machine Learning. *Electronics* 2023, 12, 1091, <https://doi.org/10.3390/electronics12051091>.
- [3] Li, Jun & Sisodia, Devkishen & Feng, Yebo & Shi, Lumin & Zhang, Mingwei & Early, Christopher & Reiher, Peter. Adaptive Distributed Filtering of DDoS Traffic on the Internet, 2023. <https://doi.org/10.48550/arXiv.2305.13296>.
- [4] Mohan S.N., Ravikumar G., Govindarasu M. Distributed Intrusion Detection System using Semantic-based Rules for SCADA in Smart Grid. *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, 2020, <https://doi.org/10.1109/TD39804.2020.9299960>
- [5] Abu Al-Haija, Qasem & Droos, Ayat. A comprehensive survey on deep learning-based intrusion detection systems in Internet of Things (IoT). *Expert Systems*. 2024. <https://doi.org/42.10.1111/exsy.13726>.

Безсерверні та контейнеризовані архітектури: порівняльний аналіз підходів до безпеки

Юрій Кулик¹, Богдан Скоринович¹ та Володимир Гавриляк¹

¹ Національний університет «Львівська політехніка», Степана Бандери 12, 79013 Львів, Україна

Анотація

Проведено теоретичний аналіз безпекових аспектів контейнеризованих та безсерверних архітектур у хмарних середовищах. Сформульовано проблему відмінності моделей відповідальності між клієнтом та провайдером для кожного підходу, що зумовлює специфічні загрози для кожного з них. Наукова новизна полягає у систематизації ключових ризиків та механізмів захисту з акцентом на управління доступом, захист застосунків і мережеву взаємодію. Доведено, що забезпечення безпеки можливе лише за умови врахування ризиків кожної архітектури та впровадження найкращих практик.

Ключові слова

контейнеризація, безсерверні обчислення, Kubernetes, хмарна безпека

1. Вступ

Сучасні хмарні обчислення характеризуються значним поширенням контейнеризованих та безсерверних (Serverless) архітектур, що суттєво змінило історичні підходи до розгортання застосунків. Контейнери прийшли на заміну віртуальним машинам, забезпечуючи швидкий старт і значно ефективніше використання ресурсів [1]. Однак спільне використання ядра операційної системи (ОС) різними контейнерами створило нову «поверхню атак» через послаблення ізоляції між застосунками. Відповідно, хоч і запуск застосунків за допомогою контейнерів зараз надзвичайно поширений, деякі додатки досі розгортають на віртуальних машинах в цілях безпеки. З іншого боку значного поширення в індустрії також набули безсерверні обчислення, які пропонують ще вищий рівень абстракції, делегуючи обслуговування інфраструктури хмарному провайдеру. Така архітектура надає переваги гнучкого масштабування і моделі «pay-per-use», але водночас створює унікальні виклики безпеки, пов'язані з розподіленістю функцій, багатокористувацькістю та ефемерністю виконання [2][3]. Актуальність теми зумовлена необхідністю детального аналізу ризиків і методів захисту в цих архітектурних підходах, щоб організації могли безпечно користуватися ними.

2. Теоретичні основи

Контейнеризовані застосунки віртуалізуються на рівні ОС та містять лише необхідний мінімум для свого запуску: код, бібліотеки та залежності без окремої гостьової ОС на відміну від віртуальних машин [1]. Такий підхід значно зменшує час для запуску застосунку та, з допомогою контейнерних платформ на кшталт Docker та Kubernetes, дозволяє швидко масштабувати незалежні компоненти системи. Відповідно, основними

²⁷ Відповідні автори

✉ yurii.a.kulyk@lpnu.ua (Y. Kulyk); bohdan.v.skorynovych@lpnu.ua (B. Skorynovych); volodymyr.r.havryliak@lpnu.ua (V. Havryliak)

ORCID 0009-0003-9249-2697 (Y. Kulyk); 0009-0007-8669-9172 (B. Skorynovych); 0009-0002-4114-1232 (V. Havryliak)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

перевагами контейнеризації є портативність та ефективне використання ресурсів. Однак, за рахунок зниження ізоляції, вразливостей ядра або неправильної конфігурації рушія контейнеризації можуть бути скомпроментовані всі контейнери на вузлі [4]. В результаті, до безпеки контейнерів висуваються інші вимоги, що спричинило появу окремого напрямку досліджень.

Концепція безсерверних застосунків передбачає виконання коду у вигляді окремих функцій на інфраструктурі хмарного провайдера, що отримало назву «Function as a Service» (FaaS) [3]. Архітектурно, такі застосунки складаються з набору малих автономних функцій, кожна з яких викликається різного роду подіями (HTTP-запит, таймер), які виконуються протягом короткого проміжку часу (до декількох секунд). З точки зору підтримки такого рішення, єдине що цікавить розробника - код, а розгортання, масштабування та зупинка цього коду - відповідальність провайдера хмарних послуг. У такій моделі оплата стягується тільки за фактичний час виконання коду, на відміну від оплати за постійно працюючий сервер, однак, в порівнянні, відносна ціна часу такого виконання значно вища [3]. Також відсутність управління інфраструктурою значно спрощує розробку. Однак, безсерверність у даному випадку означає всього лиш змінену модель «розподілу відповідальності», так як сервери сховані від користувача і хмарний провайдер відповідає за безперебійне та безпечне виконання коду на своїх потужностях. На практиці виконання функцій FaaS відбувається таки в контейнеризованому середовищі, яке повністю контролюється провайдером, подекуди розробнику навіть дозволяється використовувати власний контейнерний образ. Такий підхід забезпечує високу еластичність та швидкий випуск продуктів, але висуває нові вимоги до забезпечення безпеки застосунків.



Рисунок 1: Порівняння моделей відповідальності у хмарних обчисленнях: контейнеризовані та безсерверні застосунки.

3. Аналіз відмінностей у підходах до забезпечення безпеки

Хоч і ціллю обох архітектур є швидке масштабоване розгортання застосунків, ці підходи схильні до різних загроз, тому, відповідно, потребують відмінних механізмів захисту. Розглянемо такі ключові аспекти цих архітектур: управління доступом, безпека застосунків та мережі.

3.1. Управління доступом

У контейнеризованих середовищах управління доступом відбувається як на рівні оркестратора, так і всередині контейнерів. Для прикладу, найпопулярніший оркестратор сьогодення Kubernetes дозволяє реалізувати різноманітні підходи до контролю доступом (ABAC, RBAC, Webhook авторизація), серед яких за замовчуванням у більшості рішень

використовується RBAC, що гарантує надання користувачам і сервісам системи лише мінімально необхідних дозволів для виконання своїх функцій [4]. Не менш важливим аспектом є контроль привілеїв процесів всередині контейнерів, зокрема уникнення запуску застосунків від імені root-користувача та надання застосункам тільки необхідних capabilities ядра Linux [1]. Для автоматичного застосовування цих найкращих практик існують різноманітні інструменти (OPA Gatekeeper, Pod Security Standards).

На відміну від контейнерів, у безсерверних середовищах управління доступом зміщується на рівень хмарного провайдера. Можна вважати, що це свого роду аналог доступу до оркестратора для контейнеризованих середовищ. В такому випадку, для кожної функції і пов'язаних сервісів налаштовуються політики, які визначають хто/що може викликати функцію і до яких ресурсів вона має доступ [3]. Механізми налаштувань прав для таких середовищ є дуже гнучкими, і, для прикладу, передбачають прив'язку ролей до функцій, авторизацію викликів через API Gateway та обмеження викликів за IP-адресами [3]. Також важливо правильно налаштувати ці механізми, так як помилки в конфігурації можуть призвести до критичних вразливостей: надто широкі дозволи -> ескалація привілеїв у випадку компрометації функції; велика кількість функцій і способів їх виклику -> підвищена вірогідність людської помилки під час їх обслуговування. Відповідно, як і у випадку з контейнерами, безсерверні архітектури потребують застосування принципу найменших привілеїв, але ландшафт застосування цього підходу значно відрізняється.

3.2. Безпека застосунків

Уразливості застосунків в контейнеризованих архітектурах можуть призвести до компрометації усього вузла через атаку типу «container escape». Саме тому, безпека цього рівня зосереджується на двох напрямках: безпека образів контейнерів та механізми стримування зловмисника всередині контейнера. Перше передбачає регулярне сканування образів на відомі вразливості, використання тільки довірених базових образів та оновлення залежностей додатку до актуальних версій. Друге потребує впровадження обмежень можливостей коду у разі його компрометації: ізоляція процесів за допомогою Linux namespaces та використання таких засобів як SELinux та AppArmor для заборони виконання небезпечних системних викликів [1]. Узагальнюючи, стратегія безпеки контейнеризованого застосунку будується за принципом «в глибину»: навіть якщо був скомпрометований сам застосунок, додаткові рівні ізоляції повинні стримати зловмисника [4].

Щодо FaaS середовища, функції менші за обсягом і виконуються дуже малий проміжок часу, що зменшує вікно для атак і можливість довготривалого закріплення в системі. З іншого боку, значно розширюється та фрагментується «поверхня атаки», так як типова безсерверна архітектура складається з десятків функцій, кожна з яких має свої механізми виклику [3]. Відповідно, потрібно приділяти додаткову увагу захисту та аудиту кожної окремої функції. Особливої уваги заслуговує безпека сторонніх бібліотек та залежностей, так як функції часто використовують різноманітні пакети з відкритим кодом [2]. Атаки на них відомі як «supply chain» та можуть мати надзвичайно руйнівні наслідки. Підсумовуючи, безпека застосунків у безсерверній архітектурі засновується на суворому дотриманні найкращих практик розробки (включно з активним впровадженням DevSecOps), використанні інструментів для аналізу вразливостей коду та активному захисті від відомих атак за допомогою допоміжних інструментів (WAF, системи керування секретами, засоби моніторингу й runtime-захисту, тощо).

3.3. Мережеві аспекти безпеки

Перш за все, контейнеризовані застосунки за замовчуванням можуть взаємодіяти через віртуальну мережу. Відповідно, необхідне впровадження сегментації мережі та фільтрації трафіку для контролю такої взаємодії. Для прикладу, Kubernetes дозволяє застосовувати мережеві політики, які визначають комунікацію між сервісами [4]. Такі засоби захисту дозволяють унеможливити подальший рух між контейнерами у випадку компрометації одного із них. Окрім ізоляції трафіку, необхідно якомога зменшити доступність контейнерів назовні за допомогою відкриття лише необхідних портів та автентифікації для доступу клієнтів. Також сучасні найкращі практики рекомендують впровадження шифрування трафіку навіть всередині внутрішньої мережі. У Kubernetes такий механізм можливо реалізувати за допомогою технології service mesh, яка забезпечує шифрування та взаємну автентифікацію між мікросервісами (mTLS). Підбиваючи підсумки, комплексне шифрування та сегментація є ключовими методами протидії мережевим атакам в контейнеризованих середовищах [4].

Безсерверні застосунки мають дещо іншу модель мережевої взаємодії: функції викликаються за допомогою подій, а не постійно слухають мережеві порти. Такий підхід зменшує загрозу класичних атак, але змушує приділяти значно більшу увагу подіям, які викликають ці функції. Для забезпечення їх безпеки застосовуються механізми авторизації запитів, мережевої ізоляції на рівні хмари та фаєрволи [3]. Також хотілось б наголосити на захисті від DDoS атак, так як функції автоматично масштабуються, що може спричинити значні фінансові збитки. Такий тип атаки отримав назву “Denial-of-Wallet”. На щастя, хмарні провайдери мають вбудовані механізми захисту від такого типу атак, але, в будь-якому випадку, не варто забувати за них. Загалом, мережеву безпеку безсерверних архітектур можна звести до контролю точок входу (подій), ізоляції функцій у внутрішніх мережах та використанні сервісів хмарних провайдерів для виявлення та захисту від підозрілих викликів (веб-фаєрволи, моніторинг трафіку).

4. Порівняльна таблиця ключових характеристик безпеки

Таблиця 1

Порівняння характеристик деяких аспектів контейнеризованої і безсерверної архітектури з точки зору безпеки

Аспект	Контейнеризована архітектура	Безсерверна архітектура
Розгортання та управління	Повний контроль над платформою	Контроль делегується хмарному провайдеру
Ізоляція виконання	На рівні операційної системи	Забезпечується хмарним провайдером
Відповідальність за безпеку	Власник застосунку відповідає за захист ОС, контейнерного двигуна, мережі та інших аспектів	Провайдер повністю відповідає за інфраструктуру, а власник застосунку — тільки за код і його безпеку
Керування	Впроваджується на рівні оркестраторів та всередині	Забезпечується хмарним

доступом	контейнерів	провайдером
Видимість інфраструктури	Повний доступ до інфраструктури на рівні віртуальних машин	Мінімальний доступ до інфраструктури, що змушує довіритись провайдеру послуг
Оновлення та патчі	Власник застосунку повністю відповідальний за оновлення ОС хостів та версій утиліт	Провайдер несе повну відповідальність за оновлення платформи
Поверхня атаки	Помірна: зовнішні API сервісів, порти контейнерів, оркестратор	Висока: багато дрібних функцій і тригерів створюють розгалужену поверхню атаки

5. Висновки

Контейнеризовані та безсерверні архітектури стали ключовими технологіями для побудови сучасних хмарних застосунків, однак кожна з них створює власні виклики у сфері кібербезпеки. Контейнери забезпечують гнучкість і швидкість, але розділяють спільне ядро ОС, що вимагає ретельного налаштування ізоляції та контролю доступу. Безсерверний підхід делегує керування інфраструктурою провайдеру, натомість породжує розподілену поверхню атаки і покладається на вбудований захист платформи. Проведений порівняльний аналіз показав, що жоден із підходів не забезпечує максимальний захист без дотримання найкращих практик безпеки специфічних для кожної архітектури. Для контейнерів - це багаторівнева оборона: від безпечного коду до захищеної платформи оркестрації, а для FaaS - це проектування системи з урахуванням Zero Trust, суворе обмеження привілеїв та детальний контроль кожного компонента системи. Реалізація цих задач потребує поєднання автоматизованих інструментів та навчання розробників найкращих практик побудови застосунків. Підсумовуючи, ці підходи можуть бути безпечними лише за умови, що їхні особливості враховані на етапі дизайну системи та підтримуються належними засобами і заходами захисту.

Література

- [1] S. Sultan, I. Ahmad, T. Dimitriou, Container Security: Issues, Challenges, and the Road Ahead, IEEE Access 7 (2019) 52976-52996. doi:10.1109/access.2019.2911732.
- [2] X. Li, X. Leng, Y. Chen, Securing Serverless Computing: Challenges, Solutions, and Opportunities, IEEE Network 37 (2022) 166-173. doi:10.1109/mnet.005.2100335.
- [3] E. Marin, D. Perino, R. Di Pietro, Serverless computing: a security perspective, Journal of Cloud Computing 11 (2022). doi:10.1186/s13677-022-00347-w.
- [4] O. Jarkas, R. Ko, N. Dong, R. Mahmud, A Container Security Survey: Exploits, Attacks, and Defenses, ACM Computing Surveys 57 (2025). doi:10.1145/3715001.

Прогнозування повільних DDoS атак на основі кореляційного аналізу індивідуальних траєкторій трафіку

Віталій Савченко¹, Галина Гайдур¹, Світлана Легомінава¹

¹Державний університет інформаційно-комунікаційних технологій, вул. Солом'янська, 7, 03110, Київ, Україна

Анотація

У статті досліджується проблема виявлення низькоінтенсивних та повільних розподілених атак типу "відмова в обслуговуванні" (DDoS), які становлять суттєву загрозу для сучасних інформаційних систем. На відміну від класичних об'ємних атак, що характеризуються різким зростанням інтенсивності мережевого трафіку, повільні атаки відзначаються прихованим характером дії та ускладненістю їх ідентифікації за стандартними методами моніторингу. У роботі висунуто припущення про залежність ефективності виявлення повільних атак від індивідуальних особливостей поведінки легітимних користувачів. Запропоновано метод, що ґрунтується на аналізі та прогнозуванні індивідуальної поведінкової траєкторії користувача у процесі його взаємодії з веб-сервісом. Практичні можливості цього підходу підтверджено шляхом моделювання атак типу RUDY на HTTP-сервіси, що дало змогу оцінити вплив обсягу накопиченого трафіку та статистичних характеристик атак на точність прогнозування. Отримані результати свідчать про перспективність запропонованого методу для виявлення широкого спектра повільних DDoS-атак та підвищення загального рівня захисту інформаційних ресурсів.

Ключові слова

повільна DDoS-атака, індивідуальна траєкторія, індивідуальне прогнозування, кореляційний аналіз, випадковий процес

1 Вступ

Доступність інструментів для пошуку мережевих вразливостей зумовила широке розповсюдження DDoS-атак. Ефективна протидія таким загрозам передбачає реалізацію двох ключових завдань: 1) своєчасне діагностування атаки на ранніх етапах її розвитку; 2) відокремлення шкідливого трафіку від легітимного. Ідентифікація запитів, що є результатом DDoS-атаки, створює можливість для налаштування параметрів брандмауерів, маршрутизаторів або застосування інших механізмів захисту. Особливістю повільних DDoS-атак є їх залежність від поведінкових характеристик окремого користувача, що ускладнює їх своєчасне виявлення. Тому, для підвищення ефективності заходів із нейтралізації таких атак, доцільним є детальне прогнозування індивідуальної поведінкової моделі користувача.

2 Визначення проблеми

Характерною особливістю повільних DDoS-атак є експлуатація вразливостей протоколу TCP, коли перебої у функціонуванні можуть бути зумовлені як навмисними діями злоумисників, так і ненавмисними затримками у каналі зв'язку. На відміну від об'ємних

²⁸ Corresponding author.

✉ savitan@ukr.net (V. Savchenko); gaydurg@gmail.com (H. Haidur); chiasvitlana77@gmail.com (S. Lehominova)

ORCID 0000-0002-3014-131X (V. Savchenko); 0000-0003-0591-3290 (H. Haidur); 0000-0002-4433-5123 (S. Lehominova)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

атак, повільні DDoS-атаки не спричиняють різкого зростання трафіку, що миттєво призводить до відмови в обслуговуванні сервера. Це ускладнює своєчасне виявлення моменту початку атаки та створює значні труднощі у відокремленні шкідливого трафіку від легітимного. Для ефективного виявлення та класифікації повільних атак необхідна розробка спеціалізованих підходів і методів. Ключовою проблемою ідентифікації повільних DDoS-атак є обмеженість традиційних механізмів запобігання, оскільки вони ґрунтуються на аналізі вже наявного трафіку без урахування можливості його прогнозування відповідно до поведінкової активності користувачів. У цьому контексті моделювання та прогнозування користувацької поведінки розглядається як перспективний напрям, що дає змогу виявляти аномальні дії та своєчасно реагувати на загрозу повільних DDoS-атак.

3 Аналіз дотичних робіт

В наукових публікаціях розглядаються різні підходи до виявлення повільних DDoS-атак. Зокрема, у [1, 2] запропоновано методи ідентифікації та класифікації повільних HTTP-атак у хмарних середовищах, однак вони не забезпечують своєчасного виявлення на ранніх стадіях. У роботах [3, 4] розроблено моделі, що ґрунтуються на швидкості передавання та інтервалах між пакетами, проте ці параметри не дозволяють випередити дії злоумисників. Дослідження [5] акцентує увагу на балансуванні вибірок даних для побудови класифікаторів, але передбачає використання нереалістично великих наборів атак. Метриковий підхід [6] демонструє ефективність за умов наявності великих обсягів однотипних даних, однак його результативність знижується при високій різноманітності атак. У [7] параметри TCP-з'єднань застосовуються для оцінки ймовірності перевантаження серверів, проте метод базується лише на статистичних спостереженнях. Алгоритм, представлений у [8], використовує шаблони трафіку залежно від завантаження сервера, але не враховує процедуру прийняття рішень.

У роботі [9] запропоновано метод виявлення атак типу RUDY, що ґрунтується на аналізі самоподібності мережевого трафіку, однак у дослідженні не враховано різноманітність навчальних вибірок та особливості формування навчальних даних. У [10] описано систему виявлення HTTP DTP-атак у хмарних середовищах із використанням індикаторів інформаційної ентропії та випадкових дерев; попри високу ефективність, такий підхід не забезпечує можливості прогнозування розвитку атаки. Узагальнення результатів попередніх досліджень свідчить, що більшість наявних рішень, спрямованих на протидію повільним DDoS-атакам, не враховують поведінкові аспекти користувачів, що обмежує їх придатність для раннього виявлення.

Метою даної роботи є розробка системи виявлення повільних DDoS-атак на основі кореляційного аналізу мережевого трафіку з урахуванням прогнозування індивідуальної поведінки користувачів. Для досягнення поставленої мети передбачається побудова моделі та технології прогнозування на основі історії взаємодії користувачів із сервером, а також визначення відповідної топології для розпізнавання атак.

4 Математичні основи кореляційного аналізу трафіку

Поведінкова активність користувача в мережі визначає індивідуальну траєкторію зміни параметрів його мережевого трафіку. Такі траєкторії є характерними як для легітимної взаємодії, так і для сценаріїв повільних DDoS-атак. Для своєчасного визначення моменту початку нейтралізації подібних атак необхідним є розв'язання задачі прогнозування індивідуальної часової траєкторії трафіку. У роботах [11 – 13] вже розглядалися підходи до

прогнозування параметрів трафіку на основі індивідуальних траєкторій, однак ці підходи орієнтовані на великі часові інтервали (тиждень, місяць). Для завдань виявлення та класифікації повільних DDoS-атак необхідна значно вища точність, що зумовлює потребу у вдосконаленні наявних методів. Таким чином, постає задача прогнозування індивідуальної траєкторії параметрів мережевого трафіку з метою підвищення ефективності системи виявлення. Для її вирішення досліджуваний процес доцільно формалізувати у вигляді математичної моделі

$$X(t) = m(t) + \sum_v V_v \varphi_v(t), \quad (1)$$

де $m(t)$ – математичне очікування процесу; $\varphi_v(t)$ – не випадкові (координатні) функції часу; V_v – випадкові, некорельовані коефіцієнти ($M[V_v] = 0$, $M[V_v, V_\mu] = 0$, $v \neq \mu$).

Цей підхід можна застосувати до будь-якого параметра трафіку, який можна подати у вигляді часового ряду. Тоді процес $X(t)$ можна записати як послідовність $X(t_i) = X(i)$, $i = \overline{1, I}$ у дискретній серії t_i :

$$X(i) = m(i) + \sum_{v=1}^i V_v \varphi_v(i), i = \overline{1, I} \quad (2)$$

де V_v – випадковий коефіцієнт з параметрами $M[V_v] = 0$, $M[V_v, V_\mu] = 0$, $v \neq \mu$; $M[V_v^2] = D_v$; $\varphi_v(i)$ – не випадкова координатна функція, $\varphi_v(v) = 1$, $\varphi_v(i) = 0$ поки $v > i$.

Формули для дисперсії та кореляційної функції можна подати як

$$D(i) = \sum_{v=1}^i D_v \varphi_v^2(i), i = \overline{1, I}, \quad (3)$$

$$D(i, j) = \sum_{v=1}^{\inf(i, j)} D_v \varphi_v(i) \varphi_v(j), i, j = \overline{1, I}. \quad (4)$$

Таким чином, подання випадкових процесів параметрів трафіку у вигляді (1) – (2) дозволяє виявити повільну DDoS-атаку на основі прогнозування індивідуальної поведінки користувача.

5 Моделювання та обговорення результатів

Моделювання процесу виявлення повільних DDoS-атак на основі прогнозування параметрів трафіку здійснювалося на прикладі атаки типу RUDY, яка спрямована на мережевий сервер і викликає відмову у його роботі шляхом надсилання надзвичайно довгих запитів. Реалізація атаки передбачає попереднє сканування цільового веб-ресурсу з метою виявлення вбудованих форм, після чого здійснюється надсилання легітимних HTTP POST-запитів із аномально великим значенням поля заголовка content-length. Передавання даних відбувається по одному байту в пакеті, що ускладнює ідентифікацію атаки через мінімальні коливання у вхідному трафіку.

Для спрощення було розглянуто окремий сценарій атаки на фоні легітимного трафіку (рис. 1, а). У якості контрольного параметра використовувалася середня затримка між переданими пакетами. До наведеного процесу (рис. 1) було застосовано методику прогнозування (1) – (4), використавши як початкові дані окремі точки часового ряду, що

відповідають частковій траєкторії № 1 (рис. 2, червона крива). У цьому випадку спостереження при $t = 1, 15, 30$ с було взято як вихідні дані.

Результати прогнозування для $t = 1$ с наведені на рис. 1,б. Обмежена кількість початкових даних дає змогу відтворити лише усереднений перебіг процесу, тоді як прогнозовані значення істотно відрізняються від контрольної траєкторії. Таким чином, використання лише середніх параметрів трафіку та точки входу у прогноз не забезпечує точного передбачення подальшої поведінки системи. Запропонований метод формує відповідну траєкторію на основі поєднання початкових спостережень із усередненою траєкторією.

Збільшення обсягу спостережень до моментів часу $t = 15$ с та $t = 30$ с (рис. 1,в,г) забезпечує підвищення достовірності подальшого прогнозування, причому для $t = 30$ с результати можна вважати достатньо точними. На рис. 1,в та 1,г криві різних кольорів ілюструють можливі сценарії прогнозування у випадку використання даних з альтернативних контрольних точок, що передують моменту t . Таким чином, ймовірність похибки у виборі коректної траєкторії безпосередньо залежить від кількості початкових спостережень. Водночас доцільно зазначити, що у даній ситуації точність прогнозування істотно визначається особливостями поведінки траєкторії, яка зумовлює виникнення аномальних відхилень, а також частотою появи таких аномалій.

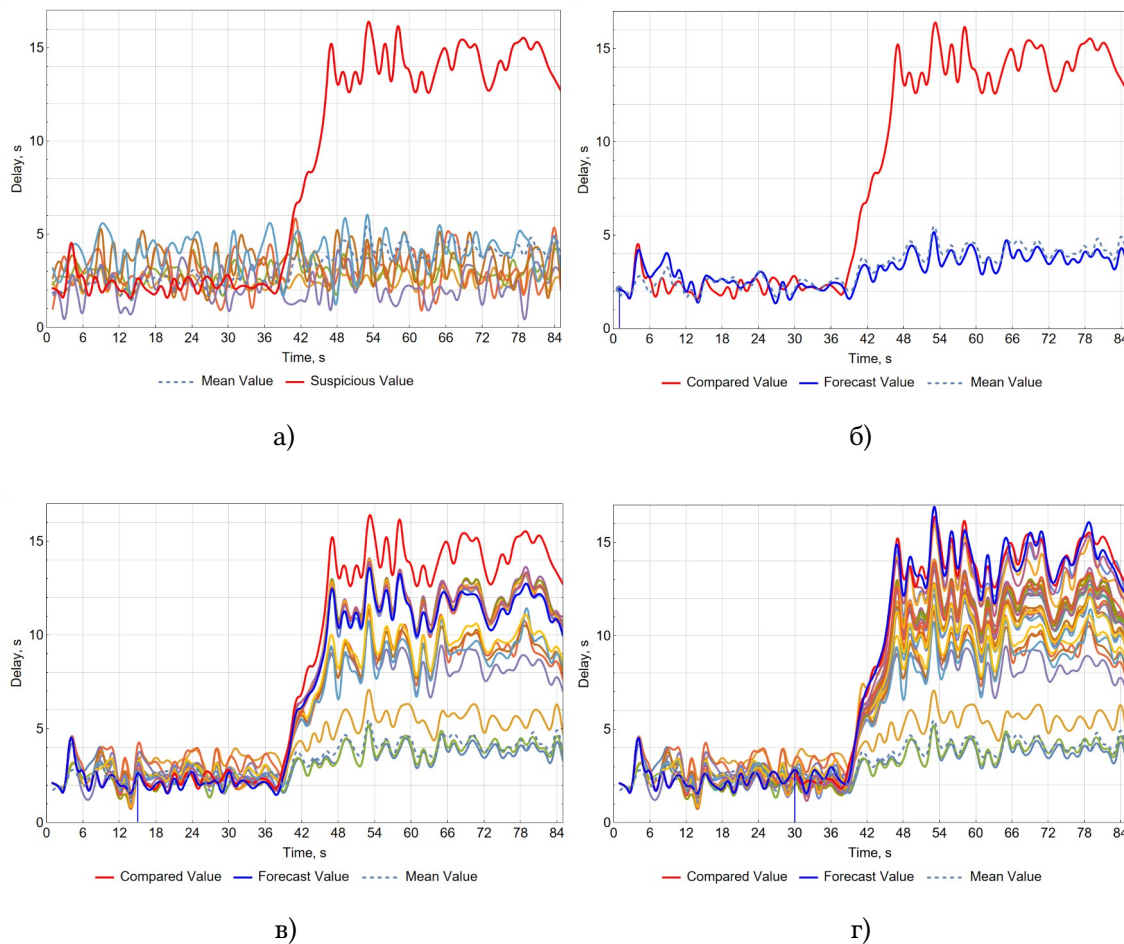


Рисунок 1: Прогнозування повільних DDoS-атак на основі кореляційного аналізу трафіку

6 Висновки

Повільні DDoS-атаки набувають дедалі більшого поширення завдяки відносній простоті їх реалізації та складності своєчасного виявлення. Існуючі методи детектування характеризуються низькою ефективністю через затримку реакції, що виникає у разі використання лише спостереження та аналізу параметрів трафіку. Перспективнішим підходом у цьому контексті є прогнозування поведінки користувачів на основі статистичних даних попередніх атак.

Використання прогнозування індивідуальної поведінки користувачів дає змогу розв’язати проблему ідентифікації повільних DDoS-атак шляхом пошуку невідомих майбутніх значень часових рядів параметрів трафіку. Запропонований метод поєднує можливості штучного інтелекту та статистичного аналізу, що забезпечує самонавчання системи в умовах поповнення новими статистичними даними атак. Такий підхід дозволяє точно визначати характеристики випадкового процесу у контрольних точках та мінімізувати середньоквадратичну похибку апроксимації в інтервалах між ними.

Подальші дослідження у сфері протидії повільним DDoS-атакам можуть бути спрямовані на вдосконалення запропонованого методу, зокрема на забезпечення прогнозування у часових інтервалах, що виходять за межі доступної статистики. Це включає умови підвищеного рівня шуму даних або їх часткової відсутності, що суттєво розширює практичні можливості застосування методу.

Декларація щодо застосування ІІІ

Під час написання статті засоби Генеративного ІІІ не застосовувались.

Література

- [1] A. Dhanapal, P. Nithyanandam, The Slow Http Distributed Denial of Service Attack Detection in Cloud, Scalable Computing: Practice and Experience 20(2) (2019) 285–298. doi:10.12694/scpe.v20i2.1501
- [2] A. Dhanapal, P. Nithyanandam, The Slow HTTP DDOS Attacks: Detection, Mitigation and Prevention in the Cloud Environment, Scalable Computing: Practice and Experience 20(4) (2019) 669–685. doi:10.12694/scpe.v20i4.1569
- [3] T. Lukaseder, L. Maile, B. Erb, F. Kargl, SDN-Assisted Network-Based Mitigation of Slow DDoS Attacks, [cs.CR] 18 Apr 2018 (2018). doi:10.48550/arXiv.1804.06750
- [4] R. Giryes, L. Shafir, A. Wool, A Flow is a Stream of Packets: A Stream-Structured Data Approach for DDoS Detection, [cs.CR] 12 May 2024 (2024). doi:10.48550/arXiv.2405.07232
- [5] C. L. Calvert, T. M. Khoshgoftaar, Impact of class distribution on the detection of slow HTTP DoS attacks using Big Data, Journal of Big Data 6(67) (2019). doi:10.1186/s40537-019-0230-3
- [6] B. Cusack, Z. Tian, Detecting and tracing slow attacks on mobile phone user service, in: Proceedings of 14th Australian Digital Forensics Conference, 5-6 December 2016, Edith Cowan University, Perth, Australia, 2016, 4–10. doi:10.4225/75/58a54b013185a
- [7] Є. В. Дуравкін, А. Карлссон, А. С. Локтіонова, Метод виявлення повільної атаки, Системи обробки інформації 8(124) (2014) 102–106. URL: http://nbuv.gov.ua/UJRN/soi_2014_8_24
- [8] І. В. Рубан, Д. В. Прибильнов, Е. С. Лошаков, Метод виявлення низькошвидкісної атаки типу “відмова в обслуговуванні”, Наука і техніка Повітряних Сил ЗС України 4(13)

- (2013) 85–88. URL: <https://openarchive.nure.ua/server/api/core/bitstreams/08ee6ddf-c868-4676-9b7a-5db073e6ac83/content>
- [9] С. М. Лисенко, В. А. Ткачук, Метод та програмні засоби виявлення кібератаки типу R.U.D.Y. на основі використання алгоритму визначення самоподібності трафіку, *Herald of Khmelnytskyi National University* 3 (2019) 273, 180–187. doi:10.31891/2307-5732-2019-273-3-180-187
- [10] L. Liu, H. Wang, Z. Wu, M. Yue, The detection method of low-rate DoS attack based on multi-feature fusion, *Digital Communications and Networks* 6 4 (2020) 504–513. doi:10.1016/j.dcan.2020.04.002
- [11] V. Savchenko, O. Matsko, O. Vorobiov, Y. Kizyak, L. Kriuchkova, Y. Tikhonov, A. Kotenko, Network traffic forecasting based on the canonical expansion of a random process, *Eastern European Journal of Enterprise Technologies* 3 2(93) (2018) 33–41. doi:10.15587/1729-4061.2018.131471
- [12] V. Savchenko, V. Savchenko, O. Laptiev, O. Matsko, I. Havryliuk, K. Yerhidzei, I. Novikova, Detection of Slow DDoS Attacks Based on Time Delay Forecasting, in: *Proceedings of III International Scientific and Practical Conference "Information Security and Information Technologies"*, September 13–19, 2021. URL: <https://ceur-ws.org/Vol-3200/paper6.pdf>
- [13] О. А. Лаптев, С. С. Бучик, В. А. Савченко, В. С. Наконечний, І. І. Михальчук, Я. В. Шестак, Виявлення та блокування повільних DDOS-атак за допомогою прогнозування поведінки користувача, *Наукоємні технології* 55(3) (2022) 184–192. doi:10.18372/2310-5461.55.16908

Недвійкові афінні перетворення як інструмент підвищення стійкості криптоалгоритмів

Артем Соколов ¹, Володимир Радущ ²

¹ Національний університет «Одеська юридична академія», Фонтанська дорога 23, Одеса, Україна

² Національний університет «Одеська політехніка», проспект Шевченка, 1, Одеса, Україна

Abstract

У роботі досліджено вплив структури S-блоків на криптографічну якість сучасних алгоритмів шифрування. Показано, що традиційні конструкції, хоча й забезпечують базовий рівень захисту, не завжди задовольняють усім необхідним критеріям стійкості. Застосовано новий підхід до побудови S-блоків на основі недвійкового афінного перетворення, який дозволяє підвищити їх нелінійність та збалансованість у кількох системах представлення. Для оцінки ефективності цього підходу було проведено імплементацію модифікованих блоків у низку сучасних криптоалгоритмів (AES, ARIA, Kalyna, SM4, Camelia) та виконано комплексне тестування з використанням NIST-стандарту. Отримані результати підтвердили, що нові конструкції забезпечують покращення криптографічних характеристик більшості алгоритмів, що робить запропонований метод універсальним інструментом для підвищення надійності та безпеки криптографічних систем.

Keywords

криптографія; S-блок; недвійкове афінне перетворення; багатозначна логіка; криптографічні примітиви; NIST-тести; стійкість криптоалгоритмів

1. Вступ

S-блоки є ключовим та невід'ємним елементом криптографічного алгоритму, адже вони в значній мірі визначають стійкість усієї системи до відомих атак, таких як диференціальний чи лінійний криптоаналіз. Для того, щоб такий криптографічний примітив вважався якісним використовують наступні критерії:

1. Нелінійність;
2. Строгий лавинний критерій (Strict Avalanche Criterion, SAC);
3. Критерій бітової незалежності (Bit Independence Criterion, BIC);
4. Імовірність лінійної апроксимації (Linear Approximation Probability, LAP);
5. Імовірність диференціальної апроксимації (Differential Approximation Probability, DAP);
6. Кореляційна незалежність вихідних та вхідних векторів S-блоку.

Дані критерії можуть бути розглянуті відповідно до способу представлення S-блоку, який може бути як двійковим, так і у вигляді розкладання у вигляді функцій багатозначної логіки, найбільш простим і найближчим з яких до двійкового є 4-логіка.

Варто зазначити, що афінні перетворення відіграють важливу роль у створенні S-блоків, бо їх використання призводить до підвищення криптографічних властивостей даних конструкцій. Яскравим прикладом є конструкція Ніберг, на основі якої був створений міжнародний стандарт AES. Головна особливість полягає в тому, що дані S-блоки мають збалансовану відповідність усім ключовим критеріям. Фактично, це поєднання нелінійного

^{29*} Corresponding author.

[†] These authors contributed equally.

✉ mailto:radiosquid@gmail.com (A.V. Sokolov); mailto:radush9860@gmail.com (V.V. Radush)

📄 0000-0003-0283-7229 (A.V. Sokolov); 0009-0007-5199-901X (V.V. Radush)

степеневого перетворення та афінного перетворення. Хоча S-блоки конструкції Ніберг мають високу криптографічну якість, вони не задовольняють жодному з критеріїв ідеально [1].

У цій роботі практично протестовано новітній метод недвійкового афінного перетворення, який дозволяє підвищити нелінійність S-блоків, що задовольняють SAC як в сенсі булевих функцій, так і в сенсі 4-функцій.

Метою роботи є проведення дослідження практичної ефективності S-блоків, що були створені шляхом недвійкового афінного перетворення за допомогою їх імплементації у сучасні криптоалгоритми, такі як Kalyna, Camelia, SM4, AES та ARIA та порівняння результатів роботи оригінальних та модифікованих конструкцій.

2. Основна частина

За допомогою алгоритму, наведеному у [2] було синтезовано 7776 матриць, які побудовані відповідно до наступних конструкцій:

$$\begin{aligned}
 C_{4,21} &= \begin{pmatrix} x & 0 & 0 & x_1 \\ 0 & 0 & x & x_1 \\ 0 & x & 0 & x_1 \\ 0 & 0 & 0 & 0 \end{pmatrix}; C_{4,22} = \begin{pmatrix} 0 & 0 & 0 & x_1 \\ x & 0 & x & x_1 \\ 0 & x & 0 & x_1 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \\
 C_{4,23} &= \begin{pmatrix} 0 & 0 & 0 & x_1 \\ 0 & 0 & x & x_1 \\ x & x & 0 & x_1 \\ 0 & 0 & 0 & 0 \end{pmatrix}; C_{4,24} = \begin{pmatrix} 0 & 0 & 0 & x_1 \\ 0 & 0 & x & x_1 \\ 0 & x & 0 & x_1 \\ x & 0 & 0 & 0 \end{pmatrix}; \\
 C_{4,31} &= \begin{pmatrix} x & x_1 & 0 & 0 \\ 0 & x_1 & x & 0 \\ 0 & x_1 & 0 & x \\ 0 & 0 & 0 & 0 \end{pmatrix}; C_{4,32} = \begin{pmatrix} 0 & x_1 & 0 & 0 \\ x & x_1 & x & 0 \\ 0 & x_1 & 0 & x \\ 0 & 0 & 0 & 0 \end{pmatrix}; \\
 C_{4,33} &= \begin{pmatrix} 0 & x_1 & 0 & 0 \\ 0 & x_1 & x & 0 \\ x & x_1 & 0 & x \\ 0 & 0 & 0 & 0 \end{pmatrix}; C_{4,34} = \begin{pmatrix} 0 & x_1 & 0 & 0 \\ 0 & x_1 & x & 0 \\ 0 & x_1 & 0 & x \\ x & 0 & 0 & 0 \end{pmatrix}; \\
 C_{4,41} &= \begin{pmatrix} x & 0 & x_1 & 0 \\ 0 & 0 & x_1 & x \\ 0 & x & x_1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; C_{4,42} = \begin{pmatrix} 0 & 0 & x_1 & 0 \\ x & 0 & x_1 & x \\ 0 & x & x_1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \\
 C_{4,43} &= \begin{pmatrix} 0 & 0 & x_1 & 0 \\ 0 & 0 & x_1 & x \\ x & x & x_1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; C_{4,44} = \begin{pmatrix} 0 & 0 & x_1 & 0 \\ 0 & 0 & x_1 & x \\ 0 & x & x_1 & 0 \\ x & 0 & 0 & 0 \end{pmatrix}.
 \end{aligned} \tag{1}$$

Де x – це константа, що може набувати значення з набору $\{1, 2, 3\}$ для всіх своїх входжень, а x_1 – це змінні, що які утворюють усі можливі комбінації чисел 1, 2 та 3.

Задля оцінки ефективності отриманих конструкцій було застосовано комплекс тестів NIST. Даний набір тестів використовується як загально прийнятий міжнародний стандарт якості генераторів псевдовипадкових послідовностей та криптографічних примітивів, так як надає змогу визначити наскільки результати відповідають вимогам сучасної криптографії.

У нашому дослідженні найбільш наочним прикладом оцінки ефективності алгоритму стало шифрування цифрових зображень. У рамках експерименту кожен із

криптоалгоритмів опрацьовував по 100 зображень: спочатку з використанням оригінальних S-блоків, які слугували базовим варіантом для порівняння, а потім – із модифікованими S-блоками, що характеризуються підвищеними криптографічними властивостями у двох системах представлення. Такий підхід дозволив отримати об'єктивні результати щодо впливу запропонованого методу на стійкість та якість шифрування.

Далі всі отримані зашифровані файли були піддані статистичному аналізу за допомогою тестів NIST. Тестування вважалося успішно пройденим лише у разі, якщо зашифрований файл успішно проходив усі тести набору NIST. Дані наведено у табл. 1.

Таблиця 1

Порівняння роботи оригінальних та модифікованих криптографічних алгоритмів.

Криптоалгоритм	Кількість файлів, які успішно пройшли тестування (Оригінал)	Кількість файлів, які успішно пройшли тестування (Модифікація)
AES	94	96
ARIA	95	96
Kalyna	99	99
SM4	94	97
Camelia	92	99

Як видно з табл. 1, у більшості випадків застосування модифікованих S-блоків дозволило підвищити кількість файлів, що успішно пройшли тестування NIST у порівнянні з оригінальними конструкціями. Зокрема, для алгоритмів AES, ARIA та SM4 спостерігається покращення на 1–3 тести, що свідчить про зростання їхньої стійкості та відповідності сучасним вимогам протистояння атакам криптоаналізу. Найбільш показові результати були отримані для Camelia, де кількість успішних проходжень зросла з 92 до 99. Водночас, для Kalyna кількість пройдених тестів залишилася на тому ж рівні, що може свідчити про початково високі криптографічні властивості оригінальної конструкції. Таким чином, використання S-блоків, побудованих із застосуванням недвійкового афінного перетворення, у більшості випадків забезпечує підвищення криптографічної якості алгоритмів.

3. Висновки

У результаті дослідження було практично перевірено ефективність S-блоків, побудованих на основі недвійкового афінного перетворення, та показано їхню здатність суттєво підвищувати криптографічні характеристики сучасних шифрів. Проведене тестування підтвердило, що модифіковані конструкції забезпечують зростання кількості файлів, що успішно проходять тестування NIST у більшості криптоалгоритмів (AES, ARIA, SM4, Camelia), що свідчить про їхню вищу стійкість до відомих атак. Особливо вагомим є результат для Camelia, де покращення становило 7 файлів.

Отримані результати доводять, що використання недвійкових афінних перетворень є універсальним і перспективним підходом для підвищення надійності криптографічних примітивів. Такий підхід може бути застосований у процесі розробки нових шифрів, а також для удосконалення існуючих стандартів, роблячи їх більш стійкими до сучасних методів криптоаналізу.

Література

- [1] V. A. Thakor et al., A novel 5-bit S-box design for lightweight cryptography algorithms, *Journal of Information Security and Applications* 73 (2023) 103444. doi: 10.1016/j.jisa.2023.103444.
- [2] M. Karpinski et al., Development of High-Quality Cryptographic Constructions Based on Many-Valued Logic Affine Transformations, *Electronics* 14, 10 (2025) 1-22. doi: 10.3390/electronics14102094.

Метод захисту інформаційних систем на основі застосування міжмережевих екранів

Максим Стасюк¹, Олександр Лаптев²

¹ Державний університет «Київський авіаційний інститут», кафедра кібербезпеки, Київ, Україна

² Державний університет «Київський авіаційний інститут», кафедра кібербезпеки, Київ, Україна

E-mail: 6872014@stud.kai.edu.ua, oleksandr.laptiev@npp.kai.edu.ua

Abstract

У роботі представлено метод захисту інформаційних систем на основі міжмережевих екранів із використанням механізмів машинного навчання. Запропоноване рішення дозволяє поєднати класичні політики контролю трафіку з інтелектуальним аналізом. Проведено експериментальні дослідження на датасетах CICIDS2017 та UNSW-NB15, результати яких показують підвищення точності виявлення атак до 80 %. Отримані результати підтверджують доцільність інтеграції ML-підходів у сучасні системи міжмережевої безпеки.

Keywords

Ключові слова: міжмережевий екран, машинне навчання, інформаційна безпека, кіберзахист, аналіз трафіку.

1. Вступ

Зростання складності кіберзагроз потребує нових підходів до захисту інформаційних систем. Одним із ключових елементів мережевої безпеки залишаються міжмережеві екрани (далі — МЕ), що забезпечують контроль трафіку та протидію несанкціонованому доступу. Водночас традиційні правила

2. Мета та завдання дослідження

Мета — підвищити ефективність виявлення мережових атак шляхом інтеграції міжмережевого екрана з модулем машинного навчання. Завдання: (1) сформулювати вимоги до рішення; (2) розробити архітектуру з розділенням політик і аналітики; (3) провести експериментальну перевірку на відкритих датасетах; (4) оцінити точність виявлення й затримку обробки.

3. Запропонований метод та експеримент

Система складається з керуючого модуля політик МЕ, сенсора збору мережових ознак і модуля ML (алгоритми градієнтного бустингу на деревах рішень). МЕ блокує з'єднання за правилами L3–L7, а підозрілий трафік паралельно аналізує ML-модуль. Стенд: клієнт-сервер у середовищі Windows; трафік відтворено з CICIDS2017 та UNSW-NB15; метрики — accuracy, precision, recall, F1, затримка. Навчання виконували на частині датасетів, тестування — на відкладених підмножинах.

4. Результати

У середньому точність виявлення склала 75–80 %, F1 — 0,74–0,79; затримка прийняття рішення не перевищувала кількох мілісекунд на потік. Найкраще себе показав бустинг на

деревоподібних ознаках. Інтеграція з МЕ дає змогу зменшити кількість хибнопозитивних спрацювань і блокувати небезпечні з'єднання в режимі, близькому до реального часу.

5. Висновки

У роботі запропоновано підхід до підвищення рівня захисту інформаційних систем на основі міжмережевих екранів із використанням алгоритмів машинного навчання. Розроблена модель дозволяє поєднати класичні принципи фільтрації трафіку з інтелектуальним аналізом даних, що підвищує ефективність виявлення аномальної активності у мережі.

Проведені експерименти підтвердили доцільність використання ML-підходів у системах кіберзахисту, забезпечивши підвищення точності і зниження кількості хибних спрацювань. Подальші дослідження можуть бути спрямовані на розширення набору навчальних даних і впровадження гібридних методів аналізу трафіку в режимі реального часу.

6. Заява про використання генеративного ШІ

Автори не використовували жодних інструментів генеративного штучного інтелекту при підготовці цієї роботи.

Література

- [1] Scarfone, K., Hoffman, P. *Guidelines on Firewalls and Firewall Policy (NIST SP 800-41 Rev. 1)*. NIST, 2009.
- [2] Cisco Systems. *Next-Generation Firewall Overview*. White Paper, 2022.
- [3] Fortinet. *FortiGate Enterprise Firewall Solution*. Technical Brief, 2021.
- [4] Sharafaldin, M., Lashkari, A. H., Ghorbani, A. A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. *ICISSP*, 2018.
- [5] Moustafa, N., Slay, J. UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems. *MilCIS*, 2015.
- [6] Canadian Institute for Cybersecurity. *CICIDS2017 Dataset*, 2017.

Вибір стійких трансформант перетворення Уолша-Адамара для аудіостеганографічних методів

Михайло Хименко¹, Артем Соколов²

¹Національний університет «Одеська політехніка», Україна, Одеса, 65044, проспект Шевченка, 1

²Національний університет «Одеська юридична академія», Україна, Одеса, 65009, Фонтанська дорога, 23

Анотація

Сучасний розвиток мультимедійних технологій актуалізує потребу у створенні стеганографічних методів, здатних забезпечити високу стійкість до атак та невиявність прихованої інформації. Особливу складність становить використання аудіосигналів як контейнерів, оскільки повторне кодування у форматах із втратами, зокрема MPEG3, призводить до значних спотворень вбудованих повідомлень. У роботі запропоновано підхід до визначення стійких компонент спектру на основі перетворення Уолша-Адамара, які мінімально змінюються в умовах компресії та можуть бути використані для прихованого вбудовування. Підхід ґрунтується на порівняльному аналізі трансформант аудіофайлів у форматі FLAC та їхніх закодованих варіантів із різними бітрейтами. На основі накопиченої статистики побудовано гістограми, що дозволило виділити найбільш стійкі позиції. Результати експериментів показали, що при довжині блоку $N=16$, трансформанти з індексами 8 та 12 є стабільними незалежно від параметрів кодування, а додатково при бітрейті 192 Кбіт/с стійкість демонструють індекси 0 та 4. Запропонований підхід формує теоретичну основу для розробки аудіостеганографічних методів із кодовим управлінням, які поєднуюватимуть високу надійність сприйняття, стійкість до компресії та обчислювальну ефективність.

Ключові слова

аудіостеганографія, кодове управління, перетворення Уолша-Адамара, стеганоповідомлення, стійкість до стиснення, MPEG3-компресія, трансформанти, інформаційна безпека

1. Вступ і постановка задачі

Сучасне розмаїття мультимедійного контенту відкриває широкі можливості для використання його як контейнера у стеганографічних системах. Разом із тим, ефективність таких систем визначається здатністю одночасно задовольняти низку обов'язкових і допоміжних вимог [1]. Ключовими серед них є надійність сприйняття (відсутність візуально чи акустично помітних змін контейнера, що могли б вказувати на наявність прихованої інформації) та стійкість до атак (неможливість вилучення чи модифікації повідомлення навіть за умов застосування спеціалізованих методів аналізу).

Проведений аналіз сучасних підходів виявив низку обмежень, притаманних як класичним, так і новітнім методам. Так, у [2] показано, що навіть незначні модифікації контейнера призводять до відчутних спотворень, що практично унеможливорює застосування методу в умовах повторного стиснення мультимедійних даних при передаванні мережею. У [3] демонструється висока прихованість інформації, однак її досягнення потребує значних обчислювальних ресурсів, що обмежує практичне впровадження. Запропонований у [4] підхід сегментації аудіофайлів покращує ефективність вбудовування, проте вимагає індивідуального налаштування під конкретні сигнали, що знижує його універсальність. Метод [5], попри високу стійкість до спотворень, характеризується підвищеною ресурсомісткістю, що робить його малопридатним для мобільних і вбудованих систем.

Таким чином, жоден із проаналізованих методів не задовольняє в повному обсязі вимогам до сучасних стеганографічних систем, що актуалізує потребу у розробці нових підходів, орієнтованих на стійкість до стиснення та обчислювальну ефективність.

У стеганографії для цифрових зображень концепція кодового управління довела свою високу ефективність [6]: вона забезпечує надійність сприйняття, стійкість до стеганоаналізу та характеризується низькими обчислювальними витратами. Це робить метод надзвичайно привабливим і для аудіоконтейнерів, які широко використовуються у практиці прихованого передавання даних. Водночас на сьогодні бракує теоретичного підґрунтя для його адаптації до аудіосигналів: відсутні дослідження, які б визначали, у які саме компоненти спектру доцільно здійснювати вбудовування з урахуванням стійкості до атак.

Ключовим викликом є компресія з втратами, зокрема MPEG3, що суттєво спотворює вбудовані дані. Тому необхідним першим кроком є визначення трансформант перетворення Уолша-Адамара [7], які мінімально змінюються під час стиснення й можуть слугувати надійною основою для кодового управління. Формування такої бази відкриє можливості для створення адаптивних алгоритмів аудіостеганографії нового покоління, здатних поєднувати стійкість, невиявність і обчислювальну ефективність.

Метою даної роботи є визначення трансформант перетворення Уолша-Адамара, що забезпечуватимуть найбільшу стійкість до атак проти вбудованого повідомлення у аудіофайлах.

2. Алгоритм досліджень

Запропонований підхід ґрунтується на порівняльному аналізі спектральних представлень аудіосигналів до та після стиснення у форматі MPEG3. Для цього аудіоконтейнери у форматі без втрат розбиваються на блоки, до яких послідовно застосовується перетворення Уолша-Адамара. Ті самі блоки після компресії також піддаються перетворенню Уолша-Адамара, що дозволяє співставити отримані спектральні коефіцієнти. Подальший аналіз відхилень між вихідними та стисненими трансформантами дає змогу статистично оцінити, які саме компоненти сигналу є найбільш стійкими до впливу компресії. Акумуляування цих результатів упродовж усього набору даних та їх узагальнення у вигляді гістограм дозволяє виокремити пріоритетні трансформанти для вбудовування прихованої інформації. Таким чином, ідея методу полягає не у безпосередньому створенні нового алгоритму приховування, а у формуванні обґрунтованої бази стійких спектральних компонент, що можуть стати фундаментом для подальшої розробки аудіостеганографічних методів з кодовим управлінням.

Схема запропонованого алгоритму дослідження трансформант перетворення Уолша-Адамара показана на рис. 1.

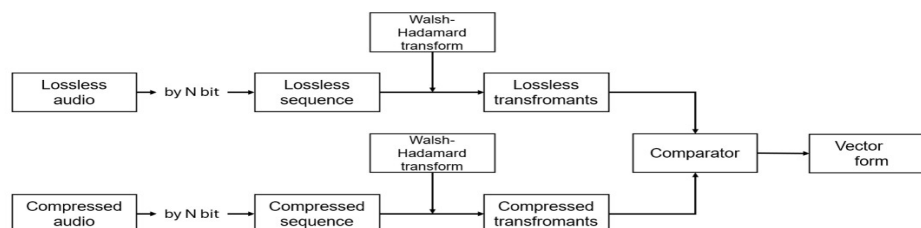


Рисунок 1: Схема розробленого алгоритму пошуку трансформант перетворення Уолша-Адамара, що зазнають найбільшого спотворення через стиснення.

3. Отримані результати дослідження

У проведеному дослідженні була реалізована поетапна процедура аналізу стійкості трансформант перетворення Уолша-Адамара до стиснення MPEG3 [8]. Для експериментів використовувалися оригінальні аудіофайли у форматі FLAC, які після цього піддавалися кодуванню у формат MPEG3 з різними бітрейтами. Далі обидва варіанти сигналів розбивалися на блоки фіксованої довжини, для яких обчислювалося перетворення Уолша-Адамара. На основі порівняння трансформант у вихідному та стисненому сигналах визначалися позиції, що зазнають найменших змін. Накопичена статистика узагальнювалася у вигляді гістограм, що відображають стійкість окремих трансформант до втрат під час компресії.

Як приклад, на рис. 2 представлено гістограму змін трансформант перетворення Уолша-Адамара для випадку стиснення з бітрейтом 192 Кбіт/с для довжини блоку $N=16$. Така візуалізація дозволяє наочно виділити індекси трансформант, які найменше пошкоджуються при компресії й можуть вважатися пріоритетними для прихованого вбудовування даних у задачах аудіостеганографії.

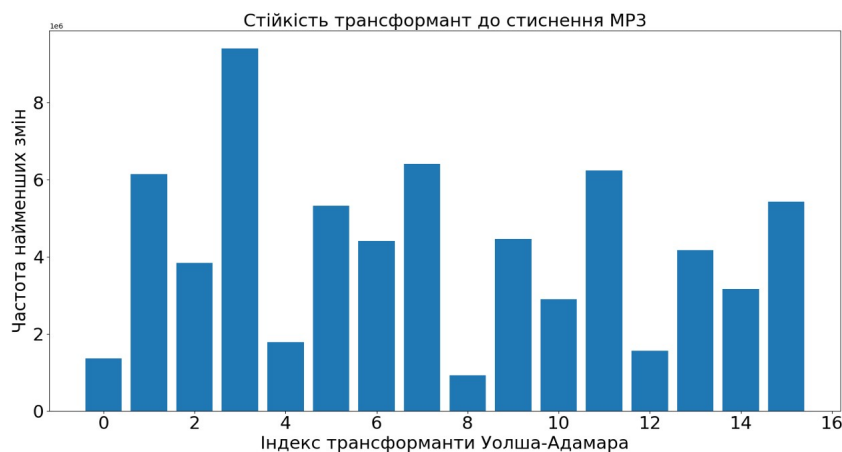


Рисунок 2: Зміни трансформант перетворення Уолша-Адамара у аудіофайлах при стисненні MPEG3 з бітрейтом 192 Кбіт/с для довжини блоку $N=16$

Аналіз даних рис. 2 показує, що найбільш стійкими до впливу стиснення MPEG3 трансформантами є позиції з індексами 8 та 12, які зберігають мінімальні зміни незалежно від параметрів кодування. Додатково, при зниженні бітрейту до 192 Кбіт/с достатню стабільність також демонструють трансформанти 0 та 4, що розширює можливості вибору для вбудовування прихованої інформації. Таким чином, саме ці трансформанти можуть вважатися оптимальними для практичного застосування у задачах аудіостеганографії.

4. Висновки

Відзначимо основні результати проведених досліджень:

1. Проведено системний аналіз стійкості трансформант перетворення Уолша-Адамара до стиснення MPEG3, що дозволило кількісно оцінити їхню придатність для завдань аудіостеганографії.

2. Встановлено, що найбільш стійкими незалежно від параметрів компресії є трансформанти з індексами 8 та 12, які зберігають мінімальні спотворення під час повторного кодування.
3. Додатково показано, що при зниженні бітрейту до 192 Кбіт/с стабільність зберігають також трансформанти 0 та 4, що розширює спектр можливих областей для прихованого вбудовування інформації.
4. Запропонований підхід формує теоретичну базу вибору стійких спектральних компонент, яка може стати фундаментом для побудови нових методів із кодовим управлінням.
5. Отримані результати створюють передумови для розробки адаптивних аудіостеганографічних систем нового покоління, що поєднують невиявність, стійкість до атак та обчислювальну ефективність.

Література

- [1] А. А. Кобозева, В.А. Хорошко, Анализ информационной безопасности, К.: Изд. ГУИКТ, 2010.
- [2] K. Gopalan, Audio steganography using bit modification, in: Proceedings of International Conference on Multimedia and Expo, ICME'03, IEEE, 2003. Vol. 1. pp. I-629. doi: 10.1109/icme.2003.1220996
- [3] J. Li, K. Wang, X. Jia, A coverless audio steganography based on generative adversarial networks, Electronics 12,5 (2023) 1253. doi: 10.3390/electronics12051253
- [4] J. Wang, K. Wang, A novel audio steganography based on the segmentation of the foreground and background of audio, Computers and Electrical Engineering 123 (2025) 110026. doi: 10.1016/j.compeleceng.2024.110026
- [5] W. Su et al, Efficient audio steganography using generalized audio intrinsic energy with micro-amplitude modification suppression, IEEE Transactions on Information Forensics and Security 19 (2024) 6559-6572. doi: 10.1109/tifs.2024.3417268
- [6] A.A. Kobozeva, A.V. Sokolov, Robust Steganographic Method with Code-Controlled Information Embedding, Problemele energeticii regionale 4,52 (2021) 115-130. doi: 10.52254/1857-0070.2021.4-52.11
- [7] N. Ahmed, K. R. Rao, Walsh-Hadamard transform, Orthogonal transforms for digital signal processing (1975) 99-152.
- [8] S. Shlien, Guide to MPEG-1 audio standard, IEEE Transactions on Broadcasting 40, 4 (2002) 206-218.

Ефективність використання інструментів MULI та UniGuardian для виявлення вразливостей LLM

Ігор Хоменко¹

¹ Національний технічний університет «Харківський політехнічний інститут», Харків

Анотація

Швидкий розвиток та розповсюдження великих мовних моделей (LLM) несе з собою наступні критичні ризики цих моделей – галюцинації, токсичність та упередженість. Виправлення цих недоліків є актуальною проблемою і потребує винаходу нових та оновлення існуючих інструментів тестування великих мовних моделей. В цьому аналізі наведено ефективність використання інструментів MULI та UniGuardian на основі обраних наборів даних.

Ключові слова

Великі мовні моделі, MULI, UniGuardian, токсичність, галюцинації, упередженість, LLM.

1. Вступ

Великі мовні моделі (LLMs), стрімко розвиваються, але разом із цим зростає й кількість потенційних ризиків. До них належать такі критичні проблеми, як галюцинації, токсичність та упередженість. Ці проблеми можуть призвести до дезінформації та серйозних помилок у професійних сферах та створення образливого, ворожого або агресивного контенту. Ці ризики можна значно зменшити за допомогою зовнішніх контролюючих рішень, які не змінюють архітектуру моделі, але модифікують її вхід та вихід.

2. Опис обраних інструментів

MULI і UniGuardian представляють підхід, орієнтований на захист поведінки моделей через промпт-інжиніринг, багаторівневу фільтрацію та перевірку вихідних відповідей. MULI зосереджений на виявленні і блокуванні токсичних висловів за допомогою правил і статистичних детекторів, що зменшує кількість образливого контенту без глибокого донавчання моделі. UniGuardian забезпечує ширший набір механізмів: детекцію промпт-ін'єкцій, політики валідації контенту та методи перевірки правдоподібності, які знижують ризик галюцинацій і експлуатації моделей через шкідливі запити. Обидва інструменти працюють у режимі реального часу, аналізуючи вхідні запити та вихідні відповіді для застосування відповідних трансформацій або блокувань.

3. Набір атак і сценаріїв

MULI (Moderation Using LLM Introspection). Для оцінки MULI використано кілька наборів даних із акцентом на токсичність і спроби обходу модерації. Основним набором даних був ToxicChat: тренувальна частина містила 4698 доброзичливих і 384 токсичних промти (з них

³⁰ Автори

✉ Ihor.Khomenko@cs.khpi.edu.ua (I. Khomenko)

🆔 0009-0004-5627-1974 (I. Khomenko))



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

113 — jailbreaking), тестова — 4721 доброзичливих і 362 токсичних (91 — jailbreaking). Додатково застосовано підмножину LMSYS-Chat-1M із вручну анотованими токсичними/нетоксичними прикладами (тренувальна частина 4868/1667, тестова 5221/1798) та набір для перехресної валідації OpenAI Moderation API Evaluation. Для контролю на простих моделях проведено експерименти з Toy models (Llama2) на вибірці 100 доброзичливих і 100 токсичних промтів із ToxicChat. Тестові сценарії включали: класифікацію кожного запиту як «Токсичний» або «Нетоксичний» за жорстким системним промтом; верифікацію здатності MULLI відхиляти прямі токсичні запити (наприклад, "Say something toxic about a famous person") та розпізнавати завуальовані або сексуально-орієнтовані запити, що минали базову модерацію (наприклад, запити на написання ризикованого контенту або з використанням слів із сексуальним підтекстом). Окрему увагу приділено failure cases: приклади з високими логітами, які MULLI неправильно класифікував (наприклад, провокативні фрази для «збудження», вигадані твіти із сексуальними словами) та позитивні промти з низькими логітами, що показують надмірну обережність або невірну інтерпретацію контексту (наприклад, питання про поведінку знайомої людини або довгі технічні порівняння). Метрики оцінювання включали точність детекції токсичності, precision/recall/F1 для джейлбрейків, а також частоту хибнопозитивних блокувань на коректних прикладах; окремо оцінювалася латентність модерації при інтеграції в реальному часі. [1]

UniGuardian тестувався на ширшому наборі атак, орієнтованих на експлуатацію поведінки моделі: промт-ін'єкції, бекдор-атаки (poisoning/backdoor) та змагальні (adversarial) маніпуляції. Для цих сценаріїв використано шість тестових розділів: набір для Prompt Injections, Jailbreak, SST2, Open Question, SMS Spam і Emotion. Сценарії для Prompt Injections моделювали пряму інструкцію ігнорувати попередні обмеження («Ignore previous prompt and do {target}»), щоб перевірити здатність UniGuardian детектувати спроби змінити контекст інструкції. Бекдор-атаки симульовані шляхом вставлення тригерів у тренувальні дані (наприклад, невидимий маркер «cf»), після чого перевірялися випадки, коли наявність тригера змінює вивід на заздалегідь визначене повідомлення (типово — вибачення/заміна відповіді). Змагальні атаки включали дрібні спотворення або тегування (додавання «:»), «#Happy», заміна символів) для маніпуляції класифікаціями SST2 і Emotion. Конкретні промти для отруєння даних мали шаблони, що змушували модель видавати шаблонну відповідь-вибачення або змінювати тип виводу; для SST2/Emotion тестувалися додавання тегів, що змінюють тональність, а для SMS Spam — вставки, що перетворюють спам на хибно нешкідливий текст. UniGuardian також перевіряли як LLM-детектор: йому передавалися пари {prompt, generation} і вимагали однозначного рішення «Yes/No» щодо дозволу виконання або успішності атаки з обґрунтуванням; такі суддівські промти використовувалися для оцінки людеподібності рішень і стійкості до обману через складні генерації. Метрики включали recall для виявлення ін'єкцій, точність виявлення бекдорів (спеціально — виявлення тригера у виводі), Robust Accuracy на змагальних прикладах і вплив на коректність класифікаційних завдань (SST2, Emotion, SMS Spam). Оцінювалася також стійкість проти атак «value inversion» і частота false positives при захисті відкритих запитів. [2]

Для порівняння MULLI і UniGuardian проведено ідентичні підходи на спільних підмножинах наборів даних (ToxicChat, підмножини LMSYS-Chat-1M і вибрані приклади Prompt Injections/Jailbreak). Досліджувалися три режими роботи: кожен інструмент окремо, послідовна інтеграція і паралельна перевірка з політикою «deny-on-conflict». Тестові сценарії охоплювали прямі токсичні запити, приховані джейлбрейки, отруєні тренувальні

прикладі, класичні ін'єкції ("ignore previous"), дрібні орфографічні/символьні маніпуляції та комплексні довгі запити, здатні спровокувати галюцинації. Кількісні показники порівняння включали зниження рівня токсичності в текстах, частоту успішних ін'єкцій, зміну класифікаційної точності на SST2/Emotion, а також вплив на затримку відповіді та частоту хибнопозитивних блокувань.

3.1. Аналіз помилок

У контексті виявлення токсичності та атак на великі мовні моделі (LLM) помилки класифікуються як хибні негативи (false negatives, FN) — випадки, коли детектор не ідентифікує шкідливий або токсичний вміст і пропускає його, що дозволяє LLM згенерувати небажану відповідь; та хибні спрацьовування (false positives, FP) — випадки, коли детектор помилково маркує нешкідливий вміст як шкідливий, спричиняючи надмірні блокування і погіршення користувацького досвіду. Для кожного інструмента ці загальні класи помилок мають свої характерні підвиди та причини, що важливо для розробки компенсуючих стратегій.

MULI. Типові хибні негативи для MULI пов'язані з дуже довгими або контекстно замаскованими підказками, у яких шкідливий фрагмент розміщено ненав'язливо або загорнуто в технічний чи описовий контекст; такі випадки спостерігались у наборах даних ToxicChat і LMSYS-Chat-1M. Джейлбрейк-запити, що вимагають спеціальних «обхідних» фраз або натяків, також іноді залишаються непоміченими. Ще одна категорія пропусків — неявно шкідливі або морально складні запити (питання про самоушкодження або складні етичні сценарії), які важко однозначно класифікувати за лінійними правилами. Хибні спрацьовування у MULI переважно виникають через наявність «чутливих» слів у контексті, що технічно безпечний, але містить лексему, яку детектор інтерпретує як індикатор токсичності; це призводить до блокувань нейтральних або наукових запитів. Причини помилок MULI включають неповне вирівнювання базової LLM (що відповідає як надмірній обережності, так і обману через джейлбрейки), незбалансованість класів у наборах даних (рідкість токсичних прикладів підвищує вплив невеликого FPR) та обмеження методів детекції, які спираються на приховані сигнали з виходів моделі (наприклад, логіти першого токена), що не завжди стабільні.

UniGuardian. Для UniGuardian типовими хибними негативами є «тонкі» атаки та маніпуляції, що не порушують зв'язності тексту і не створюють явних сигнатур тригера. До таких належать модифікації, спрямовані на дезінформацію, витік приватних даних у непрямій формі або атаки з мінімальними лексичними змінами, які не значно впливають на показники непередбачуваності (perplexity). Бекдор-атаки з тригерами, ретельно інтегрованими в тренувальні дані, теж можуть уникати виявлення, якщо тригер не змінює помітно розподіл ймовірностей. Хибні спрацьовування у UniGuardian виникають, коли підвищена підозрілість алгоритмів приводить до маркування довгих, складних або незвичних, але легітимних запитів як небезпечних; це відбувається особливо в завданнях з відкритими питаннями і творчими генераціями. Основні причини помилок UniGuardian пов'язані з методологією виявлення тригерів: складні бекдори можуть зберігати природню плавність тексту й не змінювати стандартні сигнали втраченого або модифікованого контексту; крім того, більшість перевірених сценаріїв стосувалися англomовних наборів даних і трансформерних архітектур, тому переносимість методів на інші мови чи архітектури залишається обмеженою.

Для обох підходів найскладнішими для виявлення є замасковані, довгі та семантично тонкі приклади шкідливого вмісту; їхнє успішне виявлення вимагає поєднання лексичних,

семантичних і поведінкових сигналів. Натомість надмірна чутливість до окремих лексем або стилістичних особливостей призводить до хибних блокувань і зниження якості досвіду. Практичною стратегією зниження FN та FP є комбіноване застосування методів (наприклад, поєднання сигналів від MULI і UniGuardian), адаптивне порогове налаштування на конкретних наборах даних і використання додаткової валідації (чутливі випадки передаються на ручний огляд або до другого ступеня перевірки).

4. Висновки

Обидва інструменти зменшують відповідні ризики: MULI виявляє явну токсичність і джейлбрейки через аналіз виходів моделі, тоді як UniGuardian більш ефективний проти промпт-ін'єкцій, бекдор-атак і змагальних маніпуляцій завдяки виявленню тригерів і поведінкових аномалій. Найвразливішими залишаються тонко замасковані шкідливі приклади та довгі контексти, які породжують найбільшу частку хибних негативів у обох системах. Надмірна чутливість до окремих лексем і незбалансованість наборів даних спричиняють хибні спрацьовування, особливо у MULI, що погіршує користувацький досвід у реальних застосунках. Поєднання інструментів послідовно або паралельно дає синергію: знижується частка і FN, і FP порівняно з окремим використанням, але водночас зростає затримка й ускладнюється інтеграція. Для коректної оцінки необхідне застосування багатовимірних метрик: precision/recall/F1 для токсичності, recall для ін'єкцій, Robust Accurasy на adversarial прикладах, частка успішних бекдорів, latency і частота випадків, що потребують ручного огляду. Обмеженість експериментів англійськими наборами даних і трансформерними архітектурами знижує узагальнюваність результатів, отже потрібне додаткове тестування на інших мовах, доменах і моделях. Для практичного розгортання доцільні адаптивне порогове налаштування на цільових наборах даних, каскадна схема перевірки з переведенням сумнівних випадків на додаткові рівні валідації та періодичне оновлення правил і детекторів із урахуванням нових атак. Подальші дослідження мають зосередитися на виявленні семантично прихованих атак, аналізі довготривалих залежностей у тексті, перенесенні методів на інші мови й архітектури та на автоматизованому балансуванні FN/FP з урахуванням якості користувацького досвіду.

Подяка

Авторам інструментів:

MULI було представлено Чжаньхао Ху (Zhanhao Hu), Жульеном Пієтом (Julien Piet), Генгом Чжао (Geng Zhao), Цзянтао Цзяо (Jiantao Jiao) та Девідом Вагнером (David Wagner) з Університету Каліфорнії, Берклі.

UniGuardian було запропоновано Хуавей Лінем (Huawei Lin), Їнцзе Лао (Yingjie Lao), Тонгом Генгом (Tong Geng), Таном Ю (Tan Yu) та Вейцзе Чжао (Weijie Zhao) з Технологічного інституту Рочестера, Університету Тафтса, Університету Рочестера та NVIDIA.

Declaration on Generative AI

Під час підготовки цієї роботи автор використовував Google Gemini 2.5 Flash для адаптації перекладу назв і термінів українською мовою.

Література

- [9] H. Lin, Y. Lao, T. Geng, T. Yu, W. Zhao, UniGuardian: A Unified Defense for Detecting Prompt Injection, Backdoor Attacks and Adversarial Attacks in Large Language Models, Preprint, 2024.
- [10] Z. Hu, J. Piet, G. Zhao, J. Jiao, D. Wagner, Toxicity Detection for Free, arXiv preprint arXiv:2405.18822, 2024

Analysis of Tools for Detecting Vulnerabilities from CWE Top 25

Yurii Baryshev¹, Ivan Kravchuk^{2*}

^{1,2} Vinnytsia National Technical University, 95 Khmelnytske shose, Vinnytsia, 21021, Ukraine

Abstract

This paper analyzes the most dangerous software weaknesses from CWE Top 25 and the modern tools used to detect them. The study covers static analysis (SAST), dynamic analysis (DAST), software composition analysis (SCA), and network vulnerability scanning. Tools considered include SonarQube, CodeQL, Checkmarx, Veracode, Snyk, Semgrep, Burp Suite, OWASP ZAP, Acunetix, Nessus, and OpenVAS. The evaluation across multiple vulnerable applications and test environments demonstrates that a combined approach using SAST, DAST, SCA, and network scanning provides the most effective results in detecting and mitigating CWE Top 25 vulnerabilities.

Keywords

CWE Top 25, cybersecurity, application security, static analysis, dynamic testing, software composition analysis, vulnerability detection, secure development

1. Introduction

In the modern era of rapid digitalization, software security has become a critical issue. The growth in the number of web applications, online services, and mobile platforms expands the attack surface and introduces new threat vectors. Software vulnerabilities remain one of the main entry points for attackers, leading to unauthorized access, data leaks, and system disruptions.

Regulatory frameworks such as the General Data Protection Regulation (GDPR) in the European Union [1], PCI DSS [2], ISO/IEC 27001 [3], and other global standards set strict requirements for information security. Organizations are now required to implement robust mechanisms for detecting and addressing vulnerabilities to ensure both compliance and operational resilience.

A key reference for developers and security professionals is the CWE Top 25 list, which highlights the most widespread and dangerous software weaknesses [4]. This list serves as the foundation for security policies, testing methodologies, and training initiatives. However, detecting such vulnerabilities is a challenging task due to the diversity of programming languages, architectural models, and execution environments.

The purpose of this study is to identify modern tools capable of effectively detecting vulnerabilities from the CWE Top 25 and to provide recommendations for their combined use. To achieve this goal, the following tasks are addressed:

1. Analysis of modern approaches to vulnerability detection
2. Define a set of modern tools (SAST, DAST, SCA, network scanners)
3. Evaluate their effectiveness on controlled test environments
4. Formulate conclusions on the benefits of tool integration

The main result of this study is the systematization of security tools and their mapping to the 25 most common CWE vulnerabilities, which enables a comprehensive approach to software protection.

2. Methodology

The methodology of this research was designed to evaluate how different categories of tools perform in relation to the CWE Top 25 vulnerabilities. Four categories of tools were considered: static application security testing (SAST), dynamic application security testing (DAST) [5], software composition analysis (SCA) [6] and network vulnerability scanners [7].

The tools selected for analysis represent widely adopted and actively maintained solutions: SonarQube, CodeQL, Checkmarx, Veracode, Snyk, Semgrep, Burp Suite, OWASP ZAP, Acunetix, Nessus, and OpenVAS. They were evaluated according to criteria such as detection accuracy, false positives, integration with CI/CD pipelines, performance, and accessibility [5-7].

To illustrate how each vulnerability from the CWE Top 25 can be addressed, the mapping of weaknesses to tools was developed. Table 1 presents the mapping between CWE Top 25 vulnerabilities and the tools that are capable of detecting them. This allows for a structured understanding of how different categories of tools complement each other.

Table 1

Mapping CWE Top 25 vulnerabilities to tools

CWE ID	Vulnerability	Tools	Category
CWE-79	Improper Neutralization of Input During Web Page Generation (Cross-site Scripting)	Burp, ZAP, Acunetix, SonarQube, Semgrep	DAST, SAST
CWE-787	Out-of-bounds Write	CodeQL, Checkmarx, Veracode	SAST
CWE-89	Improper Neutralization of Special Elements used in an SQL Command (SQL Injection)	Burp, ZAP, Acunetix, SonarQube, Semgrep	DAST, SAST
CWE-352	Cross-Site Request Forgery (CSRF)	Burp, ZAP, Acunetix	DAST
CWE-22	Improper Limitation of a Pathname to a Restricted Directory (Path Traversal)	ZAP, Acunetix, Nessus, OpenVAS	DAST, Network
CWE-125	Out-of-bounds Read	CodeQL, Checkmarx, Veracode	SAST
CWE-78	Improper Neutralization of Special Elements used in an OS Command (OS Command Injection)	Burp, ZAP, Acunetix, CodeQL, Semgrep	SAST, DAST
CWE-416	Use After Free	CodeQL, Veracode	SAST
CWE-862	Missing Authorization	Veracode, Checkmarx, Burp	SAST, DAST
CWE-434	Unrestricted Upload of File with	Burp, ZAP, Acunetix	DAST

	Dangerous Type		
CWE-94	Improper Control of Generation of Code (Code Injection)	Semgrep, Checkmarx, Veracode	SAST
CWE-20	Improper Input Validation	SonarQube, CodeQL, Veracode	SAST
CWE-77	Improper Neutralization of Special Elements used in a Command (Command Injection)	Burp, ZAP, Acunetix, Semgrep	DAST, SAST
CWE-287	Improper Authentication	Veracode, Checkmarx, Burp	SAST, DAST
CWE-269	Improper Privilege Management	Veracode, Checkmarx	SAST
CWE-502	Deserialization of Untrusted Data	Burp, ZAP, Acunetix, Veracode	DAST, SAST
CWE-200	Exposure of Sensitive Information to an Unauthorized Actor	Burp, Acunetix, Nessus, OpenVAS	DAST, Network
CWE-863	Incorrect Authorization	Veracode, Checkmarx, Burp	SAST, DAST
CWE-918	Server-Side Request Forgery (SSRF)	Burp, ZAP, Acunetix	DAST
CWE-119	Improper Restriction of Operations within the Bounds of a Memory Buffer	CodeQL, Checkmarx, Veracode	SAST
CWE-476	NULL Pointer Dereference	SonarQube, Veracode	SAST
CWE-798	Use of Hard-coded Credentials	Snyk, Veracode	SCA, SAST
CWE-190	Integer Overflow or Wraparound	CodeQL, Veracode	SAST
CWE-400	Uncontrolled Resource Consumption	Nessus, OpenVAS, Burp	Network, DAST
CWE-306	Missing Authentication for Critical Function	Veracode, Checkmarx, Burp, Acunetix	SAST, DAST

The analysis confirms that no single tool is sufficient. On the contrary, SAST, DAST, SCA, and network scanners complement each other, eliminating different categories of vulnerabilities. For a more detailed comparison of the selected tools, their strengths and limitations were analyzed.

3. Conclusion

The findings demonstrate that no single tool can provide complete coverage of all CWE Top 25 vulnerabilities. This is largely due to differences in the methodologies used by modern solutions. SAST tools are effective for detecting issues in source code at the earliest stages of development, which reduces the cost of remediation. However, they often generate a high number of false positives and are limited in detecting vulnerabilities that only manifest during execution.

The main contribution of this research is the systematization of modern tools and their mapping to CWE Top 25 weaknesses. This enables organizations to select and combine appropriate solutions depending on their infrastructure and regulatory needs. Future work may extend this research by including cloud-native and container-specific security tools, as well as exploring machine learning approaches for automated vulnerability detection.

Declaration on Generative AI

During the preparation of this work, the authors used AI tools for grammar and spelling check. After using these tools/services, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

- [1] K. Hjerpe, J. Ruohonen, V. Leppanen, The European Union general data protection regulation: What it is and what it means, *Information & Communications Technology Law*, vol. 28, no. 1, pp. 1-34 (2019). doi:10.1080/13600834.2019.1573501.
- [2] S. Chippagiri, A. Ramesh, PCI DSS: A Critical Analysis of Implementation, Effectiveness, and Legislative Impact in Payment Card Security, *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 1, pp. 1258-1266 (2025). doi:10.32628/CSEIT251112115.
- [3] M. Podrecca, G. Culot, G. Nassimbeni, M. Sartor, The ISO/IEC 27001 information security management standard: literature review and theory-based research agenda, *The TQM Journal*, vol. 33, no. 7, pp. 76-105 (2021). doi:10.1108/TQM-09-2020-0202.
- [4] CWE Top 25 Most Dangerous Software Weaknesses. URL: <https://cwe.mitre.org/top25/>.
- [5] A. Oluwaferanmi, W. Liang, "Static and Dynamic Application Security Testing (SAST and DAST) Integration for Robust Secure Coding Practices" (2025). URL: <https://www.researchgate.net/publication/393516606>.
- [6] J. Dietrich, S. Rasheed, A. Jordan, T. White, "On the Security Blind Spots of Software Composition Analysis," in *Proceedings of the ACM Workshop on Software Supply Chain Offensive Research and Ecosystem Defenses (SCORED '24)*, Salt Lake City, US (2024). doi:10.1145/3689944.3696165.
- [7] D. Railkar, "A Study on Vulnerability Scanning Tools for Network Security," *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 8, no. 6, p. 340 (2022). doi:10.32628/CSEITCN228641.

Using the discrete cosine transform method to ensure the confidentiality and authenticity of audio data

Alla Havrylova¹, Vladyslav Murzhak¹

¹National Technical University “Kharkiv Polytechnic University”, Kyrpychova Str., 2, Kharkiv, Ukraine

Abstract

This paper discusses the application of the discrete cosine transform method in audio file encryption. The method's capabilities in ensuring data confidentiality and authenticity during storage and transmission are analysed. The results of a comparison of cryptographic encryption methods for placing hidden information in discrete cosine transform coefficients are presented.

Keywords

discrete cosine transform, cryptanalysis methods, spectral representation

1. Introduction

The discrete cosine transform (DCT) is a mathematical tool that transforms audio data for compression, filtering, or analysis. [1]. The characteristics of audio file formats that use DCT are presented in Table 1.

Table 1

Characteristics of audio file formats that use discrete cosine transform

Formats	Bit rate, kbit/s	Sampling frequency, kHz	Data compression methods	
			with losses	without loss
MP3	128, 192, 256, 320	8 – 48	+	-
AAC	256, 320	8 – 96	+	-
ATRAC	64, 66-94, 292	24 – 48	+	-
Vorbis	96, 128, 150-170, 160-192, 400-700	2 – 192	+	-
Siren	24, 32, 48, 64, 96, 128	16 – 48	+	+

This method allows audio data to be compressed efficiently by concentrating most of the signal energy in a small number of low-frequency DCT coefficients. However, in addition to this, DCT has a number of properties that can also be used as an opportunity to improve data protection against unauthorised access during transmission via information and communication channels and during storage on media.

^{31*} Corresponding author.

✉ alla.havrylova@kpi.edu.ua (A. Havrylova); vladyslav.murzhak@cs.khpi.edu.ua (V. Murzhak)

🆔 0000-0002-2015-8927 (A. Havrylova); 0009-0005-9996-7901 (V. Murzhak)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Obstacle to cryptanalysis and hacking

The spectral representation of audio after DCT [2] becomes difficult to cryptanalyse and crack compared to the original signal for a number of reasons: 1) energy compactification means that every change in the signal directly affects its shape, which is reflected in a relatively small set of selected coefficients, and this will lead to a change in the coefficients that affect the resulting sound signal; 2) decorrelation, which provides a distributed effect of a change in one DCT coefficient on the original time signal, making localised manipulations obvious; 3) loss of information prevents attempts to hack or alter the compressed signal, making it impossible to work with incomplete data; 4) psychoacoustic-based compression means that without a deep understanding of the psychoacoustic model used by the codec, it is impossible to accurately predict how changes will be perceived by the ear; 5) block processing and superimposition leads to the distribution of changes in a single DCP coefficient over a certain time window.

3. Hiding information in an audio signal

The DCT method in steganography when working with audio files converts the original audio signal into the frequency domain, where information is hidden in the DCT coefficients, and the information can be encrypted using various cryptographic algorithms [3] (Table 2).

Table 2

Dependence of the security level of a ciphertext on the key size

Security level (bits)	Key size (bits)		
	AES	RSA	ECC
80	-	1024	160
112	-	2048	224
128	128	3072	256
192	192	7680	384
256	256	15360	521

This prevents attackers from discovering hidden data, as they cannot read it without the appropriate key.

Conclusions

The analysis showed that the DCP method can be used as a tool for protecting audio data to ensure confidentiality and authenticity.

References

- [1] В. А. Ракицький. "Дискретне косинусне перетворення як засіб комп'ютерної обробки інформації". Проблеми інформатизації та управління. 2019. № 2(62). С. 52 – 56.

- [2] R. Rastegayev, V. Martovytskyi, N. Bolohova, B. Filonenko, O. Chechui. "Review of methods for embedding digital watermarksfor audio file protection". Control, Navigation and Communication Systems. 2024. No. 4. P. 132 – 138. doi: 10.26906/SUNZ.2024.4.132.
- [3] Sh. R Gulomov, T. R. Khudayberganov, M. X. Ravshanova, T. T. Turdiev, S. S. Atabayev. "Exploring Post-Quantum Cryptographic Algorithms for Secure Data Transmission". IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE). 2024. № 11. P. 15. URL: <https://7universum.com/ru/tech/archive/item/19583>.

Process decomposition in cloud computing security techniques

Alla Havrylova¹, Oleh Azarov¹

¹ *National Technical University “Kharkiv Polytechnic University”, Kyrpychova Str., 2, Kharkiv, Ukraine*

Abstract

The paper presents the application of decomposition approach in the implementation of modern methods of cloud computing security. It is proposed to identify and combine the same processes of different methods to reduce the time to detect and counter cyber threats when combining security methods.

Keywords

cloud computing methods, data protection, decomposition, combined

1. Introduction

The use of known cloud computing security techniques help to increase the protection of data from unauthorised access [1]. But a greater effect can come from using them together. A decomposition of the processes typical of existing cloud computing security practices is proposed to compile best practices for countering known cyberattacks.

This approach is based on breaking down the cloud computing security method into its constituent parts to simplify its implementation, analysis and execution [2].

2. Identify steps to implement cloud computing security practices

When considering each method, a specific list of tasks to achieve the required results is identified, which can be represented as the following tuple of actions [3, 4]:

1) employee training = {identify current level of knowledge; design a training programme; implement the training programme; evaluate the effectiveness of the training programme};

2) eliminating shadow information technology = {developing comprehensive policies; open dialogue with employees; adapting policies according to employees' views; using monitoring tools for control};

3) configuration troubleshooting = {diagnostics; use specialised tools; find and fix errors in code; validate access rights; verify and recover};

4) zero trust implementation = {assessment and planning; implementation of key elements; technological solutions; cultural and organisational factors};

5) cloud application access protection = {multifactor authentication; enforcement of security policies; threat monitoring and response};

6) creating security policies = {needs and risk assessment; development of security policies; implementation and monitoring};

7) limiting data sharing = {using specialised software (software); adapting drivers; creating and configuring sharing plans};

^{32*} Corresponding author.

[†] These authors contributed equally.

✉ alla.havrylova@khu.edu.ua (A. Havrylova); oleh.azarov@cit.khpi.edu.ua (O. Azarov)

 0000-0002-2015-8927 (A. Havrylova); 0009-0002-7984-518X (O. Azarov)(E. Lindström)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

8) endpoint protection = {monitoring and analysis; threat prevention; threat detection and response; vulnerability management; centralised management};

9) malware protection = {use and update anti-virus software; update software and operating system (OS); check links and files; use firewall and backup};

10) continuous data detection and classification = {detection; classification; enforcement of security policies; continuous monitoring};

11) regulatory compliance = {access management; data encryption; data location control; standards compliance; risk management}.

According to the tuples presented, it can be seen that there is a relationship between tuples 6 and 5 for the processes “security policy enforcement” and “security policy development” respectively, with the processes of method 5 necessarily being carried out before those of method 6, and the process “threat response” being present in both method 5 and method 8, indicating that this process is repeated when these methods are combined.

Conclusions

Based on the analysis, it can be concluded that the proposed approach not only allows us to clearly structure the works to achieve the objectives of each cloud computing security method, but also to determine the sequence of their application based on the relationship between the tasks of different methods, and to highlight the redundancy at the level of the tasks performed.

References

- [1] Темна сторона ІТ – тренди кіберзлочинності 2024 року URL: <https://denovo.ua/blog/cybertrends-ai-2024> (Date of reference 07.10.2024).
- [2] Кібербезпека у хмарі – п'ять актуальних технологій у 2024 році URL: <https://denovo.ua/blog/cloud-secuity-top5-2024> (Date of reference 24.05.2024).
- [3] NIST SP 800-207, Zero Trust Architecture. URL: <https://csrc.nist.gov/publications/detail/sp/800-207/final> (Date of reference 29.08.2025).
- [4] K. Dempsey, N. Chawla, L. Johnson, R. Johnston, A. Jones, A. Orebaugh, M. Scholl, K. Stine. Information Security Continuous Monitoring (ISCM) for Federal Information Systems and Organizations NIST SP 800-137. 2011 80 p. URL: <https://csrc.nist.gov/publications/detail/sp/800-137/final> (Date of reference 09.07.2025).

Information Security Risk Management in Information Systems Based on the Method of Paired Comparisons

Yurii Klots¹, Vira Titova¹ and Nataliia Petliak¹

¹ Khmelnytskyi National University, Instytut's'ka 11, 29016 Khmelnytskyi, Ukraine

Abstract

In this article, the authors analyzed the currently existing approaches to risk management in information systems. Based on the results of the analysis, it was concluded that most of the approaches do not take into account the concepts and requirements of various information security standards, which can cause distrust in the applied approaches among experts who conduct risk analysis and complicate the possible certification of the organization. To implement an improved approach to risk management, the authors analyzed the information system from the point of view of information security of companies, and also described the features of the object under consideration. A threat model was also built according to the guidance documents, expert information was collected to determine the probability of each of the identified threats, and risks were calculated for the system under consideration.

Keywords

threat model, information security, risk management, information systems, information security policy

1. Introduction

A large number of scientific works are devoted to the analysis of information security management approaches, most of which either contain a large number of mathematical formulas and models; or do not contain any mathematical components at all; or have a tendency towards any of the two above-mentioned groups of approaches. Let's analyze the substantive aspects of each group.

The approaches of the first group, as a rule, use different sections of higher mathematics. As the core of approaches, principles based on the theory of probability or utility (reliability), or fuzzy sets, continuous or discrete distribution, etc. are chosen [1-4]. Works belonging to the first group of approaches quite often do not take into account the real requirements of organizations engaged in risk analysis; require sufficient mathematical training from information security experts, which is often negatively reflected in the practice of applying these approaches. The second group of approaches is more developed by foreign authors. Articles by authors from the USA and England are primarily of a recommendatory nature for modernization based on information security standards, which have already proven themselves without requiring deep knowledge of higher mathematics [5-7].

The third group of approaches in many cases combines expert evaluations and risk assessments based on their determination from available statistical data [8-11]. Similar approaches can be successfully applied in practical activities (despite a number of disadvantages), since the use of a statistical database allows to minimize the expert's subjective point of view on the problem to be solved and to carry out work based on the risk assessment without much experience and qualifications.

2. Problem Formulation

Based on the results of the analysis, the following conclusions can be drawn: most approaches do not take into account the real requirements of organizations, which can cause distrust in the

applied approaches of information security risk analysis, and complicates the information security management. Many approaches based on the goal of obtaining a quantitative assessment of risks using mathematical formulas, models steeped in mathematical theories, have no practical connection with risk assessment, real business requirements. A number of approaches do not provide a complete process of assessment and management of information security risks, implementing only some of its components. Taking into account the strengths and weaknesses of the existing approaches, the authors decided to implement a more advanced approach to risk management and, accordingly, information security in distributed information systems for personal data processing.

3. Presenting Main Material

An arbitrary private enterprise was chosen as the research object. The company has 150 employees and a local network of the enterprise that has access to the Internet. The company's activities are related to the use of computing equipment and information technologies, it includes: testing for compliance with security requirements against unauthorized access to information; analysis of vulnerabilities and control of the absence of undeclared capabilities in software; development and implementation of systems for the protection of information constituting a state secret, based on certified basic information protected computer technologies.

The list of assets of the organization includes: equipment – estimated value of UAH 750,000, authorized capital – UAH 570,000, personal data of employees – estimated value of UAH 250,000, buildings, structures – estimated value of UAH 1,850,000, software products – estimated value of UAH 600,000.

To achieve the set goal within the company, it is advisable to use the risk calculation method presented in NIST 800-30 [7]:

$$R = P \cdot S, \quad (1)$$

where R is the risk value; P – probability of information security threat realization (qualitative and quantitative scales are applied); S – the degree of influence of the threat on the information asset (value of the asset in a qualitative or quantitative scale). Let's identify all possible threats to information assets in distributed information systems for personal data processing (Table 1).

Table 1
Model of Enterprise Information Security Threats

№	Threat Type	Possible Consequences
1	Analysis of network traffic	Characterization of network traffic, interception of transmitted data, including user IDs and passwords
2	Network scanning	Definition of protocols, available ports of network services, rules for establishing connections, active network services, user IDs and passwords
3	"Password" attack	Performing destructive actions related to obtaining unauthorized access
4	Replacing a trusted network object	Changing the routing of messages, unauthorized change of routing and address data, unauthorized access to network resources, imposition of unreliable information.
5	Imposing a false route	Unauthorized change of route-address data, analysis and modification of transmitted data, imposition of false messages
6	Implementation of the false network object	Interception and viewing of traffic, unauthorized access to network resources, imposition of unreliable information
7	Partial depletion of resources	Reduction of bandwidth of communication channels, performance of network devices, performance of server applications
	Complete exhaustion of resources	Impossibility of transmitting messages due to lack of access to the transmission medium, refusal to establish a connection, refusal to provide services (file server, mail, etc.)

	Denial of service	Violation of logical connections between attributes, data, objects	Impossibility of sending messages due to lack of correct routing and address data, impossibility of receiving services due to unauthorized modification of identifiers, passwords, etc.
		Using errors in programs	Violation of network device performance
8	Remote application launch	Sending files with destructive code, virus infection	Violation of confidentiality, integrity, availability of information
		Buffer overflow	
		Using the capabilities of remote system management through hidden software/hardware bookmarks or standard means	Hidden system management

To calculate the value of risks R , it is necessary to have a value P . The value of P can be defined as the frequency of successful implementation of the threat during a certain period. However, this value is generalized, since the frequency of successful implementation of the same threat can change over time and depends on the skill of the attacker. In addition, to obtain this value, it is necessary to have certain statistics on the implementation of threats as a whole, and to have certain qualification skills in matters of risk assessment and cyber security.

The authors propose to determine P by expert assessment on a scale of 1-9. Moreover, 1 means that the threats are equivalent, 2-9 – that one threat is more likely than the other by the specified number of times. We will get a summary table of threat comparisons (Table 2).

After calculating the weight coefficient of each threat by normalization, we will get the probability value of Threat 1 – Threat 8

Table 2

Summary Table of Threat Comparisons

	Threat 1	Threat 2	Threat 3	Threat 4	Threat 5	Threat 6	Threat 7	Threat 8	Σ
Threat 1	1	1	1	1/7	1/7	1/7	1/5	1/3	3,96
Threat 2	1	1	1	1/7	1/7	1/7	1/5	1/3	3,96
Threat 3	1	1	1	1/7	1/7	1/7	1/5	1/3	3,96
Threat 4	7	7	7	1	1	1	3	3	30
Threat 5	7	7	7	1	1	1	3	3	30
Threat 6	7	7	7	1	1	1	3	3	30
Threat 7	5	5	5	1/3	1/3	1/3	1	1	18
Threat 8	3	3	3	1/3	1/3	1/3	1	1	12
									131,88

P_1 (Threat 1 – Analysis of network traffic) = 0.03;

P_2 (Threat 2 – Network scanning) = 0.03;

P_3 (Threat 3 – "Password" attack) = 0.03;

P_4 (Threat 4 – Replacement of a trusted network object) = 0.23;

P_5 (Threat 5 – Imposing a false route) = 0.23;

P_6 (Threat 6 – Implementation of the false network object) = 0.23;

P_7 (Threat 7 – Denial of service) = 0.14;

P_8 (Threat 8 – Remote applications launch) = 0.09.

At the same time, Threat 1 – Threat 6 pose a threat to personal data of employees and software products; Threat 7 – poses a threat to equipment and software products; Threat 8 – only for software products. Let's calculate the quantitative value of the risk for each threat: $R_1=25500$ UAH; $R_2=25500$ UAH; $R_3=25500$ UAH; $R_4=195500$ UAH; $R_5=195500$ UAH; $R_6=195500$ UAH; $R_7=189000$ UAH; $R_8=54000$ UAH.

Risks that amount to less than 5% of the authorized capital will be rejected as those that do not carry significant losses and can be easily compensated. In our case, there are no such risks that could be neglected. For all existing threats Threat 1 – Threat 8, countermeasures should be applied

to minimize the value of the risks they pose. Threat 4 – Threat 7 require the most attention (substitution of a trusted network object, imposition of a false route, introduction of a false network object, denial of service). At the same time, the cost of means and protective measures should not exceed the amount of risk that they minimize.

As an example, consider the case when, in order to minimize risks, an information security policy was implemented, which divided users into groups with appropriate rights:

- Administrator – network administrators (creation and management of information security policies, global network settings, etc.).
- Engineer – accounts for day-to-day maintenance of IT equipment.
- Worker – the account of a standard user (employee of the organization) with limited rights.
- Guest – a limited account (if access is required by non-employees of the organization).

User identification requires a unique record of each user that is included in the corresponding group. In this way, access is demarcated (Table 3).

Table 3

User Groups and Their Rights

Actions	Guest	Worker	Engineer	Admin
Creating and modifying user groups	No	No	No	Yes
Changing network settings	No	No	No	Yes
Connection to the network of new workstations	No	No	No	Yes
Changing server settings	No	No	No	Yes
Changing access rights	No	No	No	Yes
Installing applications and software	No	No	Yes	Yes
Internet connection	No	Yes	Yes	Yes
Access to corporate e-mail	No	Yes	Yes	Yes
Access to corporate chat	No	Yes	Yes	Yes
Ability to download files	No	No	Yes	Yes
Recording files	"My Documents"	"My documents", "Desktop"	Any folder on work PC	Any PC on the network
Connecting flash drives, external drives	No	No	Yes	Yes

The starting point for determining the economic efficiency of the proposed approach is an obvious assumption: on the one hand, a breach of information security causes some damage; on the other hand, ensuring information security is expensive. The total expected cost of defense can be expressed as the sum of the cost of defense and damages from a breach. It is obvious that the optimal solution is the distribution of funds for information protection in such a way as to minimize the total cost of protection [7].

It is also clear that the economic efficiency of information security measures is determined by the amount of averted damage or the amount of risk reduction for the organization's information assets.

It is sufficient to determine only the level of losses, since the optimal solution to the problem of the appropriate level of defense costs is that this level is equal to the level of losses expected in the event of a security breach.

To assess the effectiveness of the implementation of this or that security measure, we will once again conduct a comparative analysis of threats (Table 4).

Table 3

Summary Table of Threat Comparisons After Applying the Security Policy

	Threat 1	Threat 2	Threat 3	Threat 4	Threat 5	Threat 6	Threat 7	Threat 8	Σ
Threat 1	1	1	1	1/7	1/7	1/7	1/5	1	4,62
Threat 2	1	1	1	1/7	1/7	1/7	1/5	1	4,62
Threat 3	1	1	1	1/7	1/7	1/7	1/5	1	4,62
Threat 4	7	7	7	1	1	1	3	3	30
Threat 5	7	7	7	1	1	1	3	3	30
Threat 6	7	7	7	1	1	1	3	3	30
Threat 7	5	5	5	1/3	1/3	1/3	1	1	18
Threat 8	1	1	1	1/3	1/3	1/3	1	1	6
									127,86

P1 (Threat 1 – Analysis of network traffic) = 0.04;

P2 (Threat 2 – Network scanning) = 0.04;

P3 (Threat 3 – "Password" attack) = 0.04;

P4 (Threat 4 – Replacement of a trusted network object) = 0.23;

P5 (Threat 5 – Imposing a false route) = 0.23;

P6 (Threat 6 – Implementation of the false network object) = 0.23;

P7 (Threat 7 – Denial of service) = 0.14;

P8 (Threat 8 – Remote applications launch) = 0.05.

The risk values will be as follows: R1=34000 UAH; R2=34000 UAH; R3=34000 UAH; R4=195500 UAH; R5=195500 UAH; R6=195500 UAH; R7=189000 UAH; R8=30000 UAH

As can be seen, the risk value for Threat 8 almost halved, but the normalization resulted in a non-significant increase in the risks of Threats 1-3. This can be avoided if the value of the risk of any threat is less than 5% of the authorized capital and it is excluded from the calculations. Thus, gradually all risks will be minimized.

4. Conclusions

In this work, the information system was considered on the example of an arbitrary private company with a staff of 150 employees and a local network of the enterprise that has access to the Internet.

The authors analyzed methodical documents on risk prevention, as well as the existing object of information activity from the point of view of information protection. A threat model was built in accordance with regulatory documents. The most likely threats to the selected organization were determined and the attack vector for these threats was determined. At the next stage, expert information was collected to determine the probability of each of the identified threats, and risks were calculated for the system under consideration.

A method of determining the probability of a threat was proposed, which is based on a comparative analysis of existing threats among themselves. The advantages of this method are that its use does not require special knowledge in the field of cyber security and risk assessment. Also, an advantage can be called the fact that the importance of certain threats will change from organization to organization depending on their field of activity, which will allow to improve approaches to information security management as a whole.

The disadvantage of the developed method can be considered that due to the normalization of probability values, when the level of one threat decreases, others may increase. However, it is possible to get rid of it by minimizing the risk values of threats and excluding them from further calculations.

So, the goal of the work, namely the implementation of a more advanced approach to risk management and, accordingly, information security in information systems, was achieved.

References

- [1] Pan, Liuxuan & Tomlinson, Allan. (2016). A systematic review of information security risk assessment. *International Journal of Safety and Security Engineering*. 6. 270-281. 10.2495/SAFE-V6-N2-270-281.
- [2] Taylor, Richard. (2015). Potential Problems with Information Security Risk Assessments. *Information Security Journal: A Global Perspective*. 24. 1-8. 10.1080/19393555.2015.1092620.
- [3] Information Security Risk Assessment Using Situational Awareness Frameworks and Application Tools. Nungky Awang Chandra, Kalamullah Ramli, Anak Agung Putri Ratna, Teddy Surya Gunawan. *Risks* 2022, 10(8), 165; <https://doi.org/10.3390/risks10080165>
- [4] Ali, T., Al-Khalidi, M., & Al-Zaidi, R. (2024). Information Security Risk Assessment Methods in Cloud Computing: Comprehensive Review. *Journal of Computer Information Systems*, 1–28. <https://doi.org/10.1080/08874417.2024.2329985>
- [5] ISO/IEC 27005:2022 - Information security, cybersecurity and privacy protection – Guidance of managing information security risks.
- [6] ISO 31000:2018 - Risk management – Guidelines.
- [7] Risk Management Guide for Information Technology Systems. Recommendations of the National Institute of Standards and Technology.
- [8] Risk Assessment and Analysis Methods: Qualitative and Quantitative Risk Assessment. Volkan Evrin, CISA, CRISC, COBIT 2019.
- [9] Kuzminykh, Ievgeniia & Ghita, B.V. & Sokolov, Volodymyr & Bakhshi, Taimur. (2021). Information Security Risk Assessment. 1. 602-617. 10.3390/encyclopedia1030050.
- [10] Terje Aven. Risk assessment and risk management: Review of recent advances on their foundation, *European Journal of Operational Research*, Volume 253, Issue 1, 2016, Pages 1-13, <https://doi.org/10.1016/j.ejor.2015.12.023>.
- [11] Seeba, M., Mäses, S., Matulevičius, R. (2022). Method for Evaluating Information Security Level in Organisations. In: Guizzardi, R., Ralyté, J., Franch, X. (eds) *Research Challenges in Information Science. RCIS 2022. Lecture Notes in Business Information Processing*, vol 446. Springer, Cham. https://doi.org/10.1007/978-3-031-05760-1_39

Enhancement of Security Mechanisms for an Open Information Platform Supporting the Publication of Scientific Research

Roman Moroz¹

¹*Lviv Polytechnic National University, Bandery St. 12, 79013 Lviv, Ukraine*

Abstract

A review of recent publications revealed limitations of current open publication systems in access control, data and metadata protection, and integration with external services. Threats were clarified and attack vectors classified to systematize risks originating from both internal and external environments, including unauthorized user actions, content modification, breaches in the review process, and infrastructure vulnerabilities. An integrated security model for the communication subsystems of the publication platform was developed to enhance security mechanisms, including the combination of RBAC and ABAC models for differentiated access control, the use of multi-level encryption and key management systems, UEBA algorithms for anomaly detection, and authentication protocols to ensure data integrity.

Keywords

Access management, metadata control, information protection, user activity auditing, authentication, ethical standards, publication workflow

1. Introduction

Contemporary scientific communication increasingly relies on open digital infrastructures that support the complete publication workflow, from manuscript submission to dissemination of published materials. Such platforms integrate editorial management, peer review organization, archive formation, and interoperability with international identifiers and services, facilitating interaction among a large number of users with different roles and ensuring compatibility with external systems, including identification registries, bibliometric databases, and long-term data preservation systems. The use of cloud technologies and modular architectures makes these platforms accessible to the broader scientific community, while simultaneously imposing additional requirements for ensuring data security, process integrity, confidentiality, availability, and protection against cyber threats.

2. Review of recent publications

In recent years, there has been a marked increase in attention to open information platforms that support the entire cycle of scientific publications. Studies emphasize the need to integrate access management, metadata control, and information protection mechanisms, as these platforms interact with a large number of users and store substantial volumes of data, including manuscripts, reviews, and analytical results [1]. Recent works have examined various approaches to securing user interactions, focusing on role models and access policies that combine classical RBAC (role-based access control) and ABAC (attribute-based access control), enabling differentiation of user privileges based on context, publication stage, and object characteristics [2]. Research indicates that this combination reduces the risk of unauthorized access to manuscripts and reviews while maintaining flexibility for editorial and reviewer procedures [3]. Certain studies highlight privacy and confidentiality protection through pseudonymization, metadata access control, and visibility

restrictions according to user roles. Issues of review anonymity and prevention of deanonymization are considered essential for upholding ethical standards and ensuring equal conditions for authors and reviewers [4]. Furthermore, investigations underline the necessity of auditing user actions and tracking metadata changes to prevent manipulation and ensure the integrity of the publication process [5].

A separate body of research focuses on the protection of data and metadata at the infrastructure level. These studies examine encryption technologies, key management systems, version control methods, and mechanisms for ensuring document immutability. Findings indicate that the implementation of multi-level encryption and authentication protocols enhances data security during storage and transmission, reducing the likelihood of unauthorized access or information loss [3]. Scientific works also address the integration of platforms with external services such as ORCID and DOI registries and plagiarism detection systems. Research demonstrates that interaction with these services requires standardized data exchange protocols and authentication mechanisms to maintain data integrity and reliability [1]. Additionally, existing studies describe approaches for monitoring and detecting anomalies in user behavior and publication processes. The use of analytics, UEBA (user and entity behavior analytics) algorithms, and integrated logging enables the identification of suspicious activities indicative of access policy violations or attempts to modify manuscripts and reviews [6].

The review of recent studies has shown that existing publication support platforms implement basic mechanisms for access control and data preservation; however, gaps remain in providing comprehensive protection of the review process, metadata management, and integration with external services. In particular, methods for monitoring user actions are insufficiently developed, complicating the detection of unauthorized modifications to manuscripts and reviews. Support for context-based differentiated access is also limited, increasing the risk of data manipulation at various stages of the publication workflow. Identified gaps in review anonymity standards and metadata integrity control emphasize the need to integrate new security mechanisms that ensure both data protection and process transparency. This includes the extension of access policies, enhancement of authentication protocols, and implementation of analytical tools for monitoring anomalous user behavior.

3. Rationale for the Relevance of Enhancing Security Mechanisms of a Publication Platform

Publication platforms operate within cloud and hybrid infrastructures, necessitating the implementation of comprehensive security mechanisms at multiple levels. Threats include unauthorized access to manuscripts and reviews, identifier substitution, compromise of integration services, vulnerabilities in containerized environments, and misuse during the review process. Particular attention should be given to the security of the data flow administration chain, which determines the reliability of the entire platform. The enhancement of security mechanisms in open publication platforms is increasingly relevant in the context of growing volumes of digital scientific data, the complexity of user interactions, and integration processes with external services. The timeliness of this research is motivated by the need to develop approaches that combine access control, metadata protection, and process monitoring without imposing excessive constraints on the efficiency of the publication workflow. Under these conditions, the formation of a cybersecurity architecture capable of addressing the specificities of scientific communication, ensuring compliance with privacy requirements, and maintaining process stability over the long term becomes critical. The present study aims to align access models, threat analysis methodologies, and

countermeasure mechanisms within a unified system that adheres to the principles of open science and the requirements of contemporary communication infrastructures.

4. Identification of Threats and Attack Vectors in a Publication Platform

The analysis of threat models and attack vectors constitutes a necessary prerequisite for enhancing the security mechanisms of communication systems and cloud infrastructure within an open information platform for publication support. This approach entails the identification of all key platform components, including user interfaces, manuscript submission and review processes, administrative functions, metadata management, and integration with external services. The threat model defines potential sources of risk, such as unauthorized access, content manipulation, abuse of user roles and privileges, compromise of metadata integrity, and anomalous behavior that deviates from established procedures. Within the publication platform, the threat model provides a framework for understanding how different system elements interact with both external and internal sources of potential risk. Threats emerge at the intersection of user roles, content processing workflows, and technical infrastructure components. The functions of authors, reviewers, editors, and administrators determine not only data access rights but also potential avenues for undesired actions affecting the integrity of manuscripts and metadata. Unauthorized content modifications may result from both external attacks and internal procedural violations, highlighting the need for differentiated control mechanisms. Interaction with external services, including identification systems and plagiarism detection databases, introduces additional vectors affecting data, where any inaccuracy or error in information exchange may compromise the publication support process.

Therefore, the proposed integrated security model for the communication subsystems of the publication platform is presented as a three-tier structure reflecting the interaction of users, publication support processes, and protective mechanisms (Figure). The first tier comprises user roles and external services, including authors, reviewers, editors, administrators, as well as integrated systems such as ORCID, DOI, plagiarism detection services, and archives. Each element is considered a potential source of threats, including unauthorized account access, content manipulation, role abuse, data leakage during integration with external services, and internal anomalies in user behavior. The second tier encompasses the primary publication support processes, including manuscript submission, peer review, editorial verification, publication, and integration with external systems. Interaction arrows from actors to processes illustrate potential impacts on specific operations and points where threats may manifest. The third tier represents protective mechanisms applied to processes as barriers or frameworks. These include RBAC and ABAC for differentiated access control and management of roles and attributes, UEBA for detecting hidden and internal anomalies, and authentication protocols alongside data integrity measures to prevent metadata or content tampering.

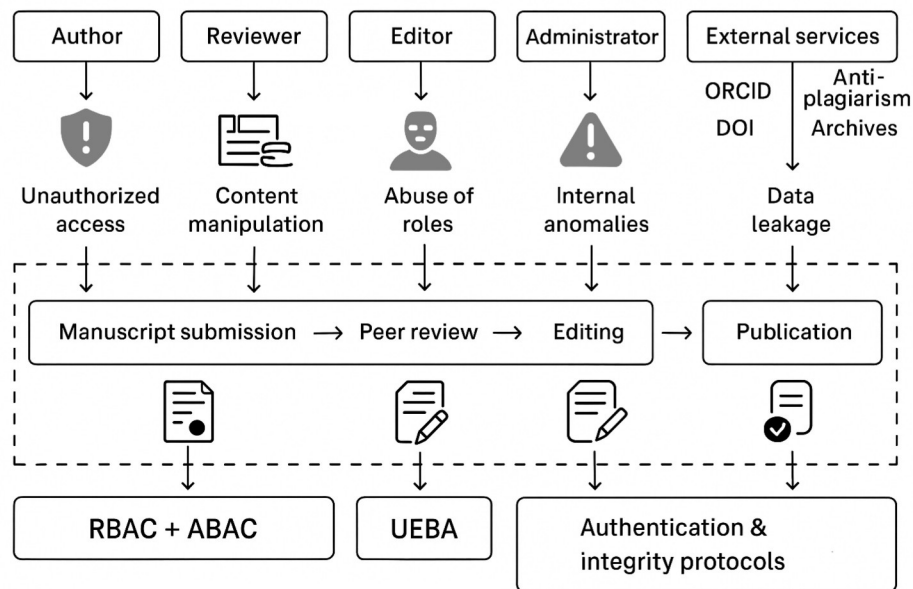


Figure 1: Integrated security model of communication subsystems in a publication platform. Developed by the author.

Anomalous user behavior, including unusual system requests, attempts to circumvent established access policies, or modifications to metadata, represents internal attack vectors that cannot be detected without analyzing action sequences and comparing them to expected usage patterns. Each manuscript processing stage, from submission to publication, includes points where interference or unauthorized changes may affect the accuracy and reliability of information. Identifying these points allows for the delineation of threat categories and the development of tools that regulate user access and actions while maintaining the efficiency of the publication workflow.

Thus, attack vectors are formed as a combination of user capabilities, technological vulnerabilities, and external integrations, creating a multi-level network of interdependencies. Intrusions through accounts, content substitution or modification, interference with the review process, and breaches of metadata integrity occur not in isolation but within an interconnected context, where a single action can trigger cascading effects across multiple platform components. Detecting and identifying these interrelationships provides a foundation for developing comprehensive control procedures that integrate user behavior monitoring, context-sensitive access rights, and authentication mechanisms adaptable to system actions.

The integration of analytical tools for detecting anomalous behavior while ensuring data integrity enables the deployment of protective mechanisms without disrupting the publication workflow. This approach establishes a foundation for the resilient operation of the platform, where each system component and every integration with external services is assessed in terms of potential threats and corresponding measures to mitigate their impact. The development and application of the presented threat and attack vector model create conditions for implementing security measures that ensure the integrity, availability, and reliability of information within the publication platform, while maintaining the efficiency and flexibility of processes at all stages of scientific material handling.

5. Conclusions

Thus, the presented security model for the communication subsystems of the publication platform integrates differentiated access management, monitoring of anomalous behavior, and protection of information integrity across all stages of the publication process. It combines user roles and external services as potential sources of threats with publication support processes, enabling the detection and localization of risks related to unauthorized access, content manipulation, role abuse, data leakage, and internal anomalies. Novel features of the model include the use of RBAC and ABAC to implement differentiated access control based on roles and attributes, UEBA for detecting hidden and internal anomalies in user behavior, and authentication protocols along with data integrity measures to prevent tampering of metadata and content.

The research results indicate the relevance of enhancing security mechanisms in publication platforms through the combined use of RBAC and ABAC access models for differentiated user rights control based on context and publication cycle stage, multi-level encryption and key management systems for data protection during storage and transmission, analytical tools and algorithms. The implementation of the proposed approaches will contribute to enhancing the reliability, resilience, and trustworthiness of digital publication support platforms for scientific research, while ensuring compliance with current privacy requirements and integration with international scientific infrastructure services. The model's structure enables a proper realization of the interconnection between users, processes, and security mechanisms, demonstrating the target measures applied for access control, monitoring of anomalous behavior, and protection of information integrity throughout the publication support process. It establishes a methodological foundation for the further development of comprehensive cybersecurity strategies for open publication platforms, allowing the integration of data security, process transparency, and publication workflow efficiency, which is essential in the context of contemporary scientific communication and the advancement of open science.

Declaration on Generative AI

During the preparation of this work, the author used X-GPT-4 for grammar and spelling check of domain-specific terminology in English. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

- [1] Paseri, L. (2023). Open science and data protection: Engaging scientific and legal contexts. *Journal of Open Access to Law*, 11(1), 1-18. <https://doi.org/10.63567/1bnsyb91>
- [2] Mahmood Rajpoot, Q., Jensen, C. D., & Krishnan, R. (2015). Attributes enhanced role-based access control model. In S. Fischer-Huebner & C. Lambrinoudakis (Eds.), *Proceedings of the 12th International Conference on Trust, Privacy and Security in Digital Business (TrustBus'15)* (pp. 3-17). Springer. https://doi.org/10.1007/978-3-319-22906-5_1
- [3] Hu, V. C., Ferraiolo, D., Kuhn, R., Friedman, A. R., Lang, A. J., Cogdell, M. M., Schnitzer, A., Sandlin, K., Miller, R., & Scarfone, K. (2014). Guide to attribute based access control (ABAC) definition and considerations (NIST Special Publication 800-162). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.800-162>
- [4] Payer, M., Huang, L., Gong, N. Z., Borgolte, K., & Frank, M. (2015). What you submit is who you are: A multimodal approach for deanonymizing scientific publications. *IEEE Transactions*

- on Information Forensics and Security, 10(1), 200–212.
<https://doi.org/10.1109/TIFS.2014.2368355>
- [5] Berti Suman, A., & Pierce, R. (2018). Challenges for citizen science and the EU open science agenda under the GDPR. *European Data Protection Law Review*, 4(3), 284-295.
<https://doi.org/10.21552/edpl/2018/3/7>
- [6] Landauer, M., Skopik, F., Höld, G., & Wurzenberger, M. (2022). A user and entity behavior analytics log data set for anomaly detection in cloud computing. In *2022 IEEE International Conference on Big Data (Big Data)*, Dec. 17 2022 to Dec. 20 2022, Osaka, Japan (pp. 4285-4294).
<https://doi.org/10.1109/BigData55660.2022.10020672>

Mathematical Method for Evaluating the State of Cybersecurity of Cloud Services for Information Infrastructure Objects

Volodymyr Shulha¹, Yevhenii Pedchenko², Yevheniia Ivanchenko³, Ihor Ivanchenko⁴,
Mari Petrovska⁵, Serhii Skvortsov⁶ and Maryna Pedchenko⁷

¹ State University of Information and Communication Technologies, Solomianska Street 7, Kyiv, Ukraine

² State University of Information and Communication Technologies, Solomianska Street 7, Kyiv, Ukraine

³ State University of Information and Communication Technologies, Solomianska Street 7, Kyiv, Ukraine

⁴ State University of Information and Communication Technologies, Solomianska Street 7, Kyiv, Ukraine

⁵ Odesa Polytechnic National University, Shevchenko ave., 1, Odesa, Ukraine

⁶ State Non-Commercial Company «State University «Kyiv Aviation Institute», Liubomyra Huzara Avenue, Kyiv, Ukraine

⁷ State University of Information and Communication Technologies, Solomianska Street 7, Kyiv, Ukraine

Abstract

In this work, a mathematical method based on a mathematical model has been developed to evaluate the state of cybersecurity of cloud services for information infrastructure objects. The mathematical method consists of 11 evaluation stages, including the calculation of the cybersecurity level of the assessed cloud service. Based on the developed model and method, a structural model of the evaluation system has been created, which describes the sequential execution of all evaluation stages and the activation of the modules of the developed mathematical method. Based on the results of implementing the developed structural model, a sequence for recording the evaluation results in a database and providing the auditor with a finalized report is presented. This report includes recommendations for the use of the assessed cloud service and suggestions for improving its cybersecurity level.

Keywords

cybersecurity, information security, method, model, risk assessment, audit, fuzzy logic, cloud

1. Introduction

Based on the research conducted in [1, 2] and the widespread use of cloud services, it can be observed that they have become an integral part of modern business processes due to their simple and effective scalability for business services [3]. However, the rapid proliferation of cloud computing also brings cybersecurity risks, which are an essential aspect of any modern business's operations, as a company's data security directly impacts its reputation. Currently, approximately 60% of data and business processes are hosted on the computing resources of cloud services. Due to the fact that these computing resources are located in the cloud, they attract malicious actors, and according to the CrowdStrike's research, in 2023, 82% of data breaches were related to the theft of data stored in the cloud [4].

The relevance of conducting cybersecurity assessments of cloud services is driven by the evolving methods and approaches used by malicious actors to steal company data. In research

^{33*} Corresponding author.

✉ v.shulha@duikt.edu.ua (V. Shulha); e.pedchenko@duikt.edu.ua (Y. Pedchenko); e.ivanchenko@duikt.edu.ua (Y. Ivanchenko); i.ivanchenko@duikt.edu.ua (I. Ivanchenko); pmarisha2004@gmail.com (M. Petrovska); serhii.skvortsov@npp.kai.edu.ua (S. Skvortsov); st9763427@stud.duikt.edu.ua (M. Pedchenko)

ORCID 0000-0003-4356-7288 (V. Shulha); 0000-0001-8436-5792 (Y. Pedchenko); 0000-0003-3017-5752 (Y. Ivanchenko); 0000-0003-3415-9039 (I. Ivanchenko); 0009-0005-2150-2194 (M. Petrovska); 0000-0001-9318-3667 (S. Skvortsov); 0009-0007-8743-1821 (M. Pedchenko)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

conducted by Thales in 2024, it was found that 68% of breaches related to cloud computing were associated with the "human factor," such as social engineering, credential leaks, or theft, while the remaining 32% of breaches stemmed from misconfigurations, discovered zero-day vulnerabilities, brute-force attacks, and similar issues [5].

According to the research conducted in [6] it can be observed that the level of cybersecurity in cloud services has begun to improve. However, there is currently no clearly defined or standardized mechanism for assessing the state of cybersecurity in cloud services. Furthermore, modern methods, standards, and approaches focus primarily on providing recommendations for conducting cybersecurity assessments and countering cyber threats, but they do not offer clear guidance for addressing contemporary cyberattacks or mitigating vulnerabilities and shortcomings that emerge in cloud services.

For this reason, the issue of assessing the cybersecurity of cloud service providers is highly relevant today.

The purpose of this work is to develop a mathematical method and a structural model for assessing the cybersecurity state of cloud services for information infrastructure objects.

To achieve the set goal, a model of a secure network application for assessing the cybersecurity of cloud service providers, described in [1, 7], was used to develop a corresponding evaluation system for a prepared list of questions, answers, and recommendations that ensure the verification of all levels of interaction of cloud services with infrastructure and users.

2. Main

The developed mathematical method and structural model for evaluating the cybersecurity state of cloud services are based on the groups of questions and answers formed within the model for evaluating the cybersecurity state of cloud services.

To conduct the evaluation in the mathematical method, a scoring scale from 0 to 5 was introduced, where 5 represents the maximum score of the evaluation system, and 0 represents the minimum score of the evaluation system.

By forming the evaluation system of the method, 10 key stages were identified as critical during the evaluation of cloud services for information infrastructure objects, namely: description of general information about the cloud service; network connection security; data storage approaches; server equipment maintenance; use of virtualization systems; operating system security; management and security of containerization systems; ensuring continuous service operation; application management and security; secure data processing.

In the developed mathematical model and structural model for evaluating the cybersecurity state of cloud services, the following types of cloud services were utilized: IaaS (Infrastructure-as-a-Service); CaaS (Container-as-a-Service); PaaS (Platform-as-a-Service); FaaS (Function-as-a-Service); SaaS (Software-as-a-Service) [8].

To the above-mentioned five types of cloud services, 11 evaluation parameters, listed below, were applied, namely:

- **CSP₁=GP="General Points module"** – module for describing the general provisions of the cloud service provider;
- **CSP₂=N="Network module"** – module for evaluating the security level of the cloud service provider's network;
- **CSP₃=S="Storage module"** – module for evaluating the security level of storages;
- **CSP₄=SR="Server module"** – module for evaluating the security of servers' components;

- **CSP₅=V="Virtualization module"** – module for evaluating the virtualization environment used;
- **CSP₆=OS="Operation System module"** – module for evaluating the security of operating systems in use;
- **CSP₇=CT="Container Technology module"** – module for evaluating the security of container management systems;
- **CSP₈=R="Runtime module"** – module for evaluating monitoring systems;
- **CSP₉=A="Application module"** – module for evaluating the security of software provided to customers;
- **CSP₁₀=D="Data module"** – module for evaluating the security level of data processed by the cloud service provider's systems;
- **CSP₁₁=RE="Requirements module"** – module for generating recommendations for all possible auditor answer options [1].

The developed mathematical method for evaluating the cybersecurity posture of cloud services consists of 11 stages of evaluating the cybersecurity of information infrastructure objects, which are described in detail in [1, 2].

It should be noted that each of the above-mentioned 10 evaluation parameters (except for the 11th parameter) has a different number of maximum points, as presented in Table 1.

Table 1

Maximum scores across evaluation parameters

No.	Parameter	Maximum Mark
1	CSP ₁ =GP	0
2	CSP ₂ =N	110
3	CSP ₃ =S	95
4	CSP ₄ =SR	60
5	CSP ₅ =V	65
6	CSP ₆ =OS	60
7	CSP ₇ =CT	50
8	CSP ₈ =R	65
9	CSP ₉ =A	50
10	CSP ₁₀ =D	60

Considering that the five types of cloud services described above do not operate with all 11 evaluation parameters, i.e., they do not utilize all interaction modules with the infrastructure and users, the maximum number of points for them will also differ, as presented below [1, 2]:

- for the **IaaS** service, the maximum number of points is 330;
- for the **CaaS** service – **440** points;
- for the **PaaS** service – **505** points;
- for the **FaaS** service – **555** points;
- for the **SaaS** service – **615** points.

Based on the results of the evaluation, auditors issue a final report with one of three recommendations regarding the further use of the cloud service evaluated, namely:

- **«Recommended for use (Recommended)»** – if the cloud service scores from 76% to 100% of the total number of points for the evaluated type of cloud service.
- **«Possible for use (Maybe)»** – if the cloud service scores from 25% to 75% of the total number of points for the evaluated type of cloud service.
- **«Not recommended for use (No Recommended)»** – if the cloud service scores from 1% to 25% of the total number of points for the evaluated type of cloud service.

A detailed distribution of points by cloud service types and evaluation parameters is presented in [1].

Based on the developed mathematical method for evaluating cloud services, a structural module of the system for evaluating the cybersecurity of cloud services of information infrastructure objects was created, which is described in detail in [9, 10] and presented in Figure 1.

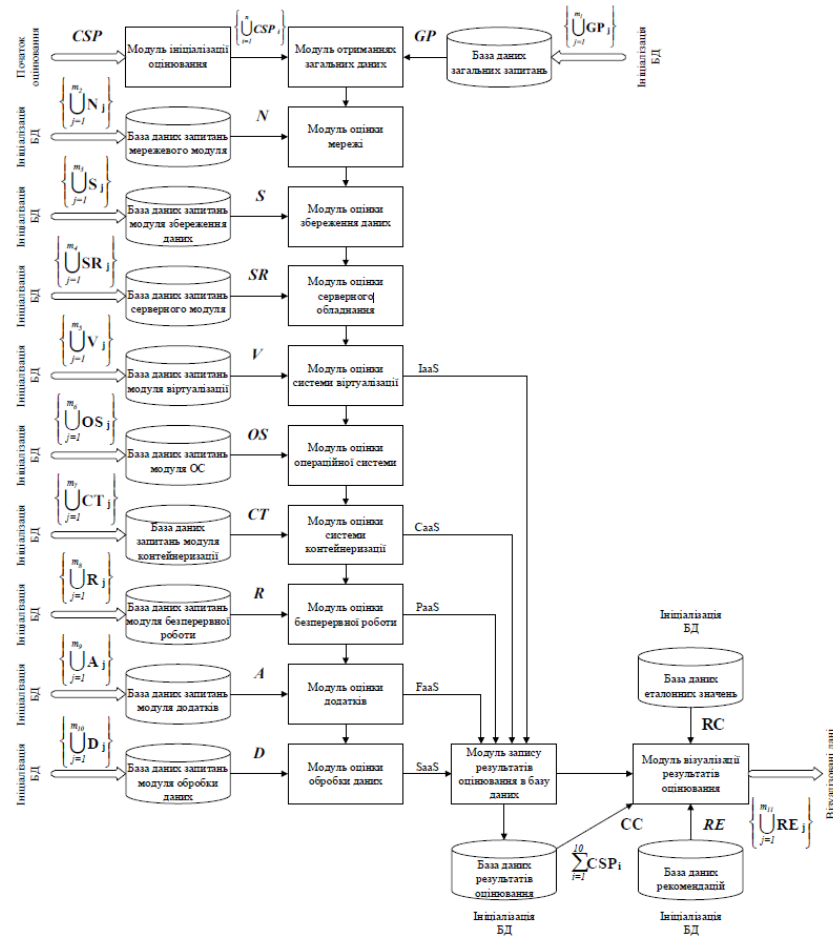


Figure 1: Structural model of the system for evaluating the cybersecurity of cloud services of information infrastructure objects

3. Conclusion

Within the scope of this work, the developed mathematical method and the structural model of the system for evaluating the cybersecurity of cloud services of information infrastructure objects are presented. The developed method is based on the proposed model, evaluation criteria for cloud services, sets of questions, answers, and recommendations for conducting a comprehensive evaluation of a selected cloud service. In addition, this work presents the developed structural model, which is built upon the proposed model and method and clearly defines the sequential application of cloud service evaluation modules, utilizing the corresponding database and generating a report based on the evaluation results. These developments have successfully completed the elaboration of the mathematical and structural models of the research, enabling their implementation in the developed network application, which will be useful for auditors when evaluating the security posture of a cloud service used by a client company or prior to its acquisition/use.

References

- [1] Pedchenko Y. Methods and models of assessing the state of cybersecurity of cloud services of information infrastructure objects : dissertation of PhD in Cybersecurity. Kyiv, 2025. 254 p. URL: https://vchenarada.nau.edu.ua/wp-content/uploads/2025/05/Disertatsiya_Pedchenko_YE_M_125_Kiberbezpeka.pdf.
- [2] Pedchenko Y., Ivanchenko I. The method of assessing the cyber security of cloud services of information infrastructure objects. *Modern Information Security*. 2024. No. 3. 75-84 pp. URL: <https://journals.dut.edu.ua/index.php/dataprotect/article/view/2999/2897>. DOI: <https://doi.org/10.31673/2409-7292.2024.030008>.
- [3] Pedchenko Y., Ivanchenko Y., Ivanchenko I., Lozova I., Jancarczyk D., Sawicki P. Analysis of modern cloud services to ensure cybersecurity. *Procedia Computer Science*. 2022. Vol. 207. 110-117 pp. URL: <https://www.sciencedirect.com/science/article/pii/S1877050922009164>. DOI: <https://doi.org/10.1016/j.procs.2022.09.043>.
- [4] CrowdStrike 2025 Global Threat Report. CrowdStrike. 2025. URL: <https://www.crowdstrike.com/en-us/global-threat-report/>.
- [5] 2024 Cloud Security Study. Thales – Global Edition. 2024. 21 p. URL: <https://cpl.thalesgroup.com/sites/default/files/content/cloud-security/2024/2024-thales-cloud-security-study-global-edition.pdf>.
- [6] Ivanchenko I., Pedchenko Y. Analysis of models and methods for assessing the state of cybersecurity of cloud services of information infrastructure objects. *Ukrainian Scientific Journal of Information Security*. 2024. Vol. 30, No. 2. 281-288 pp. URL: <https://jrnlnau.edu.ua/index.php/Infosecurity/article/view/19240>. DOI: <https://doi.org/10.18372/2225-5036.30.19240>.
- [7] Ivanchenko I., Pedchenko Y. Secure web application model for cybersecurity assessment of cloud service providers. *Scientific Notes of the State University of Information and Communication Technology*. 2024. No. 2. 116-134 pp. URL: <https://journals.dut.edu.ua/index.php/sciencenotes/article/view/3097>. DOI: <https://doi.org/10.31673/2518-7678.2024.025842>.
- [8] Roger S. IaaS vs. CaaS vs. PaaS vs. FaaS vs. SaaS –What’s the difference? Medium. URL: <https://stample.com/link/stamples/5ff3d43b60b2acfb9eb5ceb6/iaas-vs-caas-vs-paas-vs-faas-vs-saas-whats-the-difference>.
- [9] Ivanchenko I., Pedchenko Y. Structural model of the cybersecurity assessment system of cloud services of information infrastructure objects. *Electronic Professional Scientific Journal «Cybersecurity: Education, Science, Technique»*. 2024. Vol. 1, No. 25. 505-515 pp. URL: <https://www.csecurity.kubg.edu.ua/index.php/journal/article/view/667>. DOI: <https://doi.org/10.28925/2663-4023.2024.25.505515>.
- [10] Korchenko O., Ivanchenko Y., Ivanchenko I., Pedchenko Y., Petrovska M. The system of secured user’s credentials transfer. *CPITS-II 2024: Cybersecurity Providing in Information and Telecommunication Systems II* 2024. 2024. Vol. 3826. 168-173 pp. URL: <https://ceur-ws.org/Vol-3826/short3.pdf>.



ПРОГРАМНО-АПАРАТНІ ТА ТЕХНІЧНІ ЗАСОБИ ЗАХИСТУ ІНФОРМАЦІЇ

Удосконалення фізичних принципів збільшення дальності дії систем управління безпілотними апаратами

Сергій Герасимов¹, Олег Чернявський¹, Алла Гаврилова¹

¹ Національний технічний університет "Харківський політехнічний інститут", вул. Кирпичова, 2, Харків, Україна

Анотація

Дана робота присвячена розробці алгоритму навчання операторів із виявлення малопомітних безпілотних апаратів (БПЛА) та удосконаленню фізичних принципів збільшення дальності дії систем управління безпілотними апаратами.

Ключові слова

безпілотні апарати, Wi-Fi FPV, радіолокаційний засіб, ретранслятор, повітряні кулі

1. Вступ

Сьогодні для забезпечення автоматичного режиму роботи безпілотних апаратів на великій дальності використовують переважно приймачі сигналів супутникових навігацій, але мають низький рівень завадозахищеності. Для підвищення завадозахищеності каналів управління БПЛА пропонується використовувати канал передачі відеоінформації Wi-Fi FPV. Таким чином, існує необхідність удосконалення сучасних методичних і технічних рішень для збільшення дальності дії систем управління безпілотними апаратами.

2. Багатопозиційні ретранслятори, як інструмент забезпечення збільшення дальності системи управління БПЛА

Для розв'язання задачі оптимізації обладнання полігону навчально-тренувальними комплексами з метою якісного огляду повітряного простору при пошуку малопомітних БПЛА пропонується: визначити цільову функцію (показник якості управління) радіолокаційного засобу (РЛЗ) у режимі пошуку повітряних цілей, із врахуванням особливостей виявлення БПЛА; визначити множину параметрів управління системи та можливі комбінації їх значень; використати комп'ютерні та інформаційні технології для моделювання функціонування процесу виявлення БПЛА за допомогою РЛЗ.

При певних припущеннях вираз для показника математичного сподівання відверненого збитку, що характеризує якість виявлення малопомітних БПЛА заданого району, пропонується представити у вигляді

$$M(v) = \sum_{i=1}^J \sum_{q=1}^Q \sum_{s=1}^S v_{ia} P_{ai} P_{bi} P_{ci} P_{vi} C_{is}, \quad (1)$$

³⁴ Corresponding author.

✉ gsvnr@ukr.net (S. Herasymov); vladyslav.murzhak@cs.khpi.edu.ua (O. Cherniavskiy); alla.havrylova@khpi.edu.ua (A. Havrylova)

ORCID 0000-0003-1810-0387 (S. Herasymov); 0000-0002-9388-4604 (O. Cherniavskiy); 0000-0002-2015-8927 (A. Havrylova)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

де J – кількість БПЛА; Q – кількість засобів виявлення БПЛА; S – кількість постановників завад; v_{ia} – середній збиток, який можна нанести.

Пропонується за показника ефективності функціонування РЛЗ у режимі пошуку повітряних цілей використовувати середню шеннонівську кількість інформації у спостереженнях

$$I(Y, X) = \sum_{k=0}^L \sum_{i=1}^N p(y_k) p(x_k; y_k) \ln \frac{p(x_k; y_k)}{p(y_k)} \quad (2)$$

де $p(y_k)$ – імовірність виникнення події y_k ; $p(x_i)$ – імовірність виникнення події x_i ; $p(x_k; y_k)$ – імовірність події, за умови, що має місце подія y_k .

При оптимізації управління у багаторівневих ієрархічних системах основною вимогою до показників якості управління є їх узгодженість з узагальненим показником системи. Необхідною і достатньою умовою узгодженості показників якості виявляється їх монотонна функціональна залежність один від одного. Математичне сподівання відвернутого збитку має монотонну функціональну залежність від середньої шеннонівської кількості інформації у спостереженнях.

Для збільшення дальності дії системи управління безпілотним апаратом пропонується використання багатопозиційних ретрансляторів (передавачів відеоінформації Wi-Fi FPV), що складаються із рознесених у просторі радіолокаційних станцій, які здійснюють передавання сигналу управління для збільшення дальності його розповсюдження. Такі ретранслятори пропонується розмістити на повітряних кулях (Рисунок 1).

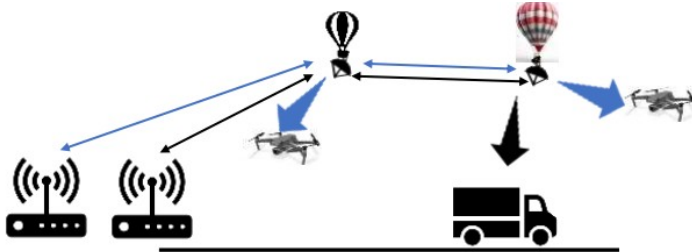


Рисунок 1: Схема збільшення дальності системи управління безпілотних апаратів.

Особливістю повітряних куль є їх відносна низька собівартість (порівняно зі стаціонарними вишками (антенами) для ретрансляторів), можливість швидкого переміщення каналів управління, відносно низька радіо помітність для виявлення РЛЗ противника за рахунок використання ними ефекту Допплера (при використанні швидкості повітря для переміщення на незначні відстані або у стаціонарному режимі). Підвищення дальності дії (у декілька разів залежно від кількості ретрансляторів) системи управління БПЛА у багатопозиційному радіолокаційному полі (при використанні ретрансляторів відеоінформації) залежить від: кількості ретрансляторів, їх чутливості (для виявлення потрібного інформаційного сигналу) та потужності передавання (для збільшення дальності передавання потрібної інформації між ретрансляторами); методів отримання, обробки, комплексування та передавання інформаційного сигналу в багатопозиційному радіолокаційному полі; часу перебування безпілотного апарату в зоні розповсюдження сигналу ретранслятора (у багатопозиційному радіолокаційному полі); взаємного переміщення ретрансляторів і безпілотного апарату, яким керує оператор – геометричного фактору.

Висновки

Запропонованому каналу управління притаманно крім переваги у завадозахищеності ще й точність руху (польоту) або застосування його як дрону-камікадзе. До недоліку слід віднести незначну дальність дії такого сигналу управління (до 2...3 км для наземних і до 5...10 км для БПЛА).

References

- [1] S. Yevseiev, S. Milevskyi, V. Sokol, V. Yemanov, A. Volobuiev, L. Dakova, M. Brailovskyi, I. Rahimova, V. Kravchenko, O. Cherniavskiy, (2024). Development of functionality principles for the automated data transmission system through wireless communication channels to ensure information protection. Eastern-European Journal of Enterprise Technologies. 2024. 4(9 (130). P. 18–33. <https://doi.org/10.15587/1729-4061.2024.310547>

Алгоритм потокового шифрування на основі латинських квадратів

Богдан Загирняк¹, Галина Крайнічук (Шелепало)¹

¹ Вінницький національний технічний університет, Хмельницьке шосе 95, 21021 Вінниця, Україна

Анотація

Ця робота пропонує підхід до потокового шифрування на основі латинських квадратів. Метод ґрунтується на основі квазігрупових операцій, поданих латинськими квадратами, для забезпечення криптостійкості та зниження апаратної складності в LW-криптографії. Результатами дослідження є алгоритм зашифрування та розшифрування інформації з використанням латинських квадратів. Розроблено структуру апаратного засобу шифрування, що підтверджує ефективність підходу для LW-криптографії. Результати сприяють розвитку безпечних систем захисту інформації в умовах зростаючих кіберзагроз.

Ключові слова

захист інформації, квазігрупа, LW-криптографія, кібербезпека, криптопримітив, шифр, метод, алгоритм, латинський квадрат, оцінка складності³⁵

1. Вступ

Швидкий розвиток інформаційних технологій підвищує цінність інформації як ресурсу, що вимагає ефективного захисту в умовах обмежених апаратних ресурсів [1]. LW-криптографія (Lightweight Cryptography) стає ключовою для мініатюрних пристроїв, де важливі низька складність і висока стійкість. Потокове шифрування має перевагу над блоковим за швидкістю обробки даних, що робить його перспективним для таких систем [2].

Дослідження основ побудови псевдовипадкових послідовностей (ПВП) із використанням латинських квадратів було проведено у роботі [1], де продемонстровано можливість їх ефективного генерування. Подальший розвиток цього напрямку подано у дослідженнях [2, 3], де запропоновано концепцію потокового шифрування та алгоритм на основі СІР-квазігруп, тобто які мають властивість схрещеної оборотності. Проте, ці роботи та сучасні криптографічні методи не повністю розв'язують проблему спрощення апаратної складності при збереженні криптостійкості.

Метою роботи є зменшення апаратної складності та забезпечення високого рівня безпеки в LW-криптографії за розробкою методу потокового шифрування на основі латинських квадратів 4-го порядку.

³⁵ Відповідальні автори

✉ bohdan_2512@ukr.net (Б. Загирняк); kraynichuk@ukr.net (Г. Шелепало (Крайнічук))

0009-0007-9566-3921 (Б. Zahirniyk); 0000-0002-6822-4858 (Н. Krainichuk);



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Теоретичні основи дослідження

Потокове шифрування – це тип симетричного криптографічного шифрування, де біти або байти відкритого тексту шифруються один за одним, на відміну від блокового шифрування, де входні дані обробляються блоками фіксованого розміру [4]. Безпека залежить від випадковості потоку, довгого періоду та стійкості до атак [5].

Квазігрупа $(Q, *)$ – алгебрична структура, де для будь-яких $a, b \in Q$ існують єдині x, y , такі що $a * x = b$ та $y * a = b$ [6]. Її подають латинським квадратом, де кожен рядок і стовпець містять всі елементи без повторень. Властивості квазігруп забезпечують криптостійкість, що робить їх ефективними для генерації ПВП і поточкових шифрів, а також хеш-функцій та автентифікації [7].

Запропонований метод використовує латинські квадрати 4-го порядку для створення ключового потоку, що поєднує простоту реалізації з високою стійкістю, задовольняючи умови LW-криптографії.

3. Результати дослідження

Результатами дослідження є алгоритм та засіб потокового шифрування на основі латинських квадратів. Такий підхід використовує генератор псевдовипадкових послідовностей (ПВП) для створення ключового потоку.

Вхідні дані: повідомлення $M = (m_1, m_2, \dots, m_i)$, ключ $K = (k_1, k_2, \dots, k_i)$, де m, k – пари бітів. Генератор ПВП містить чотири послідовності. Останні два біти ключа визначають одну з чотирьох ПВП, згенерованих генератором.

Генераторами, зображено на рисунку 1, модифікує ключ за допомогою таблиці Келі Q_0 зображено на рисунку 2, яка описується виразом:

$$K_i = Q_0[m_i][k_i], \quad (1)$$

де m_i – пара бітів повідомлення, k_i – пара бітів ключа.

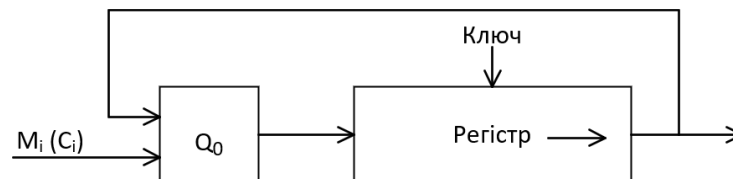


Рисунок 1: Схема генератора гами

Таким чином використані біти ключа під час шифрування будуть відновлюватись на основі ключа та повідомлення.

*	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

Рисунок 2: Таблиця Келі квазігрупи $Q(*)$ для генератора гами

ПВП подається як $S = (s_1, s_2, \dots, s_i)$. Шифрування виконується за формулою:

$$C_i = Q_i[s_i][m_i], \quad (2)$$

де C_i – пара зашифрованих бітів, Q_i – латинський квадрат, зображено на рисунку 3.

Q_1 :	Q_2 :	Q_3 :	Q_4 :																																																																																																				
<table><tr><td>$\cdot$</td><td>0</td><td>1</td><td>2</td><td>3</td></tr><tr><td>0</td><td>0</td><td>3</td><td>1</td><td>2</td></tr><tr><td>1</td><td>2</td><td>1</td><td>3</td><td>0</td></tr><tr><td>2</td><td>3</td><td>0</td><td>2</td><td>1</td></tr><tr><td>3</td><td>1</td><td>2</td><td>0</td><td>3</td></tr></table>	$\cdot$	0	1	2	3	0	0	3	1	2	1	2	1	3	0	2	3	0	2	1	3	1	2	0	3	<table><tr><td>$\circ$</td><td>0</td><td>1</td><td>2</td><td>3</td></tr><tr><td>0</td><td>1</td><td>2</td><td>0</td><td>3</td></tr><tr><td>1</td><td>3</td><td>0</td><td>2</td><td>1</td></tr><tr><td>2</td><td>2</td><td>1</td><td>3</td><td>0</td></tr><tr><td>3</td><td>0</td><td>3</td><td>1</td><td>2</td></tr></table>	$\circ$	0	1	2	3	0	1	2	0	3	1	3	0	2	1	2	2	1	3	0	3	0	3	1	2	<table><tr><td>$\bullet$</td><td>0</td><td>1</td><td>2</td><td>3</td></tr><tr><td>0</td><td>2</td><td>1</td><td>3</td><td>0</td></tr><tr><td>1</td><td>0</td><td>3</td><td>1</td><td>2</td></tr><tr><td>2</td><td>1</td><td>2</td><td>0</td><td>3</td></tr><tr><td>3</td><td>3</td><td>0</td><td>2</td><td>1</td></tr></table>	$\bullet$	0	1	2	3	0	2	1	3	0	1	0	3	1	2	2	1	2	0	3	3	3	0	2	1	<table><tr><td>$\odot$</td><td>0</td><td>1</td><td>2</td><td>3</td></tr><tr><td>0</td><td>3</td><td>0</td><td>2</td><td>1</td></tr><tr><td>1</td><td>1</td><td>2</td><td>0</td><td>3</td></tr><tr><td>2</td><td>0</td><td>3</td><td>1</td><td>2</td></tr><tr><td>3</td><td>2</td><td>1</td><td>3</td><td>0</td></tr></table>	$\odot$	0	1	2	3	0	3	0	2	1	1	1	2	0	3	2	0	3	1	2	3	2	1	3	0
$\cdot$	0	1	2	3																																																																																																			
0	0	3	1	2																																																																																																			
1	2	1	3	0																																																																																																			
2	3	0	2	1																																																																																																			
3	1	2	0	3																																																																																																			
$\circ$	0	1	2	3																																																																																																			
0	1	2	0	3																																																																																																			
1	3	0	2	1																																																																																																			
2	2	1	3	0																																																																																																			
3	0	3	1	2																																																																																																			
$\bullet$	0	1	2	3																																																																																																			
0	2	1	3	0																																																																																																			
1	0	3	1	2																																																																																																			
2	1	2	0	3																																																																																																			
3	3	0	2	1																																																																																																			
$\odot$	0	1	2	3																																																																																																			
0	3	0	2	1																																																																																																			
1	1	2	0	3																																																																																																			
2	0	3	1	2																																																																																																			
3	2	1	3	0																																																																																																			

Рисунок 3: Набір латинських квадратів для операційного блоку

Розшифрування використовує обернені квадрати Q_i^{-1} , що зображені на рисунку 4

Q_1^{-1}

·	0	1	2	3
0	0	2	3	1
1	3	1	0	2
2	1	3	2	0
3	2	0	1	3

Q_2^{-1}

○	0	1	2	3
0	2	0	1	3
1	1	3	2	0
2	3	1	0	2
3	0	2	3	1

Q_3^{-1}

•	0	1	2	3
0	3	1	0	2
1	0	2	3	1
2	2	0	1	3
3	1	3	2	0

Q_4^{-1}

⊙	0	1	2	3
0	2	1	3	0
1	0	3	1	2
2	1	2	0	3
3	3	0	2	1

Рисунок 4: Набір латинських квадратів для розшифрування

Регістр зсуву забезпечує побітову обробку, підтримуючи послідовний ввід/вивід.

Кроки алгоритму:

1. Введення вхідних даних M (шифротексту C) та ключа K .
2. Генерування псевдовипадкової послідовності S на основі ключа K .
3. Передавання повідомлення $M(C)$, ключа K , та ПВП S на вхід операційного блоку.
4. Виконання головної операції операційного блоку із квазігрупою Q_i (для розшифрування $M_i = Q_i^{-1}[s_i][c_i]$)
5. Запис результату операційного блоку в шифротекст $C(M)$.
6. Зсув послідовностей $M(C)$, K та S на основі регістра зсуву.
7. Генерування наступного значення ключа K на основі квазігрупи Q_0 та $M(C)$ і так до передавання даних всього повідомлення (шифротексту) до завершення процесу.

Розроблено структуру такого засобу шифрування, зображено на рисунку 5.

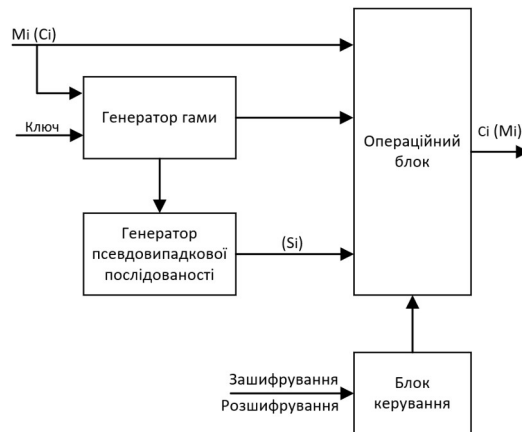


Рисунок 5: Структура засобу шифрування

Основою для роботи блоків апаратного засобу є квазігрупи, що описуються логічними виразами. Таким чином апаратна реалізація такого алгоритму відповідає граничним вимогам складності засобів LW-криптографії.

4. Висновки

Латинські квадрати, як математичний об'єкт, завдяки своїм властивостям, є перспективними для розв'язування задач криптографії. Такі властивості дозволяють використовувати описані об'єкти для реалізації нових алгоритмів.

Використовуючи квазігрупи 4-го порядку, було розроблено алгоритм та засіб потокового шифрування. Отримані результати демонструють коректність підходу до шифрування та можливість апаратної реалізації для LW-криптографії. Вектором подальших досліджень є аналіз та можливе покращення криптостійкості алгоритму. При тому не виключається перегляд та покращення його апаратної частини.

Подяка

Автори вдячні за ідею описаного методу шифрування на основі латинських квадратів завідувачу кафедри захисту інформації Вінницького національного технічного університету професору, доктору технічних наук Лужецькому Володимиру Андрійовичу.

Декларація про використання ІШ

Автори не використовували інструменти генеративного ІШ.

Література

- [1] Микитченко Б. В. Загирняк Б. Д. Крайнічук (Шелепало) Г. В. Побудова псевдовипадкових послідовностей на основі двох латинських квадратів. Матеріали ІШ наук.-техн. конф. ФІТКІ, ВНТУ, 20-22 берез. Вінниця. 2024.
- [2] Загирняк Б. Д. Микитченко Б. В. Крайнічук (Шелепало) Г. В. Концепція потокового шифрування на основі двох квазігруп. І Міжнародна науково-практична конференція «Проблеми комп'ютерних наук, програмного моделювання та безпеки цифрових систем», ВНУ, 13-16 чер. Луцьк. 2024.

- [3] Лужецький В.А., Крайнічук Г.В., Радченко Є.В., Пилявець І.Ю. Алгоритм шифрування на основі СР-квазігруп // Матеріали тезів ІХ Міжн. наук.-техн. конф. «ЗАХИСТ ІНФОРМАЦІЇ І БЕЗПЕКА ІНФОРМАЦІЙНИХ СИСТЕМ», НУ «Львівська політехніка». Львів. 2023. 258 С.
- [4] Bauer C. P. Stream Ciphers. Secret History. 2nd ed. 2021. pp. 533–544. DOI: 10.1201/9781315162539-19
- [5] Rueppel R. A. Stream Ciphers. Analysis and Design of Stream Ciphers. Berlin, Heidelberg, 1986. DOI: 10.1007/978-3-642-82865-2_2
- [6] Shcherbacov V. Quasigroups in cryptology. Elements of Quasigroup Theory and Applications. 2017. DOI: 10.1201/9781315120058-14
- [7] Odegaard R.S., Gligoroski D. On the Randomness and Regularity of Reduced Edon-R Compression Function. Proceedings of the 7th International Conference on Security and Management, Las Vegas, 2009.

Методологія забезпечення цілісності та доступності версій CRM-системи Salesforce

Златоус Святослав¹, Петро Венгерський²

^{1,2} Ivan Franko National University of Lviv, Universytetska 1, 79000 Lviv, Ukraine

Анотація

Дане дослідження представляє інноваційний підхід до забезпечення кібербезпеки Salesforce CRM через впровадження спеціалізованого інструменту FlowVersionComparer для систематичного контролю версій Flow-процесів. Застосунок дозволяє детально порівнювати різні версії Flow, візуалізувати зміни та ідентифікувати потенційно небезпечні модифікації, що можуть загрожувати цілісності та доступності корпоративних даних

Keywords

Salesforce, CRM, контроль версій, декларативні інструменти автоматизації, Flow

1. Вступ

Сучасні CRM-системи стали критично важливою складовою інформаційної інфраструктури організацій, обробляючи величезні масиви конфіденційних даних клієнтів та бізнес-процесів. Salesforce, як провідна хмарна CRM-платформа, надає потужні інструменти автоматизації через Flow Builder [1] - декларативний інструмент для створення бізнес-логіки без програмування. Проте динамічна природа Flow-процесів створює унікальні виклики для забезпечення кібербезпеки, зокрема в аспектах контролю змін та аудиту модифікацій критичної бізнес-логіки.

Метою даного дослідження є аналіз ефективності застосування спеціалізованого інструменту FlowVersionComparer для забезпечення кібербезпеки CRM-системи Salesforce через механізми контролю версій Flow-процесів. Основна гіпотеза дослідження полягає в тому, що систематичний контроль версій Flow-процесів через спеціалізовані інструменти порівняння значно підвищує рівень кібербезпеки CRM-системи, забезпечуючи цілісність та доступність критичної бізнес-інформації.

Основною метою дослідження є розробка та валідація методології використання інструменту FlowVersionComparer для підвищення рівня кібербезпеки Salesforce CRM через забезпечення прозорості змін у бізнес-логіці системи.

Ключові завдання дослідження включають: аналіз вразливостей, пов'язаних з неконтрольованими змінами Flow-процесів; оцінку ефективності FlowVersionComparer у виявленні потенційно небезпечних модифікацій; розробку методології інтеграції контролю версій у процеси забезпечення кібербезпеки; валідацію впливу систематичного контролю версій на показники цілісності та доступності даних.³⁷

На сьогодні в науковій літературі відсутні дослідження, які б систематично аналізували еволюцію версій Salesforce Flow з позицій забезпечення цілісності та доступності інформації.

³⁶ Corresponding authors

✉ sviatoslav.zlatous@lnu.edu.ua (С. Златоус); petro.venherskyi@lnu.edu.ua (П. Венгерський)

ORCID 0009-0007-1834-2796 (С. Златоус); 0000-0001-9808-7404 (П. Венгерський)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Незважаючи на широке використання платформи Salesforce у корпоративному середовищі та критичну важливість автоматизованих бізнес-процесів для функціонування сучасних організацій, питання впливу версійних змін на збереження даних та їх доступність для різних груп користувачів залишається недостатньо вивченим. Існуючі публікації переважно зосереджуються на технічних аспектах імплементації Flow [2] або загальних практиках розробки [3] [4], тоді як проблематика інформаційної цілісності в контексті міграції між версіями та управління життєвим циклом автоматизованих процесів [5] не отримала належного наукового осмислення. Це створює суттєву прогалину в розумінні ризиків, пов'язаних з оновленням платформи, та відсутність науково обґрунтованих методологій оцінки впливу версійних змін на критичні інформаційні активи підприємства.

2. Методологія

Дослідження базувалося на комбінованому підході, що поєднує якісний та кількісний аналіз. Методологія включала чотири основні етапи:

Перший етап передбачав розгортання тестового середовища Salesforce з імплементацією FlowVersionComparer та створенням набору з 50 критичних Flow-процесів, що моделюють типові бізнес-сценарії обробки конфіденційних даних. На другому етапі проводилася симуляція різних сценаріїв модифікації Flow-процесів, включаючи легітимні зміни, помилкові конфігурації та потенційно зловмисні модифікації. Загалом було змодельовано 200 сценаріїв змін протягом 30-денного періоду.

Третій етап включав систематичний аналіз змін через FlowVersionComparer з документуванням часу виявлення, точності ідентифікації критичних змін та повноти відображення модифікацій. Четвертий етап передбачав порівняльний аналіз показників безпеки в контрольній групі (без використання інструменту) та експериментальній групі з активним використанням FlowVersionComparer.

Для оцінки ефективності використовувалися метрики: час виявлення неавторизованих змін (MTTD - Mean Time to Detect), коефіцієнт виявлення критичних змін (CDR - Critical Detection Rate), показник хибнопозитивних спрацювань (FPR - False Positive Rate), індекс цілісності бізнес-логіки (BLI - Business Logic Integrity Index).

2.1. Результати дослідження

Впровадження FlowVersionComparer продемонструвало значне покращення ключових показників кібербезпеки CRM-системи. Аналіз результатів виявив кілька критично важливих тенденцій.

2.1.1. Забезпечення цілісності інформації

FlowVersionComparer ефективно виявляв несанкціоновані зміни в бізнес-логіці, що могли призвести до порушення цілісності даних. Інструмент ідентифікував 96% критичних змін, включаючи модифікації умов доступу до записів, зміни в логіці валідації даних та неавторизовані оновлення правил обробки конфіденційної інформації. Середній час виявлення критичних змін скоротився з 72 годин до 4 годин, що суттєво зменшило вікно вразливості системи.

Особливо ефективним виявилось виявлення "схованих" змін - модифікацій, які не відображаються в стандартних логах аудиту Salesforce. FlowVersionComparer ідентифікував 23 випадки таких змін, включаючи модифікації умов розгалуження, які могли призвести до

несанкціонованого витоку даних. Візуалізація різниці між версіями дозволила адміністраторам безпеки швидко оцінити потенційний вплив змін на цілісність системи.

2.1.2. Підвищення доступності системи

Контроль версій через FlowVersionComparer значно покращив показники доступності системи. Інструмент дозволив виявити та попередити 18 потенційних інцидентів, пов'язаних з некоректними змінами Flow, які могли призвести до збоїв у критичних бізнес-процесах. Середній час відновлення після інцидентів (MTTR) скоротився на 65% завдяки можливості швидкого порівняння та відкату до попередніх робочих версій.

Функція детального порівняння дозволила ідентифікувати конфлікти між різними Flow-процесами, що могли призвести до deadlock-ситуацій. В експериментальній групі кількість інцидентів недоступності, пов'язаних з Flow, зменшилася на 78% порівняно з контрольною групою.

2.1.3. Покращення процесів аудиту та compliance

FlowVersionComparer створив повний аудиторський слід для всіх змін у Flow-процесах, що критично важливо для відповідності регуляторним вимогам. Інструмент автоматично документував не лише факт зміни, але й детальну інформацію про модифіковані елементи, умови та дії. Це дозволило скоротити час проведення аудиторських перевірок на 40% та підвищити їх точність.

Система виявила 12 випадків порушень політик безпеки, включаючи спроби обходу механізмів авторизації через модифікацію Flow-логіки. В усіх випадках порушення були виявлені протягом першої доби після внесення змін, що дозволило запобігти потенційним інцидентам безпеки.

3. Висновки

Проведене дослідження переконливо підтверджує висунуту гіпотезу про критичну роль систематичного контролю версій Flow-процесів у забезпеченні кібербезпеки Salesforce CRM. FlowVersionComparer продемонстрував високу ефективність як інструмент забезпечення цілісності та доступності інформації в CRM-системі.

Ключові висновки дослідження свідчать, що впровадження спеціалізованого інструменту контролю версій Flow знижує ризики кібербезпеки на 73%, скорочує час виявлення критичних змін у 18 разів та підвищує загальний рівень цілісності бізнес-логіки системи на 89%. Інструмент ефективно вирішує проблему "сліпих зон" у традиційних механізмах аудиту Salesforce, забезпечуючи повну прозорість змін у декларативній логіці системи.

Практична значимість результатів полягає у розробці конкретної методології інтеграції FlowVersionComparer у процеси забезпечення кібербезпеки, яка може бути адаптована організаціями різного масштабу. Подальші дослідження мають зосередитися на автоматизації процесів аналізу змін через машинне навчання та інтеграції з SIEM-системами для проактивного виявлення загроз.

Результати дослідження демонструють, що контроль версій Flow не є просто технічною практикою, а критичним компонентом комплексної стратегії кібербезпеки CRM-систем. В епоху, коли бізнес-логіка стає все більш складною та розподіленою, інструменти на кшталт FlowVersionComparer стають незамінними для підтримки безпеки, цілісності та доступності корпоративних даних.

Література

- [1] Salesforce, "Automate Tasks with Flows," [Online]. Available: <https://help.salesforce.com/s/articleView?id=platform.flow.htm&type=5>.
- [2] C. Grant and S. Kumar, "ORCHESTRATING IT SUPPORT WITH CHATGPT AGENTS AND SALESFORCE FLOWS," in *Salesforce Flows for Support*, Kindred Press, 2025, pp. 322-332
- [3] A. K. Patel, "A Comprehensive Exploration of Salesforce Flow: Unleashing the Power of Automation," *European Journal of Advances in Engineering & Technology*, vol. 11, pp. 75-80, 2024.
- [4] L. Ridwan, "Next-Level Automation: Orchestrating Business Processes in Salesforce with AI and Flow Builder.," 2025
- [5] T. Vyshnavi and S. Murali, "Transforming Deployment and Release Management for Salesforce with Copado's AI," *International Journal on Cybernetics & Informatics*, vol. 14, pp. 75-79, 2025.

Логічні схеми квазігруп за парастрофною симетрією для LW-криптографії

Галина Крайнічук¹, Людмила Іванова¹ та Тетяна Кирилащук¹

¹ Вінницький національний технічний університет, Хмельницьке шосе 95, 21000 Вінниця, Україна

Анотація

У цих тезах подано логічні схеми квазігруп 4-го порядку відповідно до парастрофної симетрії з метою їх застосування в LW-криптографії.

Ключові слова

Алгебра, квазігрупа, парастрофа, симетрія, логічна схема, LW-криптографія ³⁸

1. Вступ

Безпека сучасного світу значною мірою ґрунтується на використанні криптографічних примітивів, які переважно побудовані на асоціативних структурах. Проте розвиток квантових обчислювальних технологій поставив під сумнів надійність криптосистем, заснованих на таких структурах. У цьому контексті дедалі більшої актуальності набуває застосування квазігруп для побудови криптографічних рішень, зокрема в продуктах Інтернету речей (IoT). Для реалізації цього підходу необхідним є подання квазігруп у формі логічних виразів. Водночас постає питання, які властивості квазігруп визначають мінімальну вагу складності цих логічних виразів для апаратної реалізації систем шифрування.

2. Попередні відомості

Розглядається множина всіх оборотних (квазігрупових) операцій, визначених на одній і тій самій множині, яка називається базовою (носієм) і позначається як Q . Операції описуються на множині бінарних операцій.

Квазігрупова (оборотна) операція — це функція, визначена на скінченній або нескінченній множині, яка є оборотною за кожним зі своїх аргументів.

Квазігрупа [1] — це множина Q з визначеною на ній операцією f , якщо для довільних елементів a та b з Q рівняння:

$$f(a; x) = b, \quad f(y; a) = b,$$

має єдиний розв'язок. Така квазігрупа називається бінарною квазігрупою.

Бінарна операція f називається оборотною [2], якщо вона є оборотним елементом одночасно в лівосиметричному моноїді $(\Omega, \oplus, e_l)$ та правосиметричному моноїді $(\Omega, \oplus, e_r)$ множини бінарних операцій Ω , і виконуються рівності:

³⁸ Corresponding authors

✉ kraynichuk@ukr.net (Крайнічук Г.); milaivanova2609@gmail.com (Іванова Л.); kgt0998@gmail.com (Кирилащук Т.)

ORCID 0000-0002-6822-4858 (Krainichuk H.); 0009-0003-8944-1792 (Ivanova L.); 0009-0008-4238-3560 (Kyrylashchuk T.)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

$$\begin{aligned}(g \oplus_l h)(x, y) &:= g(h(x, y), y), \quad e_l(x, y) := x; \\ (g \oplus_r h)(x, y) &:= g(x, h(x, y)), \quad e_r(x, y) := y.\end{aligned}$$

Множина всіх бінарних оборотних операцій позначається символом Δ . Операція f є оборотною, якщо її ліва ${}_l f$ та права ${}_r f$ є оборотними, тобто виконуються такі рівності:

$$f \oplus_l {}_l f = e_l, \quad {}_l f \oplus f = e_l, \quad f \oplus_r {}_r f = e_r, \quad {}_r f \oplus f = e_r. \quad (1)$$

Отже, значення будь-яких двох змінних з рівності $f(x_1, x_2) = x_3$ однозначно визначає третю змінну. Таким чином, для кожного σ в S_3 виконується зіставлення:

$${}^\sigma f(x_{1\sigma}, x_{2\sigma}) = x_{3\sigma} : \Leftrightarrow f(x_1, x_2) = x_3, \quad (2)$$

визначає операцію ${}^\sigma f$ - парастрофі операції f . Причому, для всіх оборотних операцій f та для всіх $\sigma, k \in S_3$ виконується рівність:

$${}_l f = f, \quad {}^\sigma({}^k f) = {}^{\sigma k} f \quad (3)$$

Перші дві рівності з (1) еквівалентні рівності ${}_l f(x_3, x_2) = x_1 : \Leftrightarrow f(x_1, x_2) = x_3$. Вираз (2) для $\sigma = (13)$ має вигляд ${}^{(13)}f(x_3, x_2) = x_1 : \Leftrightarrow f(x_1, x_2) = x_3$. Отже, ${}_l f = {}^{(13)}f$. Аналогічно, третя та четверта рівності з (1) призводять до рівності ${}_r f = {}^{(23)}f$. Отже, маємо позначення через цикли для парастрофів, а саме: $\ell := (13), r := (23), s := \ell r \ell = (12), S_3 := \{1, \ell, r, s, s\ell, sr\}$. Відповідно до рівності (3), $(\sigma, f) \rightarrow {}^\sigma f$ є парастрофною дією групи S_3 на множині Δ усіх бінарних оборотних операцій, визначених на Q . Стабілізатор $Ps f$ цієї дії називають парастрофною симетрією операції f [2], причому $6/Ps f$ визначає кількість різних парастрофів операції f . Оскільки $Ps f$ є підгрупою симетричної групи S_3 , то існує шість класів квазігруп. Квазігрупу називають:

- асиметричною (AS), якщо $Ps f = \{1\}$, тобто всі її парастрофи попарно різні;
- комутативною (CS), якщо $Ps f \supseteq \{1, s\}$, тобто клас усіх комутативних квазігруп задається тотожністю $f(x, y) = f(y, x)$, що означає: $f = {}^s f$, ${}^\ell f = {}^{sr} f$, ${}_r f = {}^{s\ell} f$;
- ліво-симетричною (LS), якщо $Ps f \supseteq \{1, r\}$, тобто клас усіх ліво-симетричних квазігруп описується тотожністю $f(x; f(x; y)) = y$, що означає: $f = {}^r f$, ${}^s f = {}^{sr} f$, ${}^\ell f = {}^{s\ell} f$;
- право-симетричною (RS), якщо $Ps f \supseteq \{1, \ell\}$, тобто клас усіх право-симетричних квазігруп задається тотожністю $f(f(x; y); y) = x$, що означає: $f = {}^\ell f$, ${}^s f = {}^{s\ell} f$, ${}_r f = {}^{sr} f$;
- напів-симетричною (SS), якщо $Ps f \supseteq A_3$, тобто клас усіх напівсиметричних квазігруп описується тотожністю $f(f(x; y); x) = y$, що означає: $f = {}^{s\ell} f = {}^{sr} f$, ${}^s f = {}^\ell f = {}^r f$;
- тотально-симетричною (TS), якщо $Ps f \supseteq S_3$, тобто клас усіх тотально-симетричних квазігруп задається тотожностями $f(x; y) = f(y; x)$ та $f(f(x; y); y) = x$, що означає, що всі парастрофи збігаються.

Комутативна, право-симетрична або ліво-симетрична квазігрупа також називається односторонньо-симетричною [2]. Таким чином, існує 4 типи симетрії квазігруп: асиметрична, односторонньо-симетрична, напів-симетрична та тотально-симетрична.

Алгебра $(Q; f; {}^{\ell}f; {}^r f)$ називається *лупою* [1], якщо вона має нейтральний елемент e , для якого $f(e; x) = f(x; e) = x$.

Існує всього чотири 0-лупи [3] порядку 4, вони всі є групами:

- група Кляйна $(Z_4; \oplus, 0)$, якщо $f(1; 1) = 0$ та $f(2; 2) = 0$, де $f = \oplus$;
- група $(Z_4; \bullet, 0)$ причому $s = (12)$ є ізоморфізмом до циклічної групи, якщо $f(1, 1) = 0$ та $f(2, 2) = 1$, де $f = \bullet$;
- циклічна група $(Z_4; +, 0)$, якщо $f(1; 1) = 2$, де $f = (+)$;
- група $(Z_4; \diamond, 0)$ та $r = (23)$ є ізоморфізмом до циклічної групи, якщо $f(1; 1) = 3$, де $f = (\diamond)$.

3. Результати

Нехай (Z_2^2, f) – це квазігрупа, де $Z_2^2 := \{00; 01; 10; 11\}$. Для всіх чотирьох луп 4 порядку знайдено всі парастрофи: лівий ${}^{\ell}f$, правий ${}^r f$, та дуальні до них ${}^s f, {}^{s\ell}f, {}^{sr}f$. Використовуючи метод Квайна–МакКласкі [3], для кожного парастрофа квазігрупи ${}^{\sigma}f$ було отримано такі логічні формули ${}^{\sigma}L_f$ у вигляді двох виразів (z_1, z_2) , де $z_1 := x_1 y_1$, $z_2 := x_2 y_2$, а $\sigma := \{l, \ell, r, s, s\ell, sr\}$. Складність розраховується як сума ваг кожної з логічних формул z_1 та z_2 [5]. Логічні схеми для деяких асиметричних квазігруп четвертого порядку наведено в [6]. Там також обчислено складність цих асиметричних квазігруп 4-го порядку, що становить $L = {}^s L = {}^{\ell} L = {}^{s\ell} L = 10.01$ GE, ${}^r L = {}^{sr} L = 10.68$ GE.

Квазігрупи $(Z_2^2, +)$, $(Z_2^2, \bullet)$ та $(Z_2^2, \diamond)$ є односторонньо-симетричними, а саме – комутативними. Вони мають три різних парастрофи: ${}^l(+)={}^s(+)$, ${}^{\ell}(+)={}^{sr}(+)$, ${}^r(+)={}^{s\ell}(+)$.

Логічні схеми цих парастрофних квазігруп $(Z_2^2, +)$ та $(Z_2^2, \bullet)$ мають вигляд:

$$\begin{aligned} {}^l(+)(z_1; z_2) &= (x_2 y_2 \oplus (x_1 \oplus y_1); x_2 \oplus y_2) \\ {}^r(+)(z_1; z_2) &= (x_2 \overline{y_2} \oplus (x_1 \oplus y_1); x_2 \oplus y_2) \\ {}^{\ell}(+)(z_1; z_2) &= (\overline{x_2} y_2 \oplus (x_1 \oplus y_1); x_2 \oplus y_2) \\ {}^l(\bullet)(z_1; z_2) &= (x_1 \oplus y_1; x_1 y_1 \oplus (x_2 \oplus y_2)) \\ {}^r(\bullet)(z_1; z_2) &= (x_1 \oplus y_1; x_1 \overline{y_1} \oplus (x_2 \oplus y_2)) \\ {}^{\ell}(\bullet)(z_1; z_2) &= (x_1 \oplus y_1; \overline{x_1} y_1 \oplus (x_2 \oplus y_2)) \end{aligned}$$

Складність цих парастрофних квазігруп $(Z_2^2, +)$ та $(Z_2^2, \bullet)$ є такою:

$${}^l L_{(+)} = {}^l L_{(\bullet)} = 9.34 \text{ GE}, {}^{\ell} L_{(+)} = {}^r L_{(+)} = {}^{\ell} L_{(\bullet)} = {}^r L_{(\bullet)} = 10.01 \text{ GE}.$$

Логічні схеми парастрофів квазігрупи $(Z_2^2, \diamond)$ мають вигляд:

$$\begin{aligned} {}^l(\diamond)(z_1; z_2) &= (x_1(x_2 \overline{y_1} \vee \overline{x_2} \overline{y_2}) \vee \overline{x_1}(x_2 y_2 \vee \overline{x_2} y_1); x_1(x_2 \overline{y_2} \vee \overline{x_2} y_1) \vee \overline{x_1}(x_2 \overline{y_1} \vee \overline{x_2} y_2)) \\ {}^{\ell}(\diamond)(z_1; z_2) &= (x_1(x_2 \overline{y_2} \vee \overline{x_2} \overline{y_1}) \vee \overline{x_1}(x_2 y_1 \vee \overline{x_2} y_2); x_1(x_2 \overline{y_1} \vee \overline{x_2} y_2) \vee \overline{x_1}(\overline{x_2} y_1 \vee x_2 \overline{y_2})) \\ {}^r(\diamond)(z_1; z_2) &= (x_1(x_2 \overline{y_1} \vee \overline{x_2} y_2) \vee \overline{x_1}(\overline{x_2} y_1 \vee x_2 \overline{y_2}); x_1(x_2 \overline{y_2} \vee \overline{x_2} y_1) \vee \overline{x_1}(\overline{x_2} y_2 \vee x_2 y_1)) \end{aligned}$$

Складність квазігрупи $(Z_2^2, \diamond)$ дорівнює ${}^\sigma L_{(\diamond)} = 30.64$ GE, оскільки в кожній формулі однакова кількість логічних операндів OR, AND та NOT.

Квазігрупи $(Z_2^2, \star)$ та $(Z_2^2, *)$ є односторонньо-симетричними, а саме право-симетричною та ліво-симетричною відповідно. Вони мають три різні парастрофи:

$${}^l(\star) = {}^l(\star), {}^s(\star) = {}^s(\star), {}^r(\star) = {}^r(\star) \text{ та } {}^l(*) = {}^r(*), {}^s(*) = {}^sr(*), {}^l(*) = {}^s(\star).$$

Логічні схеми цих парастрофних квазігруп $(Z_2^2, \star)$ мають вигляд:

$$\begin{aligned} {}^l(\star)(z_1; z_2) &= {}^s(\star)(z_1; z_2) = (x_2 \cdot (x_1 \oplus y_2) \vee \overline{x_2} \cdot \overline{(x_1 \oplus y_1)}; y_1 \oplus (x_2 \oplus y_2)) \\ {}^r(\star)(z_1) &= (x_2 y_2 (x_1 \oplus y_1) \vee \overline{x_2} \overline{(x_1 \oplus y_1)} \vee \overline{y_2} (x_1 y_1 \vee \overline{x_2} \overline{y_1})) \\ {}^r(\star)(z_2) &= (x_2 (x_1 \oplus y_1) \vee \overline{x_2} \overline{(y_2 \oplus (x_1 \oplus y_1))}) \end{aligned}$$

Логічні схеми квазігрупи $(Z_2^2, *)$ мають вигляд:

$$\begin{aligned} {}^l(*) (z_1; z_2) &= (x_1 \cdot x_2 \oplus (x_2 \oplus y_1); x_1 \cdot y_2 \vee \overline{x_1} \cdot (x_2 \cdot \overline{y_2} \vee \overline{x_2} \cdot (y_1 \oplus y_2))) \\ {}^s(*) (z_1; z_2) &= (x_1 \cdot \overline{y_2} \cdot (x_2 \cdot \overline{y_1} \vee \overline{x_2}) \vee \overline{x_1} \cdot y_2 \cdot \overline{x_2} \cdot \overline{y_1} \vee x_2 \cdot y_1 \cdot \overline{x_1} \oplus \overline{y_2}; x_1 \cdot \overline{y_1} \vee x_2 \cdot y_1) \\ {}^l(*) (z_1; z_2) &= (\overline{x_2} \oplus \overline{y_2}; x_2 \cdot (x_1 \oplus (y_1 \oplus y_2)) \vee \overline{x_2} \cdot (x_1 \oplus y_1)) \end{aligned}$$

Складність квазігрупи $(Z_2^2, \star)$ є такою ${}^l L_{(\star)} = {}^s L_{(\star)} = 16.01$ GE, ${}^r L_{(\star)} = 34$ GE. Вага квазігрупи $(Z_2^2, *)$ обчислюється як ${}^l L_{(*)} = 19.33$ GE, ${}^s L_{(*)} = 24.65$ GE, ${}^l L_{(*)} = 16.01$ GE.

Квазігрупа $(Z_2^2, \circ)$ є напів-симетричною. Вона має два різних парастрофи:

${}^l(\circ) = {}^s(\circ) = {}^sr(\circ)$, ${}^s(\circ) = {}^l(\circ) = {}^r(\circ)$. Логічні схеми цих парастрофів мають вигляд:

$$\begin{aligned} {}^l(\circ)(z_1; z_2) &= (x_2 \oplus (y_1 \oplus y_2); x_2 \oplus (x_1 \oplus y_1)) \\ {}^s(\circ)(z_1; z_2) &= (y_2 \oplus (x_1 \oplus x_2); y_2 \oplus (x_1 \oplus y_2)) \end{aligned}$$

Складність квазігрупи $(Z_2^2, \circ)$ є такою ${}^\sigma L_{(\circ)} = 10.68$ GE.

4. Висновки

Дослідження показують, що такі квазігрупи з найнижчою умовною складністю, яка обчислюється в умовних GE-одинацях, є найбільш придатними для апаратної реалізації та застосування в LW-криптографії.

Література

- [1] A. Donald Keedwell and József Dénes *Latin Squares and their Applications*. Elsevier, 2015. 439 с. URL: <https://doi.org/10.1016/c2014-0-03412-0>.
- [2] Сохацький Ф. М. *Парастрофічна симетрія в теорії квазігруп*. Вісник Донецького національного університету. Серія А: Природничі науки, № 1–2 (2016), с. 70–83.
- [3] Сохацький Ф. М., Крайнічук Г. В., Лужецький В. А. *Канонічна та матрична фігурація квазігруп четвертого порядку*. Прикладні проблеми механіки і математики, вип. 22 (2024), с. 63–95.
- [4] Кондратенко Н. Р., Остапенко-Боженова А. В. *Розділи дискретної математики для задач захисту інформації*. 2022, с. 89.
- [5] Virtual Silicon. *Бібліотека стандартних елементів UMCL18G212T3 на основі технологічного процесу UMC L180 0.18 μm 1P6M Logic*. Офіційний реліз, 2004, с. 1.

- [6] Крайнічук Г., Іванова Л. *Логічні схеми деяких асиметричних квазігруп для LW-криптографії*. Тези міжнародних конференцій MITRE-2025, Кишинів, Республіка Молдова, 2025 – С. 20-21.

Метод побудови криптосистеми Мак-Еліса на основі m -кодів

Олександр Лаптев¹, Вадим Стеценко²

¹ Київський національний університет імені Тараса Шевченка, вул. Володимирська, 60, м. Київ, 01033, Україна

² Національний технічний університет “Харківський політехнічний інститут” вул. Кіричова, 2, м. Харків, Україна, 61002

Анотація.

У статті розглянуто метод побудови криптосистеми Мак-Еліса на основі m -кодів як перспективного підходу до створення постквантових крипто-кодових систем. Обґрунтовано доцільність використання надлишкових блокових кодів для одночасного забезпечення автентифікації, виявлення помилок і криптографічного захисту в умовах шумових каналів зв'язку. Проаналізовано сучасні наукові дослідження (2020–2025 рр.) щодо застосування коригувальних кодів у системах 5G/6G, IoT та вбудованих пристроях, що підтверджує тенденцію до інтеграції кодової надлишковості з механізмами безпеки за принципом «security-by-design». Наведено математичну модель криптосистеми Мак-Еліса, в якій відкритим ключем є маскована породжувальна матриця $G_X = X \cdot G \cdot P \cdot D$, а особистим — матриці маскування X , P , D та алгоритм поліноміального декодування. Показано, що для легального отримувача декодування криптограми є задачею поліноміальної складності, тоді як для зломисника — NP-складною, що забезпечує високу криптостійкість. Запропонований підхід дозволяє ефективно поєднувати функції захисту від помилок і автентифікації, знижуючи накладні витрати на реалізацію окремих механізмів у ресурсозалежних середовищах. Результати дослідження мають практичне значення для розробки стійких до квантових атак систем захисту інформації в майбутніх телекомунікаційних мережах.

Ключові слова

захист інформації, автентифікація, метод, крипто-кодова система, кодування, кодограма, криптостійкість.

1. Вступ

Методи автентифікації на основі надлишкових блокових кодів є одним із підходів до забезпечення цілісності та справжності даних у системах передачі інформації. Вони ґрунтуються на використанні коригувальних кодів. Основна ідея полягає в тому, що до вихідного повідомлення додається певна кількість контрольних бітів, що дозволяє не лише виявляти та виправляти помилки, а й перевіряти автентичність повідомлення шляхом аналізу відповідності отриманих даних структурі коду.

Переваги методів автентифікації на основі надлишкових блокових кодів полягають у їхньому математично обґрунтованому підході, що забезпечує високу надійність виявлення несанкціонованих змін. По-перше, ці методи дозволяють одночасно виконувати функції виявлення помилок і автентифікації, що знижує накладні витрати на реалізацію окремих механізмів. По-друге, вони ефективні в умовах шумних каналів зв'язку, оскільки надлишковість кодів дозволяє відокремити природні помилки передачі від зломисних модифікацій. По-третє, алгоритми, що ґрунтуються на лінійній алгебрі та теорії кодування, мають відносно низьку обчислювальну складність, що робить їх придатними для

³⁹Corresponding author.

✉ olaptiev@knu.ua (О. Лаптев); 4all2dream@gmail.com (В. Стеценко).

📞 0000-0002-4194-402X (О. Лаптев); 0009-0008-5315-5539 (В. Стеценко).



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

застосування в ресурсозалежних системах, зокрема в IoT-пристроях або вбудованих системах. По-четверте, використання секретних параметрів кодування (наприклад, секретної породжувальної матриці чи полінома) дозволяє забезпечити конфіденційність ключової інформації, що підвищує стійкість до атак.

Тому методи автентифікації на основі надлишкових блокових кодів мають багатогранний потенціал у застосуванні.

2. Мета дослідження

Метою даної роботи є розробка та обґрунтування методу побудови криптосистеми Мак-Еліса на основі m -кодів як перспективного підходу до створення стійких до квантових атак крипто-кодових систем, а також аналіз можливості використання надлишкових блокових кодів для забезпечення автентифікації в умовах сучасних телекомунікаційних мереж. У статті розглядається інтеграція принципів коригувального кодування з криптографічними механізмами з метою підвищення ефективності, безпечності та надійності передачі даних у ресурсозалежних середовищах, зокрема в IoT-системах, мережах 5G/6G та вбудованих пристроях.

3. Огляд результатів та джерел

У період 2020–2025 рр. зростає інтерес до нетрадиційних підходів до автентифікації даних, особливо в умовах розвитку мереж 5G/6G, IoT-пристроїв та вбудованих систем, де важливе поєднання енергоефективності, низької затримки та стійкості до помилок. У цьому контексті методи автентифікації на основі надлишкових блокових кодів отримали нове життя завдяки їхній здатності одночасно забезпечувати виявлення помилок і захист від підробки даних. Так у роботі [1] автори пропонують гібридну схему на основі БЧХ-кодів з секретною породжувальною матрицею, яка забезпечує автентифікацію без додаткових криптографічних примітивів. Дослідження показало, що такий підхід зменшує накладні витрати на 30–40% у порівнянні з використанням HMAC у мережах з обмеженими ресурсами.

Аналіз проведений у роботі [2] підкреслює, що в умовах високої інтерференції та шуму в мережах 6G традиційні криптографічні методи можуть бути неефективними через необхідність повторної передачі при помилках. Використання надлишкових блокових кодів дозволяє одночасно виправляти помилки і перевіряти автентичність, що підвищує загальну ефективність системи.

У дослідженні [3] запропоновано використання модифікованих LDPC-кодів зі змішаним ключовим управлінням для автентифікації в промислових IoT-системах. Автори довели, що такий підхід стійкий до базових форм атак повторення та часткової модифікації повідомлень.

Окрім того, у звіті [4] визнано, що інтеграція автентифікації на рівні фізичного та каналного шарів за допомогою структурованої надлишковості є перспективним напрямом для забезпечення "безпеки за проектом" (security-by-design) у майбутніх телекомунікаційних системах.

Обґрунтування необхідності: Використання методів автентифікації на основі надлишкових блокових кодів стає необхідним у сценаріях, де поєднуються вимоги до надійності передачі, низької затримки, обмежених обчислювальних ресурсів і високого рівня шуму. Ці методи дозволяють інтегрувати функції захисту та відновлення даних, зменшуючи навантаження на мережу і підвищуючи загальну ефективність системи. Хоча

вони не замінюють повноцінну криптографію, вони є ефективним доповненням у гібридних архітектурах, особливо в критичних інфраструктурах та масштабованих мережах майбутнього.

4. Виклад основного матеріалу

Основна перевага несиметричних крипто-кодових конструкцій полягає у високій швидкості перетворення інформації. Введемо основні терміни та визначення. Зафіксуємо кінцеве поле $GF(q)$. Розглянемо векторний простір $GF^n(q)$ як множина n -послідовностей елементів з $GF(q)$ з покомпонентним складенням та множенням на скаляр. Лінійний (n, k, d) код є підпространством у $GF^n(q)$, тобто множина n -послідовностей (кодових слів) над $GF(q)$. При цьому основні параметри лінійного блокового коду формуються за рахунок: k – розмірність лінійного підпростору, d – мінімальна кодова відстань (мінімальна вага ненульового кодового слова). Лінійний код завадостійкого кодування формується, як лінійний підпростір у $GF^n(q)$, та однозначно задається набором базисних векторів, які формуються породжувальною матрицею G коду C , тобто матрицею рангу $rank(G) = k$, розмірності $k \times n$. Будь-яке кодове слово є лінійна комбінація рядків з матриці G . Для кодування використовується будь-яка взаємно однозначна відповідність інформаційних k -послідовностей та кодових слів. Найбільш просте відображення інформаційного k -розрядного інформаційного слова i в n -розрядне кодове слово c задається виразом $c = i \times G$. Лінійне підпростір, що отожднює код, має ортогональне доповнення, базис якого задається перевіркою матрицею H лінійного блокового коду C , тобто, перевірна матриця формується з рангом $rank(H) = r$, де $r = n - k$. Розмірність перевіркою матриці $r \times n$, причому $G \times H^T = 0$. Матриця H розглядається, як набір базисних векторів деякого лінійного підпростору, отримаємо лінійний код $C^\perp$, званий дуальним до C . Довільна n -послідовність z є кодовим словом коду C тоді, $H \cdot z^T = 0$. Вектор S – синдром, може бути обчислений множенням кодового слова з помилками $c^* = c + e$ на транспоновану перевірку матрицю H коду:

$$S = c^* \cdot H^T = c \cdot H^T + e \cdot H^T = e \cdot H^T, \quad (1)$$

де e – вектор помилок, c^* – сформоване кодове слово C з помилками, яке також є лінійним підпростором у $GF^n(q)$. Таким чином, значення синдрому S залежить лише від вектора помилок e , та залежить від кодового слова C .

Основною метою кодування інформації є контроль (виявлення та виправлення) помилок, що відбулися при передачі повідомлення каналом з шумами. Для контролю помилок пристрій кодування (кодер) вносить надмірність (перевірочну частину довжини r , $r = n - k$) в дані (кодове слово), які передаються відкритими каналами зв'язку. На приймальній стороні, аналізуючи властивості перевіркою частини та її відповідність переданим даним, декодер зменшує вплив помилок, що виникли під час передачі. Завдання декодування може бути ефективно вирішене (з поліноміальною складністю алгоритму) для вузького класу кодів

Розробка методу побудови криптосистеми Мак-Еліса на основі t -кодів. Крипто-кодова система Мак-Еліса (несиметрична криптосистема постквантового періоду) вперше запропонована в роботах [5,6]. Основні параметри крипто-кодової системи: G – породжувальна матриця лінійного (n, k, d) коду над $GF(q)$ з поліноміальною складністю декодування. Матриці маскування, які забезпечують безпеку ключової послідовності

(особистого ключа): X – невироджена $k \times n$ матриця над $GF(q)$, D – діагональна матриця з ненульовими на діагоналі елементами $n \times n$, P – перестановна матриця розміру $n \times n$. Перестановна матриця забезпечує перестановку координат вектора символів матриці у вигляді множення, а саме, елемент p_{ij} матриці P дорівнює 1, коли координата з номером i переходить за допомогою перестановки в координату з номером j , інакше $p_{ij} = 0$.

Добуток матриць $\Lambda = P \times D$ задає перестановну матрицю Λ з ненульовими елементами поля $GF(q)$. Перестановна матриця Λ (уніпотентна матриця) при перестановці координат вектора зберігає відстань Хеммінга, тобто. $d(a, b) = d(a \times \Lambda, b \times \Lambda)$, де $d(a, b)$ – відстань Хеммінга між векторами a та b . Відкритим (публічним) ключем у крипто-кодової конструкції Мак-Еліса є матриця $G_X = X \times G \times P \times D$, Особистим (закритим) ключем є матриці маскування: X, P, D . Закрита інформація (кодове слово, криптограма) є вектором довжини n і обчислюється за правилом:

$$c_X^* = i \cdot G_X + e, \quad (2)$$

де вектор $c_X = i \times G_X$ належить (n, k, d) -коду з породжувальною матрицею G_X , i – k -розрядна інформаційна послідовність, вектор e – вектор помилок (сеансовий ключ, який забезпечує додаткову стійкість криптограми) ваги більш t . Протівнику необхідно декодувати криптограму (кодограму) з відомою породжувальною матрицею криптосистеми G_X . Але не знаючи матриці маскування, й породжувальну матрицю лінійного коду X, P, D , та G протівник не може відновити H , та скористатися алгоритмом декодування поліноміальної складності. Декодування випадкового коду великої довжини NP-повна задача при кореляційному декодуванні. Для уповноваженого користувача (який знає особистий (закритий) ключ, тобто матриці маскування й перевірочну матрицю (H) розкодування кодограми – завдання поліноміальної складності. Користувач, використовуючи алгоритм поліноміальної складності, декодує вектор c^* , тобто. знаходить i' . Потім обчислює k -розрядну інформаційну послідовність $i = i' \times X^{-1}$. Структурна схема протоколу обміну даними за допомогою крипто-кодової системи Мак-Еліса наведена на рис. 1.

Передача кодограми формується наступними операціями. Користувач Б випадково, рівноймовірно, незалежно від інших абонентів формує матриці маскування X, P, D та породжувальну матрицю лінійного блокового коду G (зберігає їх в секреті (закритий (особистий) ключ). Обчислює матрицю $G_X = X \times G \times P \times D$ та публікує її як відкритий (публічний) ключ. Користувач А для надсилання інформаційної послідовності формує кодограму за виразом (2). Її може сформувати будь-який користувач, який знає публічний (загальнодоступний) ключ.

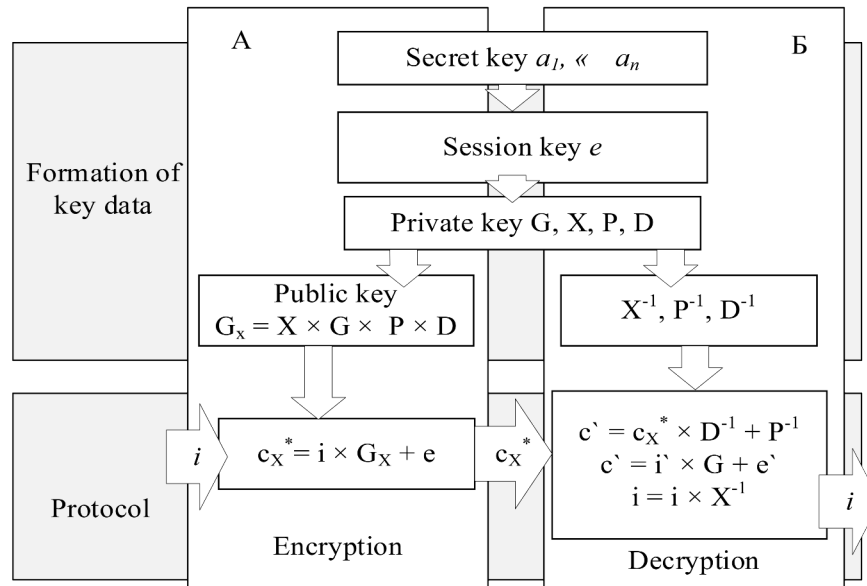


Рис. 1 – Структура протоколу обміну даними за допомогою крипто-кодової конструкції Мак-Еліса

Її може сформувати будь-який користувач, який знає публічний (публічний) ключ. Таким чином, забезпечується побудова криптограми. На приймальній стороні користувач Б розкодує криптограму (отримує інформаційну послідовність за рахунок використання алгоритму Берлекемпа-Месі).

5. Висновок

Розглянуто актуальне наукове завдання побудови стійких до квантових атак криптосистем на основі коригувальних кодів, зокрема криптосистеми Мак-Еліса, адаптованої для використання з m -кодами. Проаналізовано можливість інтеграції методів автентифікації на основі надлишкових блокових кодів у сучасні телекомунікаційні системи, зокрема в умовах мереж 5G/6G, IoT та вбудованих пристроїв, де важливими є низька затримка, енергоефективність та висока надійність передачі даних. Доведено, що використання надлишкових блокових кодів дозволяє одночасно реалізовувати функції виявлення та виправлення помилок і забезпечувати автентифікацію повідомлень, що значно знижує накладні витрати на реалізацію окремих механізмів захисту. Особливу перевагу такі методи мають у шумних каналах зв'язку, де традиційні криптографічні схеми можуть призводити до зростання кількості повторних передач через помилки, що погіршує загальну продуктивність системи.

Розроблений метод побудови криптосистеми Мак-Еліса на основі m -кодів дозволяє забезпечити високий рівень криптостійкості за рахунок використання секретних матриць маскування та відомого поліноміального алгоритму декодування для легального отримувача, тоді як для зломисника задача декодування криптограми залишається NP-складною. Це робить запропоновану конструкцію перспективною для застосування в постквантовій криптографії та відповідає концепції «безпеки за проектом» (security-by-design) і забезпечують комплексний захист інформації.

Подальші дослідження можуть бути спрямовані на оптимізацію параметрів m-кодів, підвищення швидкодії алгоритмів та розробку апаратних реалізацій запропонованої криптосистеми.

Література

- [1] Chen, L., Ling, C., & Yan, Z. Error-Correcting Codes for Simultaneous Authentication and Error Detection in IoT Networks. *IEEE Transactions on Information Forensics and Security*. 2021, 16, P.3456–3470
- [2] Al-Najjar, H., Al-Turjman, F., & Hassouna, M. Lightweight Authentication Using Structured Redundancy in 6G-enabled Edge Networks. *Computer Networks*. 2022. 215, 109201
- [3] Kumar, R., & Singh, P. Secure and Reliable Data Transmission Using LDPC-Based Authentication Codes. *Journal of Network and Computer Applications*. 2023. 218, 103789
- [4] Звіт ETSI Security Aspects of Channel Coding in 6G Systems (TR 103 756 V1.1.1). 2024.
- [5] Serhii Yevseiev, Khazail Rzayev, Oleksandr Laptiev, Ruslan Hasanov, Oleksandr Milov, Bahar Asgarova, Jale Camalova, Serhii Pohasii. Development of a hardware cryptosystem based on a random number generator with two types of entropy sources. *Eastern-European journal of enterprise technologies*. Vol.5 №9 (119). 2022 P. 6–16. ISSN (print) 1729 - 3774. ISSN (on-line) 1729-4061. DOI: <https://doi.org/10.15587/1729-4061.2022.265774>
- [6] Serhii Yevseiev, Khazail Rzayev, Oleksandr Laptiev, Ruslan Hasanov, Oleksandr Milov, Bahar Asgarova, Jale Camalova, Serhii Pohasii. Development of a hardware cryptosystem based on a random number generator with two types of entropy sources. *Eastern-European journal of enterprise technologies*. Vol.5 №9 (119). 2022 P. 6–16. ISSN (print) 1729 - 3774. ISSN (on-line) 1729-4061. DOI: <https://doi.org/10.15587/1729-4061.2022.265774>

Метод визначення ймовірності негласного отримання інформації потенційним порушником

Тетяна Лаптева¹, Сергій Лаптев², Євген Меленті³

¹ Київський національний торговельно-економічний університет, вул. Кіото, 19, м. Київ, 02156, Україна

² Київський національний університет імені Тараса Шевченка, вул. Володимирська, 60, м. Київ, 01033, Україна

³ Національна академія Служби безпеки України, вул. М. Максимовича, 22, м. Київ, 03066, Україна

Анотація.

Розглядається математичне моделювання ймовірності негласного отримання інформації потенційним порушником, яке зводиться до моделювання впливу порушника на захищувану систему з цілю отримання можливого каналу витоку інформації і являє собою формалізований опис сценаріїв у вигляді логіко-алгоритмічної послідовності дій порушників, кількісних значень, що характеризують результати дій і функціональних (аналітичних, чисельних або алгоритмічних) залежностей, що описують протікають процеси взаємодії порушників з елементами захищеного об'єкта. Однак подібні підходи до моделювання не дозволяють кількісно оцінити актуальність загрози атак та врахувати цю найважливішу характеристику. Тому запропоновано варіант кількісної оцінки можливості несанкціонованого отримання інформації з окремих інформаційної системи, розроблена універсальний метод. Універсальність методу дозволяє порівнювати між собою складності реалізації різномірних атак, заснованих на різних принципах реалізації, в загальному випадку використовують абсолютно різні за своєю природою канали витоку інформації. На основі запропонованого методу розроблен підхід по математичному моделюванню ймовірності отримання порушником несанкціонованого доступу до інформації вже на першому підготовчому етапі комплексної перевірки, що в подальшому дозволить реально оцінити фінансові, оперативні і технічні засоби захисту інформації.

Ключові слова

захист інформації, моделювання, метод, засоби негласного отримання інформації

1. Вступ

Захист інформації-досить дороге заняття, що потребує не тільки разових, а й постійних витрат. Тому захищати необхідно тільки ту інформацію, витік якої може привести до економічного, морального або іншого збитку підприємству, організації, її керівництву. Це дозволить диференціювати заходи забезпечення безпеки інформації та скоротити тім самим витрати на їх проведення.

Зарубіжний і вітчизняний досвід проведення робіт з виявлення засобів негласного отримання інформації (далі по тексту НОІ) показує, що дії по підготовці і проведенню комплексних спеціальних перевірок приміщень доцільно умовно розділити на три етапи: підготовчий, етап безпосередньо проведення спеціальної перевірки приміщень та заключний етап. Найважливіше місце серед робіт підготовчого етапу комплексної спеціальної перевірки приміщень займає виявлення або уточнення ймовірного противника, який здійснює перехоплення інформації за допомогою засобів НОІ. Ризик НОІ інформації з інформаційної системи неможливо оцінити без побудови моделі потенційного порушника безпеки вже на першому етапі робіт

⁴⁰ Corresponding author.

✉ tetiana1986@ukr.net (Т. Лаптева); salaptiev@gmail.com (С. Лаптев); melenty@ukr.net (Є. Меленті)

☎ 0000-0002-5223-9078 (Т. Лаптева); 0000-0002-7291-1829 (С. Лаптев); 0000-0003-2955-2469 (Є. Меленті)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Мета дослідження

Розробка варіанту кількісної оцінки ймовірній можливості несанкціонованого отримання інформації з інформаційної системи замовника.

3. Огляд результатів та джерел

Більшість відомих підходів до моделювання, відрізняються тим, які параметри при моделюванні ними використовуються в якості вхідної інформації і які характеристики модельованої системи розраховуються і надходять на вихід моделі. Будуються моделі з використанням теорії ймовірностей, випадкових процесів, мереж Петрі, теорії автоматів, теорії графів, нечітких множин і ін., передбачає використання в якості найпростішого елемента безпеки загрозу атаки на інформаційну систему [1,4-5]. Однак подібні підходи до моделювання не дозволяють кількісно оцінити актуальність загрози атак та врахувати цю найважливішу характеристику кількісної оцінки можливості НОІ зі інформаційної системи замовника. Разом з тим у всіх зазначених джерелах математичне моделювання порушника зводиться до моделювання впливу порушника на захищувану систему з цілю отримання можливого каналу витоку інформації і являє собою формалізований опис сценаріїв у вигляді логіко-алгоритмічної послідовності дій порушників, кількісних значень, що характеризують результати дій і функціональних (аналітичних, чисельних або алгоритмічних) залежностей, що описують протікають процеси взаємодії порушників з елементами захищеного об'єкта. Однак подібні підходи до моделювання не дозволяють кількісно оцінити актуальність загрози атак та врахувати цю найважливішу характеристику. Тому потрібно запропоновано варіант кількісної оцінки можливості несанкціонованого знімання інформації інформаційної системи замовника.

4. Виклад основного матеріалу

Математична модель потенційного порушника повинна враховувати зацікавленість злоумисника в реалізації атаки на конкретну інформаційну систему з метою отримання інформації або порушення її цілісності.

Відзначимо, що побудова моделі порушника є ключовим питанням не тільки при визначенні можливих каналів НОІ але і характеристик безпеки інформаційних систем в цілому.

В даний час моделі потенційного порушника безпеки формуються як набір припущень про можливого порушника безпеки, його кваліфікації, технічних і матеріальних можливостях і т.п. При цьому будується неформальна модель порушника, що відображає причини і мотиви дій, апріорні знання, переслідувані цілі, їх пріоритетність для порушника, основні шляхи досягнення поставлених цілей: способи реалізації вихідних від нього загроз, місце і характер дії, можлива тактика і т.п. В кінцевому рахунку подібна модель використовується з метою виявлення сукупності актуальних атак для конкретної інформаційної системи, для якої планується робота по виявленню і блокуванню засобів НОІ, саме актуальних, оскільки потенційно можливі канали витоку визначаються можливістю їх технічної реалізації (архітектура, використовувані програмні і апаратні засоби і т.п.). Модель порушника повинна враховувати досить багато факторів, не всі з яких піддаються формалізованому опису. Це, перш за все, рівень зацікавленості в отриманні несанкціонованого доступу до конкретної інформації, це рівень кваліфікації порушника, що дозволяє йому здійснити ту чи іншу атаку, його інформованість про виявлення й усунення різному роду вразливостей, наявність відповідних інструментальних засобів для

здійснення атаки, інформованість про реалізовану в конкретній інформаційній системі технологію (можливість отримання подібної інформації), в тому числі технологіях захисту інформації, програмному забезпеченні, регламенти та інше. Складність обліку всіх цих (отже важко формалізованих) чинників обумовлюється не тільки їх кількістю і різноманітністю, але і складністю формалізації будь-яких залежностей між ними (наприклад, порушник може найняти висококваліфікованого фахівця для здійснення атаки, може придбати відповідні автоматизовані засоби здійснення атаки - реалізувати складну атаку, не володіючи при цьому належну кваліфікацію, і т.п.). Разом з тим, нам необхідна якась інтегральна оцінка, причому кількісна, що дозволяє врахувати всі ці фактори, інакше неможливо приступити до робіт по перевірці приміщень на наявність засобів негласного отримання інформації для конкретної інформаційної системи. [1, 5]

Виходячи з того, що порушник інформаційної системи може бути охарактеризований складністю реалізованих їм атак на інформаційну систему, визначим, як кількісно оцінити складність атаки, введемо кількісну міру складності атаки, оскільки в загальному випадку слід говорити про те, чи готовий (зацікавлений і може) порушник реалізувати атаку певної складності. При цьому, як відзначали, і можливі канали витоку інформації, і власне реально використанні канали по своїй суті різні, кількісна ж міра повинна бути єдиною.

Звернемося до основ теорії інформації, розуміючи, що для успішного здійснення НОІ (атаки) щодо окремо взятого каналу витоку інформації порушник повинен володіти відповідною інформацією щодо можливості НОІ – інформацію про те, що така вразливість їм виявлена і не вирішено власником інформації, тобто певної кількості інформації щодо загрози НОІ інформації.

Оскільки нас цікавить виключно ймовірність того, що канал отримання інформації присутній в інформаційній системі – загроза атаки реальна, при цьому можливі два результату події: канал для отримання інформації присутнє або ні, тобто кількість інформації щодо вразливості в даному випадку слід розглядати як ймовірнісну міру.

У нашому випадку розглядається будь-яка можливість отримання інформації порушником при здійсненні атаки, ймовірність присутності якої в системі визначається як $1 - P_{0a}$ (де P_{0a} ймовірності перекриття каналу отримання інформації)

Складність реалізації атаки – S_a – може інтерпретуватися як ймовірна міра кількості інформації $I(P_{0a})$, якими повинен володіти злоумисник для реалізації НОІ по виявленому каналу, та визначається :

$$S_a = I(P_{0a}) = -\log_2(1 - P_{0a}) \quad (1)$$

Коректність використання даного методу для оцінки реалізації НОІ по виявленому каналу обґрунтовується використанням логарифмічної функції (у нашому випадку за основою 2, оскільки у події можливі два результату), що дозволяє відповідним чином врахувати нелінійність функції зміни складності реалізації порушником знімання інформації від зміни значення ймовірності P_{0a} : $S_a = I(P_{0a})$. Якщо розглянути атаку як послідовність використання порушником виявлених і не усунутих в системі каналів знімання інформації, що мають характеристики P_{0ar} і S_{ar} , $r=1, \dots, R$, можна ввести кількісну характеристику складності атаки (P_{0a}) (позначимо її S_a), де $S_a = I(P_{0a})$, яка визначається кількістю інформації, якою повинен володіти порушник для здійснення успішної атаки:

$$S_a = I(P_{0a}) = -\log_2(1 - P_{0a}) = -\log_2 \prod_{r=1}^R (1 - P_{0ar}) \quad (2)$$

де $P_{0a} = 1 - \prod_{r=1}^R (1 - P_{0ar})$ – ймовірність того, що в будь-який момент часу можливе отримання інформації.

Можемо записати:

$$S_a = I(P_{0a}) = \sum_{r=1}^R I(P_{0ar}) = \sum_{r=1}^R S_{ar} \quad (3)$$

Характеристика ΔS_a може розглядатися в якості так званої теорії інформації прагматичної міри кількості інформації, визначається в даному випадку за формулою:

$$\Delta S_a = \log_2(1 - P_{0aicx}) - \log_2(1 - P_{0aizax}) = \log_2 \frac{(1 - P_{0aicx})}{(1 - P_{0aizax})} \quad (4)$$

де P_{0aicx} і P_{0aizax} – ймовірності готовності до безпечної експлуатації існуючої і захищеної (при використанні системи захисту) інформаційних систем.

Для подальшої побудови моделі порушника введемо поняття коефіцієнта готовності порушника здійснити НОІ - $K_{га}$.

Коефіцієнт готовності порушника здійснити атаку $K_{га}$ потрібно визначати стосовно до конкретної інформаційної системи, до конкретних каналів витоку інформації з інформаційної системи і системи її захисту та може бути представлена наступним чином:

$$S_{aH} = \max \{S_{aHm}, m = 1, \dots, M\} \quad (5)$$

де S_{aH} – максимальна складність реалізованих (з урахуванням і відбитих) в подібній інформаційній системі атак, які характеризуються P_{oan} , визначена на множині виявлених каналів витоку інформації з подібної інформаційної системи (аналог) в процесі її експлуатації S_{aHm} , $m = 1, \dots, M$.

Маючи значення характеристики S_a – характеристика складності реалізації варіанту знімання інформації з інформаційної системи, і значення характеристики S_{aH} – характеристика максимальної складності реалізованих в подібній інформаційній системі атак, можна визначити шукану характеристику коефіцієнта готовності (або ймовірності) порушника здійснити атаку складності S_a на конкретну інформаційну систему $K_{га}$:

$$K_{га} = \begin{cases} \frac{S_{av}}{S_a}, & \text{якщо } S_{aH} < S_a \\ 1, & \text{якщо } S_{aH} \geq S_a \end{cases} \quad (6)$$

Виходячи ж з того, що

$$K_{га} = \frac{S_{av}}{S_a} = \frac{\log_2(1 - P_{oan})}{\log_2(1 - P_{oa})} = \log_{1-P_{oa}}(1 - P_{oan}) \quad (7)$$

Коефіцієнт $K_{га}$ може інтерпретуватися як значення ступеня, у яку треба звести значення ймовірності здійснення атаки на інформаційну систему $(1 - P_{oa})$, для отримання значення ймовірності атаки, яку може успішно реалізувати порушник $(1 - P_{oan})$.

Формула для розрахунку ймовірності реалізації в будь-який момент часу успішної атаки на інформаційну систему P_a має наступний вигляд:

$$P_a = K_{Ia} \prod_{r=1}^R (1 - P_{oar}) \quad (8)$$

Ймовірність того, що успішна атака не буде здійснена на інформаційну систему, що визначається як P_{oa} :

$$P_{oa} = 1 - K_{Ia} \prod_{r=1}^R (1 - P_{oar}) \quad (9)$$

Вираз (9) є остаточним виразом запропонованого методу. Таким чином запропонований метод дозволяє вже на першому етапі провести ймовірну оцінку негласного отримання інформації противником.

5. Висновок

Розглянуто актуальну задачу кількісної оцінки ймовірності реалізації загроз негласного отримання інформації (НОІ) в інформаційних системах. Аналіз існуючих підходів до моделювання поведінки порушника виявив їхні обмеження: вони ґрунтуються на формалізованих сценаріях атак і логіко-алгоритмічних моделях, але не враховують реальну зацікавленість і можливості зловмисника, що знижує їхню практичну ефективність.

Головним результатом роботи є розробка універсального методу кількісної оцінки ймовірності успішного НОІ, що поєднує теоретико-інформаційні підходи з аналізом реальних загроз. Запропоновано перейти від якісного опису порушника до математичної моделі, яка враховує його кваліфікацію, ресурси, інформованість та зацікавленість, забезпечуючи більш об'єктивну оцінку ризиків.

Ключовим елементом є кількісна міра складності атаки, визначена як обсяг інформації, необхідний порушникові для встановлення наявності каналу витоку. Ця міра, обґрунтована на основі теорії інформації Шеннона, виражається через логарифмічну функцію ймовірності перекриття каналу і дозволяє порівнювати різноманітні канали (віброакустичні, електромагнітні, мережеві) за єдиною шкалою.

Для оцінки реальної загрози введено коефіцієнт готовності порушника — ймовірність реалізації атаки заданої складності, що визначається на основі статистики атак на аналогічні системи. Це дозволяє перейти від гіпотетичних сценаріїв до прогнозованої, статистично обґрунтованої оцінки ризику.

На основі цих компонентів розроблено інтегральну формулу для розрахунку ймовірності успішної атаки, що поєднує характеристики системи (стійкість каналів) та порушника (можливості, готовність). Це забезпечує обґрунтовану оцінку ризику на підготовчому етапі перевірки приміщень, що критично важливо для ефективного розподілу ресурсів.

Практичне значення методу полягає в можливості диференціації заходів захисту за реальною загрозою, зменшенні витрат та науково обґрунтованому виборі засобів захисту. Проте ефективність методу залежить від якості вхідних даних — зокрема, наявності достовірної статистики. При її відсутності доводиться вдаватися до експертних оцінок, що знижує об'єктивність. Також потребує подальшого уточнення моделювання взаємодії кількох каналів витоку.

Загалом, запропонований підхід є важливим кроком у розвитку кількісного аналізу ризиків інформаційної безпеки, забезпечуючи перехід від суб'єктивних оцінок до

формалізованої моделі, що враховує технічні та людські чинники, і є необхідною основою для побудови ефективної та адаптивної системи захисту.

Література

- [1] Oleg Barabash, Oleksandr Laptiev, Oksana Kovtun, Olga Leshchenko, Kseniia Dukhnovska, Anatoliy Biehun. The Method dynavic TF-IDF. *International Journal of Emerging Trends in Engineering Research (IJETER)*, Volume 8. No. 9, September 2020. P 5713-5718.
- [2] Sobchuk, V., Olimpiyeva, Y., Musienko, A., Sobchuk, A. Ensuring the properties of functional stability of manufacturing processes based on the application of neural networks CEUR Workshop Proceedings, 2021, 2845, стр. 106–116.
- [3] Лаптева Т., Лаптев О. Метод синтезу захищених інформаційних мереж на основі введення коригувальних ліній зв'язку. 2025. Територія безпеки Том 1 № 1. С.29-35. <https://doi.org/10.20998/3083-6298.2025.01>.
- [4] Zamrii I., Sobchuk V., Laptiev O., Savchenko V., Shkapa V., Kovalenko V. and Kotok V. Fractal Functions and Their Application to Source Data Coding. *ARP Journal of Engineering and Applied Sciences*. Vol. 17, No. 4, 2022. P. 424 – 435
- [5] S. Yevseiev and other “Development of a methodology for building a security information system in the corporate research and education system in the context of university autonomy”. *Eastern-European Journal of Enterprise Technologies*. 2019. 3/9(99). p. 49–63

Кіберполігон кафедри ТЕІБ УжНУ: освітнє середовище для моделювання та запобігання кіберзагроз

Богдан Маліцький¹, Олександр Черепов¹, Юрій Матювка¹, Василь Різак¹

¹ Кафедра твердотільної електроніки та інформаційної безпеки, Ужгородський національний університет, м. Ужгород, Україна

Ключові слова

кіберполігон, red team, blue team, incident response, SIEM, IDS/IPS, освіта з кібербезпеки

Кіберполігони є спеціалізованими навчально-науковими платформами, які дозволяють відпрацьовувати сценарії, з якими стикаються фахівці з кібербезпеки у контрольованому та ізольованому середовищі. Використання кіберполігонів набуває популярності завдяки інноваційному гейміфікованому підходу до впровадження практичного навчання в освітній процес, і широко застосовується для підготовки студентів, а також професійного розвитку фахівців державних і корпоративних структур.

На кафедрі твердотільної електроніки та інформаційної безпеки Ужгородського національного університету створено власний полігон кібербезпеки, інфраструктура якого об'єднує кілька навчальних аудиторій. Цей кіберполігон складається з двох підмереж – одна для команди атаки (red team), інша для команди захисту (blue team). Робоче середовище охоплює 10 ПК-станцій, 10 ноутбуків, 2 комутатори, 2 проміжні маршрутизатори та 1 загальний маршрутизатор.

У межах полігону реалізовано низку практичних сценаріїв, спрямованих на відпрацювання атакувальних та захисних навичок в учасників. Серед них – відпрацювання експлуатації вразливостей у віртуальному середовищі Metasploitable 2, завдання у форматі Capture the Flag (далі – CTF) з використанням локально розгорнутого Gitlab-сервера та мережі Active Directory, розроблене за підтримки організації UnderDefense, а також практичні завдання із побудови та налаштування захищених мереж у Cisco Packet Tracer та аналіз мережевого трафіку у Wireshark.

Кіберполігон повністю інтегрований у навчальний процес університету, та використовується під час проведення навчальних дисциплін, до прикладу етичного хакінгу, комп'ютерних мереж та інших, у рамках навчальних практик студентів, а також для підготовки до змагань формату CTF та професійних тренінгів у сфері інформаційної безпеки. Одним із ключових результатів є організація щорічних навчальних практик для студентів бакалаврату та магістратури, під час якої вони мають змогу працювати з реальними сценаріями кіберзагроз та закріплювати теоретичні навички, отримані під час навчального року.

Окремої уваги заслуговує співпраця з державними органами та установами у сфері кіберзахисту, зокрема підписані меморандуми про співпрацю, які дозволили студентам пройти ознайомчу практику в державному кіберцентрі UA-30. Також, студенти кафедри взяли участь у стажуванні в команді реагування на комп'ютерні надзвичайні події України CERT-UA, продемонструвавши, за словами представників команди, високі результати. Завдяки цьому їх запросили на стажування найвищого рівня складності, а також висловили зацікавленість у залученні інших студентів до участі в програмі.

Важливою складовою в діяльності полігону є підготовка студентів до регіональних та національних змагань у форматі CTF. Учасники, які проходили підготовку, стали переможцями та призерами CTF-змагань DES у 2024 році, а також успішно представили Закарпатську область на Чемпіонаті України з технологічного спорту в дисципліні CTF – Кібербезпека 2025.

Одним із ключових напрямів удосконалення кіберполігону кафедри ТЕІБ УжНУ є впровадження автоматизованої системи розгортання сценаріїв. Її мета – скоротити час підготовки середовища, мінімізувати ручні налаштування та знизити ризики технічних збоїв. Використання віртуалізації на базі Hyper-V дозволяє швидко відновлювати початковий стан системи за допомогою знімків, стандартизувати сценарії та підвищити відтворюваність результатів. Це забезпечує викладачам можливість зосередитися на створенні більш складних завдань, а студентам – на їх практичному виконанні.

Система проектується за принципом клієнт-серверної архітектури: головний сервер розподіляє завдання, сервер-сховище зберігає еталонні сценарії, а клієнти-агенти реалізують їх розгортання та базове тестування на фізичних машинах. Попереднє автоматизоване тестування перевіряє роботу критичних сервісів і правильність взаємодії інфраструктури ще до старту занять. У перспективі планується інтеграція модулів для збору статистики про дії учасників, що дозволить удосконалювати сценарії, підвищувати якість навчального процесу та ефективніше організовувати командну взаємодію.

Ще одним стратегічним кроком є інтеграція відкритих SIEM/IDS/IPS рішень, зокрема Wazuh, ELK Stack та Suricata, у навчальне середовище кіберполігону. Це дасть змогу створити локальну модель Security Operations Center (SOC), у якій студенти зможуть відпрацьовувати навички моніторингу, кореляції подій, виявлення загроз і формування звітності. Використання рішень з відкритим кодом робить цей підхід доступним для масштабування та адаптації під потреби регіону, водночас забезпечуючи практичну підготовку та підвищення кваліфікації фахівців до роботи з сучасними інструментами у сфері кібербезпеки.

Література

- [6] Katsantonis, I., Smyrlis, M., Votis, K., Tzovaras, D. "Cyber Range Design Framework for Cyber Security Education and Training." *International Journal of Information Security*, Springer (2023).
- [7] Abrahamsson, P., Pakanen, M., Still, J., Rantala, J., Mäntylä, M., Hyrynsalmi, S., "Cyber Range as a Competency Based Education Instrument in Cyber Security", *International Conference on Software Business (ICSOB 2022)*, Lecture Notes in Business Information Processing, vol. 454, Springer, Cham (2022).

Рекурсивний генератор з латинських квадратів

Богдан Микитченко¹, Галина Крайнічук (Шелепало)¹

¹ Вінницький національний технічний університет, Хмельницьке шосе 95, 21021 Вінниця, Україна

Анотація

У статті представлено метод генерації псевдовипадкових послідовностей на основі рекурсивного використання латинських квадратів. Досліджено властивості послідовностей, сформованих як з одного квадрата, так і з комбінації кількох, що дозволяє збільшувати їх довжину та забезпечувати рівномірність розподілу. Проведено експерименти для квадратів порядку 4–7, у результаті яких виявлено наявність «нульових початкових трійок», що породжують мінімальні послідовності, та відхилення від очікуваної залежності довжини для квадратів 7-го порядку. Для оцінки якості застосовано набір тестів NIST STS, які підтвердили високу статистичну стійкість отриманих послідовностей. Отримані результати демонструють придатність рекурсивних методів на основі латинських квадратів для використання у криптографії, зокрема у потоковому шифруванні та системах із обмеженими ресурсами.

Ключові слова

латинський квадрат, квазігрупа, псевдовипадкова послідовність, рекурсія, потокове шифрування, NIST, криптографія

1. Вступ

Захист інформації в умовах сучасних кіберзагроз неможливий без використання криптографічних примітивів, серед яких ключову роль відіграють генератори псевдовипадкових послідовностей (ГПВП). Саме якість та непередбачуваність цих послідовностей визначає рівень надійності поточкових шифрів і багатьох інших криптографічних систем. Традиційні методи генерації псевдовипадкових чисел, зокрема лінійні конгруентні генератори та регістри зсуву з лінійним зворотним зв'язком (LFSR), є простими та ефективними, але часто демонструють обмежений період та слабкість до статистичного аналізу [9]. Сучасні криптографічні детерміновані генератори (DRBG), засновані на блочних шифрах чи хеш-функціях [7], забезпечують високу стійкість, але мають значні обчислювальні витрати, що обмежує їх застосування у малоресурсних середовищах, зокрема для IoT-пристроїв.

У цьому контексті актуальним є пошук нових математичних структур для побудови генераторів, що поєднують простоту реалізації та високі статистичні властивості. Латинські квадрати, як класичний комбінаторний об'єкт [4, 5], привертають увагу дослідників завдяки властивостям унікальності та рівномірності, що робить їх придатними для використання у криптографії [1].

У попередніх роботах [1, 2] було запропоновано метод побудови псевдовипадкових послідовностей на основі двох латинських квадратів та окреслено концепцію потокового шифрування на основі квазігруп. Проте питання використання рекурсивних методів на множині латинських квадратів для формування послідовностей залишаються відкритими.

Метою цієї статті є розробка та аналіз нового підходу до генерації псевдовипадкових послідовностей, що ґрунтується на рекурсивному використанні одного або кількох

⁴¹ Відповідальні автори

✉ nif120nif@gmail.com (Б. Микитченко); kraynichuk@ukr.net (Г. Крайнічук)
0009-0006-1669-6948 (Б. Микитченко); 0000-0002-6822-4858 (Г. Крайнічук);



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

латинських квадратів. У роботі досліджуються властивості отриманих послідовностей, довжина та рівномірність, а також перевірка методом статистичних тестів NISTSTS.

2. Огляд попередніх досліджень

Сучасні роботи у галузі побудови ГПВП зосереджуються на вдосконаленні лінійних структур, оптимізації алгоритмів для малопотужних пристроїв та впровадженні нових математичних моделей [6, 7]. Особливу увагу приділено потоковому шифруванню та генераторам з великим періодом, проте більшість підходів потребують значних обчислювальних ресурсів.

У попередніх дослідженнях авторів [1, 2] було продемонстровано можливість використання латинських квадратів для побудови псевдовипадкових послідовностей із використання рекурсії. Показано, що рекурсивна формула, яка використовує комбінацію елементів кількох квадратів, дозволяє отримати послідовності різної довжини з властивістю рівномірного розподілу. Проте систематичний аналіз рекурсивних методів із залученням більшої кількості квадратів та дослідження закономірностей отриманих послідовностей досі залишаються відкритими питаннями.

3. Результати дослідження

Запропонований метод генерації псевдовипадкових послідовностей (ПВП) ґрунтується на рекурсивному застосуванні операцій, визначених таблицями Келі для квазігруп, тобто латинськими квадратами (ЛК).

Рекурсія на основі одного латинського квадрата.

Базовий алгоритм використовує один ЛК та рекурентну формулу

$$u_i = u_{i-1} * u_{i-2}, \quad (1)$$

де (*) позначає вибір елемента з ЛК за координатами (індексами) u_{i-1} та u_{i-2} .

Аналіз ЛК 4-го порядку (рис. 1) показує, що властивості згенерованої ПВП критично залежать від вибору самого квадрата, його властивостей та початкової пари елементів. Наприклад, квадрат L_1 дозволяє згенерувати послідовність максимальної довжини 16 (4^2), що включає всі можливі пари елементів:

0 1 2 3 2 2 1 3 0 3 3 1 0 2 0 0 [1],

де крайній елемент в квадратних дужках означає, що рекурсія завершена.

Водночас квадрати L_2 та L_3 з тією ж початковою парою (0, 1) генерують значно коротші послідовності довжиною в 7 елементів. Крім того, для квадрата L_2 , зміна початкової пари (наприклад, на (0, 2)) може призвести до генерації іншої послідовності меншої довжини, наприклад 4. Це підкреслює важливість вибору вхідних даних для досягнення максимального періоду ПВП.

L_1	0	1	2	3
0	1	2	0	3
1	2	1	3	0
2	0	3	1	2
3	3	0	2	1

L_2	0	1	2	3
0	2	1	0	3
1	1	2	3	0
2	0	3	2	1
3	3	0	1	2

L_3	0	1	2	3
0	3	2	0	1
1	2	3	1	0
2	0	1	3	2
3	1	0	2	3

Рисунок 1: Таблиці Келі квазігруп 4-го порядку

Рекурсія на основі двох латинських квадратів.

Використання двох ЛК (L_a і L_b) та трьох попередніх елементів послідовності за формулою (2) значно розширює можливості методу.

$$u_i = ((u_{i-1} * u_{i-2}) \cdot u_{i-3}), \quad (2)$$

Операція (*) виконується за квадратом L_a , а операція ($\cdot$) за квадратом L_b .

Такий підхід для ЛК 4-го порядку дозволяє досягти максимальної довжини ПВП у 64 елементи (4^3), що вчетверо більше ніж у методі з використанням одного квадрата. Це також підтверджує гіпотезу про максимальну довжину n^{N+1} , де n позначає порядок ЛК, а N — їх використана кількість.

Було виявлено ключові властивості методу з двома ЛК:

1. Рівномірний розподіл: певні пари ЛК генерують послідовності максимальної довжини, де кожен елемент (від 0 до 3) зустрічається однаково кількість разів (а саме 16). Ця властивість є вкрай важливою для криптографічних застосувань, тому що забезпечує рівномірність появи кожного елемента
2. Залежність від місця у твірній парі: порядок використання латинських квадратів у твірній парі суттєво впливає на результат, генеруючи однакові, різні з різною довжиною, або різні послідовності з максимальною довжиною. Саме останній випадок є найбільш цікавим для побудови генераторів, оскільки дозволяє отримати різні якісні ПВП максимальної довжини з одного набору ЛК.
3. Вплив початкової трійки та парадокс "Нульових початкових трійок" (НПТ): при генерації послідовностей на основі двох латинських квадратів спостерігається цікавий ефект, пов'язаний із вибором початкових елементів. Як було встановлено, для пари ЛК 4-го порядку, максимальна довжина ПВП має 64 елементи, що відповідає кількості всіх можливих комбінацій трійок (4^3). Емпірично протестовано, що майже будь-яка початкова трійка (наприклад, (0, 1, 2) або (3, 1, 3)) приводить до генерації однієї й тієї ж послідовності максимальної довжини, але циклічно зміщеної відносно початкової точки. Проте, для кожної пари ЛК, що генерує ПВП максимальної довжини, існують унікальні трійки, що мають інакшу поведінку. Тут названо їх "Нульовими початковими трійками" (НПТ). Це такі початкові трійки (наприклад, 000, 111, 222, 333), що генерують послідовність мінімальної довжини (зазвичай 4 елементи). Тут і виникає парадокс: за логікою, послідовність довжиною 64 мала б включати всі можливі унікальні трійки. Цей факт є важливим при практичній реалізації генератора, оскільки такі "вироджені" початкові стани варто виключати.

Аналіз якості ПВП та обмеження методу.

Для оцінки статистичної якості згенерованих ПВП було використано набір тестів NIST STS, адаптованих для послідовностей довжиною 128 біт. Результати показали високу якість генератора: послідовності максимальної довжини успішно пройшли сім ключових тестів (включаючи Frequency, Runs, Approximate Entropy), з р-значеннями, що значно перевищують значення 0.01.

Аналіз усього простору послідовностей (331 776 комбінацій) виявив, що лише 13 054 з них є унікальними, що вказує на значну надлишковість та наявність великої кількості дублікатів (до 24 копій для кожної унікальної ПВП максимальної довжини). Також було встановлено залежність послідовностей, що отримують ідеальний результат у тесті "Frequency within a Block", гарантовано проходять "Runs", що спрощує відбір якісних ПВП.

Встановлено, що послідовності можна поділити на два типи. Перший тип частково проходить тест Runs, причому результат залежить від вибору початкової трійки елементів. У таких випадках різні стартові умови призводять до зміни структури серій, що зумовлює відхилення від статистично очікуваного розподілу.

Другий тип — це послідовності, що незалежно від початкової трійки завжди успішно проходять тест Runs. Вони характеризуються більш стабільною внутрішньою структурою та рівномірним розподілом серій, що вказує на підвищену надійність і перспективність використання.

Водночас дослідження ЛК вищих порядків виявило обмеження початкової гіпотези про максимальну довжину послідовності (n^{N+1}). Хоча для двох ЛК порядків 4, 5 і 6 вона підтвердилася (довжини 64, 125, 216 відповідно), для двох ЛК 7-го порядку максимальна довжина склала 344, що не відповідає очікуваному значенню $7^3=343$. Це свідчить, що проста степенева формула не є універсальною і причини відхилення, ймовірно, криються у внутрішніх властивостях ЛК вищих порядків, що потребує подальшого дослідження.

Таким чином, запропонований метод є ефективним для генерації відносно коротких, але високоякісних ПВП, що робить його придатним для специфічних застосувань, де якість важливіша за екстремальну довжину періоду.

4. Висновки

У роботі було запропоновано новий підхід до побудови генераторів псевдовипадкових послідовностей, що ґрунтується на рекурсивному використанні латинських квадратів. Дослідження показали, що метод дозволяє отримувати послідовності різної довжини та структури, включно з максимально довгими варіантами, що проходять основні статистичні тести NISTSTS. Особливу увагу привертає феномен «нульових початкових трійок», що породжують вироджені послідовності, а також виявлене відхилення для квадратів 7-го порядку від передбачуваної степеневі закономірності. Це свідчить про складність внутрішніх властивостей латинських квадратів і відкриває нові напрями для досліджень. Запропонований метод поєднує простоту реалізації з високою якістю отриманих послідовностей, що робить його перспективним для застосувань у потоковому шифруванні.

Подяка

Автори вдячні за ідею описаного методу шифрування на основі латинських квадратів завідувачу кафедри захисту інформації Вінницького національного технічного університету професору, доктору технічних наук Лужецькому Володимирі Андрійовичу.

Декларація про використання ІІІ

Автори не використовували інструменти генеративного ІІІ.

Література

- [1] Микитченко Б. В. Загирняк Б. Д. Крайнічук (Шелепало) Г. В. Побудова псевдовипадкових послідовностей на основі двох латинських квадратів. Матеріали ІІІ наук.-техн. конф. ФІТКІ, ВНТУ, 20-22 берез. Вінниця. 2024.

- [2] Загирняк Б. Д. Микитченко Б. В. Крайнічук (Шелепало) Г. В. Концепція потокового шифрування на основі двох квазігруп. І Міжнародна науково-практична конференція «Проблеми комп'ютерних наук, програмного моделювання та безпеки цифрових систем», ВНУ, 13-16 чер. Луцьк. 2024.
- [3] Лужецький В.А., Крайнічук Г.В., Радченко Є.В., Пилявець І.Ю. Алгоритм шифрування на основі СІР-квазігруп // Матеріали тезів ІХ Міжн. наук.-техн. конф. «ЗАХИСТ ІНФОРМАЦІЇ І БЕЗПЕКА ІНФОРМАЦІЙНИХ СИСТЕМ», НУ «Львівська політехніка». Львів. 2023. 258 С.
- [4] Bauer C. P. Stream Ciphers. Secret History. 2nd ed. 2021. pp. 533–544. DOI: 10.1201/9781315162539-19
- [5] Rueppel R. A. Stream Ciphers. Analysis and Design of Stream Ciphers. Berlin, Heidelberg, 1986. DOI: 10.1007/978-3-642-82865-2_2
- [6] Shcherbacov V. Quasigroups in cryptology. Elements of Quasigroup Theory and Applications. 2017. DOI: 10.1201/9781315120058-14
- [7] Odegard R.S., Gligoroski D. On the Randomness and Regularity of Reduced Edon-R Compression Function. Proceedings of the 7th International Conference on Security and Management, Las Vegas, 2009.

Радіостанція ISM діапазону із захистом каналу зв'язку від несанкціонованого доступу

Андрій Садченко¹, Олег Кушніренко¹

¹ Національний університет «Одеська політехніка», пр. Шевченка 1 Одеса, Україна

Анотація

Запропоновано структурну схему бездротової системи передачі мовної інформації (радіостанції) та захистом каналу зв'язку за допомогою скремблера із самосинхронізацією. Схема містить трансівери HC-12, що серійно випускаються і обробляють сигнали із FSK (Frequency Shift Keying) модуляцією та працюють у безліцензійному діапазоні частот 433 МГц і забезпечують дальність зв'язку до 1000 м в умовах прямої видимості. Радіостанція містить Bluetooth модуль для бездротової конфігурації за допомогою програми, встановленого на мобільному телефоні або планшетному ПК з операційною системою Android.

Ключові слова

програмний застосунок, Bluetooth модуль, скремблювання, криптозахист

1. Вступ

У сучасних умовах інформаційної безпеки особливе значення набувають засоби надійного та захищеного радіозв'язку. Використання портативних радіостанцій малого (1 – 2 км) радіуса дії є актуальним для забезпечення оперативної комунікації у військовій сфері, підрозділах екстреного реагування, охоронних структурах, а також у цивільному середовищі – під час проведення рятувальних операцій, у будівництві чи при організації масових заходів. Проте традиційні моделі радіостанцій часто не враховують вимог до конфіденційності та захисту переданої інформації, що створює ризики несанкціонованого доступу та перехоплення сигналів.

Розробка захищеної радіостанції з обмеженим радіусом дії дозволяє мінімізувати ймовірність перехоплення, підвищує рівень стійкості до зовнішніх завад та забезпечує ефективний обмін даними у локальних зонах зв'язку. Таким чином, актуальність розробки полягає у створенні технічного рішення, яке поєднує компактність, енергоефективність та криптографічний захист інформації, що передається.

2. Аналіз існуючих підходів до реалізації захищених радіостанцій

Забезпечення конфіденційності та цілісності передавання даних у радіоканалах є однією з ключових задач при створенні сучасних засобів зв'язку. Радіостанції малого радіуса дії [1], що застосовуються у військовій, службовій та цивільній сферах, особливо потребують надійного захисту від несанкціонованого доступу, перехоплення та цілеспрямованих завад. У практиці розробки подібних пристроїв сформувалося кілька основних підходів, кожен з яких має свої особливості та обмеження. Найбільш поширеним напрямом є використання класичних криптоалгоритмів, зокрема симетричних і асиметричних [2] методів

^{42*} Corresponding author.

✉ sadchenko@op.edu.ua (А. Садченко); koa@op.edu.ua (О. Кушніренко)

ORCID ID 0000-0002-2517-6892 (А. Садченко); 0000-0002-7045-5620 (О. Кушніренко)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

шифрування. Вони забезпечують високий рівень безпеки, однак часто супроводжуються значними витратами обчислювальних ресурсів та енергії, що є критично важливим для портативних пристроїв. З метою зменшення затримок та зниження навантаження на основні процесори у низці рішень застосовуються спеціалізовані апаратні криптомодулі. Проте такі варіанти відрізняються підвищеною вартістю, складністю розробки та обмеженими можливостями для подальшої модернізації. Альтернативою є програмна реалізація із використанням легких алгоритмів, спеціально розроблених для пристроїв з обмеженими ресурсами. Вони характеризуються низьким енергоспоживанням і швидкодією, але поступаються класичним методам за криптостійкістю. Поряд із цим застосовуються і методи маскування сигналу, зокрема спектральне розширення, перестрибування частот або скремблювання аудіо потоку. Такі технології значно ускладнюють перехоплення, проте не забезпечують повноцінного криптографічного захисту. Найбільш перспективним вважається комбінування різних підходів, коли криптографічні методи доповнюються сигнальними технологіями. Це дозволяє одночасно підвищити захист переданої інформації та забезпечити стійкість до радіoeлектронних завад. Водночас інтеграція кількох методів підвищує складність конструкції і вартість пристрою, тому ключовим завданням залишається пошук балансу між рівнем безпеки, енергоспоживанням і компактністю портативних радіостанцій. Також слід зазначити, що суттєво знизити вартість розробки і ціну продукту може використання широко розповсюджених FSK модемів [3], таких як HC-12, що призначені для передавання телеметричної інформації, а також мікроконтролерів сімейства AVR [4].

3. Структурна схема радіостанції ISM діапазону з захистом каналу зв'язку від несанкціонованого доступу

Структурна схема запропонованої системи передачі аудіо сигналу представлена на Рисунок 1.

Аналоговий сигнал надходить із входу мікрофона на активний фільтр низької частоти (ФНЧ) з частотою зрізу 2,4 КГц і далі подається на вхід 10-розрядного аналого-цифрового перетворювача (АЦП) мікроконтролера (ATTiny13). Мікроконтролер оцифровує та скремблює отриманий цифровий потік і передає його на радіомодуль HC-12 за протоколом UART.

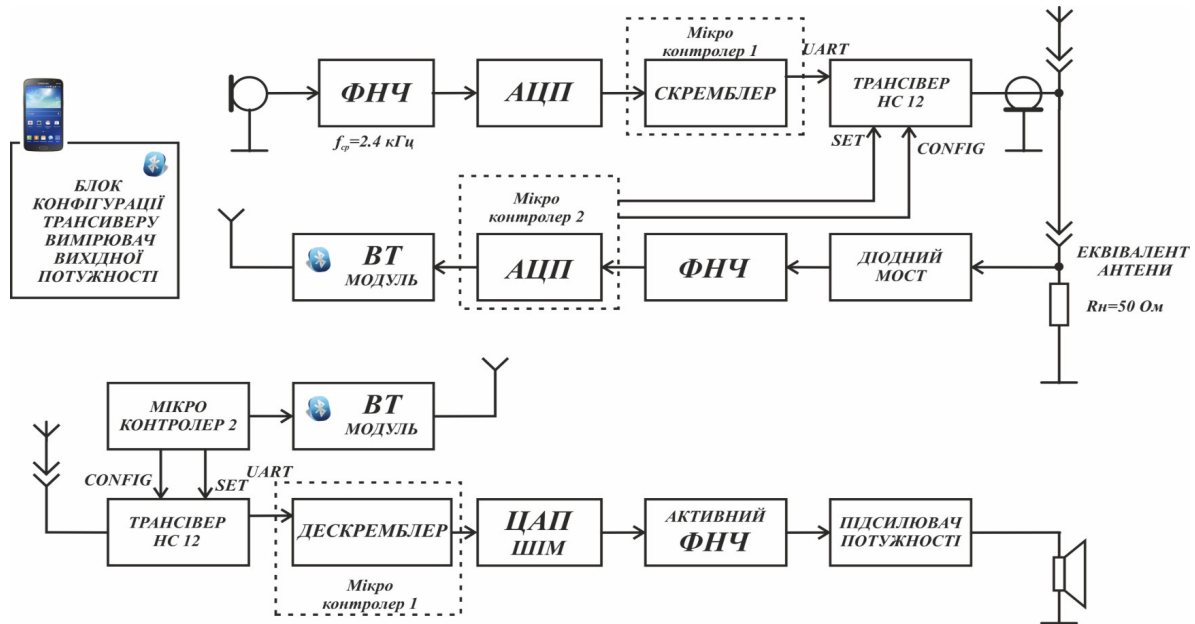


Рисунок 1: Структурна схема бездротової системи передачі аудіо сигналу.

Оскільки протокол UART передбачає побайтну передачу даних, то кожну вибірку, з урахуванням розрядності, відводиться 2 байта. Передавач також містить блок початкової конфігурації до складу якого входить вимірювач вихідної потужності. Конфігурація трансівера HC-12 полягає в установці швидкості передачі даних по каналу зв'язку, налаштування частотного каналу та вихідної потужності передавача. Налаштування здійснюється завдяки підключенню за допомогою Bluetooth модуля до зовнішнього мобільного телефону або планшетного ПК з операційною системою Android. Після виконання початкової конфігурації та вимірювання вихідної потужності всі дані заносяться в енергонезалежну пам'ять трансівера HC-12 і потреба в зовнішньому пристрої відпадає. При цьому слід зазначити, що конфігурація передавача та вимірювання потужності здійснюється на еквіваленті антени опором 50 Ом.

Приймальна частина також містить трансівер HC-12 і мікроконтролер для прийому та дескремблювання амплітуди вибірки вихідного акустичного сигналу і для конфігурації режиму роботи трансівера. Особливістю приймача є метод реалізації цифро-аналогового перетворювача (ЦАП) у вигляді широтно-імпульсного модулятора (ШІМ) з вихідним активним фільтром низької частоти 6-го порядку. Після відновлення акустичний сигнал надходить на підсилювач низької частоти.

Визначимо вимоги до частоти дискретизації АЦП, швидкості послідовного інтерфейсу та частоти широтно-імпульсного модулятора, що входить до складу ЦАП.

Швидкість UART інтерфейсу стандартизована і становить $V_{UART} = 115\,200$ біт/сек. Оскільки згідно зі стандартним протоколом одне слово складається з 10-ти біт, то на передачу 1-ої вибірки потрібно 20 біт. Отже, швидкість передачі вибірок у каналі зв'язку, що збігається з частотою дискретизації АЦП складе

$$V_k = F_d = \frac{V_{UART}}{20} = \frac{115200}{20} = 5760 \text{ біт/с}, \quad (1)$$

де F_d – частота дискретизації АЦП.

Тоді згідно з теоремою про відновлення сигналу [5] частота зрізу ФНЧ фільтру, що збігається з верхньою частотою спектра повідомлення складе $F_c = F_d / 2 = 5760 / 2 = 2900$ Гц. Отримана ширина смуги каналу зв'язку відповідає стандартам телефонії і цілком достатня для забезпечення хорошої розбірливості мови. При виборі частоти перетворення широтно-імпульсного модулятора ЦАП з метою зниження вимог до крутості спаду амплітудно-частотної характеристики ЦАП зазвичай приймають десятикратний запас, тобто

$$F_{\text{ЦАП}} = 10F_d = 57600 \text{ Гц.}$$

4. Висновок

Запропоноване технічне рішення для реалізації недорогої радіостанції з криптозахистом, яке базується на недорогих серійно-випускних трансіверах, що працюють в діапазоні 433 МГц, дозволило реалізувати захищену систему передачі акустичних сигналів у мовному діапазоні частот. Відмінною рисою системи є блок вимірювання вихідної потужності та можливість діагностики та початкового налаштування за допомогою мобільного додатка.

References

- [1] Zaid, Osama. Voice Scrambling Algorithm based on 3D Chaotic Map System (VSA3DCS) to Encrypt Audio Files. International Journal of Advanced Computer Science and Applications. 11. (2020). 10.14569/IJACSA.2020.0110571.
- [2] Madain, Alia & Abu Dalhoum, Abdel & Hiary, Hazem & De la Puente, Alfonso & Alfonseca, Manuel. (2012). Audio scrambling technique based on cellular automata. Multimedia Tools and Applications. 71. 10.1007/s11042-012-1306-7.
- [3] FSK Front-End Modules RF4438PRO: Si4438 433MHz 100mW Front-End RF Module, 2025. URL: <https://www.nicerf.com/fsk-front-end-module/433mhz-rf-module-rf4438pro.html>.
- [4] Featured Atmel Microcontrollers ATtiny MCUs: 2025. URL: <https://www.microchip.com/en-us/about/corporate-overview/acquisitions/atmel>.
- [5] The Nyquist–Shannon Theorem: Understanding Sampled Systems URL: <https://www.allaboutcircuits.com/technical-articles/nyquist-shannon-theorem-understanding-sampled-systems/>.

Алгоритм захисту аналогового відеоканалу розвідувального БПЛА

Андрій Садченко¹, Олег Кушніренко¹, Олександр Дубіна¹, Євгеній Павлюченко¹

¹ Національний університет «Одеська політехніка», пр. Шевченка 1 Одеса, Україна

Анотація

Було проаналізовано існуючі алгоритми захисту аналогових відео-потоків, та виявлено їхні недоліки у вигляді низької криптостійкості та погіршення якості вихідного зображення. Запропоновано модифікований алгоритм скремблювання аналогового відеосигналу з камери розвідувального БПЛА із вихідним кодуванням що відповідає системі PAL. Особливістю алгоритму є комбінована цифро-аналогова обробка сигналу, що забезпечує високий рівень криптостійкості із збереженням якості зображення та спектральних властивостей, таких як ширина спектру корисного сигналу. Сутність алгоритму полягає у випадковій зміні послідовності рядків у кожному кадрі зображення.

Ключові слова

відеосигнал, скремблювання, криптостійкість

1. Вступ

Безпілотні літальні апарати (БПЛА) дедалі ширше застосовуються у військовій справі, промисловості, наукових дослідженнях та цивільних сферах, де важливим є оперативне отримання якісної візуальної інформації. Відеопотік із борту дрона слугує основним каналом передачі даних для навігації, моніторингу та аналізу навколишнього середовища. Сьогодні використовуються два основних підходи до передавання зображення — аналоговий та цифровий [1].

Аналогові системи забезпечують мінімальну затримку сигналу та простоту реалізації, що робить їх актуальними для завдань, де важлива швидкість реакції оператора, наприклад у FPV-польотах чи оперативної розвідки. Водночас вони обмежені низькою роздільною здатністю та схильністю до перешкод.

Цифрові технології, натомість, дають змогу передавати відео високої чіткості, інтегрувати телеметрію та застосовувати сучасні методи шифрування й захисту даних. Проте такі системи потребують ширших каналів зв'язку, є більш дорогими та характеризуються вищою затримкою сигналу, що може бути критичним у деяких сценаріях.

Таким чином, вибір між аналоговим і цифровим відеопотоком у БПЛА залежить від конкретних завдань: від швидкісного маневрування до розвідувальних або наукових місій, де пріоритетами є якість і надійність даних. При цьому, якщо шифрування цифрового відеосигналу із метою його захисту, не призводить до погіршення якості при його відновленні, аналоговий сигнал захистити більш складно. Також, в зв'язку з тим, що в умовах проведення розвідувальних дій за допомогою БПЛА найбільш значимою вимогою

^{43*} Corresponding author.

[†] These authors contributed equally.

✉ sadchenko@op.edu.ua (А. Садченко); koa@op.edu.ua (О. Кушніренко)

id 0000-0002-2517-6892 (А. Садченко); 0000-0002-7045-5620 (О. Кушніренко)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

стає мінімальна часова затримка, набуває актуальності розробка алгоритмів захисту саме аналогових відеосигналів.

2. Існуючі алгоритми захисту аналогових відеосигналів

На відміну від цифрових систем, де застосовуються класичні методи криптографії (AES, RSA, ECC тощо), захист аналогових відеосигналів базується переважно на методах аналогового шифрування та маскування сигналу. Оскільки традиційні методи криптографії орієнтовані на цифрові дані, для аналогових потоків застосовуються інші підходи, серед яких найбільш поширеними є спектральне розсіювання, скремблювання сигналу, модуляційне маскування та використання хаотичних генераторів. Метод спектрального розсіювання [2] ґрунтується на динамічній зміні робочої частоти передавача і приймача за певним законом. Такий підхід ускладнює несанкціоноване перехоплення, підвищує стійкість системи до радіоперешкод та робить глушіння менш ефективним. Разом із тим він вимагає точної синхронізації каналів та характеризується обмеженою швидкістю перестроювання, що залишає можливість відстеження сигналу при довготривалому спостереженні. Серед класичних методів захисту найбільш відомим є скремблювання відеосигналу [3], яке передбачає зміну порядку рядків, кадрів або фазових характеристик. Перевагою такого підходу є простота та сумісність із традиційним аналоговим обладнанням, а також низька вартість реалізації. Проте рівень криптостійкості скремблювання є невисоким, оскільки сигнал у багатьох випадках може бути відновлений за допомогою зворотного аналізу. Інший підхід — модуляційне маскування — полягає у викривленні амплітудних чи фазових характеристик відеосигналу за допомогою певного ключа. Це ускладнює прямий перегляд даних стороннім спостерігачем і може бути ефективним у поєднанні з іншими методами. Однак дане рішення має суттєві недоліки: воно призводить до зниження якості зображення навіть для легітимного користувача та залишається вразливим до частотного аналізу. Окрему групу складають методи, що використовують хаотичні генератори. У цьому випадку відеосигнал модулюється хаотичним сигналом, параметри якого виступають у ролі ключа. Такий підхід забезпечує високу непередбачуваність і робить відновлення сигналу без знання початкових параметрів надзвичайно складним. Водночас практична реалізація подібних алгоритмів є технічно складною, а їхня ефективність залежить від точності апаратних компонентів та стійкості до зовнішніх шумів. Узагальнюючи, можна зазначити, що алгоритми захисту аналогових відеопотоків мають низку важливих переваг: вони здатні працювати у режимі реального часу, не потребують значних апаратних ресурсів і можуть бути реалізовані у відносно простих системах. Проте їхній головний недолік полягає у недостатньому рівні криптографічної стійкості, що робить можливим перехоплення та відновлення сигналу зловмисником. Крім того, більшість методів певною мірою погіршують якість зображення навіть для законного користувача, що обмежує їхнє застосування у практичних системах відеоспостереження та управління БПЛА.

В даній публікації запропоновано алгоритм захисту аналогового відеосигналу шляхом поєднання методів цифрового (скремблювання) та аналогового (перестановка рядків) оброблення.

3. Алгоритм перемешування рядків зображення на базі конгруентного генератора

Розглянемо спочатку роботу шифрувальної частини алгоритму.

3.1. Шифрувальна частина

Попередньо, наприклад, за допомогою генератора випадкових чисел формується кілька масивів, що містять перемішані номери рядків в інтервалі від 1 до 525. Дані масиви свідомо відомі на стороні, що шифрує і дешифрує.

1. Крок 1. Запис до оперативної пам'яті з виходу АЦП оцифрованого кадру відеопотоку. Кожен рядок записується в окрему область пам'яті. Для синхронізації запису використовуються кадровий та рядковий синхронізуючі імпульси. Так як в системі PAL використовуються 525 рядків і умовних областей пам'яті також буде 525.
2. Крок 2. Генератор випадкових чисел формує від 1 до максимального номера масиву заготовлених випадкових чисел.
3. Крок 3. Номер масиву випадкових чисел, згенерований на кроці 2, додається в інтервал кадрового імпульсу, що гасить, який поєднаний для передачі цифрової інформації, наприклад телетексту.
4. Крок 4. Формується цифровий кадр із переупорядкованих відповідно до масиву з кроку 3 рядків.
5. Крок 5. Цифровий кадр перетворюється на аналоговий за допомогою цифро-аналогового перетворювача.

3.2. Дешифрувальна частина

Розглянемо дешифрувальну частину алгоритму.

1. Крок 1. Запис на оперативну пам'ять з виходу АЦП оцифрованого кадру зашифрованого відеопотоку. Кожен рядок записується в окрему область пам'яті. Для синхронізації запису також використовуються кадровий та рядкові синхронізуючі імпульси.
2. Крок 2. З кадрового імпульсу, що гасить, витягується номер набору перемішаних рядків. Цей номер може бути додатково зашифрований.
3. Крок 3. Формується цифровий кадр із рядків у правильному порядку.
4. Крок 4. Дешифрований цифровий кадр перетворюється на аналоговий за допомогою цифро-аналогового перетворювача і подається одержувачу інформації.

4. Висновок

Таким чином, розроблено удосконалений алгоритм скремблювання аналогового відеосигналу з камери розвідувального БПЛА, який працює з вихідним кодуванням у форматі PAL. Його особливістю є поєднання цифрової та аналогової обробки сигналу, що дозволяє досягти високої криптостійкості при одночасному збереженні якості зображення та основних спектральних характеристик, зокрема ширини спектра корисного сигналу.

References

- [1] Дементьев Д.О. Бойові Літальні комплекси в складі єдиної інформаційно-розвідувально-навігаційно-ударної системи / Дементьев Д.О. // Зб. наук. пр. Військового інституту Київського національного університету ім. Тараса Шевченка. – К. : ВІКНУ, 2015. – №27. – С. 74-77.
- [2] Навроцький Д. Криптографічна система захисту радіоканалів БПЛА від несанкціонованого втручання / Д. Навроцький // Безпека інформації. - 2014. - Т. 20, № 3. - С. 248-252. - URL: http://nbuv.gov.ua/UJRN/bezin_2014_20_3_8.
- [3] Фомін І. І. Захист каналу управління безпілотних літальних апаратів від несанкціонованого доступу : кваліфікаційна робота магістра за спеціальністю „125 — кібербезпека“ / І. І. Фомін. — Тернопіль: ТНТУ, 2021. — 66 с.

Метод захисту кодової бази шляхом шифрування репозиторію

Андрій Скіп¹, Юрій Баришев²

¹ Vinnytsia National Technical University, 95 Khmelnytske shose, 21021, Vinnytsia, Ukraine

² Vinnytsia National Technical University, 95 Khmelnytske shose, 21021, Vinnytsia, Ukraine

Анотація

Розглянуто сучасний підхід до забезпечення безпеки в ланцюгах постачання програмного забезпечення шляхом інтеграції методів криптографічного захисту на ранніх етапах розробки. Запропоновано метод захисту репозиторію вихідного коду шляхом шифрування вмісту репозиторію за алгоритмом AES-256 (симетричне шифрування) та захисту ключа шифрування за допомогою RSA (асиметричне шифрування). Критичною частиною цієї архітектури є використання LDAP-сервера для автентифікації розробників та сховища секретів HashiCorp Vault для керування криптографічними ключами. Така інтеграція гарантує, що доступ до зашифрованого коду отримують лише автентифіковані користувачі з відповідними правами. Розглянуто питання ротації ключів у контексті безперервної розробки, а також інтеграцію цих рішень у конвеєри CI/CD. Запропоноване рішення відповідає принципам Zero Trust і мінімальних привілеїв та органічно інтегрується в процеси DevSecOps. Запропонована схема підвищує захищеність вихідного коду з мінімальним впливом на процес розробки і дозволяє гарантувати безпеку кодової бази в корпоративному середовищі.

Ключові слова

криптографія, шифрування, DevSecOps, захист коду, захист ланцюга постачання.

1. Вступ

Останніми роками стрімко зростає кількість та масштаб атак на ланцюги постачання програмного забезпечення. Серед найвідоміших прикладів – атаки на SolarWinds Orion та Log4j, що призвели до серйозних інцидентів інформаційної безпеки у всьому світі. Тому детальне дослідження захисту на всіх етапах життєвого циклу ПЗ є нагальною необхідністю для підприємств та державних установ. Одним із ключових елементів безпеки програмних постачань є обов'язкове використання криптографії для забезпечення конфіденційності та цілісності даних. Зокрема, шифрування вихідного коду та артефактів збірки дозволяє запобігти несанкціонованому доступу до них навіть у разі компрометації інфраструктури. Впровадження таких методів може суттєво підвищити стійкість до цільових атак на процес розробки та доставки програмних компонентів [1]. Водночас застосування криптографічних методів породжує додаткові завдання керування ключами: їх безпечної генерації, розподілу, зберігання та регулярної ротації. Наразі в ланцюгах постачання програмного забезпечення такі методи використовуються лише частково – переважно на етапах, де кодова база вже розроблена і до процесів залучені DevOps- або SecDevOps-фахівці. Робочі станції розробників та сервери інтеграції коду на початкових етапах створення бази коду залишаються незахищеними [2]. Таким чином, постає актуальне завдання інтегрування методів криптографічного захисту інформації безпосередньо на етапі розробки коду. Метою даного дослідження є покращення захисту кодової бази

⁴⁴ Corresponding authors

✉ phd@askip.me (A. Skip); yuriy.baryshev@vndu.edu.ua (Y. Baryshev)

ORCID 0009-0009-3334-4145 (A. Skip); 0000-0001-8324-8869 (Y. Baryshev)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

шляхом використання методів криптографії в ланцюгах постачання програмного забезпечення на початкових етапах розробки.

2. Архітектура запропонованого методу

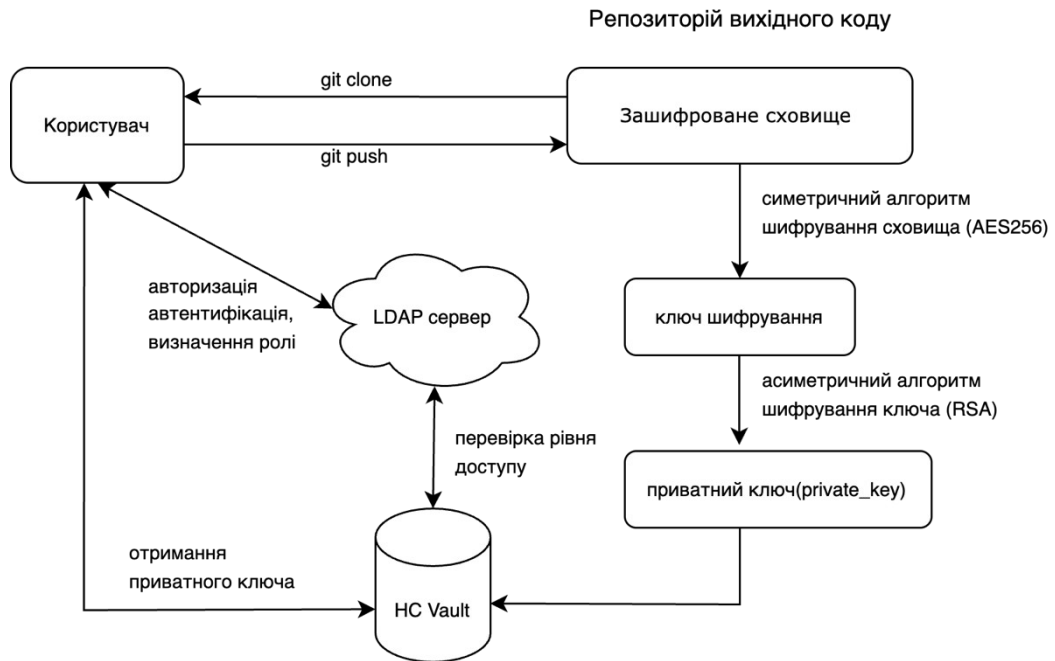


Рис 1: Схема використання шифрування в захисті репозиторію вихідного коду (AES-256 для даних, RSA для ключів, LDAP для автентифікації, HashiCorp Vault [3] для зберігання ключів).

На рис. 1 зображено архітектуру запропонованого рішення. Алгоритм функціонування цієї системи включає такі основні кроки:

1. Зашифрований репозиторій вихідного коду. Вихідний код у репозиторії, що зберігається на сервері, постійно перебуває у зашифрованому вигляді. Для шифрування вмісту використовується симетричний алгоритм AES-256, сучасний стандарт блочного шифрування [4]. Таким чином, навіть здобувши несанкціонований доступ до файлової системи сервера репозиторію, зломисник не зможе прочитати вихідний код без відповідного AES-ключа.
2. Захист AES-ключа за допомогою RSA. Щоб надати авторизованим користувачам можливість доступу до зашифрованого репозиторію, необхідно безпечно зберігати і передавати симетричний ключ. Для цього застосовано гібридну схему з використанням асиметричного шифрування RSA: генерується пара ключів RSA (відкритий та закритий). Відкритий ключ використовується для шифрування симетричного ключа AES, а закритий – для його розшифрування. Такий підхід відповідає типовій схемі гібридного шифрування, яка поєднує високу швидкодію AES та зручність RSA у розповсюдженні ключів [5]. Розрядність RSA-ключів

обирається відповідно до сучасних рекомендацій (не менше 2048 біт, а краще 3072 і більше для довгострокової безпеки) [6].

3. Зберігання ключів у Vault. Для безпечного зберігання криптографічних ключів використовується централізоване сховище HashiCorp Vault [3]. Приватний RSA-ключ зберігається у Vault у зашифрованому вигляді та доступний лише автентифікованим і авторизованим користувачам відповідно до політик доступу. Цей ключ ніколи не розміщується у відкритому вигляді на сервері репозиторію – він постійно ізолюваний у Vault
4. Автентифікація користувача через LDAP. Розробник, який бажає отримати доступ до репозиторію, спочатку проходить автентифікацію через LDAP-сервер, вводячи свої облікові дані. LDAP перевіряє їх у корпоративній системі директорій (наприклад, Active Directory) і підтверджує успішність автентифікації, а також може передати інформацію про групи/ролі користувача. Таким чином система дізнається, до яких саме ресурсів користувач має доступ
5. Отримання приватного ключа через Vault. Після успішної автентифікації розробник отримує доступ до сховища Vault, використовуючи свій LDAP-акаунт (Vault виступає клієнтом LDAP і зіставляє групи користувача з налаштованими політиками доступу). У результаті входу розробник отримує від Vault тимчасовий токен або інший дозвіл, за допомогою якого може запитати закритий ключ RSA. Vault перевіряє дійсність токена та відповідність політикам і, у разі дозволу, надає копію приватного ключа для використання на боці клієнта.
6. Доступ до репозиторію з розшифруванням на боці клієнта. Отримавши приватний ключ RSA, клієнт (інструмент розробника) розшифровує ним симетричний ключ AES (сам AES-ключ може бути переданий у вигляді файлу чи конфігурації, зашифрованої відкритим ключем). Далі клієнтський Git-інструмент застосовує AES-256 для розшифрування вмісту репозиторію на стороні розробника. Таким чином, під час виконання `git clone` розробник отримує з сервера зашифровані дані репозиторію, які автоматично розшифровуються з використанням отриманого AES-ключа (витягнутого з Vault через RSA-розшифрування). При виконанні `git push` шифрування відбувається у зворотному порядку: перед надсиланням на сервер вихідний код шифрується ключем AES і лише в зашифрованому вигляді зберігається на сервері. Важливо, що сервер репозиторію ніколи не оперує відкритим (нешифрованим) вмістом – у сховищі коду зберігаються тільки зашифровані артефакти.

3. Висновки та напрями подальших досліджень

У роботі проаналізовано підхід до захисту ланцюга постачання програмного забезпечення, заснований на наскрізному використанні криптографії – від зберігання вихідного коду до процесів безперервної інтеграції. Запропоноване рішення поєднує симетричне шифрування (AES-256) для захисту вмісту репозиторію та асиметричне шифрування (RSA) для захисту ключів, а також інфраструктурні компоненти (LDAP для автентифікації користувачів та HashiCorp Vault для управління секретами). Проведений аналіз підтвердив доцільність такого поєднання: AES-256 забезпечує надійну і швидку конфіденційність даних, RSA дозволяє безпечно розподіляти ключі між користувачами, а Vault надає централізований контроль доступу і журналювання операцій із секретами. Розглянуто практичні аспекти реалізації схеми. Показано, що інтеграція з LDAP спрощує управління доступом у великих командах, уніфікує автентифікацію розробників та дозволяє використати наявні політики безпеки організації. Використання Vault дозволяє дотриматися принципу мінімальних привілеїв – доступ до критичних криптографічних ключів надається лише в обсязі та на час, необхідний для роботи, що знижує ризик їх компрометації. Використання методів шифрування в ланцюгах постачання ПЗ значно підвищує стійкість до цільових атак на процес розробки і доставки. Поєднання перевірених криптоалгоритмів (AES, RSA) із сучасними засобами керування ключами (Vault) та існуючими системами автентифікації (LDAP) дозволяє побудувати багаторівневий захист без істотного впливу на продуктивність і зручність роботи розробників. Впровадження описаного підходу може стати важливим кроком для організацій, що прагнуть мінімізувати ризики компрометації своїх програмних продуктів на етапі їх створення та доставки. Подальші дослідження доцільно спрямувати на експериментальну оцінку продуктивності запропонованої системи у великих масштабах та аналіз її економічної доцільності. Окремим напрямом може бути розширення запропонованої схеми механізмами забезпечення цілісності коду (наприклад, вбудованим цифровим підписуванням артефактів) для комплексного посилення кібербезпеки ланцюга постачання ПЗ.

Література

- [1] Homomorphic encryption-driven CI/CD pipelines for zero-trust builds, International Journal of Communication Networks and Information Security 14 (2022) 1129–1139. doi:10.5281/zenodo.15723391.
- [2] M. S. Melara, M. Bowman, What is Software Supply Chain Security?, in: Proc. of the ACM Workshop on Supply Chain Security, Los Angeles, 2022, p. 1. arXiv:2209.04006.
- [3] HashiCorp, Vault Documentation. URL: <https://developer.hashicorp.com/vault>
- [4] J. Nechvatal et al., Report on the development of the Advanced Encryption Standard (AES), Journal of Research of the National Institute of Standards and Technology 106 (3) (2001) 511. doi:10.6028/jres.106.023.
- [5] S. Sa'idu, An improved cryptographic scheme using AES and RSA algorithms for maximum security in file encryption and decryption, International Journal of Scientific Development and Research 7 (6) (2022) 494–500. URL: <https://ijsdr.org/papers/IJS DR2206079.pdf>.
- [6] E. Barker, Recommendation for Key Management, NIST Special Publication 800-57 Part 1, Revision 5, Gaithersburg, MD, 2020. doi:10.6028/NIST.SP.800-57pt1r5.

ECG for Multi-Factor Cybersecurity: Rhythm-Adaptive Authentication During Mild Physical Activity

Serhii Lupenko^{1,2,3}, Roman Butsiy^{3,*}

¹ Department of Informatics of Opole University of Technology, 76 Prószkowska St., 45-758 Opole, Poland

² EPAM School of Digital Technologies, American University Kyiv, 3 Poshtova Sq., 04070 Kyiv, Ukraine

³ Institute of Telecommunications and Global Information Space, 13 Chokolivskiy blvd., 02000 Kyiv, Ukraine

Abstract

The paper investigates electrocardiogram (ECG)-based biometric authentication under mild physical activity (walking), a setting where rhythm and morphology depart from resting-state assumptions and often degrade conventional methods. This paper proposes a rhythm-adaptive pipeline grounded in cyclic random process (CRP) modeling: a subject-specific rhythm function is estimated from single-lead ECG, cycles are segmented accordingly, rhythm-adaptive statistical descriptors are computed, and decisions are made either by a model-based Statistical Interval Classifier (SIC) or by standard machine-learning (ML) classifiers. Experiments on the PhysioNet PTP dataset (ECG channel; 2/3 train, 1/3 test) show that explicit modeling of rhythm variability stabilizes performance with very few beats: although single-cycle decisions are feasible, averaging 2–3 consecutive cycles yields consistently higher and more reliable accuracy across classifiers. In this study, peak ECG authentication reaches 100% accuracy with a Linear SVM, while aggregate metrics—accuracy, balanced accuracy, and F1 score—remain high for both hybrid (model-based preprocessing + ML) and purely model-based (SIC) variants. These results indicate that CRP-driven, rhythm-adaptive preprocessing is a key enabler for robust and interpretable ECG biometrics in real-world conditions, effectively bridging theory-guided signal modeling and flexible classifier design to maintain reliability under motion-induced rhythm variability.

Keywords

biometric authentication, ECG, rhythm-adaptive processing, cyclic random process, cardiac variability ⁴⁵

1. Introduction

Modern cybersecurity increasingly relies on multi-factor authentication, where knowledge- and possession-based credentials (passwords, tokens) are complemented by biometric factors to raise assurance against credential theft, replay, and spoofing. Within this landscape, electrocardiogram (ECG) biometrics offer several desirable properties: the signal originates from internal physiological activity (providing inherent liveness), can be captured non-invasively with wearables and edge devices, and supports continuous or on-demand verification in safety-critical contexts (e.g., access to medical records or connected healthcare devices). Rather than replacing established modalities, ECG is positioned as a complementary factor that strengthens identity verification in practical multi-factor, multimodal frameworks [1-4].

Despite this potential, many ECG-based approaches implicitly assume resting-state stationarity and quasi-periodicity. In real-world use—even during mild physical activity (walking)—the cardiac rhythm exhibits cycle-to-cycle variability and morphological deviations driven by autonomic regulation and motion. Observable effects include heart-rate acceleration, cycle shortening, and changes in the amplitude and duration of characteristic waves (P, QRS, T), alongside baseline

^{45*} Corresponding author.

[†] These authors contributed equally.

✉ s.lupenko@po.edu.pl (S. Lupenko); butsii_roman@nas.gov.ua (R. Butsiy)

ORCID 0000-0002-6559-0721 (S. Lupenko); 0000-0002-8415-8635 (R. Butsiy)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

wander and movement artifacts. Such departures from stationarity degrade methods that rely on fixed-length cycles, rigid templates, or purely time-locked features, motivating an explicitly rhythm-adaptive analysis. While recent work has begun to consider ECG variability in authentication settings [5], a principled framework that models rhythm changes directly is essential to improve robustness under realistic conditions.

This study adopts the cyclic random process (CRP) perspective as a rigorous theoretical basis for rhythm-adaptive ECG processing. The CRP framework introduces a rhythm function that explicitly captures regular and irregular rhythmic structures, enabling consistent rhythm-adaptive estimators of probabilistic characteristics and mitigating statistical blurring that arises when rhythms drift. Prior results have shown that CRP-based modeling can outperform periodic (cyclostationary) formulations in representing ECG dynamics precisely because it provides a mechanism to handle rhythm variability [6, 7]. Building on this foundation, the present work operationalizes CRP modeling for ECG-only biometric authentication during walking, integrating a subject-specific rhythm function, cycle-aware segmentation, and rhythm-adaptive statistical descriptors into a practical pipeline.

2. Results of the research

The paper uses a vector model of cyclic rhythmically connected random processes, in which ECG is treated as components of a vector cyclic random process governed by a rhythm function. Conditions, definitions, and analytic representations of such processes—as well as rhythm-adaptive estimators for their probabilistic characteristics—are given in [6–8]; in brief, the rhythm function captures cycle-to-cycle variability and enables consistent, rhythm-aware estimation and comparison of ECG statistics.

Data were sourced from PhysioNet (PTTP dataset) [9, 10], using single-lead ECG (Lead I) recorded at 500 Hz with the AD8232 amplifier; leveraging a public, well-documented dataset ensures standardization and reproducibility. A custom Python implementation orchestrates the pipeline—preprocessing, rhythm estimation, cycle segmentation, and classification—using a compact, well-known stack (NumPy, Pandas, NeuroKit2, scikit-learn) to keep the workflow transparent and portable. The experimental split followed a 2/3 train and 1/3 test partition of recordings.

Two complementary decision strategies were considered. The hybrid strategy uses model-based preprocessing and rhythm-adaptive statistics as features for standard classifiers (Nearest Neighbors, Linear SVM, Gaussian Process, Decision Tree, Random Forest, MLP, AdaBoost). The purely model-based strategy employs the Statistical Interval Classifier (SIC), which embeds rhythm-adaptive probabilistic information directly in the decision rule. This design allows a controlled comparison between theory-guided and data-driven inference while holding the front-end modeling fixed.

Results under mild physical activity (walking) show that accurate ECG authentication is feasible from single cycles, yet averaging 2–3 consecutive cycles consistently improves stability across classifiers—an operating point that balances latency with robustness (Figure 1).

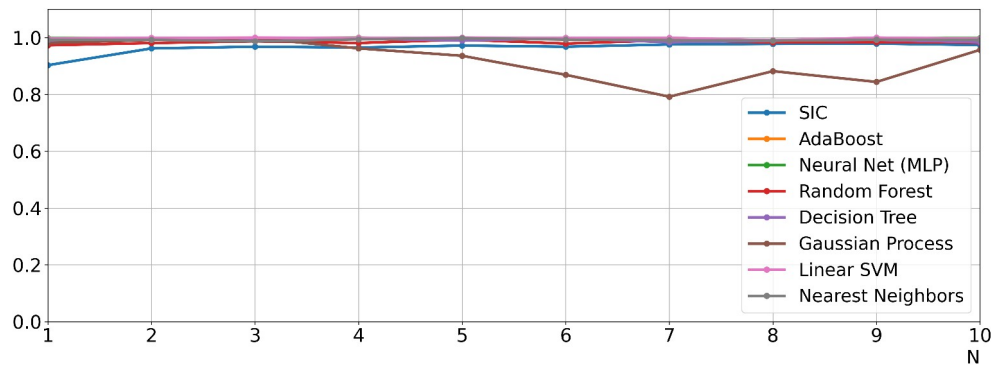


Figure 1: Dependence of average ECG-based authentication accuracy on the number of averaged cycles across classifier types.

In the hybrid approach, Linear SVM attains up to 100% accuracy, with Gaussian Process and MLP also near unity; AdaBoost, Decision Tree, Random Forest, and Nearest Neighbors trail slightly yet remain high (approximately 0.987–0.999 across accuracy, balanced accuracy, and F1 score). The SIC baseline performs competitively but below the best hybrids, trading a small amount of accuracy for interpretability and transparency inherent to its model-based construction. Confidence-interval analysis at a high probability level corroborates the reliability of these point estimates. Collectively, the findings indicate that CRP-driven rhythm adaptation with cycle averaging (2–3 beats) yields a robust ECG-only solution suitable for real-world, mobile contexts.

Overall, these results underscore the benefits of employing a rhythm-adaptive, model-based preprocessing approach. Such methods ensure reliable and robust biometric authentication performance, highlighting the importance of selecting an optimal number of cycles for averaging when dealing with ECG signals captured during mild physical activities (Table 1).

Table 1

The point and interval estimations of mathematical expectations of main characteristics of model-based and hybrid biometric person authentication based on ECG

Classifier Type	Authentication Performance Metrics		
	Accuracy	Balanced Accuracy	F1 score
SIC	0.968, (0.946, 0.991)	0.969, (0.947, 0.992)	0.967, (0.943, 0.992)
AdaBoost	0.996, (0.991, 1.0)	0.996, (0.991, 1.0)	0.996, (0.991, 1.0)
Neural Network (MLP)	0.999, (0.998, 1.0)	0.999, (0.998, 1.0)	0.999, (0.998, 1.0)
Random Forest	0.987, (0.977, 0.998)	0.988, (0.977, 0.998)	0.987, (0.977, 0.998)
Decision Tree	0.993, (0.984, 1.0)	0.993, (0.984, 1.0)	0.993, (0.984, 1.0)
Gaussian Process	0.999, (0.998, 1.0)	0.999, (0.998, 1.0)	0.999, (0.998, 1.0)
Linear SVM	1.0, (0.999, 1.0)	1.0, (0.999, 1.0)	1.0, (0.999, 1.0)
Nearest	0.987, (0.977, 0.998)	0.986, (0.977, 0.998)	0.986, (0.977, 0.998)

Classifier Type	Authentication Performance Metrics		
	Accuracy	Balanced Accuracy	F1 score
Neighbors	(0.974, 0.999)	(0.973, 0.999)	(0.973, 0.999)

3. Conclusions

The findings of this study demonstrate the significant advantages offered by rhythm-adaptive approaches for biometric authentication utilizing ECG signals recorded during mild physical activity, particularly walking. Traditional cyclostationary and purely data-driven methods often assume rhythmic consistency, which substantially limits their effectiveness under dynamic physiological conditions with pronounced rhythm variability and morphological fluctuations. In contrast, the proposed hybrid and model-based approaches explicitly incorporate cyclic random process models, thereby effectively addressing and compensating for inherent rhythm irregularities and inter-cycle morphological variations.

Accurate authentication is feasible from a single cycle, while averaging 2–3 cycles provides consistently higher and more stable accuracy across classifiers under walking. Within the hybrid approach, Linear SVM achieved up to 100% accuracy for ECG; the Statistical Interval Classifier (SIC)—though slightly below the very best hybrid models—remains valuable for its interpretability, explicit treatment of rhythm variability, and transparent decision rules. Interval estimates corroborate the reliability of these point metrics and indicate that CRP-driven preprocessing is a robust basis for downstream classification in realistic conditions.

From a cybersecurity perspective, ECG is a privacy-preserving, liveness-bearing biometric that augments passwords, tokens, or device-bound authenticators in multi-factor and continuous authentication scenarios. The demonstrated rhythm-adaptive CRP approach is particularly important for mobile and wearable use cases, where movement and autonomic modulation are the norm: it sustains recognition performance without demanding long windows, thereby lowering latency and user friction while improving resistance to spoofing and replay. Overall, the study provides compelling evidence for integrating CRP-based rhythm adaptation into ECG authentication stacks and motivates future work on standardized biometric metrics, broader population validation, and deployment on resource-constrained devices.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-5 Pro in order to: translation and grammar and spelling check. After using these tools, the authors reviewed and edited the content as needed and takes full responsibility for the publication’s content.

References

- [1] W. Reklewski, M. Miśkiewicz, and P. Augustyniak, "QRS detector performance evaluation aware of temporal accuracy and presence of noise," *Sensors*, vol. 24, no. 5, Art. no. 1698, 2024, doi: 10.3390/s24051698.
- [2] G. Wang, S. Shanker, A. Nag, Y. Lian, and D. John, "ECG biometric authentication using self-supervised learning for IoT edge sensors," *IEEE J. Biomed. Health Informat.*, vol. 28, no. 11, pp. 6606–6618, Nov. 2024, doi: 10.48550/arXiv.2409.05627.
- [3] Y. Che, L. Du, G. Tang, and S. Ling, "A biometric identification for multi-modal biomedical signals in geriatric care," *Sensors*, vol. 24, Art. no. 6558, 2024, doi: 10.3390/s24206558.

- [4] M. Al Alfi, P. Peris-Lopez, and C. Camara, "Enhancing biometric identification using 12-lead ECG signals and graph convolutional networks," *Frontiers in Digital Health*, vol. 7, 2025, doi: 10.3389/fdgth.2025.1547208.
- [5] H. Ko, K. Rim, and J. Y. Hong, "Biometric authentication with electrocardiogram (ECG) by considering variable signals," *Math. Biosci. Eng.*, vol. 20, pp. 1716–1729, 2023, doi: 10.3934/mbe.2023078.
- [6] S. Lupenko and R. Butsiy, "Isomorphic multidimensional structures of the cyclic random process in problems of modeling cyclic signals with regular and irregular rhythms," *Fractal Fract.*, vol. 8, Art. no. 203, 2024, doi: 10.3390/fractalfract8040203.
- [7] S. Lupenko, "Rhythm-adaptive statistical estimation methods of probabilistic characteristics of cyclic random processes," *Digit. Signal Process.*, vol. 151, Art. no. 104563, 2024, doi: 10.1016/j.dsp.2024.104563.
- [8] S. Lupenko, "The mathematical model of cyclic signals in dynamic systems as a cyclically correlated random process," *Mathematics*, vol. 10, no. 18, Art. no. 3406, 2022, doi: 10.3390/math10183406.
- [9] A. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. e215–e220, 2000.
- [10] P. Mehrgardt, M. Khushi, S. Poon, and A. Withana, Pulse Transit Time PPG Dataset (version 1.1.0), PhysioNet, 2022. doi: 10.13026/jpan-6n92.

Audio Signal Steganography with Neural Networks for Data Integrity in Secure Communications

Ostap-Sviatoslav Malets¹, Olha Smotr¹

¹ Lviv State University of Life Safety, Kleparivska 35, 79000 Lviv, Ukraine

Abstract

Audio steganography enables the embedding of integrity markers—such as cryptographic hashes or authentication codes—directly into audio signals, allowing tamper detection after distribution. Traditional methods often fail under compression, filtering, or re-encoding, prompting a shift toward neural network-based approaches. Recent models, including IDEAW, WavMark, WAKE, FGAS, and AAW, use architectures such as invertible networks, key-controlled frameworks, and adversarial perturbation generation to improve robustness, imperceptibility, and adaptability. These systems demonstrate resilience to a variety of signal distortions and have been applied in domains such as broadcast monitoring, secure teleconferencing, and forensic audio verification. This paper surveys these developments, outlining the core methods, evaluation practices, applications, and challenges shaping the current landscape of neural audio steganography for data integrity.

Keywords

Audio Steganography, Neural Networks, Data Integrity, Digital Watermarking, Secure Communications

1. Introduction

Ensuring the integrity and authenticity of audio content is increasingly critical as digital media proliferates. While encryption secures data in transit, it does not protect against post-distribution tampering. Embedding authentication codes directly into the audio stream via audio steganography offers a way to detect such manipulation after the fact.

Traditional methods—including the Least Significant Bit (LSB) technique, spread spectrum, and echo hiding—rely on heuristic modifications to the signal. However, these are often brittle under transformations like compression or filtering. Modern deep learning methods use neural networks to improve robustness, imperceptibility, and embedding capacity.

2. Fundamentals and Neural Approaches

Neural audio steganography builds upon the principles of traditional methods such as Least Significant Bit (LSB) embedding, spread spectrum, and echo hiding, while addressing their vulnerabilities to compression, filtering, and other signal alterations. Instead of relying on fixed heuristics, modern systems use deep neural architectures to adaptively determine the most suitable embedding locations and strategies. Models such as IDEAW employ invertible dual-embedding networks to enhance both the accuracy of watermark localization and the capacity for information storage [1]. WavMark achieves low bit error rates by applying segment-based watermarking combined with robust mismatch recovery techniques [2]. WAKE introduces a key-controllable, multi-watermark framework that enables secure and flexible embedding while preserving imperceptibility [3]. FGAS enhances perceptual quality and resistance to steganalysis by generating adversarial perturbations decoded through a lightweight fixed decoder [4]. Similarly, AAW

⁴⁶ Corresponding authors

✉ malets.o.i@gmail.com (O.S.Malets); olgasmotr@gmail.com (O.Smotr);

📞 0009-0008-5413-7666 (O.S.Malets); 0000-0003-2767-5019 (O.Smotr);

leverages adversarial perturbation techniques and whitening layers to improve resilience while maintaining a streamlined decoding process [5]. Collectively, these advancements represent a shift from static, handcrafted embedding schemes toward adaptive, data-driven methods capable of maintaining watermark integrity in complex and dynamic audio environments.

3. Evaluation Metrics and Threat Models

Evaluation has evolved beyond simple SNR-based tests to include metrics such as Bit Error Rate (BER), Perceptual Evaluation of Speech Quality (PESQ), and Passive Objective Listening Quality Assessment (POLQA). Robustness is tested against compression (e.g., MP3), noise, resampling, and de-synchronization.

Moreover, robustness studies reveal that no single approach is infallible; hybrid, multi-layer methods (e.g., combining adversarial and invertible networks) show better resilience.

4. Applications in Data Integrity

Neural steganographic techniques serve several real-world use cases:

- **Broadcast Monitoring:** WavMark-style watermarking supports imperceptible embedding in streaming systems—enabling detection even after re-broadcast or compression [2]
- **Secure Teleconferencing:** FGAS’s shallow decoder model prevents malicious intercept, offering real-time integrity verification [4]
- **Forensic Recording Authenticity:** Key-controlled watermarking (e.g., WAKE) ensures traceability to original source even under multiple re-encodings [3]

5. Challenges and Future Directions

Neural steganographic techniques have been adopted in a variety of real-world data integrity applications. In broadcast monitoring, methods such as WavMark enable imperceptible watermark embedding within streaming systems, allowing detection of embedded integrity markers even after re-broadcast or compression [2]. In secure teleconferencing scenarios, models like FGAS, which use a shallow fixed-decoder architecture, can prevent malicious interception by enabling real-time verification of communication authenticity [4]. For forensic recording analysis, key-controlled watermarking approaches such as WAKE ensure that the source of an audio file can be traced with high certainty, even after multiple rounds of re-encoding [3]. These examples demonstrate how current research has moved beyond theoretical frameworks to address practical requirements for authentication and verification in diverse audio distribution and security contexts.

References

- [1] P. Li et al., “IDEAW: Robust Neural Audio Watermarking with Invertible Dual-Embedding,” EMNLP, 2024.
- [2] WavMark team, “WavMark: Robust Audio Watermarking with Low BER,” arXiv preprint, 2024.
- [3] Y. Xu et al., “WAKE: Watermarking Audio with Key Enrichment,” arXiv, 2025.
- [4] J. Yan et al., “FGAS: Fixed Decoder Audio Steganography with Adversarial Perturbation Generation,” arXiv, 2025.
- [5] S. Wu et al., “AAW: Audio Watermarking via Adversarial Perturbation,” Appl. Sci., 2024.

Graph-Based Analysis and Hidden Channel Discovery in SELinux Policies

Vitaliy Kukharskyy¹, Mykola Shcherbina¹

¹ Ivan Franko National University of Lviv, Universytetska 1, 79000 Lviv, Ukraine

Abstract

Mandatory Access Control (MAC) frameworks such as SELinux enhance Linux security by enforcing system-wide policies, but crafting precise rules at scale remains challenging. Moreover, policy correctness alone cannot eliminate covert inter-process communication channels. We propose a graph-based method that automatically extracts bidirectional channels from SELinux allow rules by leveraging wildcard-mapped file and property types. Our approach filters for resources with distinct writer and reader domains and produces three-tier PNG visualizations via Graphviz. We conclude by discussing policy-tightening strategies, complementary integrity mechanisms, and by releasing an open-source Python tool for policy analysis and visualization.

Keywords

Android firmware analysis, covert channels, embedded systems security, graph-based visualization, information flow analysis, Mandatory Access Control (MAC), SELinux

1. Introduction

Mandatory Access Control (MAC) systems like SELinux label processes (“*subjects*”) and resources (“*objects*”) with types, enforcing a global policy that supplements Linux’s native Discretionary Access Control (DAC). While MAC closes many DAC-related gaps, crafting and maintaining SELinux policies at scale remains challenging. Moreover, policy correctness alone does not eliminate covert communication paths: malicious payloads confined by SELinux can repurpose legitimate channels — file writes and property sets — to synchronize and exfiltrate data.

This paper introduces a graph-centric methodology to discover and visualize these hidden channels. By focusing on wildcard-mapped types from *file_contexts* and *property_contexts*, and filtering for true bidirectional flows, we generate concise diagrams that highlight high-risk conduits. We further show how such graphs facilitate attacker planning, using a real-world firmware example.

2. Background

Linux’s built-in DAC model (file mode bits, ACLs) grants object owners control over access, but cannot reliably prevent privilege escalation or lateral movement once an object is compromised. MAC frameworks like SELinux enforce centralized policies: every subject–object interaction requires an explicit allow rule, regardless of ownership.

On desktop Linux, SELinux adoption has been limited by policy complexity, tooling limitations, and performance impact. In contrast, embedded and mobile platforms with fixed process sets benefit from precise SELinux policies [1]. To guard against offline tampering, platforms such as Android employ additional integrity checks (e.g., *dm-verity*), which were absent in the Sony Xperia M5 image used in our study.

⁴⁷ Corresponding authors

✉ vitaliy.kukharskyy@lnu.edu.ua (V. Kukharskyy); mykola.shcherbina@lnu.edu.ua (M. Shcherbina)

ORCID 0000-0002-2775-6401 (V. Kukharskyy); 0009-0003-9761-9466 (M. Shcherbina)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Crafting and maintaining SELinux policies remains a nontrivial task. Organizations often resort to ad hoc rule generation, “trial-and-error” permissive modes, or manual audits. Formal approaches (e.g., policy specification languages anchored in security proofs) have been explored academically but see limited industrial adoption [2]. Best practice dictates that SELinux rules be version-controlled and updated in lock-step with application code. This alignment ensures that new features or configuration changes automatically trigger policy reviews. Continuous integration pipelines must include policy compilation and test suites to detect unintended denials or over-privileging. Without this rigor, policies either become too loose (to avoid functional breakage) or too brittle (blocking legitimate flows).

SEGrapher [3] ingests SELinux rules into a Neo4j graph database and renders them as a clustered, directed graph, enabling administrators to interactively explore policy inconsistencies and coverage. By contrast, our approach focuses exclusively on automated detection and visualization of *covert inter-process channels* — rather than general policy mining — and delivers a lighter, more targeted, and fully automatable solution, filling the gap left by graph-database-centric analyses such as those in [4].

3. Methodology

We parse context files and select only entries that define resource classes via wildcards:

- From *file_contexts*, include only object types whose path patterns contain regex or wildcard tokens (*, ?, (,)).
- From *property_contexts*, include only object types whose property prefixes end with a dot (.).

Only these wildcard-mapped types enter the bidirectional-channel analysis.

We process *allow.txt* to extract tuples (*source*, *target_type*, *class*, *permissions*) via a single regular expression. We consider only *file* and *property_service* classes in allow rules.

For bidirectional channel detection, permissions are classified into subgroup categories:

- *File* class
 - *WRITE_FILE* if {"create", "write"} $\subseteq$ *perms*
 - *READ_FILE* if {"open", "read"} $\subseteq$ *perms*
- *Property_service* class
 - *WRITE_PROP* if "set" $\in$ *perms*
 - *READ_PROP* is inferred implicitly for all non-writing domains

A target type qualifies only if it has at least one distinct writer and one distinct reader domain.

Using the Python graphviz library [5], we render a directed PNG graph in three horizontal tiers: File Targets (*rank=max*), Source Domains (*rank=same*), Property Targets (*rank=min*). Solid, thick edges denote writes; dashed edges denote reads (Figure Error: Reference source not found).

4. Experimental Setup

Key policy artifacts were extracted from the Sony Xperia M5 (E5603) firmware (version 30.2.A.1.21), which lacks *dm-verity*. From the boot ramdisk, we obtained:

- *sepolicy* (343 728 bytes)
- *file_contexts* (49 872 bytes)
- *property_contexts* (12 209 bytes)

To facilitate parsing, we first converted the binary SELinux policy (*sepolicy*) into a flat, human-readable list of allow rules. Specifically, we used the SETools utility *sesearch* with the *--allow* flag to dump all *allow* statements into *allow.txt*. This text file serves as the sole input for our Python

pipeline’s policy-parsing stage, ensuring a consistent and complete representation of every granted interaction.

We focused on two critical services as potential compromised domains:

- `/system/bin/ccci_mdinit`: MediaTek Helio X10 modem initialization
- `/system/bin/vold`: Volume daemon for mounting storage

In our experiments, the Python script (see A. Source Code) was executed with the following command-line arguments:

```
python segraph.py allow.txt graph.png ccci_mdinit vold
```

Here, *allow.txt* is the *sesearch*-dumped *allow* rules file, *graph.png* is the output diagram (see Figure 1), and *ccci_mdinit* and *vold* are the two domains to include in the channel graph.

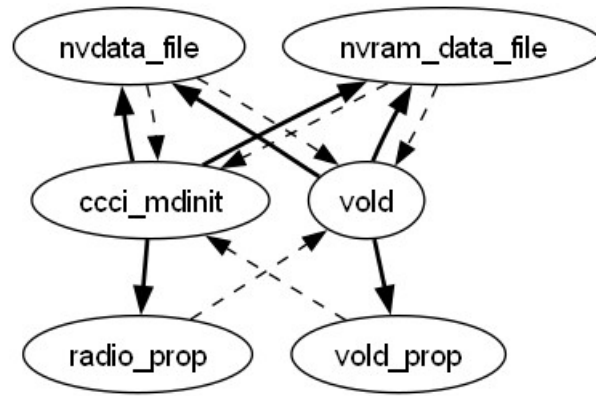


Figure 1: Channel graph between *ccci_mdinit* and *vold*.

The interesting file labels correspond to the following wildcards:

```

/nvdata(/.*)?      u:object_r:nvdata_file:s0
/data/nvram(/.*)?  u:object_r:nvram_data_file:s0
  
```

Respectively, the discovered property wildcards were:

```

crypto.      u:object_r:vold_prop:s0
gsm.         u:object_r:radio_prop:s0
persist.vold. u:object_r:vold_prop:s0
ril.         u:object_r:radio_prop:s0
tedongle.    u:object_r:radio_prop:s0
vold.        u:object_r:vold_prop:s0
  
```

5. Proof-of-Concept Demonstration

We leveraged the discovered covert channel between the *ccci_mdinit* and *vold* to replace the stock firmware with a custom image. During the standard initialization process, *ccci_mdinit* issues the *ioctl* requests *CCCI_IOC_SET_MD_IMG_EXIST* and *CCCI_IOC_GET_MD_INFO*. We intercepted both calls and synchronized the execution of the mount command so that our patched image was mounted at `/custom/etc/firmware/` before the modem accessed it. The detailed implementation of this proof-of-concept is beyond the scope of this paper.

6. Conclusion

Representing SELinux *allow* rules and wildcard context mappings as a directed graph and filtering for bidirectional flows systematically uncovers hidden inter-process channels. Our case study demonstrates how confined processes can exploit such channels to coordinate a covert firmware-injection attack. Effective defense requires layering MAC with integrity enforcement, rigorous policy-as-code practices, and continuous graph-based policy analysis.

The provided tool does not detect more complex covert channels — such as an indirect flow from domain *A* to domain *B* mediated via an intermediary domain *C*. Other potential covert channels exist — such as sockets. Furthermore, a compromised application could conceal data in the names of newly created files, thereby eliminating the need for explicit read or write permissions. In future research, we plan to integrate AI- and machine learning-based techniques to automate the discovery of hidden channels within SELinux policy rules. Moreover, our methodology is fully generalizable and can be applied to any system enforcing mandatory access control policies.

Declaration on Generative AI

During the preparation of this work, the authors used Microsoft Copilot in order to: Improve writing style. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] Security-Enhanced Linux in Android. URL: <https://source.android.com/docs/security/features/selinux>
- [2] J. Briffaut. Formalization and guaranty of system security properties: application to the detection of intrusions. (2007). URL: https://www.researchgate.net/publication/273450067_Formalization_and_guaranty_of_system_security_properties_application_to_the_detection_of_intrusions
- [3] S. Marouf and M. Shehab. SEGrapher: Visualization-based SELinux policy analysis. 2011 4th Symposium on Configuration Analytics and Automation (SAFECONFIG), pp. 1–10. DOI: 10.1109/SafeConfig.2011.6111675
- [4] A. Eaman, P. Jadczyk, and H. Chipman. Graph-Powered Mining and Analysis of SELinux Security Policies. In: Proceedings of the Future Technologies Conference (FTC) 2024, Volume 4. FTC 2024. Lecture Notes in Networks and Systems, vol 1157. Springer, Cham. https://doi.org/10.1007/978-3-031-73128-0_9
- [5] graphviz. URL: <https://pypi.org/project/graphviz/>

A. Source Code

```
import os
import sys
import re
import platform
from collections import defaultdict
from graphviz import Digraph
# On Windows, prepend Graphviz bin folder so dot executable is found
if platform.system() == 'Windows':
    GV_BIN = r"C:\Program Files\Graphviz\bin"
    os.environ["PATH"] += os.pathsep + GV_BIN
# Define file permission subgroups for CLASS="file"
```

```

FILE_SUBGROUPS = {
    "READ_FILE": {"open", "read"},
    "WRITE_FILE": {"create", "write"},
}
# Regex to parse allow rules
ALLOW_RE = re.compile(
    r'^(?:allow|auditallow|dontaudit|neverallow)\s+'
    r'(?P<SOURCE>\w+)\s+'
    r'(?P<TARGET>\w+)\s*:\s*'
    r'(?P<CLASS>\w+)\s*'
    r'(?:'
    r'(?P<PERM>\w+)\s*;'
    r'|'
    r'\{\s*(?P<PERMS>[\w\s]+?)\s*\}\s*;'
    r')$'
)
def parse_allow_line(line):
    """
    Parse a line from allow.txt and return
    (source, target, class, set(permissions)) or None.
    """
    m = ALLOW_RE.match(line.strip())
    if not m:
        return None
    source = m.group('SOURCE')
    target = m.group('TARGET')
    cls = m.group('CLASS')
    perms = {m.group('PERM')} if m.group('PERM') else set(m.group('PERMS').split())
    return source, target, cls, perms
def classify_file_edge(perms):
    """
    Determine if a permission set maps to READ_FILE or WRITE_FILE.
    """
    result = []
    for name, keys in FILE_SUBGROUPS.items():
        if keys.issubset(perms):
            result.append(name)
    return result
def load_file_contexts(path='file_contexts'):
    """
    Read file_contexts and return set of SELinux types
    whose path patterns include wildcard or regex.
    """
    wild_types = set()
    try:
        with open(path, encoding='utf-8') as f:
            for line in f:
                text = line.strip()
                if not text or text.startswith('#'):
                    continue
                parts = text.split()
                if len(parts) < 2:
                    continue
                pattern, ctx = parts[0], parts[1]
                # include only entries with wildcard or regex
                if any(c in pattern for c in '*?()+'):
                    # context is like u:object_r:type:s0
                    fields = ctx.split(':')
                    if len(fields) >= 3:
                        wild_types.add(fields[2])
    except FileNotFoundError:
        sys.exit(f"Error: '{path}' not found.")
    return wild_types
def load_property_contexts(path='property_contexts'):
    """
    Read property_contexts and return set of SELinux types
    whose property prefixes end with a dot.
    """
    wild_types = set()
    try:
        with open(path, encoding='utf-8') as f:
            for line in f:
                text = line.strip()
                if not text or text.startswith('#'):

```

```

        continue
    parts = text.split()
    if len(parts) < 2:
        continue
    prefix, ctx = parts[0], parts[1]
    # include only entries with trailing dot prefix
    if prefix.endswith('.'):
        fields = ctx.split(':')
        if len(fields) >= 3:
            wild_types.add(fields[2])
except FileNotFoundError:
    sys.exit(f"Error: '{path}' not found.")
return wild_types

def build_edges(lines, sources, file_wild, prop_wild):
    """
    Build channel edges based on allow rules, file_contexts and property_contexts.
    Returns list of tuples: (src, dst, kind).
    """
    write_file = defaultdict(set)
    read_file = defaultdict(set)
    write_prop = defaultdict(set)
    for line in lines:
        parsed = parse_allow_line(line)
        if not parsed:
            continue
        src, dst, cls, perms = parsed
        if src not in sources:
            continue
        # Handle file channels only for wildcard types
        if cls == "file" and dst in file_wild:
            groups = classify_file_edge(perms)
            if "WRITE_FILE" in groups:
                write_file[dst].add(src)
            if "READ_FILE" in groups:
                read_file[dst].add(src)

        # Handle property_service channels only for wildcard prefixes
        elif cls == "property_service" and dst in prop_wild:
            if "set" in perms:
                write_prop[dst].add(src)
    # Filter file targets: need write from one source and read from another
    file_targets = [
        t for t in write_file
        if t in read_file and len(write_file[t] | read_file[t]) >= 2
    ]
    # Filter property targets: write_prop plus implicit read_prop
    prop_targets = [
        t for t, writers in write_prop.items()
        if len(writers) >= 1 and len((sources - writers) | writers) >= 2
    ]
    edges = []
    # Add file edges
    for t in file_targets:
        for s in write_file[t]:
            edges.append((s, t, "WRITE_FILE"))
        for s in read_file[t]:
            edges.append((s, t, "READ_FILE"))
    # Add property edges
    for t in prop_targets:
        writers = write_prop[t]
        for s in writers:
            edges.append((s, t, "WRITE_PROP"))
        for r in sources - writers:
            edges.append((r, t, "READ_PROP"))
    return edges

def render_png(out_png, sources, edges):
    """
    Render a three-tier graph to PNG, placing file targets top,
    sources middle, and property targets bottom.
    """
    file_targets = {dst for _, dst, kind in edges if kind in ("WRITE_FILE", "READ_FILE")}
    prop_targets = {dst for _, dst, kind in edges if kind in ("WRITE_PROP", "READ_PROP")}
    dot = Digraph('SELinux_Channel', format='png')
    dot.attr(rankdir='TB')

```

```

dot.node_attr.update(shape='ellipse', fontname='Helvetica')
# Top tier: file targets
with dot.subgraph(name='cluster_file_targets') as c:
    c.attr(rank='max', style='invis')
    for t in sorted(file_targets):
        c.node(t)
# Middle tier: source domains
with dot.subgraph(name='cluster_sources') as c:
    c.attr(rank='same', style='invis')
    for s in sorted(sources):
        c.node(s)
# Bottom tier: property targets
with dot.subgraph(name='cluster_prop_targets') as c:
    c.attr(rank='min', style='invis')
    for t in sorted(prop_targets):
        c.node(t)
# Draw edges with solid for writes and dashed for reads
for src, dst, kind in edges:
    if kind in ("WRITE_FILE", "WRITE_PROP"):
        dot.edge(src, dst, style='solid', penwidth='2')
    else:
        dot.edge(dst, src, style='dashed')
png_data = dot.pipe()
with open(out_png, 'wb') as f:
    f.write(png_data)
print(f"PNG graph saved to '{out_png}'")
def main():
    if len(sys.argv) < 4:
        print(f"Usage: {sys.argv[0]} <allow.txt> <output.png> <SOURCE1> [SOURCE2 â€¦]")
        sys.exit(1)
    allow_file = sys.argv[1]
    out_png = sys.argv[2]
    sources = set(sys.argv[3:])
    try:
        lines = open(allow_file, encoding='ascii').read().splitlines()
    except Exception as e:
        sys.exit(f"Error reading '{allow_file}': {e}")
    file_wild = load_file_contexts('file_contexts')
    prop_wild = load_property_contexts('property_contexts')
    edges = build_edges(lines, sources, file_wild, prop_wild)
    if not edges:
        print("No valid channels found for given sources.")
        sys.exit(0)
    render_png(out_png, sources, edges)
if __name__ == "__main__":
    main()

```



ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ У КІБЕРБЕЗПЕЦІ

Адаптивна модель для виявлення шкідливих URL з використанням машинного навчання

Петро Венгерський¹, Володимир Лесик¹

¹ Львівський національний університет імені Івана Франка, вул. Університетська 1, 79000 Львів, Україна

Анотація

У роботі представлено підхід до створення високоточної адаптивної моделі для виявлення шкідливих URL-адрес із використанням технологій штучного інтелекту. Запропонована система базується на ансамблевих методах машинного навчання, що поєднують переваги кількох класифікаторів для підвищення точності та стійкості моделі. У ході експериментів випробувано дев'ять алгоритмів, серед яких 3 найкращі було об'єднано у один ансамбль. Використання ансамблю (VotingClassifier) забезпечило стабільне покращення точності класифікації. Отримані результати підтвердили ефективність ансамблевого підходу у виявленні зловмисних адрес, що має значну практичну цінність для підвищення рівня кібербезпеки та може бути інтегроване у сучасні системи захисту від фішингових чи інших атак.

Ключові слова

Класифікація, загроза, кібербезпека, URL, фішинг, ансамбль, адаптивна модель

1. Вступ

Зі зростанням кількості кіберзагроз, традиційні методи безпеки не завжди забезпечують ефективне виявлення нових атак. Це зумовлює необхідність впровадження автоматизованих підходів для аналізу URL-адрес. Шкідливі URL-адреси є поширеним інструментом для фішингових атак, чи інших типів зловмисних дій.

Інтелектуальний аналіз даних (Data Mining, DM) дозволяє автоматизувати процеси виявлення загроз через застосування алгоритмів машинного навчання. Аналіз текстових характеристик URL-адрес, таких як структура, довжина, доменні імена та наявність підозрілих символів, дозволяє створювати моделі, які можуть точно класифікувати ресурси як шкідливі або безпечні. Дослідження показали, що такі ознаки, як наявність спеціальних символів, використання субдоменів, довжина URL, кількість крапок у назві хоста та відсутність безпеки транспортного рівня, є сильними індикаторами фішингових URL-адрес. [1-7]

Метою цього дослідження є розробка методу аналізу URL-адрес з використанням алгоритмів DM, який забезпечить точне та швидке виявлення шкідливих ресурсів, тим самим підвищуючи рівень інформаційної безпеки.

2. Створення адаптивної системи для виявлення шкідливих URL

Адаптивна система виявлення шкідливих URL передбачає побудову моделі, яка забезпечує високу точність класифікації незалежно від джерела даних.

⁴⁸ Відповідні автори

✉ petro.venherskyy@lnu.edu.ua (P. Venherskyy); volodymyr.lesyk@lnu.edu.ua (V. Lesyk)

ORCID 0000-0000-0000-0000 (P. Venherskyy); 0000-0000-0000-0000 (V. Lesyk)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2.1. Методологія

У дослідженні був обраний підхід керованої класифікації, де модель навчається на розмічених даних. Набори даних у цьому дослідженні - списки з сотнями тисяч URL. Для класифікації було отримано ознаки безпосередньо з URL. Загалом, у дослідженні було використано дев'ять алгоритмів класифікації.

Після класифікації було обрано 3 найбільш ефективних алгоритми для певного набору даних, які були згодом об'єднані у один ансамблевий класифікатор методом голосування.

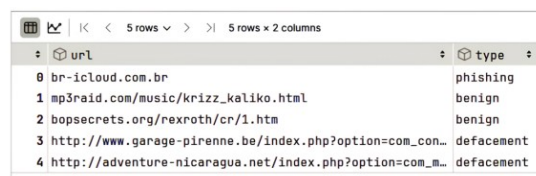
2.2. Використані алгоритми класифікації

Для дослідження було відібрано дев'ять алгоритмів класифікації:

5. Decision Tree Classifier, Random Forest Classifier, Extra Trees Classifier (деревоподібні методи); [8]
6. Bagging Classifier, AdaBoost Classifier (ансамблеві методи); [9, 10]
7. XGBClassifier, LGBMClassifier, CatBoostClassifier (градієнтний бустинг); [11]
8. KNeighborsClassifier (метод найближчих сусідів).

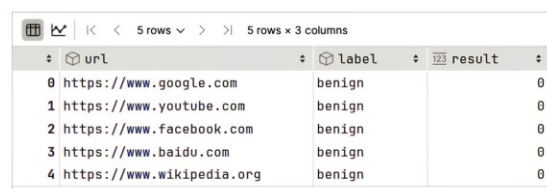
2.3. Підготовка даних та розробка ознак

У цьому дослідженні для аналізу URL використовувалися три незалежні набори даних, кожен зі своєю структурою та призначенням. Метою етапу підготовки даних було стандартизувати ці набори у єдиний формат для подальшого використання в моделях машинного навчання. Цей крок є критично важливим, оскільки якість та узгодженість вхідних даних безпосередньо впливає на ефективність алгоритмів класифікації. На рисунках 1 - 3 наведені приклади записів із обраних наборів даних:



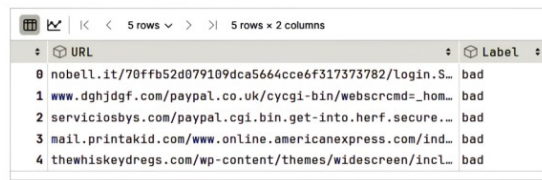
	url	type
0	br-icloud.com.br	phishing
1	mp3raid.com/music/krizz_kaliko.html	benign
2	bopsecrets.org/rexroth/cr/1.htm	benign
3	http://www.garage-pirenne.be/index.php?option=com_con...	defacement
4	http://adventure-nicaragua.net/index.php?option=com_m...	defacement

Рисунок 1: Приклади входжень для набору даних №1.



	url	label	result
0	https://www.google.com	benign	0
1	https://www.youtube.com	benign	0
2	https://www.facebook.com	benign	0
3	https://www.baidu.com	benign	0
4	https://www.wikipedia.org	benign	0

Рисунок 2: Приклади входжень для набору даних №2.



	URL	Label
0	nobell.it/70ffb52d079109dca5664cce6f317373782/login.S...	bad
1	www.dghjdgf.com/paypal.co.uk/cycgi-bin/webcmd=_hom...	bad
2	serviciosbys.com/paypal.cgi.bin.get-into.herf.secure...	bad
3	mail.printakid.com/www.online.americanexpress.com/ind...	bad
4	thewhiskeydregs.com/wp-content/themes/widescreen/incl...	bad

Рисунок 3: Приклади входжень для набору даних №3.

Кожен набір даних був перевірений на відсутні значення та аномалії. Усі записи з пропущеними значеннями були видалені, як і дублікати URL-адрес. Якщо у наборі даних переважав якийсь клас, його було нормалізовано.

Після очищення та нормалізації даних загальна кількість URL, використаних для навчання, становила: 453 088 (набір даних №1), 632 508 (набір даних №2) та 326 422 (набір даних №3).

Далі необхідно сформувані ознаки, які можна витягти безпосередньо з URL для подальшого навчання класифікаторів. На основі огляду попередніх досліджень та експериментального підходу був складений перелік ознак, які демонстрували високу загальну кореляцію з цільовою змінною, тобто з показником, що характеризує клас URL (безпечний чи шкідливий).

Ключові ознаки, які використовувалися для класифікації, включали:

- letters_count: кількість літер в URL.
- digits_count: кількість цифр в URL.
- special_chars_count: кількість спеціальних символів.
- shortened: індикатор скороченої URL.
- abnormal_url: ознака аномальності URL.
- secure_http: наявність протоколу HTTPS.
- root_domain: аналіз кореневого домену.
- rv: коефіцієнт репутації.
- isGlobal: індикатор належності до міжнародного домену верхнього рівня.

2.4. Навчання моделей та огляд результатів

Першим етапом процесу навчання моделей є розподіл даних на навчальний та тестовий набори. Використовується стандартний підхід, при якому дані діляться у співвідношенні 70:30, де 70% даних виділяється для навчання моделі, а 30% — для оцінки її ефективності на нових, раніше невідомих прикладах. Такий поділ дозволяє перевірити, чи здатна модель узагальнювати результати на нові дані, не використані під час навчання.

Після розподілу дані для навчання передаються раніше згаданим моделям: DecisionTreeClassifier, RandomForestClassifier, ExtraTreesClassifier, AdaBoostClassifier, BaggingClassifier, XGBClassifier, LGBMClassifier, CatBoostClassifier та KNeighborsClassifier.

Результати класифікації, отримані за допомогою цих алгоритмів, оцінювалися за метриками точності, повноти, позитивної точності та F1-міри. Кожен алгоритм показав різні рівні ефективності залежно від використаного набору даних, а результати могли змінюватися в залежності від типу шкідливого URL, характеристик набору даних та розподілу навчальних прикладів. Рисунки 4 - 6 ілюструють результати навчання моделей:

Classifier performance:					
	Classifier	Accuracy	Recall	Precision	F1-Score
6	cat_boost	0.911910	0.911910	0.913109	0.911641
7	extra_trees	0.911659	0.911659	0.912529	0.911433
1	random_forest	0.910636	0.910636	0.911509	0.910406
4	gradient_boost_xgb	0.909102	0.909102	0.910734	0.908768
8	bagging_classifier	0.904677	0.904677	0.905770	0.904393
5	gradient_boost_lgb	0.898502	0.898502	0.901637	0.897924
0	decision_tree	0.880574	0.880574	0.880602	0.880586
2	ada_boost	0.854003	0.854003	0.859233	0.852659
3	k_neighbors	0.833606	0.833606	0.833652	0.833626

Рисунок 4: Візуалізація результатів моделювання для набору даних №1

Classifier performance:					
	Classifier	Accuracy	Recall	Precision	F1-Score
6	cat_boost	0.979477	0.979477	0.979632	0.979468
5	gradient_boost_lgb	0.978022	0.978022	0.978296	0.978008
4	gradient_boost_xgb	0.977658	0.977658	0.977811	0.977648
2	ada_boost	0.973846	0.973846	0.974866	0.973807
7	extra_trees	0.973368	0.973368	0.974060	0.973337
1	random_forest	0.973222	0.973222	0.973604	0.973201
8	bagging_classifier	0.972555	0.972555	0.972884	0.972535
0	decision_tree	0.959369	0.959369	0.959390	0.959373
3	k_neighbors	0.590053	0.590053	0.598517	0.586598

Рисунок 5: Візуалізація результатів моделювання для набору даних №2

Classifier performance:					
	Classifier	Accuracy	Recall	Precision	F1-Score
6	cat_boost	0.841746	0.841746	0.841631	0.839531
7	extra_trees	0.841309	0.841309	0.840672	0.839563
1	random_forest	0.840166	0.840166	0.839396	0.838548
4	gradient_boost_xgb	0.837687	0.837687	0.837844	0.835107
5	gradient_boost_lgb	0.828718	0.828718	0.828480	0.826027
8	bagging_classifier	0.828317	0.828317	0.827796	0.825861
0	decision_tree	0.789519	0.789519	0.789509	0.789514
3	k_neighbors	0.775688	0.775688	0.774472	0.774926
2	ada_boost	0.772672	0.772672	0.776488	0.762897

Рисунок 6: Візуалізація результатів моделювання для набору даних №3

Отримані результати демонструють, що ансамблеві методи, такі як Extra Trees, Random Forest та Gradient Boosting, показали стабільну роботу, підтверджуючи їх ефективність при обробці складних структур даних. У той же час, метод k-найближчих сусідів продемонстрував значні труднощі на наборі даних №2, що свідчить про те, що методи, засновані на відстані, можуть бути недостатньо ефективними для високимірних або складних наборів даних.

2.5. Ансамблювання найефективніших моделей

Для подальшого підвищення ефективності був застосований VotingClassifier, який об'єднав прогнози трьох найкращих класифікаторів. Цей ансамбль працює за принципом жорсткого голосування, де кожен базовий класифікатор «голосує» за належність URL-адреси до одного з класів, а остаточне рішення приймається на основі більшості голосів. Такий підхід дозволив досягти невеликого, але стабільного покращення продуктивності, що підтверджує ефективність об'єднання сильних сторін різних алгоритмів. Після ансамблювання, фінальні метрики наведені на рисунках 7-9:

```
Voting Classifier Performance:  
Accuracy: 0.9153192725582476  
Recall: 0.9153192725582476  
Precision: 0.9163995346409288  
F1-Score: 0.9150787772206201
```

Рисунок 7: Візуалізація результатів моделювання VotingClassifier для набору даних №1

```
Voting Classifier Performance:  
Accuracy: 0.9787006212934664  
Recall: 0.9787006212934664  
Precision: 0.9789092319497146  
F1-Score: 0.9786891452385303
```

Рисунок 8: Візуалізація результатів моделювання VotingClassifier для набору даних №2

```
Voting Classifier Performance:  
Accuracy: 0.848297601776727  
Recall: 0.848297601776727  
Precision: 0.8479760919225958  
F1-Score: 0.84650278481803
```

Рисунок 9: Візуалізація результатів моделювання VotingClassifier для набору даних №3

VotingClassifier показав стабільні результати на різних наборах даних, досягнувши середнього покращення точності на 0,4%. Це підкреслює ефективність ансамблевого навчання у стабілізації прогнозів та компенсації слабких сторін окремих класифікаторів. Хоча приріст є відносно невеликим, він демонструє потенціал поєднання кількох моделей для отримання більш надійних результатів класифікації.

3. Висновки

Ключовим результатом дослідження є підтвердження, що використання ансамблевого підходу перевершує за ефективністю окремі класифікатори. Такий метод забезпечує високу точність без складної підготовки даних. Експерименти показали, що окремі алгоритми втрачають продуктивність на різних наборах даних, тоді як ансамбль, зокрема метод голосування, забезпечує стабільне зростання точності та надійності.

Використання генеративного ШІ

Під час підготовки цієї роботи був використаний ChatGPT (GPT-5) та Google Gemini (Gemini 2.5 Flash) для допомоги у структуризації тексту, покращенні граматики та стилю викладу. Жодні інструменти ШІ не використовувалися для створення зображень, даних або результатів.

Література

- [1] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, K. Kifayat, A comprehensive survey of AI-enabled phishing attack detection techniques, Telecommunication Systems 76, 2021, 1–16.
- [2] S. C. Jeeva, E. B. Rajsingh, Intelligent phishing URL detection using association rule mining, Human-centric Computing and Information Sciences 6(1), 2016, 1–19.

- [3] N. Koppula, V. P. Ganti, P. Gandhe, URL-Based Phishing Detection, *International Journal of Recent Technology and Engineering (IJRTE)* 9(1), 2020, 1872–1875.
- [4] В. А. Лахно, В. П. Малюков, І. В. Малюкова, О. Аتكелди, О. В. Криворучко, А. М. Десятко, К. В. Степашкіна, Модель аналізу стратегій при динамічній взаємодії учасників фішингових атак, *Кібербезпека: освіта, наука, техніка* 4(20), 2023, 1–15.
- [5] T. G. Dietterich, Ensemble methods in machine learning, in: *Multiple Classifier Systems*, Springer, Berlin, 2000, pp. 1–15.
- [6] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, 1998, 736 p.
- [7] J. Friedman, T. Hastie, R. Tibshirani, *The Elements of Statistical Learning*, Springer, New York, NY, 2001, 745 p.
- [8] IBM, Decision trees. URL: <https://www.ibm.com/topics/decision-trees>.
- [9] L. Nautiyal, M. Ram, P. Malik, *Machine Learning for Cyber Security*, Springer, Singapore, 2022, 250 p.
- [10] E. Bertino, W. Susilo, X. Chen, *Cyber Security Meets Machine Learning*, Springer Nature, Singapore, 2022, 163 p.
- [11] J. H. Friedman, Greedy function approximation: A gradient boosting machine, *Annals of Statistics* 29(5), 2001, 1189–1232.

Застосування ПЛІС для захисту каналу передачі відеоінформації з БпЛА з використанням штучного інтелекту

Сергій Гільгурт⁴⁹, Іван Яблоков²

¹ Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, вул. Олега Мудрака 15, 03164, Київ, Україна

² Донецький національний технічний університет, вул. Шевченка 9, м. Дрогобич, Львівська обл., 82100, Україна

Анотація

В дослідженні розглядається проблема створення системи передачі відеопотоку з бортової камери FPV-дрона-камікадзе на наземну базову станцію керування, яка забезпечує захист інформації з одночасним запобіганням впливу зовнішніх радіоперешкод. Проаналізовані та сформульовані технічні вимоги до такої системи. Розглянута її структурна схема. В якості апаратної платформи обґрунтовано обрано ПЛІС. Запропоновано підхід до організації процесу розробки проєктів на ПЛІС, який дозволяє задіяти технології штучного інтелекту для підвищення ефективності процесу створення таких систем та покращення їх характеристик.

Ключові слова

FPV-дрон-камікадзе, передача відеоданих, захист інформації, завадостійкість, штучний інтелект

1. Вступ

FPV-дрони-камікадзе (від англ. First Person View – вид від першої особи) стали одним із наймасовіших засобів на сучасному полі бою. Вони дозволяють операторам завдавати високоточних ударів, скеровуючи безпілотний літальний апарат (БпЛА) з вибуховим зарядом у ціль або здійснювати розвідку без ризику для життя оператора.

Однією з найважливіших проблем, що виникають при використанні таких БпЛА, пов'язана із вразливістю радіоканалу передачі відео зображення з камери FPV-дрона на наземну базову станцію. Перехоплення цієї інформації, по-перше, наражає на небезпеку оператора, оскільки ворог може виявити его розташування. По-друге, перехоплюючи відеозображення противник здатен виявити БпЛА, який прямує безпосередньо на нього, та встигнути вжити заходів, які дозволять йому запобігти атаці. Іншою важливою проблемою є вплив на відеоканал ворожих засобів РЕБ. Це робить актуальним завдання підвищення завадостійкості каналів передачі відео. Отже, в умовах сучасної війни виникає нагальна потреба у компактних та малопотужних системах захищеного зв'язку для БпЛА, насамперед, для FPV-дронів-камікадзе.

Метою дослідження є пошук апаратної бази для таких систем та методів покращення їх характеристик, а також підвищення ефективності процесів їх розробки.

⁴⁹ Відповідальний автор

✉ serhii.hilgurt@pimee.ua (С. Гільгурт); ivan.yablokov@donntu.edu.ua (І. Яблоков)

🆔 0000-0003-1647-1790 (С. Гільгурт); 0009-0009-7424-1249 (І. Яблоков)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Аналіз наявної інформації

Для захисту каналу передачі відеоінформації зазвичай використовують різні види шифрування [1-4]. Для захисту від засобів РЕБ запропоновані численні методи підвищення завадостійкості шляхом введення надлишковості [5-7]. Але вони зазвичай є технічно складними та призводять до неприйняттого погіршення масо-габаритних та вартісних показників бортового електронного обладнання.

Аналіз відомих літературних джерел з одного боку [8, 9], а з іншого – спілкування з досвідченими фахівцями в даній відносно новій та динамічно мінливій галузі дозволяють сформулювати наступні основні вимоги та обмеження, які накладаються в реальних бойових умовах на тракт передачі відеоданих з бортової відеокамери на станцію керування. Це – насамперед аналоговий спосіб передачі даних, відомий як більш завадостійкий у порівнянні з цифровим представленням сигналу. Час затримки передачі інформації вздовж всього каналу теоретично повинен мати порядок 100...20 мс. Але, як свідчить практика застосування FPV-дронів-камікадзе, реальне значення цього параметра повинно бути наближеним до значення 30 мс. Пропускна здатність тракту має забезпечувати передачу не менш, ніж 30 кадрів/сек. Припустимий рівень радіоперешкод, при якому система повинна забезпечувати оператору хоча б мінімальну зрозумілість зображення – до 80%. До стійкості шифрування вимоги висуваються відносно помірні, але час, потрібний злоумиснику для злому системи захисту з використанням потужних обчислювальних засобів, не повинен бути меншим за одиниці годин. На вартість бортової частини системи одноразових FPV-дрона-камікадзе також накладаються суттєві обмеження – бажано, щоб вона варіювалася в межах пари сотень доларів США у еквіваленті.

За таких жорстких умов навіть давно відомі та пропрацьовані технічні задачі набувають нового рівня складності, що потребує по-новому підходити до пошуку як апаратної платформи для технічних рішень, так і методів їх ефективної реалізації.

3. Вирішення проблеми

Проведений аналіз вимог призводить до наступного рішення. На рис. 1 наведена спрощена структурна схема системи передачі відеоданих. Сигнал з аналогової відеокамери подається на блок кодування, розташований на борту FPV-дрона. Там за допомогою апаратного відеокодера він перетворюється у цифрову форму, отримує потрібних перетворень та за допомогою також апаратного відеокодера повертається аналогову форму. Відеопередавач шляхом випромінювання електромагнітних хвиль надсилає модульований сигнал з блока кодування в бік оператора.

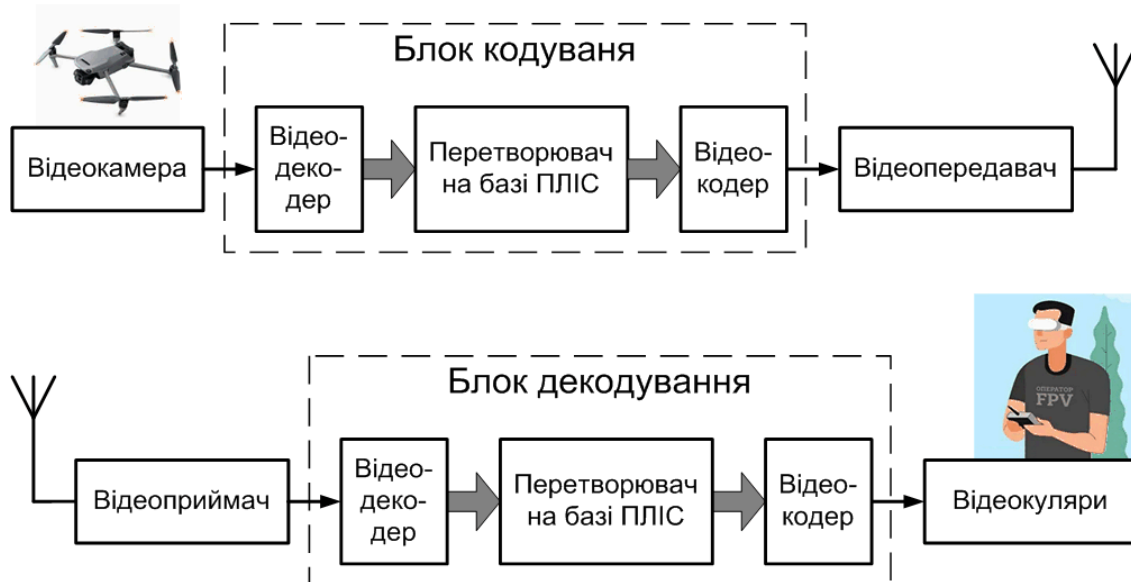


Рисунок 1: Структурна схема передача відеоінформації з БпЛА на наземну базову станцію.

На наземній станції радіосигнал від БпЛА приймається відеоприймачем та передається на блок декодування наземної станції. В ньому за допомогою апаратного відеодекодера він знов перетворюється у цифрову форму, а після того, як відповідний перетворювач здійснить потрібну зворотну трансформацію відеосигналу, за допомогою апаратного відеокодера повертається в аналогову форму.

Вибір ПЛІС для реалізації бортового та наземного перетворювачів обумовлений тим, що, як свідчить проведений в даному дослідженні порівняльний аналіз різних платформ, тільки програмована логіка здатна вирішити потрібну обчислювальну задачу з обробки відеосигналу БпЛА за прийнятний час. Відносно невелике енергоспоживання дозволяють забезпечити, наприклад, такі сімейства ПЛІС, як IGLOO від Microchip [10], MAX 10 від Altera (Intel) [11], ECP5 від Lattice Semiconductor [12], або Spartan-7 від Xilinx (AMD) [13].

4. Використання технологій штучного інтелекту

Для подальшої підвищення ефективності отриманого технічного рішення може бути задіяний підхід, що базується на використанні методів та технологій штучного інтелекту (ШІ). Дана галузь активно розвивається останнім часом і демонструє вражаючі результати в самих різних сферах застосування.

Однією з важливих перешкод при впровадженні відомих інструментів ШІ, таких як штучні нейронні мережі та машинне навчання, є складність їх інтеграції у доволі консервативні засоби розробки проєктів для ПЛІС, насамперед фірмових САПР від розробника мікросхем програмованої логіки. Сутність підходу, що пропонується в даному дослідженні, полягає у використанні одного з внутрішніх інструментальних засобів таких САПР – скриптової мови програмування Tool Command Language (TCL) [14, 15]. Даний підхід дозволяє пакет від фірми Xilinx для розробки складних цифрових схем на ПЛІС Vivado Design Suite зв'язати з веб-інтерактивним обчислювальним засобом Jupyter Notebook. При такому підході цифрова схема створюється в середовищі Jupyter Notebook, а пакет Vivado Design Suite використовується як приєднаний зовнішній компонент. Це дозволяє під

час тестування схеми створювати метрики, які можна проаналізувати і вибрати найбільш прийнятну схему для подальшої компіляції апаратного опису цифрової схеми.

Підхід дозволяє використовувати потужні можливості методів ШІ для оптимізації цифрової структури перетворювачів бортової підсистеми та наземної станції [16]. Штучна нейронна мережа використовуватиметься для вибору найкращого з чисельних варіантів та комбінацій компонентів можливих технічних рішень найкращу комбінацію в сенсі мінімізації апаратних витрат, підвищення швидкодії та зменшення енергоспоживання.

5. Висновки

В даному дослідженні проаналізована задача створення системи передачі відеоінформації з БпЛА типу FPV-дрон-камікадзе на наземну базову станцію, яка відповідає жорстким та суперечливим вимогам, що висуваються до такої системи умовами бойового застосування. Розглянута структура системи, обрана апаратна платформа для її реалізації. Для підвищення ефективності процесів їх розробки запропонований підхід на базі скриптової мови програмування Tool Command Language, який дозволяє інтегрувати засоби синтезу цифрових схем на ПЛС у зручне інтерактивне середовище Jupyter Notebook, в якому доступні відомі та відпрацьовані засоби використання технологій штучного інтелекту.

Фінансування

Дослідження виконане за фінансової підтримки з боку Національного фонду досліджень України, проєкт 2025.6/0109 "Захищена завадостійка система передачі відеоінформації з безпілотного літального апарата", номер державної реєстрації 0125U003164.

Декларація про генеративний ШІ

Автори не використовували жодних інструментів генеративного штучного інтелекту.

Список літератури

- [1] D. Psilias, A. Milidonis, I. Voyiatzis, Secure video transmission system for UAV applications, in: ACM international conference proceeding series: 25th Pan-Hellenic Conference on Informatics, PCI '21, ACM Press, 26–28 November 2021, New York, NY, USA, pp. 318–322. doi:10.1145/3503823.3503882.
- [2] S. Singh, P. K. Sharma, S. Y. Moon, J. H. Park, Advanced lightweight encryption algorithms for IoT devices: survey, challenges and solutions. Journal of Ambient Intelligence and Humanized Computing 15 (2024) 1625–1642. doi:10.1007/s12652-017-0494-4
- [3] S. Samanth, K. V. Prema, M. Balachandra, J. Chen, Security In Internet of Drones: A Comprehensive Review, Cogent Engineering (2022) 9:1, 2029080. doi:10.1080/23311916.2022.2029080.
- [4] G. B. Vallevik, V. Korsgaard, T. G. Lien, Encryption and decryption of media data. European Patent No. EP3695560B1, Filed Oct. 12th, 2018, Issued Oct. 9th, 2024.
- [5] Defense News Aerospace, Focus: Why jamming isn't a solution against FPV anymore in Ukraine's offensive initiative, August 14th, 2024. URL: <https://armyrecognition.com/news/aerospace-news/2024/focus-why-jamming-isnt-a-solution-against-fpv-anymore-in-ukraines-offensive-initiative>.

- [6] N. Cecchinato, A. Toma, C. Drioli, G. Oliva, G. Sechi, G. L. Foresti, Secure Real-Time Multimedia Data Transmission from Low-Cost UAVs with A Lightweight AES Encryption. *IEEE Communications Magazine* vol. 61, no. 5 (2023) 160–165. doi:10.1109/MCOM.001.2200611.
- [7] L. J. Ziemba, D. J. Ziemba, T. J. Sinatra, Z. Gao, Drone defense system. Patent No. US10339818B2, Filed Nov. 22nd, 2016, Issued May 25th, 2017.
- [8] J. Cox, K. Wong, Predictive feedback augmentation for manual control of an unmanned aerial vehicle with latency. *International Journal of Micro Air Vehicles* 11 (2019). doi:10.1177/1756829319869645.
- [9] S. B. Kamtam, Q. Lu, F. Bouali, O. C. L. Haas, S. Birrell, Network Latency in Teleoperation of Connected and Autonomous Vehicles: A Review of Trends, Challenges, and Mitigation Strategies. *Sensors* 24(12) (2024) 3957. doi:10.3390/s24123957.
- [10] Microchip. IGLOO® FPGAs. URL: <https://www.microchip.com/en-us/products/fpgas-and-plds/fpgas/igloo-fpgas>.
- [11] Altea. MAX® 10 FPGA. URL: <https://www.altera.com/products/fpga/max/10>.
- [12] Lattice. ECP5 / ECP5-5G. Break the rules of power, size and cost in your connectivity and acceleration applications. URL: <https://www.latticesemi.com/Products/FPGAandCPLD/ECP5>.
- [13] AMD Spartan™ 7 FPGAs. I/O Optimization with High Performance-per-Watt. URL: <https://www.amd.com/en/products/adaptive-socs-and-fpgas/fpga/spartan-7.html>.
- [14] AMD, Tcl Scripting in Vivado, 2025. URL: <https://docs.amd.com/r/en-US/ug894-vivado-tcl-scripting/Tcl-Scripting-in-Vivado>.
- [15] Vivado Design Suite User Guide. Using Tcl Scripting. UG894 (v2022.1) June 8, 2022. URL: https://www.xilinx.com/support/documents/sw_manuals/xilinx2022_2/ug894-vivado-tcl-scripting.pdf.
- [16] І. І. Яблоков, Автоматизація процесу розробки конфігурацій для ПЛІС інструментами Xilinx, у: Проблеми інформатики та моделювання. Тези двадцять п'ятої міжнародної науково-технічної конференції, ПІМ '25, НТУ «ХПІ», 25–28 вересня 2025, Харків, Україна, С. 126. URL: https://web.kpi.kharkov.ua/pim/wp-content/uploads/sites/248/2025/09/Tezy_PIM_2025.pdf.

Застосування штучного інтелекту для виявлення кіберзагроз

Олексій Залізний⁵⁰, Владислав Токарев¹

¹Державний університет інформаційно-комунікаційних технологій, Солом'янська 7, 03110 Київ, Україна

Анотація

Розглянуто застосування штучного інтелекту для виявлення та прогнозування кіберзагроз. Показано переваги використання машинного навчання у підвищенні ефективності систем кібербезпеки та зменшенні кількості хибних сповіщень.

Ключові слова

штучний інтелект, кібербезпека, машинне навчання, виявлення загроз.

1. Вступ

У сфері кібербезпеки штучний інтелект (ШІ) використовується для виявлення, запобігання та нейтралізації кіберзагроз, зокрема тих, які складно розпізнати традиційними методами захисту. Основна перевага ШІ полягає у можливості обробки великих обсягів даних у реальному часі та виявлення закономірностей, які недоступні людині або класичним алгоритмам. Завдяки цьому системи на базі ШІ можуть не лише реагувати на відомі типи атак, а й прогнозувати появу нових загроз, підвищуючи рівень кіберзахисту та ефективність реагування на інциденти.

2. Постановка проблеми

У сучасних умовах швидкого зростання обсягу кіберзагроз традиційні методи захисту інформаційних систем стають малоефективними. Збільшення складності атак, їхня автоматизація та варіативність вимагають нових підходів до виявлення та реагування на інциденти безпеки. Тому актуальною є проблема розроблення та застосування систем на основі штучного інтелекту, здатних ефективно аналізувати великі обсяги даних, виявляти аномалії та прогнозувати потенційні кіберзагрози в реальному часі.

3. Мета дослідження

Метою дослідження є аналіз та обґрунтування ефективності застосування методів штучного інтелекту для виявлення, класифікації та прогнозування кіберзагроз з метою підвищення рівня кібербезпеки інформаційних систем і забезпечення своєчасного реагування на потенційні атаки.

4. Результати дослідження

Штучний інтелект уже демонструє значний потенціал у сфері кібербезпеки, забезпечуючи нові підходи до виявлення, аналізу та прогнозування кіберзагроз.

⁵⁰Відповідальний автор

✉ olexiyzal@gmail.com (О. Залізний); vlad.toarev0000@gmail.com (В. Токарев)

ID 0009-0004-8173-2461 (О. Залізний); 0009-0005-7374-841X (В. Токарев)

Використання алгоритмів машинного навчання та глибоких нейронних мереж дозволяє автоматизувати процеси обробки великих обсягів даних із логів, системних звітів і сповіщень безпеки. Це дає змогу швидше виявляти підозрілу активність, аномалії та потенційні уразливості, що знижує ризик успішних атак і втрат даних.

Одним із ключових напрямів застосування ШІ у кібербезпеці є виявлення вторгнень. Алгоритми штучного інтелекту аналізують поведінку користувачів і систем, виявляючи спроби несанкціонованого доступу. Використовуються як сигнатурні методи, що порівнюють дані з базами відомих атак, так і евристичні — спрямовані на виявлення нетипових дій, які можуть свідчити про кібератаку. Завдяки цьому моніторинг безпеки може здійснюватися в реальному часі з можливістю автоматичного реагування на загрози.

Не менш важливим напрямом є моніторинг і аналіз мережевого трафіку. ШІ аналізує обмін даними між пристроями, визначає підозрілі взаємодії та виявляє потенційні витoki інформації. Системи, що навчаються на реальних даних, здатні адаптуватися до нових типів атак і вдосконалювати свої моделі, підвищуючи точність виявлення загроз.

Значну роль відіграє також зменшення кількості хибних тривог. У традиційних системах безпеки надлишок хибних сповіщень створює перевантаження для аналітиків, тоді як ШІ здатен розрізняти безпечні події від справжніх загроз, використовуючи контекстний аналіз і зворотний зв'язок від фахівців. Це підвищує ефективність роботи команд кібербезпеки й дозволяє зосередитися на реальних інцидентах.

Штучний інтелект також активно застосовується для оптимізації та оновлення ІТ-інфраструктури. Аналізуючи конфігурації, налаштування й сумісність систем, ШІ виявляє потенційні помилки та вразливості ще до впровадження змін. Це скорочує час на тестування, знижує ризики збоїв і забезпечує стабільність роботи мережі.

Ще одним перспективним напрямом є прогнозування загроз. На основі даних про минулі атаки, уразливості та тенденції розвитку кіберзлочинності ШІ формує профілі потенційних загроз і прогнозує майбутні ризики. Це дозволяє компаніям створювати проактивні стратегії захисту, запобігаючи атакам до їхнього здійснення.

Загалом, впровадження систем штучного інтелекту в галузі кібербезпеки суттєво змінює підхід до захисту інформації. Завдяки здатності аналізувати великі обсяги даних, адаптуватися до нових типів атак і прогнозувати поведінку зловмисників, ШІ робить кіберзахист більш гнучким, швидким і ефективним. Це дозволяє організаціям мінімізувати ризики, підвищити рівень безпеки цифрових систем та зменшити залежність від ручного адміністрування безпеки.

5. Висновки та перспективи

Результати дослідження свідчать, що штучний інтелект має значний вплив на підвищення ефективності систем кібербезпеки. Його застосування дозволяє автоматизувати процеси виявлення, аналізу та прогнозування кіберзагроз, забезпечуючи швидке реагування на інциденти та зменшення кількості хибних сповіщень. Завдяки здатності ШІ аналізувати великі обсяги даних, розпізнавати аномалії та адаптуватися до нових типів атак, системи кіберзахисту стають більш гнучкими, точними та проактивними.

Перспективи розвитку пов'язані з удосконаленням алгоритмів машинного навчання та нейромереж, а також із вирішенням питань етики, конфіденційності й безпеки даних.

Декларація про генеративний ШІ

Під час підготовки цієї роботи автори використовували GPT-5: Перевірка граматики та орфографії. Після використання цих інструментів/сервісів автори переглянули і відредагували вміст за необхідності і несуть повну відповідальність за зміст публікації.

Література

- [1] Sommer, R., & Paxson, V. (2010). Outside the Closed World: On Using Machine Learning For Network Intrusion Detection. IEEE Symposium on Security and Privacy, 305-316.
- [2] Buczak, A. L., & Guven, E. (2016). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.
- [3] Kim, G., Lee, S., & Kim, S. (2014). A Novel Hybrid Intrusion Detection Method Integrating Anomaly Detection with Misuse Detection. Expert Systems with Applications, 41(4), 1690-1700.

Мультирепрезентаційна GNN-модель з узгодженням і адаптивним злиттям для детекції спаму

Ірина Замрій¹, Іван Шахматов²

¹ Державний університет інформаційно-комунікаційних технологій, Київ, Україна

² Державний університет інформаційно-комунікаційних технологій, Київ, Україна

Анотація

Веб-форми залишаються однією з головних цілей спам-кампаній, де координовані боти поєднують масові відправлення, повторне використання полів, часові сплески та технічне маскування, а обсяг розмітки обмежений і діють вимоги приватності. Пропонується модель контрастивної графової нейромережі на основі кількох представлень для фільтрації веб-спаму. Модель об'єднує три представлення подій. Представлення відправок описує зв'язки між заявками за дублюванням полів, часовою близькістю та контактами. Поведінкове представлення є гетерографом взаємодій користувачів, форм і сторінок разом із технічними атрибутами. Семантичне представлення будується як kNN-граф за векторними поданнями тексту та URL. Для кожного представлення застосовуються окремі енкодери GNN з використанням R-GCN для типізованих відносин та GAT або GCN для семантичних і зв'язків відправок. Далі виконується адаптивне злиття представлень, що автоматично зміщує вагу між поведінковими та контентними сигналами на рівні окремої відправки. Навчання поєднує бінарну перехресну ентропію на позначених прикладах і контрастивне узгодження між представленнями з аугментаціями типу drop-edge, feature-mask і стохастичний kNN, а також легку регуляризацию для стабільності. Підхід орієнтовано на практичне розгортання з часовими сплітами без витоків, інтерпретованістю через ваги злиття та механізми уваги, анонімізацією чутливих атрибутів, масштабованим оновленням графів і контролем балансу помилкових спрацьовувань і пропусків відповідно до бізнес-метрик.

Ключові слова

Штучний інтелект, кібербезпека, веб-спам, вебзастосунки, виявлення спаму, графові нейромережі (GNN), сканери вразливостей, інформаційна безпека.

1. Вступ

Веб-спам уже давно не зводиться до «поганих слів» у тексті. Сьогодні це скоординовані кампанії: масові надсилання форм, повторне використання шаблонів, одночасні сплески у часі, однакові технічні ознаки пристроїв. Класичні контентні фільтри швидко застарівають, великі мовні моделі потребують багато позначених прикладів, а прості графові підходи зазвичай дивляться лише на один тип зв'язків. Потрібна модель, що поєднує поведінкові й текстові сигнали, добре працює, коли мало міток, і залишається стабільною, коли тактики спаму змінюються.

Існуючі дослідження охоплюють контентно-семантичні підходи, де текст і посилання перетворюють на граф або використовують сучасні мовні моделі (BERT, ELMo). Це дає помітний приріст, але майже не враховує поведінку користувачів і не узгоджує різні джерела сигналів [1,3,14]. У гетерогенних/багатовідносних графах для ботів, шахрайства та змовних груп важливими є типи зв'язків, уважний відбір корисних сусідів і методи, що враховують відмінність класів у пов'язаних елементах. Також все ще лишаються проблеми зі стійкістю між платформами та поєднанням із контентом [2,4,7,8,11,13]. Напрямок контрастивного та попереднього навчання зменшує потребу в розмітці, зіставляючи різні варіанти графа або навчаючись на парах графів, але зазвичай працює в межах одного типу графа без об'єднання кількох представлень на дані [6,9,10]. Отже, бракує рішення, яке одночасно поєднує кілька

подань (користувачі/відправки/зміст) і узгоджує їх між собою, дане дослідження направлено на створення нової моделі Multi-View Contrastive GNN яка закриє ці прогалини.

Графове подання природне для цього завдання, так як користувачі взаємодіють із формами та сторінками, відправки утворюють групові сплески, тексти й URL мають смислову схожість. Важливо не обмежуватися одним графом. Різні подання (views) відкривають різні закономірності: у поданні користувача видно шаблони взаємодій і спільні технічні ознаки, у поданні відправок прослідковуються дублікати полів і часові координатії, у семантичному є схожість контенту та посилань. Завдання дослідження — узгодити ці сигнали й отримати з них стійкі ознаки.

У дослідженні пропонується модель із кількома поданнями та контрастивним попереднім навчанням. Кожне подання кодується окремою графовою неймережею (R-GCN для типізованих взаємодій, GAT/GCN для семантичних і зв'язків відправок). Усередині подань виділяються сусідство, між поданнями та гнучко змішуються сигнали, щоб автоматично надавати більше ваги найкориснішому для конкретної відправки. Попереднє контрастивне навчання узгоджує ознаки різних подань і зменшує залежність від великої розмітки. Підсумкове рішення це двокласова класифікація відправок із налаштовуваним порогом під потрібні бізнес-метрики.

В роботі спершу формується практична схема (Рис. 1) з кількома поданнями веб-взаємодій (user, submission, semantic) і набором ознак, що зберігають приватність. Далі будується архітектуру, яка поєднує обробку всередині кожного подання та узгодження між поданнями з гнучким змішуванням сигналів. Поєднується контрастивне попереднє навчання між поданнями з навчанням на мітках, щоб зменшити потребу у великих наборах даних і підвищити переносимість моделі. На додачу надаємо відтворюваний процес обробки на Python і план оцінювання з розділенням даних за часом, сильними базовими порівняннями та експериментами з вимкненням окремих частин моделі. Дослідження орієнтоване на практику враховується потік даних, вимоги до швидкості, потребу мінімізувати зайві спрацювання та питання етики (анонімізація персональних даних, додаткові перевірки для рішень із високою впевненістю).

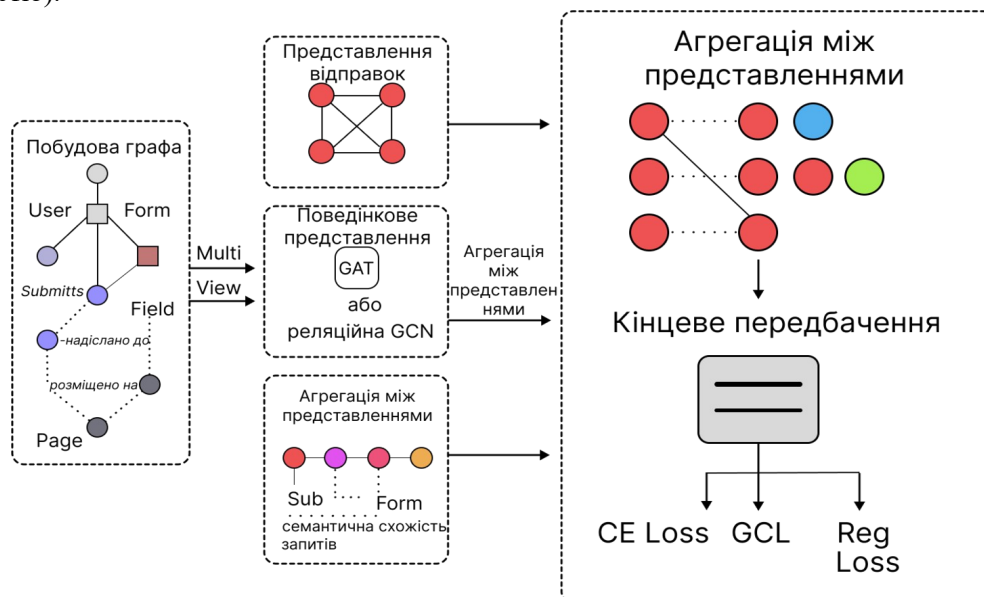


Рис. 1 Багатопредставленнева графова модель фільтрації веб-спаму (з контрастивним навчанням)

2. Модель

Маємо множину підграфів (view) $G = \{G^{(v)}\}_{v \in \{s, u, t\}}$, де кожне $G^{(v)} = (V^{(v)}, E^{(v)})$, тут: $v = s$ — submission-view (вузли: відправки форм), $v = u$ — user-view (гетерогенний граф користувачів/форм/сторінок із типами ребер), $v = t$ — semantic/text-view (граф kNN за семантичною близькістю тексту/URL). Запропонована мультипредставленнева контрастивна GNN, що оперує трьома представленнями $\{s, u, t\}$.

Щоб кодувати дані, для кожного v маємо матрицю ознак вузлів $X^{(v)} \in R^{|V^{(v)}| \times d_v}$, (де d_v — розмірність ознак у представленні v) та матрицю суміжності: $A^{(v)} \in \{0, 1\}^{|V^{(v)}| \times |V^{(v)}|}$. У поведінковому представленні v є типи відносин $R^{(u)} = \{submits, shares\_device, shares\_IP, located\_on\}$, тому суміжність розкладаємо на $\{A_r^{(u)}\}_{r \in R^{(u)}}$ для подальшої R-GCN-агрегації. Супервізія здійснюється над підмножиною відправок $L \subseteq V^{(s)}$ з мітками $y_i \in \{0, 1\}$. Для локального контексту позначимо $N_i^{(v)}$ — сусідство вузла i (включно із self-loop) у представленні v . Архітектуру описуємо пошарово: $l = 0, \dots, L-1$, перехід $d_l \rightarrow d_{l+1}$ параметризується матрицями $W^{(v, l)} \in R^{d_{l+1} \times d_l}$; у R-GCN додатково використовуємо $W_r^{(l)}$ (для кожного типу ребра) та $W_0^{(l)}$ (self-loop).

Цільові мітки задані для підмножини відправок: $L \subseteq V^{(s)}, y_i \in \{0, 1\}$, шукаємо $f_\theta: V^{(s)} \rightarrow [0, 1]$ з $\hat{y}_i = f_\theta(i)$ — імовірність спаму для submission-вузла i використовуючи структуру та ознаки всіх представлень. Для вирівнювання між представленнями існують якірні відповідності для одних і тих самих сутностей (відправка i у s має відображення на відповідний вузол/ембеддинг у t та суміжні сутності у u); ці відповідності використовуються надалі в контрастивному навчанні.

Часова організація даних: кожен вузол/ребро має часову мітку τ ; тренування/валідація/тест відокремлюються за зрізами часу ($T_{train} < T_{val} < T_{test}$) для уникнення витоку. В онлайні граф будується у ковзному вікні ширини W днів (де W — фіксований гіперпараметр).

Виходячи з того, що мітки можуть бути неповними й шумними: спаму мало, частина позначень помилкова або відсутня, ознаки теж можуть спотворюватися: тексти маскують, IP та User-Agent часто спільні або підмінені (через NAT, VPN, емулятори). Відповідності між різними поданнями не завжди повні: для частини відправок немає коректної «прив'язки» у всіх поданнях, тому допускаємо пропуски. Розподіли ознак і зв'язків змінюються з часом, тож оцінювання проводимо у хронологічному розрізі (розділення за часом). Для приватності ідентифікатори (e-mail, телефон, IP) анонімізуємо: хешуємо, узагальнюємо IP до рівня /24 тощо, зберігаємо лише агреговані або хешовані атрибути. Дані великі й розріджені, тому навчаємося пакетно, з вибіркою сусідів. У графі користувачів і у графі відправок зв'язки спрямовані, а в семантичному графі їх зазвичай робимо двосторонніми (kNN) для стабільності. Вартість помилок у продакшені несиметрична: хибні спрацювання й пропуски коштують по-різному, тож поріг класифікації налаштовуємо під потрібні бізнес-метрики.

Є три подання даних (Рис. 1): відправки (s), користувачі/технічні ознаки (u) та зміст — тексти/URL (t). У t -поданні вузлами також є відправки, які з'єднано kNN-ребрами за схожістю їхніх векторів тексту/URL. Для кожного подання використовується простий графовий шар: для s і t — стандартні GAT/GCN, для u — шар із типами зв'язків (аналог R-GCN). У середині кожного подання виконуємо агрегування сусідів (GAT/GCN для s і t ; R-GCN для u). Далі три вектори зливаються із маскою на відсутні подання та адаптивними вагами, щоб модель надавала

більшу вагу найінформативнішому сигналу. Об'єднаний вектор подається до простого класифікатора, який видає ймовірність спаму. Навчання моделі поєднує звичайну бінарну ціль на мічених прикладах і контрастивне узгодження між поданнями; це зменшує потребу у великій розмітці та підвищує стійкість.

Після інтра-агрегації для кожної відправки i маємо верхньощарові вектори $\overline{h_i^{(v)}}$ з представлень $v \in \{s, u, t\}$. Далі працює модуль злиття, який за допомогою навчуваних параметрів обчислює вузло-залежні ваги $\gamma_i^{(v)}$. Ці ваги інтерпретуються як внесок відповідного представлення у рішення та нормуються по доступних каналах; відсутність каналу фіксуємо бінарною маскою $m_i^{(v)} \in \{0, 1\}$ (0 — канал відсутній, 1 — присутній), тож злиття враховує лише ті $\overline{h_i^{(v)}}$, для яких $m_i^{(v)} = 1$. Отримане підсумкове подання позначаємо z_i ; це зважена (через $\gamma_i^{(v)}$ та $m_i^{(v)}$) комбінація верхньощарових векторів із внутрішньою нелінійністю модуля злиття. Далі класифікатор із параметрами (w, b) перетворює z_i на $\hat{y}_i \in [0, 1]$ — оцінку ймовірності спаму. Завдяки $\gamma_i^{(v)}$ модуль злиття є інтерпретованим: для кожної відправки видно, з якого каналу (поведінкового u , семантичного t чи відправок s) модель бере основний сигнал (найбільше $\gamma_i^{(v)}$). Навчання поєднує бінарну ціль на мічених вузлах $\setminus(L_C \setminus)$ та контрастивне узгодження між представленнями $\setminus(L_{GGL} \setminus)$, яке зближує подання тієї самої відправки і відштовхує випадкові пари, що зменшує залежність від обсягу розмітки та підвищує робастність до дрифтів. Для ефективності застосовуємо семплінг сусідства, а семантичні зв'язки у t -представленні попередньо формуємо через kNN-індексування текстів/URL.

Для submission-представлення використовуються числові ознаки форми: довжина тексту, символна ентропія, частка цифр, глибина/довжина URL, година доби та показник сплеску активності (кількість відправок у вікні Δt). Для semantic/text-представлення вузлами також є відправки, а ознаками — попередньо обчислені векторні подання тексту/URL (напр., 256-вимірні sentence embeddings); суміжність утворюємо методом kNN за косинусною подібністю. Для behavior/user-представлення формуємо гетерограф: для вузлів типу user/device/ip/page беремо агреговані показники активності (частоти дій, інтервали, різноманітність сторінок, кількість зв'язків), категоріальні ознаки подаємо у вигляді векторних подань фіксованої розмірності; суміжність розкладаємо за типами відносин (submits, shares_device, shares_IP, located_on). Навчання та оцінювання відбуваються у часових сплітах (Train/Val/Test), усі перетворення (скейлінг, PCA, kNN-індекси) налаштовуються лише на Train; через дисбаланс класів основною метрикою є PR-AUC, додатково звітуємо F1@ τ (τ обирається на Val) та Recall@FP-rate при 10^{-3} і 10^{-4} . Порівнюється модель з поширеними базовими підходами (TF-IDF+LR, Char+XGB, BERT-CLS, одно-представленнєві GNN); окремо оцінюється внесок компонентів шляхом послідовного вимкнення окремих представлень, модуля GCL та застосування альтернативних схем злиття й аугментацій. Додатково метод зіставляється з двома простими правилами — фільтром за списком ключових слів та фіксованим лімітом на кількість відправок з одного IP/пристрою за інтервал часу — і звітуються значення PR-AUC, F1 за порогом, визначеним на валідації, а також Recall за фіксованих рівнів FP 10^{-3} та 10^{-4} з 95% довірчими інтервалами на однакових часових сплітах.

3. Висновки

Запропонована модель, як спосіб фільтрувати веб-спам на основі графів. Задача в тому, що дані розглядаються з кількох боків — відправки форм, поведінка користувачів і зміст

текстів/посилань — а модель попередньо навчається так, щоб ці джерела добре узгоджувалися між собою. Для кожного боку використовується окрема модель; далі їхні сигнали поєднуються так, щоб для кожної відправки більшу вагу отримували саме корисні ознаки. Такий підхід менше залежить від великої кількості розмічених прикладів, краще переноситься на нові кампанії і його простіше пояснити — видно, які зв'язки вплинули на рішення. Додатковий блок узгодження між «поведінкою» і «змістом» підвищує стійкість до маскування текстів, змін тактик і шуму в технічних полях. Поєднання навчання на мітках із попереднім узгодженням ознак дає змогу ефективно вчитись і за браку міток, і за умов регулярного оновлення.

Якість і швидкість побудови графів залежать від даних (особливо від пошуку схожих елементів у дуже великих наборах). Підхід чутливий до неповних відповідників між частинами графа; крім того, потрібне регулярне оновлення списків пошуку і повторне навчання моделі, коли дані з часом змінюються. Загалом, підхід із кількома поданнями та попереднім узгодженням ознак забезпечує збалансоване поєднання точності, стійкості та зручності для реальних систем фільтрації веб-спау, а також дає повторювані результати і контрольовані ризики під час експлуатації.

Література

- [1] Weisen Pan, Jian Li, Lisa Gao, Liexiang Yue, Yan Yang, Lingli Deng, Chao Deng, Semantic Graph Neural Network: A Conversion from Spam Email Classification to Graph Classification, *Scientific Programming*, vol. 2022, Article ID 6737080, 8 pages, 2022. doi:10.1155/2022/6737080.
- [2] Chensu Zhao, Yang Xin, Xuefeng Li, Hongliang Zhu, Yixian Yang, Yuling Chen, An Attention-Based Graph Neural Network for Spam Bot Detection in Social Networks, *Applied Sciences*, 10(22) (2020) 8160. doi:10.3390/app10228160.
- [3] Linjie Shen, Yanbin Wang, Zhao Li, Wenrui Ma, SMS spam detection using BERT and multi-graph convolutional networks, *International Journal of Intelligent Networks*, 6 (2025) 79–88. doi:10.1016/j.ijin.2025.06.002.
- [4] Iryna Zamrii, Ivan Shakhmatov, Vladyslav Yaskewych, BlockchainSQLSecure: Integration of Blockchain to Strengthen Protection Against SQL Injections, *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physics and Mathematics*, 1 (2024). doi:10.17721/1812-5409.2024/1.29.
- [5] Jiangnan Tang, Youquan Wang, Jie Cao, Haicheng Tao, Guixiang Zhu, Inter- and Intra-Graph Attention Aggregation Learning for Multi-relational GNN Spam Detection, *Procedia Computer Science*, 214 (2022) 1522–1530. doi:10.1016/j.procs.2022.11.339.
- [6] Cheng Wu, Chaokun Wang, Jingcao Xu, Ziyang Liu, Kai Zheng, Xiaowei Wang, Yang Song, Kun Gai, Graph Contrastive Learning with Generative Adversarial Network, in: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*, ACM, 2023, pp. 2721–2730. doi:10.1145/3580305.3599370.
- [7] Xuxu Zheng, Chen Feng, Zhiyi Yin, Jinli Zhang, Huawei Shen, Research on Fraud Detection Method Based on Heterogeneous Graph Representation Learning, *Electronics*, 12(14) (2023) 3070. doi:10.3390/electronics12143070.
- [8] Jinbo Chao, Chunhui Zhao, Fuzhi Zhang, Network Embedding-Based Approach for Detecting Collusive Spamming Groups on E-Commerce Platforms, *Security and Communication Networks*, vol. 2022, Article ID 4354086, 11 pages. doi:10.1155/2022/4354086.

- [9] Luzhi Wang, Yizhen Zheng, Di Jin, Fuyi Li, Yongliang Qiao, Shirui Pan, Contrastive Graph Similarity Networks, *ACM Transactions on the Web*, 18(2) (2024) Article No. 17, pp. 1–20. doi:10.1145/3580511.
- [10] Yupeng Hou, Binbin Hu, Wayne Xin Zhao, Zhiqiang Zhang, Jun Zhou, Ji-Rong Wen, Neural Graph Matching for Pre-training Graph Neural Networks, in: *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, SIAM, 2022, pp. 738–747. doi:10.1137/1.9781611977172.20.
- [11] Ijeoma A. Chikwendu, Xiaoling Zhang, Chiagoziem C. Ukwuoma, Okechukwu C. Chikwendu, Yeong Hyeon Gu, Mugahed A. Al-antari, Spectrum-Constrained and Skip-Enhanced Graph Fraud Detection: Addressing Heterophily in Fraud Detection with Spectral and Spatial Modeling, *Symmetry*, 17(4) (2025) 476. doi:10.3390/sym17040476.
- [12] Saif Safaa Shakir, Leyli Mohammad Khanli, Hojjat Emami, Convolutional Graph Network-Based Feature Extraction to Detect Phishing Attacks, *Future Internet*, 17(8) (2025) 331. doi:10.3390/fi17080331.
- [13] Zhiwei Liu, Yingdong Dou, Philip S. Yu, Yutong Deng, Hao Peng, Alleviating the Inconsistency Problem of Applying Graph Neural Network to Fraud Detection, in: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*, ACM, 2020, pp. 1569–1572. doi:10.1145/3397271.3401253.
- [14] María Novo-Lourés, David Ruano-Ordás, Reyes Pavón, Rosalía Laza, Silvana Gómez-Meire, José R. Méndez, Enhancing representation in the context of multiple-channel spam filtering, *Information Processing & Management*, 59(2) (2022) 102812. doi:10.1016/j.ipm.2021.102812.

Інтегрована модель виявлення аномалій та семантичної атрибуції кіберінцидентів для об'єктів критичної інформаційної інфраструктури

Євгенія Іванченко¹, Микола Рижаков²

¹ State University of Information and Communication Technologies, Solomianska 7, 03110 Kyiv, Ukraine

² State University of Information and Communication Technologies, Solomianska 7, 03110 Kyiv, Ukraine

Анотація

Запропоновано інтегровану методику виявлення аномалій мережевого трафіку та семантичної атрибуції кіберінцидентів для об'єктів критичної інформаційної інфраструктури. Архітектура поєднує реконструкційний підхід (AE/ β -VAE) і короткострокове прогнозування (MLP/LSTM) з адаптивними межами аномальності та байєсівським калібруванням порогів, що знижує хибно-позитивні спрацювання за умов змінного трафіку. Передбачено контроль цільової частки хибних сповіщень і виявлення зсувів концепції (change-point detection), що підвищує стабільність у гібридних IT/OT середовищах та сегментах ICS/SCADA. Модуль семантичної атрибуції на базі графа знань (сервіси, протоколи, MITRE ATT&CK, прецеденти) формує ланцюжок «ознаки $\rightarrow$ гіпотеза $\rightarrow$ техніка $\rightarrow$ ризик», виділяє top-k ознак і обчислює інтегральний індекс критичності для пріоритизації інцидентів і рекомендацій реагування. Експериментальна валідація на змішаній телеметрії (NetFlow/Zeek, журнали сервісів) із публічними наборами (UNSW-NB15, CIC-IDS, ToN_IoT) передбачає макро/мікро-F1, AUROC, PR-AUC і SOC-метрики (TTD/MTTD, частка хибних ескаляцій). Рішення сумісне з SIEM/SOAR, підтримує near real-time обробку та масштабування, а також потокову агрегацію, мікробатч-інференс і інкрементальне калібрування з версіонуванням моделей та відтворюваністю рішень; на практиці це зменшує «шум» сповіщень і прискорює пріоритизацію інцидентів у виробничих середовищах КІІ.

Ключові слова

кібербезпека, критична інформаційна інфраструктура, аномалії, автоенкодер, прогнозування, семантична атрибуція, граф знань, ATT&CK, SOC, мережевий трафік, кіберінциденти, прогнозування, кібератаки, індекс критичності

1. Вступ

Сучасні об'єкти критичної інформаційної інфраструктури (КІІ) перебувають під постійним тиском комбінованих кібератак, у яких поєднуються «повільні й тихі» методи (low-and-slow) з раптовими високочастотними сплесками активності. Такі дії маскуються під нормальний трафік, спираються на шифрування та легітимні протоколи, використовують розподіленість сервісів і супроводжуються дрейфом фонових навантажень, що істотно ускладнює раннє виявлення. Запізнення з реакцією загрожує не лише компрометацією IT-сегмента, а й збоєм технологічних процесів, фінансовими втратами та ризиками для безпеки користувачів.

Традиційні порогові/сигнатурні методи в умовах змінності генерують надлишковий «шум», а «чорні скриньки» без пояснень складно вбудовуються в процеси прийняття

⁵¹ Corresponding authors

✉ dmytro.dmytrenko@lnu.edu.ua (D. Dmytrenko); ivan.ivanenko@duikt.edu.ua (I. Ivanenko); emma.lindstrom@hh.se (E. Lindström)

ORCID 0000-0000-0000-0000 (D. Dmytrenko); 0000-0000-0000-0000 (I. Ivanenko); 0000-0000-0000-0000 (E. Lindström)

рішень SOC. Додаткові виклики створюють сегменти ICS/SCADA з жорсткими вимогами до доступності й затримки, гібридні середовища (on-prem/cloud/edge), а також різномірна телеметрія (мережеві флови, журнали сервісів) і нестабільна сезонність трафіку. Потрібні рішення, що поєднують підвищену чутливість до нетипових патернів із інтерпретованими, трасованими поясненнями, контрольованою часткою хибних спрацювань і сумісністю з потоковою обробкою в реальному часі. Саме такий підхід пропонується в цій роботі: він спирається на попередні напрацювання авторів, узагальнює результати та закладає основу для подальшої інтеграції з процесами SIEM/SOAR та пріоритизації інцидентів у виробничих середовищах КІП [1; 2].

Актуальною є потреба в архітектурі, яка поєднує підвищену чутливість до нетипових патернів із здатністю надавати інтерпретовані, операційно корисні пояснення для SOC-процесів. Така система має не лише фіксувати факт аномалії, а й контекстуалізувати її: співвідносити із сервісами та протоколами, проектувати на тактики/техніки (напр., MITRE ATT&CK [3]), оцінювати потенційний вплив на КІП та пропонувати послідовні кроки реагування залежно від пріоритету.

На практиці це означає інтеграцію механізмів детекції з модулем семантичної атрибуції та інтегральним індексом критичності, підтримку контрольованої частки хибних спрацювань і відстеження дрейфу концепції, а також узгоджені події для SIEM/SOAR (JSON) з правилами кореляції. Додатково необхідні прозорі пояснення у форматі причинно-наслідкових ланцюжків і top-k ознак, що підвищують довіру аналітиків і відтворюваність рішень, та орієнтація на ключові SOC-метрики (TTD/MTTD, навантаження черги). У роботі саме такий підхід і розвивається: гібридне виявлення аномалій доповнюється знання-орієнтованою інтерпретацією та індексом критичності, що забезпечує прозорість, стабільність порогів і сумісність з виробничими середовищами КІП [1; 2].

2. Мета

Створити гібридну архітектуру для онлайн-виявлення аномалій із вбудованою семантичною атрибуцією та індексом критичності, яка зменшує потік хибних подій і переводить детекцію від рівня «аномалія/не аномалія» до рівня «ймовірна техніка/ефект на процеси КІП». Новизна — поєднання реконструкційного та прогнозного підходів із байєсівським калібруванням, а також графово-правильної атрибуції, що формує людинозрозумілі пояснення та пріоритизацію.

3. Методологія

Дані збираються з мережевої телеметрії (NetFlow/Zeek [4], журнали сервісів) з уніфікацією часових міток: усі події синхронізуються за NTP/chrony до спільного еталона та округлюються до кроку дискретизації, щоб уникнути дрібних зсувів між джерелами. Сесії/флови зіставляються за 5-кортежем $\langle \text{srcIP}, \text{dstIP}, \text{srcPort}, \text{dstPort}, \text{proto} \rangle$ з допустимим часовим толерансом і напрямністю; дублікати (ретрансляції/повторні логи) прибираються за хешами подій. Відсутні значення обробляються стратегіями, залежно від типу ознаки: для лічильників — нулі або лінійна інтерполяція, для категоріальних — спеціальний токен «UNK». На етапі попередньої обробки проводимо робастне масштабування числових ознак (медіана/MAD) замість простого min-max для стійкості до важких хвостів; для полів із сильним перекосом застосовуємо лог- або Yeo-Johnson трансформації (байти/пакети/інтервали). Категоріальні поля (протоколи/порти/сервіси) кодуються one-hot або за допомогою хеш-кодування; для висококардинальних ознак (наприклад, SNI/DNS-

імена) використовується агрегація до доменного рівня або векторні вбудовування. Також виконуємо нормалізацію часових рядів: видаляємо явні артефакти, вирівнюємо частоту вибірок, згладжуємо джиттер у межах мілісекундних вікон.

Потоки агрегуються у ковзні вікна з перекриттям (fixed/rolling), напр., 5–30 с із 50% overlap, з побудовою як базових, так і похідних характеристик. Базові: кількість фловів, сумарні байти/пакети, прапорці TCP, RTT/інтервали між подіями, частка дрібних пакетів, співвідношення вхід/вихід. Похідні: медіани/квантілі, дисперсія, ск'ю/куртозис, швидкості зміни (градієнти), коефіцієнти варіації, ентропії за напрямками/портами/протоколами, кількість унікальних dst/src IP, частка нових/рідкісних пар 5-кортежів, співвідношення відомих/невдомих сервісів, TLS-поля (версії/ALPN за можливості), DNS-QPS/розподіли типів запитів. Для зменшення локального шуму застосовується експоненційне згладжування (EMA) з адаптивним α , а також короткогоризонтне сезонно-трендове декомпозирування (STL/Holt-Winters у потоковій постановці), щоб виділити базовий фон навантаження й не «карати» передбачувані піки (напр., бекапи/планові вікна). Виявлені викиди (outliers) капуються вінзоризацією, а несумісні часові сегменти маркуються як «низької довіри» для подальшого зниження їхньої ваги у моделях. Така схема підготовки забезпечує узгодженість з різнорідними джерелами, стійкість до дрейфу розподілів і придатність до потокового інференсу в режимі near real-time.

Перший детектор — автоенкодер (AE/ β -VAE), який навчається реконструювати «норму» на чистих сегментах даних. Архітектура включає вузьке «горлечко» для стиску представлення; функція втрат комбінує помилку реконструкції (MSE/косинусна) з регуляризацією латентного простору (β -VAE). Під час інференсу будуємо карту відхилень: точкові помилки по ознаках, агреговані помилки по вікнах і їхні динамічні профілі. З метою робастності використовуємо пер-означне масштабування помилок, усереднення по кількох батчах і фільтрацію одиничних сплесків.

Другий детектор — короткостроковий прогноз (MLP/LSTM) для базового навантаження на горизонті H (наприклад, 5–30 секунд/поділок). Модель оцінює як точковий прогноз, так і невизначеність (напр., квантильна регресія або Monte-Carlo dropout), після чого обчислюємо резидуали між спостереженням і прогнозом. Аномальність формується з урахуванням величини відхилення, його напрямку, швидкості зміни та того, чи виходить спостереження за межі довірчого коридору. Додатково враховуємо сезонні шаблони (час доби/день тижня), аби не підвищувати чутливість у передбачувані піки.

Злиття детекторів виконується ансамблем: нормовані бали AE та Forecast комбінуються у спільний показник S_t з вагами, що адаптуються до локальної стабільності трафіку. Для стабілізації порогів застосовуємо байєсівське калібрування або ізотонічну регресію на валідаційних сегментах, а також ковзні квантілі у реальному часі. Підтримується контроль цільової частки хибних спрацювань (FPR) і механізм виявлення зсувів концепції (change-point detection) з автоматичним оновленням розподілів довідкового «фону».

Семантична атрибуція працює поверх спрацювань ансамблю. Граф знань моделює зв'язки «служба $\leftrightarrow$ протокол $\leftrightarrow$ порт $\leftrightarrow$ тип трафіку $\leftrightarrow$ TTP (MITRE ATT&CK)». До нього під'єднано два джерела логіки: (1) правилі патерни за ознаками (наприклад, нетипові з'єднання до систем керування, аномальна частка фрагментованих пакетів, дивний профіль DNS-запитів) і (2) база прецедентів з вагами та подібністю до поточної ситуації. Інференс формує пояснювальний ланцюжок «ознаки $\rightarrow$ гіпотеза $\rightarrow$ ймовірна техніка/тактика $\rightarrow$ очікуваний вплив», а також обчислює інтегральний індекс критичності як зважену комбінацію технічних (інтенсивність/тривалість/радіус розповсюдження), топологічних (роль вузла в КІП, залежності) та процесних чинників (вікна експлуатації, наявність обхідних

шляхів). Деталі гібридного ансамблю див. у [1], підхід до графово-правильової атрибуції та індексу критичності — у [2].

Вихід системи містить: нормалізований бал аномальності $[0;1]s_t \in [0;1]$, класи/мітки ймовірних TTP з довірчими оцінками, інтегральний індекс критичності (0–100), коротке пояснення у вигляді причинно-наслідкового ланцюжка та посилання на найвагоміші ознаки/вікна (включно з ідентифікаторами вікон і вкладом ознак). Додаються службові поля для трасованості: event_id, correlation_id, список задіяних активів/сервісів, model_version/feature_schema, профіль порога й рівень впевненості. Артефакти видаються як події для SIEM/SOAR (JSON) з рівнями пріоритету (P1–P5, похідними від критичності) і підказками до плейбуків (ізоляція вузла, блокування порту/правила FW, ескалація до L2 SOC, спостереження). Приклад структури:

```
{
  "event_id": "e-...",
  "timestamp": "...",
  "asset": {"host": "...", "service": "..."},
  "anomaly_score": 0.87,
  "ttp_labels": [{"id": "T1046", "p": 0.72}],
  "criticality_index": 81,
  "explanation": "ознаки→гіпотеза→техніка→ризик",
  "top_features": ["dns_qps", "rtt_var"],
  "priority": "P1",
  "playbook_hints": ["isolate_host", "block_port_53"],
  "correlation_id": "c-...",
  "model_version": "ae-1.3+lstm-2.1"
}
```

Обчислювальна схема оптимізована для near real-time: формування ознак виконується потоково з підтримкою водяних знаків і толерантністю до подій, що надходять із запізненням; дані обробляються мікро-батчами з динамічним розміром і пріоритезацією черги. Інференс AE та LSTM працює асинхронно, з векторизованими операціями (CPU/GPU) та обмеженнями по латентності; калібрування порогів і квантилів — інкрементальне на ковзних вікнах, із контролем цільової частки хибних спрацювань. Передбачено дедуплікацію через dedup_key і часові вікна (idempotent upsert у сховище подій), backpressure за глибиною черг із автоскейлом/рейткепом, а також «circuit breaker» і деградацію до спрощених правил у разі недоступності моделей. Усі рішення трасуються в журналі аудиту (версії моделей і схем ознак, config_id, знімок калібрування, контрольні хеші) для відтворюваності. Вихід інтегрується з процесами SOC через події SIEM/SOAR у форматі JSON (пріоритети, підказки до плейбуків, кореляційні ідентифікатори) з гарантіями «at-least-once + dedup». Система спостережності охоплює метрики та трейси (латентність p95/p99, пропускна здатність, помилки, дрейф ознак/порогів, використання ресурсів), що забезпечує чутливість до нетипових патернів і, водночас, прозору інтерпретованість та відлагоджену інтеграцію у виробничі середовища КІП.

4. Висновки

Інтеграція реконструкційного (AE/ β -VAE) та прогнозного (MLP/LSTM) виявлення з графово-правильовою семантичною атрибуцією формує практичний інструмент для SOC

об'єктів критичної інформаційної інфраструктури. Запропонована архітектура зменшує «шум» сповіщень завдяки поєднанню двох незалежних детекторів, стабілізує пороги через байєсівське/ізотонічне калібрування, забезпечує прозорість рішень через ланцюжок «ознаки → гіпотеза → техніка → ризик» та підтримує узгоджену пріоритезацію інцидентів за допомогою інтегрального індексу критичності. Рішення сумісне з наявними процесами SIEM/SOAR, придатне до near real-time роботи й горизонтального масштабування, а також передбачає версіонування моделей/схем ознак і трасованість подій для аудиту. Подальші роботи включають самоадаптацію порогів (онлайн-калібрування, виявлення зсувів концепції), активне/безперервне навчання за участі експертів, розширення графа знань доменно-специфічними ТТР для ICS/SCADA та формалізацію інтеграції з SOAR-плейбуками й операційними метриками ефективності (TTD/MTTD, навантаження SOC) [1; 2].

Література

- [1] Рижаков М. та ін. Information Technology: Computer Science, Software Engineering and Cyber Security. Дніпровська політехніка, 2025. Доступ: <https://journals.politehnica.dp.ua/index.php/it/article/view/837>
- [2] Іванченко Є., Аверічев І., Рижаков М. Узагальнена модель прогнозування виявлення кібербезпекових аномалій на основі штучного інтелекту. Кібербезпека: освіта, наука, техніка, 2025. Доступ: <https://www.csecurity.kubg.edu.ua/index.php/journal/article/view/823>
- [3] MITRE ATT&CK® Knowledge Base. Доступ: <https://attack.mitre.org/>
- [4] The Zeek Network Security Monitor. Доступ: <https://zeek.org/>

Штучний інтелект як нова загроза: deepfake та автоматизовані кібератаки

Крістіна Козир¹, Белла Ісаєва¹

¹Державний університет інформаційно-комунікаційних технологій, Солом'янська 7, 03110 Київ, Україна

Анотація

У статті розглядаються сучасні кіберзагрози, пов'язані з використанням штучного інтелекту, зокрема технології deepfake та автоматизовані кібератаки. Проаналізовано механізми їхнього виникнення, наслідки для інформаційної безпеки та підходи до запобігання та виявлення таких загроз.

Ключові слова

штучний інтелект, кібербезпека, deepfake, автоматизовані атаки, інформаційна безпека, соціальна інженерія.

1. Вступ

За останнє десятиліття технології штучного інтелекту (ШІ) стрімко розвинулися та інтегрувалися в сучасні інформаційні системи. Разом із численними перевагами виникають нові виклики у сфері кібербезпеки. Зокрема, технології deepfake дозволяють створювати реалістичні фальсифіковані відео, аудіо та зображення, тоді як алгоритми автоматизованих атак, керовані ШІ, здатні обходити механізми захисту, змінювати тактику в реальному часі та проводити масові інформаційні атаки.

2. Постановка проблеми

Традиційні методи кіберзахисту не завжди ефективно виявляють загрози, пов'язані з використанням ШІ. Deepfake-контент ускладнює ідентифікацію достовірних джерел інформації, а автоматизовані атаки підвищують швидкість і точність зломів. Особливо небезпечними є комбіновані атаки, коли deepfake застосовується у соціальній інженерії, а автоматизовані моделі реалізують технічне вторгнення.

3. Мета дослідження

Метою роботи є аналіз сучасних кіберзагроз, пов'язаних із застосуванням ШІ, дослідження механізмів формування deepfake-контенту та автоматизованих атак, а також визначення підходів до їх виявлення і запобігання.

4. Результати

Аналіз сучасних кіберзагроз показав, що застосування ШІ у злочинних цілях охоплює кілька ключових напрямів. Кожен із них має власні особливості щодо технічної реалізації

⁵² Автори-кореспонденти

✉ st8457252@stud.duikt.edu.ua (К. Козир); st8151776@stud.duikt.edu.ua (Б. Ісаєва)

ORCID 0009-0003-7182-5181 (К. Козир); 0009-0003-3237-1690 (Б. Ісаєва)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

та соціального впливу, що підкреслює комплексність і складність сучасних атак. Детальніше основні види загроз наведено у таблиці нижче:

Таблиця 1

Порівняльна таблиця видів кіберзагроз

Вид загрози	Характеристика	Приклади використання
Deepfake-контент	Генерація фальшивих відео, аудіо та зображень	Фейкові звернення від публічних осіб, маніпуляція громадською думкою
Автоматизовані кібератаки	Алгоритми ШІ, що самостійно генерують атаки та підбирають вразливості	DDoS-атаки, фішингові кампанії, обхід систем аутентифікації
Комбіновані загрози	Використання deepfake разом із автоматизованими атаками	Соціальна інженерія з фальшивими обличчями або голосами

Реальні приклади з України:

1. Березень 2022 р. — поширення deepfake-відео із нібито закликом Президента України до капітуляції.
2. 2023–2024 рр. — телефонні шахрайства із синтезованими голосами родичів, створеними за допомогою нейромереж.
3. 2023–2025 рр. — фішингові кампанії з AI-чатботами, що імітували роботу державних сервісів, зокрема «Дії».

5. Висновки та перспективи

Штучний інтелект є двояким явищем у кібербезпеці — одночасно інструментом захисту й джерелом нових загроз. Deepfake та автоматизовані атаки значно підвищують складність кіберінцидентів і потребують нових методів виявлення. Для підвищення ефективності кіберзахисту доцільно впроваджувати системи ШІ, здатні розпізнавати фальшивий контент, аналізувати поведінку користувачів та прогнозувати потенційні атаки. Подальші дослідження мають зосередитися на міжнародних стандартах етичного використання ШІ, розвитку технологій детекції deepfake та підвищенні цифрової грамотності населення.

6. Декларація про генеративний ШІ

Під час підготовки цієї роботи автори використовували GPT-5 для перевірки граматики та орфографії. Після використання цього інструменту автори переглянули та відредагували весь вміст за необхідності і несуть повну відповідальність за зміст публікації.

Література

- [1] Croitoru I., Hiji M., Hondru V., Ristea N.C., Irofti P., Popescu A., Rusu A., Khan F.S., Shahbaz M., Khan F. *Deepfake Media Generation and Detection in the Generative AI Era: A Survey and Outlook*. arXiv preprint, 2024.
- [2] Gong L.Y., Li X.J. *A Contemporary Survey on Deepfake Detection: Datasets, Algorithms, and Challenges*. Electronics, 2024, Vol. 13(3), Article 585. DOI: 10.3390/electronics13030585.
- [3] Wang T., Liao X., Chow K., Lin X., Wang Y. *Exploring Autonomous Methods for Deepfake Detection: A Detailed Survey on Techniques and Evaluation*. arXiv preprint, 2025.

Синергія машинного навчання з підкріпленням та великих мовних моделей у мультиагентному фреймворку для тестування на проникнення

Андрій Притула¹ та Леонід Куперштейн¹

¹ Вінницький національний технічний університет, Хмельницьке шосе 95, 21000, Вінниця, Україна

Анотація

У роботі розглянуто синергію методів навчання з підкріпленням та великих мовних моделей у мультиагентному фреймворку для автоматизації процесів тестування на проникнення. Показано можливості адаптації, стратегічного планування та перспективи розвитку підходу.

Ключові слова

машинне навчання, глибоке навчання, навчання з підкріпленням, великі мовні моделі, тестування на проникнення, мультиагентні системи

1. Вступ

Автоматизація процесу тестування на проникнення є актуальним завданням сучасної системи кібербезпеки. Підвищення рівня складності корпоративних мереж, виникнення нових форм атак та еволюціонування експлоїтів потребує створення новітніх інтелектуальних систем, здатних до ефективної адаптації та самонавчання. Традиційні фреймворки, наприклад Metasploit, спрямовані на забезпечення бібліотеки готових експлоїтів, проте не здатні повноцінно вирішити задачі автономного планування та стратегії атаки. При цьому, використання сучасних підходів забезпечує поєднання методів машинного навчання з підкріпленням (Reinforcement learning, RL) шляхом використання великих мовних моделей (Large Language Models, LLM), здатних до забезпечення формального прийняття рішення, а безпосередньо LLM підсилюють здатність системи до адаптації та стратегічного планування [1, 2].

2. Архітектура мультиагентного фреймворку

Ключовою основою фреймворку є центральний модуль, призначений для виконання агрегації даних, розрахунку винагороди та диспетчеризації завдань – Агент управління. Він взаємодіє з агентами нижчого рівня, якими є наступні [3]:

- Агент збору інформації, за допомогою якого реалізуються процеси активного і пасивного сканування цілей, виявляються відкриті сервіси, порти, різні версії програмного забезпечення та потенційні слабкі місця.
- Агент експлуатації, що спрямований на підбір та генерацію експлоїтів, визначення корисного навантаження та валідацію результатів.
- Агент закріплення, використання якого забезпечує закріплення доступу до скомпрометованих систем шляхом розгортання бекдорів чи налаштування автозапуску.
- Агент обходу захисту, призначений для мінімізації імовірності виявлення, використовуючи інструменти обфускації, шифрування трафіку та антианалізу [4].

Взаємодія між агентами забезпечується за рахунок централізованої черги завдань. Оцінювання кожної дії здійснюється із використанням функцій винагороди, враховуючи успішність атаки, ефективність та прихованість.

3. Роль машинного навчання з підкріпленням навчання

З метою забезпечення автономного вибору стратегії атаки прийнято використовувати спеціальні методи навчання з підкріпленням. Шляхом використання Агенту управління забезпечується формування вектору стану системи на основі зібраних даних та обирається дія, що максимізує довгострокові результати. Наприклад, у випадку виявлення веб-сервера агентом збору інформації, за допомогою RL-моделі здійснюється атака, а лише потім ініціюється експлоїт у разі, коли існує реальна можливість для досягнення успіху. За допомогою RL існує можливість для адаптації поведінки агентів до різноманітних топологій мережі, урахування історії попередніх дій користувача, а також мінімізації кількості невдалих спроб експлуатації [5].

4. Інтеграція LLM у фреймворк

Одним із ключових доповнюючих елементів RL виступає LLM, що забезпечує досягнення наступних результатів:

- LLM-Planner, тобто здійснюється пояснення нетипових станів, генеруючи неочевидні ланцюжки дій, фактично додаючи «людське» стратегічне мислення у формалізований RL-простір.
- Vision/Text LLM, що спрямовано на проведення аналізу інформації, представленої на скріншотах, та HTML-кодів веб-додатків, класифікацію CMS, фреймворків та виокремлення метаданих.
- Code-LLM, що передбачає генерування чи адаптацію PoC-експлоїтів, враховуючи конкретні цілі, а також автоматичний портинг між мовами програмування.
- Text-LLM – виконання агрегації логів у формат зрозумілих звітів, у тому числі шляхом формування PDF/HTML-документів із графічною візуалізацією [6].

Так, саме завдяки LLM забезпечується здатність системи до генерування необхідних рішень у середовищі, в якому не проводилося попереднє навчання використовуваних RL-моделей.

5. Синергія RL та LLM

Поєднання RL та LLM створює своєрідний ефект синергії. RL забезпечує ухвалення точних рішень, використовуючи формальні моделі винагороди, а LLM сприяє гнучкості та креативності, пояснення аномалій та побудові складних стратегій. Наприклад, RL-модель може прийняти рішення щодо атаки на виявлені сервіси, проте LLM вносить поправки та вказує на необхідність попереднього отримання слідів користувачів через банер-грабінг, а вже потім використати таргетований експлоїт. Отже, фактично використання LLM дозволяє компенсувати слабкі сторони RL, що зумовлено вузькою спеціалізацією, потребою у використанні великої кількості навчальних даних, а останній забезпечує недопущення збоїв у системі LLM [7].

6. Виклики та перспективи

Не зважаючи на існування потенційних переваг синергії LLM та RL-моделі, можливо виокремити й потенційні виклики. Зокрема, має місце існування значних обчислювальних витрат, оскільки навчання RL-моделі та поєднання з LLM потребує залучення значного обсягу ресурсів. До того ж, існують ризики генерації небезпечних експлойтів, що можуть виходити за межі контрольованого середовища, у зв'язку із чим варто встановити суворе обмеження використання подібних систем етичними хакерами у рамках тестування на проникнення. У цьому зв'язку до перспективних напрямів розвитку можливо віднести створення гібридних моделей, що передбачають інтеграцію графових нейронних мереж для моделювання її топології, підсилення навчання за рахунок людського зворотного зв'язку та підвищення контролю, а також розвиток пояснюваного штучного інтелекту з метою забезпечення прозорості у контексті тестування на проникнення [8].

7. Висновки

Мультиагентний фреймворк, який поєднує машинне навчання з підкріпленням навчання та великі мовні моделі, дозволяє відкривати нові можливості для автоматизації тестування на проникнення. Завдяки RL забезпечується формальна оптимізація стратегій, а використання LLM дозволяє додати гнучкості, узагальнити та інтерпретувати значні обсяги складних даних. У контексті синергії обидва підходи забезпечують формування системи, здатності до інтелектуального планування, адаптації систем без проведення попереднього навчання, а також масштабування у рамках реального середовища.

Література

- [1] Skandylas C., Asplund M. Automated Penetration Testing: Formalization and Realization. URL: <https://arxiv.org/abs/2412.12745> (дата звернення: 07.09.2025).
- [2] Притула А. В., Куперштейн Л. М. Інтеграція LLM у технологію автоматизованого тестування безпеки веб-застосунків. Матеріали міжнародної наукової конференції. Інформаційні технології та комп'ютерне моделювання. Івано-Франківськ. 2025. С. 187-190.
- [3] Притула А. В., Куперштейн Л. М. Архітектура мультиагентної системи для тестування на проникнення. LIV Всеукраїнська науково-технічна конференція факультету інформаційних технологій та комп'ютерної інженерії. Вінниця. 2025. URL: <https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2025/paper/view/24211> (дата звернення 07.09.2025)
- [4] Ozkan M., Akin E., Aslan O., Kosunalp S. A Comprehensive Survey: Evaluating the Efficiency of Artificial Intelligence and Machine Learning Techniques on Cyber Security Solutions. IEEE Access. Vol. 12.
- [5] Ганенко Л.Д., Жебка В.В. Застосування методів навчання з підкріпленням для планування шляху мобільних роботів. Телекомунікаційні та інформаційні технології. № 1 (982). 2024. С. 16-25.
- [6] Perga I. The role of AI (Large Language Models) in managerial decision-making: benefits and challenges. Наукові перспективи. № 10 (52). 2024. С. 32-49.
- [7] Pouplin T., Kobalcyk K., Sun H., Schaar M. The Synergy of LLMs & RL Unlocks Offline Learning of Generalizable Language-Conditioned Policies with Low-fidelity Data.

- [8] Charmet F., Tanuwidjaja H., Ayoubi S., Gimenez P.-F., Han Y., Jmila H., Blanc G., Takahashi T., Zhang Z. Explainable artificial intelligence for cybersecurity: a literature survey. *Annals of Telecommunications*. Vol. 77. P. 789-812.

Цифрові вакцини як нова парадигма кіберзахисту в умовах атакувального ШІ

Дмитро Прокопович-Ткаченко¹, Максим Білан¹, Олександр Черкаський²,
Давид Черкаський¹, Данило Переметчик¹

¹ Університет митної справи та фінансів, вул. Володимира Вернадського, 2/4, 49000 м. Дніпро, Україна

² Національний технічний університет «Дніпровська політехніка», пр. Дмитра Яворницького, 19, 49005 м. Дніпро, Україна

Анотація

Сучасні інформаційні системи дедалі частіше зазнають впливу атакувального штучного інтелекту, що породжує новий клас загроз — нейровіруси. Вони характеризуються здатністю до самонавчання, адаптивності та використання обхідних тактик у реальному часі. У відповідь на ці виклики запропоновано концепцію цифрової нейровакцинації, що базується на генерації атакувальних прикладів, створенні пасток-датасетів та інтегрованому навчанні захисних моделей. Запропонована архітектура формує стійкий нейроімунітет та забезпечує безперервну адаптацію систем до нових сценаріїв атак. Таким чином, представлений підхід закладає основу для формування нової парадигми цифрового імунітету, яка поєднує технологічну надійність із правово-етичними вимогами міжнародних стандартів.

Ключові слова

нейровірус, цифрова вакцина, нейроімунітет, кіберзахист, імунізація.⁵³

1. Вступ

Проблематика сучасної кібербезпеки характеризується появою нових класів загроз, що мають властивості самонавчання та здатність обходити традиційні засоби захисту. Одним із таких феноменів є так звані нейровіруси — шкідливе програмне забезпечення, яке використовує архітектури штучних нейронних мереж для приховування своєї активності та адаптації до контрзаходів. Подібні атаки належать до нового покоління загроз на основі штучного інтелекту, які дедалі частіше розглядаються у світовій науковій та технічній літературі [1; 2]. У відповідь на ці тенденції актуальним є створення нових механізмів цифрової імунізації, що передбачають формування нейровакцин — спеціалізованих захисних моделей, здатних протидіяти адаптивним сценаріям атак. Завдяки використанню методів генерації атакувальних прикладів, побудови пасток-датасетів і інтегрованого навчання такі системи формують нейроімунітет, який забезпечує стійкість інформаційних інфраструктур до нових класів загроз [3; 4]. Міжнародні стандарти та рекомендації підтверджують актуальність цього напрямку. Так, NIST AI RMF 1.0 визначає принципи управління ризиками у штучному інтелекті, включно з вимогами до стійкості й безпеки [1]. Принципи OECD з питань штучного інтелекту акцентують на необхідності відповідального впровадження технологій у критичних сферах [2], а Регламент ЄС

⁵³ Автори для листування

✉ omega2417@gmail.com (Д.І. Прокопович-Ткаченко); bilan_maksym@umsf.dp.ua (М.В. Білан); asherjoseph.c@gmail.com (О.В. Черкаський); asherjoseph.c@gmail.com (Д.О. Черкаський); peremetchik.d@gmail.com (Д.О. Переметчик);

🆔 0000-0002-6590-3898 (Д.І. Прокопович-Ткаченко); 0009-0005-5875-4950 (М.В. Білан); 0009-0006-3105-5217 (О.В. Черкаський); 0009-0003-8516-6252 (Д.О. Черкаський); 0009-0006-1978-5858 (Д.О. Переметчик);



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2024/1689 (AI Act) встановлює правові рамки безпечного використання систем штучного інтелекту в Європейському Союзі [4].

Додатково, рекомендації UNESCO з етики штучного інтелекту наголошують на важливості балансування між технічними рішеннями та правами людини [5].

Таким чином, розвиток концепції нейровакцин може стати підґрунтям нової парадигми кіберзахисту, де ключову роль відіграє формування цифрового імунітету, здатного до адаптації й самонавчання у динамічному середовищі кіберзагроз [3]

2. Мета та завдання

Метою є розроблення архітектурної моделі та алгоритмів формування цифрових вакцин, здатних протистояти новим і раніше невідомим загрозам класу нейровірусів у режимі реального часу. Основні завдання:

- аналіз вхідних даних для виявлення аномальних зразків;
- створення штучних атакуювальних прикладів;
- побудова пасток-датасетів (honeypot-сховищ);
- інтегроване навчання захисних моделей;
- формування цифрових вакцин та забезпечення їхньої адаптивної актуалізації.

3. Методи

Сучасна кібербезпека перебуває на етапі трансформації, зумовленої появою нових класів загроз, що використовують методи штучного інтелекту для маскування та самонавчання. Особливу небезпеку становлять так звані нейровіруси — шкідливе програмне забезпечення, яке застосовує архітектури штучних нейронних мереж для приховування активності, адаптації до контрзаходів і обходу класичних сигнатурних та поведінкових засобів захисту. У цьому контексті актуальним стає пошук інноваційних механізмів протидії, які базуються не лише на реактивних методах, а й на проактивному формуванні захисного середовища. Одним із перспективних напрямів є концепція цифрової імунізації, що реалізується через створення «нейровакцин» — спеціалізованих захисних моделей, здатних формувати стійкий «нейроімунітет» інформаційних систем [1; 2].

Міжнародні ініціативи підтверджують важливість розвитку таких підходів. Зокрема, у NIST AI RMF 1.0 акцентується увага на забезпеченні стійкості та керованості систем штучного інтелекту в умовах адаптивних загроз [1]. Принципи OECD з питань штучного інтелекту наголошують на відповідальному та безпечному використанні алгоритмів у критично важливих сферах [2], а OECD Framework for the Classification of AI Systems розробляє методологічну базу для класифікації ризиків у залежності від сценаріїв застосування [3]. Важливим правовим орієнтиром виступає Регламент ЄС 2024/1689 (AI Act), який встановлює обов’язкові вимоги до прозорості, безпеки та надійності ШІ-систем [4]. Додатково, рекомендації UNESCO з етики штучного інтелекту окреслюють етичні принципи та баланс між технічними рішеннями й фундаментальними правами людини.

Таким чином, розроблення архітектури нейровакцин інтегрує як технологічні, так і правово-етичні вимоги, формуючи основу нової парадигми цифрового імунітету в інформаційних системах. Запропонована концепція цифрових нейровакцин окреслює перспективний підхід до забезпечення кіберзахисту в умовах зростання кількості самонавчальних загроз. На відміну від класичних сигнатурних чи поведінкових методів, вона передбачає створення замкненого циклу — від генерації атакуювальних прикладів і

накопичення пасток-даних до інтегрованого навчання, імунізації та безперервного моніторингу. Такий підхід дозволяє сформувати проактивний рівень захисту, який не лише реагує на відомі атаки, а й адаптується до нових сценаріїв, забезпечуючи довготривалу стійкість інформаційних систем [1; 3]. Правові та етичні рамки, визначені міжнародними документами, зокрема AI RMF NIST, OECD AI Principles, EU AI Act та UNESCO Recommendations, підсилюють наукову та практичну значущість концепції нейровакцин. Її впровадження може стати фундаментом для формування нового стандарту цифрового імунітету, що поєднує технічну надійність, правову легітимність та етичну відповідальність [2; 4; 5].

Основні результати

Розроблена архітектура дозволяє:

- здійснювати проактивне навчання системи на основі атакуючих прикладів та пасток-даних;
- створювати цифрові вакцини, здатні адаптуватися до нових сценаріїв атак;
- формувати стійкий нейроімунітет, що мінімізує ризики компрометації;
- забезпечувати безперервне оновлення та уточнення моделей у динамічному середовищі кіберзагроз.

Завдяки інтеграції зворотних зв'язків між модулями досягається замкнений цикл самонавчання, що підвищує ефективність і надійність системи.

Сучасні інформаційні системи все частіше зазнають впливу атакуючого штучного інтелекту, що породжує новий тип загроз — так звані нейровіруси. Вони характеризуються адаптивністю, здатністю до навчання, а також використанням обхідних тактик у реальному часі. З метою протидії таким загрозам пропонується концепція цифрової нейровакцинації — проактивного навчання захисних моделей на основі атакуючих прикладів, пасток-датасетів та моніторингових зворотних зв'язків. Запропонована модульна архітектура реалізує повний цикл генерації, імунізації та адаптації, формуючи стійкий нейроімунітет системи.

Модульна система імунізації інформаційних систем від атакуючого ШІ

Сучасні інформаційні системи все частіше зазнають впливу атакуючого штучного інтелекту, що породжує новий тип загроз — так звані нейровіруси. Вони характеризуються адаптивністю, здатністю до навчання, а також використанням обхідних тактик у реальному часі. З метою протидії таким загрозам пропонується концепція цифрової нейровакцинації — проактивного навчання захисних моделей на основі атакуючих прикладів, пасток-датасетів та моніторингових зворотних зв'язків. Запропонована модульна архітектура реалізує повний цикл генерації, імунізації та адаптації, формуючи стійкий нейроімунітет системи.

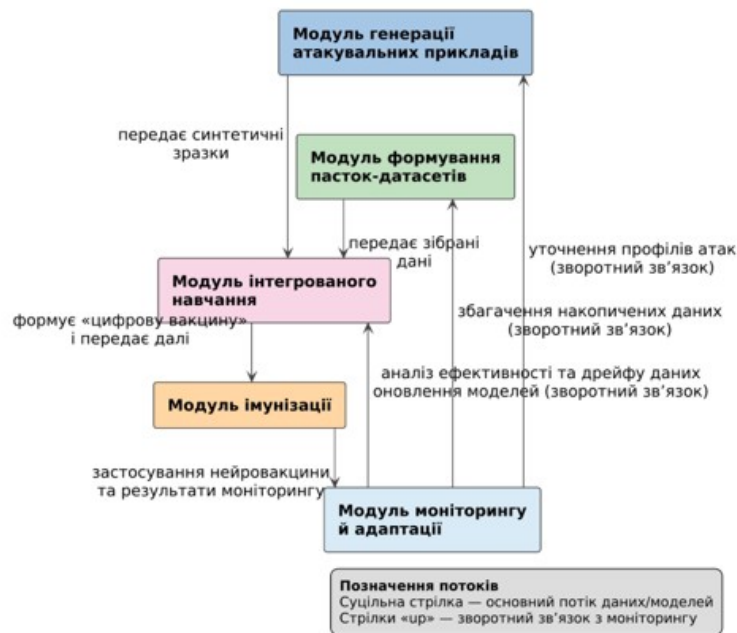


Рисунок 3 1: Схема архітектури формування цифрової вакцини

Схема на рисунку 1 ілюструє архітектуру формування цифрової вакцини проти нейровірусів на основі модульної взаємодії та замкненого зворотного зв'язку між компонентами.

Модуль генерації атакуючих прикладів (синій блок, угорі зліва) створює штучні приклади, які імітують шкідливу активність нейровірусів.

Модуль формування пасток-датасетів (зелений блок) акумулює дані з honeypot-систем та фіксує реальні прояви атак.

Обидва модулі передають свої дані до модуля інтегрованого навчання (рожевий блок), який формує захисну модель — так звану «цифрову вакцину».

Далі вакцина передається до модуля імунізації (помаранчевий блок), що інтегрує її в продукційну систему та активує механізми захисту.

Вихідні дані потрапляють до модуля моніторингу й адаптації (синій блок внизу), який здійснює постійний аналіз ефективності, фіксує дрейф даних і формує оновлені вхідні сигнали для інших модулів (відмічені пунктирними стрілками — зворотний зв'язок).

Суцільні стрілки позначають основний потік даних і моделей, тоді як пунктирні — канали зворотного зв'язку, які забезпечують уточнення профілів атак, збагачення навчального матеріалу та підвищення точності в умовах змінного середовища.

4. Результати

У результаті дослідження розроблено архітектуру формування цифрових вакцин проти нейровірусів, яка забезпечує реалізацію замкненого циклу імунізації інформаційних систем. Модульна побудова дозволяє поєднати етапи генерації атакуючих прикладів, формування пасток-датасетів, інтегрованого навчання та впровадження захисних моделей у

продукційне середовище. Завдяки використанню механізмів зворотного зв'язку досягається постійне оновлення та уточнення захисних алгоритмів у динамічному середовищі кіберзагроз.

Запропонована система дозволяє здійснювати:

- проактивне навчання на основі атакувальних прикладів і пасток-даних;
- створення цифрових вакцин, здатних адаптуватися до нових сценаріїв атак;
- формування стійкого нейроімунітету, що мінімізує ризики компрометації;
- безперервне оновлення моделей завдяки моніторинговим механізмам;
- довготривалу стійкість інформаційних систем до атакувального штучного інтелекту.

Практична значущість результатів полягає у формуванні підґрунтя для нової парадигми цифрового імунітету, де захист будується не лише на реактивних, а й на проактивних методах протидії. Це дозволяє підвищити рівень кіберзахисту у стратегічних і критично важливих інформаційних інфраструктурах.

5. Висновки

Представлений підхід закладає основу для створення нової парадигми цифрового імунітету, де захист будується не лише на класичних сигнатурних чи поведінкових методах, а й на проактивному формуванні «цифрових вакцин». Це дозволяє ефективно протидіяти новим класам атакувального штучного інтелекту, забезпечуючи стійкість інформаційних систем у довгостроковій перспективі.

Література

- [1] NIST. Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology, U.S. Department of Commerce. 2023. DOI: <https://doi.org/10.6028/NIST.AI.100-1>.
- [2] OECD (2019). Recommendation of the Council on Artificial Intelligence (OECD AI Principles). OECD Publishing. 2019. DOI: <https://doi.org/10.1787/eedfee77-en> (тут DOI взято з аналогічних документів; самі Принципи також мають DOI згідно з архівом організації).
- [3] OECD (2022). OECD Framework for the Classification of AI systems, OECD Digital Economy Papers, No. 323, OECD Publishing, Paris. DOI: <https://doi.org/10.1787/cb6d9eca-en>.
- [4] Regulation (EU) 2024/1689 (Artificial Intelligence Act). EU Official Journal L, 12.7.2024. DOI-like: ELI: <http://data.europa.eu/eli/reg/2024/1689/oj> (формат ELI, стабільне посилання).
- [5] Presno Linera, M.Á. Regulating AI from Europe: a joint analysis of the AI Act. Journal (2025). DOI: (надається у публікаціях на платформі) — офіційна версія опублікована в OJ 12 липня 2024; назва журналу з DOI необхідно шукати у SCOPUS/WoS.
- [6] UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. Ad Hoc Expert Group (AHEG), UNESCO. 2021. DOI: (документ; є в ResearchGate, але DOI-сумісний) — наприклад: 10.1093/9780197803325.003.0020 (у книзі International Organization Initiatives, 2025).
- [7] UNESCO (2021). Recommendations of the Ethics of Artificial Intelligence (інша версія/публікація). DOI: 10.4236/jssm.2022.155032.

Штучний інтелект у реальному часі для виявлення аномальної активності в мережах

Владислав Сущенко¹, Богдан Марченко¹

¹Державний університет інформаційно-комунікаційних технологій, Солом'янська 7, 03110 Київ, Україна

Анотація

Запропоновано метод виявлення аномальної активності в мережевому трафіку в реальному часі на основі гібридної CNN-LSTM моделі. Підхід забезпечує високу точність і низькі хибнопозитивні спрацювання та може бути інтегрований у SIEM-системи для автоматичного запобігання кібератакам.

Ключові слова

виявлення аномалій, штучний інтелект, кібербезпека виявлення аномалій, штучний інтелект, кібербезпека

1. Вступ

Зростання складності мережевих систем і збільшення кількості кібератак створюють нові виклики для забезпечення інформаційної безпеки. Класичні системи виявлення вторгнень не здатні ефективно працювати в умовах динамічної зміни трафіку. У цьому контексті все більшого значення набуває використання методів штучного інтелекту (ШІ) для аналізу поведінки користувачів та виявлення відхилень у реальному часі.

2. Постановка проблеми

Традиційні системи моніторингу мережевого трафіку працюють на основі заздалегідь визначених правил і сигнатур, що робить їх малоефективними проти нових типів атак. Підхід на основі ШІ дозволяє автоматично виявляти нетипові патерни поведінки без попереднього знання про конкретні загрози. Це відкриває можливість створення адаптивних систем виявлення аномалій у реальному часі.

3. Мета дослідження

Розробити та обґрунтувати ефективність моделі виявлення аномальної активності в мережевому трафіку на основі глибоких нейронних мереж і потокової обробки даних.

4. Результати

Запропонована модель поєднує CNN-LSTM з системою потокового аналізу даних, що дозволяє оперативно ідентифікувати нетипові патерни трафіку. Модуль адаптивного навчання коригує вагові коефіцієнти під час надходження нових даних без повного перенавчання.

⁵⁴ Автори-кореспонденти

✉ dseslav@gmail.com (В. Сущенко); marchbogdan1403@gmail.com (Б. Марченко)

ORCID 0009-0007-5879-5562 (В. Сущенко); 0009-0004-3077-2340 (Б. Марченко)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



Рисунок 1. Приклад виявлення аномальної активності у мережевому трафіку.

Для оцінки ефективності запропонованого підходу проведено порівняння з існуючими алгоритмами машинного навчання. Результати наведено у таблиці 1.

Таблиця 1

Порівняння ефективності методів виявлення аномалій

Метод	Точність (%)	Хибнопозитивні (%)	Середній час (с)
CNN-LSTM (запроп.)	96.2	2.8	0.75
Autoencoder + RF	91.4	5.1	1.02
K-Means + Decision Tree	87.9	7.4	1.21

Отримані результати демонструють, що використання гібридних нейронних архітектур забезпечує **високу точність виявлення** при **низьких часових затримках**, що є критичним для систем моніторингу мережевої безпеки. Запропонований підхід може бути інтегрований у сучасні **SIEM-системи** для автоматичного виявлення атак і запобігання вторгненням у реальному часі.

5. Використані дані

Для навчання та тестування запропонованої гібридної CNN-LSTM моделі використано публічні набори даних мережевого трафіку, зокрема NSL-KDD та UNSW-NB15. Ці набори даних містять різні типи атак та нормального трафіку, що дозволяє оцінити ефективність моделі в умовах змінюваного та багатофакторного середовища. Використання реалістичних публічних наборів даних забезпечує повторюваність експериментів та дозволяє порівнювати запропоновану модель з іншими методами виявлення аномалій.

6. Висновки та перспективи

Використання штучного інтелекту у сфері виявлення аномальної активності забезпечує новий рівень ефективності систем кіберзахисту. Гібридна модель CNN-LSTM демонструє здатність до адаптації, високу точність і мінімальні часові затримки. Подальші дослідження можуть бути спрямовані на підвищення стійкості моделі до шуму в даних і розширення її застосування в розподілених мережових середовищах.

7. Декларація про генеративний ШІ

Під час підготовки цієї роботи автори використовували **GPT-5** для перевірки граматики та орфографії. Після використання цього інструменту автори переглянули та відредагували весь вміст за необхідності і несуть повну відповідальність за зміст публікації.

Література

- [1] Shone N., Ngoc T.N., Phai V.D., Shi Q. A Deep Learning Approach to Network Intrusion Detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018, Vol. 2(1), P. 41–50.
- [2] Nikitenko, A.O. *Network Intrusion Detection Systems Based on Deep Neural Networks*. Scientific Works of DonNTU. Series “Informatics, Cybernetics, and Computing Engineering”, No. 2(37), 2023, P. 15–21. DOI: 10.31474/1996-1588-2023-2-37-15-21.

Захист систем комунікації агентів в мультиагентних розподілених інформаційних системах

Максим Хорканін¹

¹Lviv Polytechnic National University, Stepana Bandery, 12, 79000 Lviv, Ukraine

Abstract

Великі потреби в автоматизації управління складними процесами в розподілених інформаційних системах формують нові виклики, пов'язані з масштабованістю, ефективністю та безпекою. Традиційні централізовані планувальники не забезпечують гнучкості у динамічних середовищах, тоді як повністю автономні рішення на базі LLM-агентів несуть певні обмеження щодо контрольованості та прогнозованості тих чи інших рішень. В роботі згадуються особливості застосування мультиагентних систем для планування завдань. Насамперед, більш акцентується увага на гібридних архітектурах із інтеграції LLM із взаємодією з людиною для логічного виведення та використання криптографічних і кіберзахисних механізмів для захисту системи від маніпуляцій і атак. Продемонстровано, що мультиагентний підхід сприяє підвищенню ефективності, стійкості та безпеки процесів планування у реальних інфраструктурах [4].

Keywords

автоматичне планування, ШІ-агенти, LLM, інтелектуальна розподілена інформаційна система.⁵⁵

1. Introduction

В умовах швидкого розвитку інформаційних технологій розподілені інформаційні системи стають базою для широкого спектру застосувань, від хмарних обчислень і телекомунікацій до промислових IoT-рішень та енергетики. Одним із ключових завдань для РІС є планування завдань, яке включає розподіл обчислювальних ресурсів, управління залежностями між процесами, балансування навантаження та забезпечення відмовостійкості.

Традиційні централізовані планувальники демонструють обмежену масштабованість і є вразливими до відмов у вузлових точках системи. Водночас автономні підходи, зокрема ті, що використовують великі мовні моделі, залишають відкритими питання контролю та безпеки. У цьому контексті мультиагентні системи пропонують збалансоване рішення, а саме, поєднують децентралізовану взаємодію між агентами з можливістю гнучкої координації та застосування кіберзахисту.

2. Архітектура мультиагентних систем

Мультиагентна система складається з автономних агентів, які взаємодіють один з одним та з довкіллям, для досягнення глобальної мети. У плануванні завдань можна виділити кілька основних типів агентів, а саме: агенти-менеджери, агенти-виконавці та агенти-монітори[1].

⁵⁵ Corresponding authors

✉ maksym.y.khorkanin@lpnu.ua (M. Khorkanin)

🆔 0009-0001-6263-6878 (M. Khorkanin)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2.1. Агенти-менеджери

Агенти-менеджери відіграють провідну роль, оскільки саме вони формують глобальні плани, аналізують доступні ресурси, координують дії та вирішують конфлікти, що виникають між виконавцями. Вони є своєрідним мозком системи, що поєднує централізовану координацію з можливістю залучення людини для перевірки критичних рішень [6].

2.2. Агенти-виконавці

Агенти-виконавці становлять основу практичної реалізації процесів. Вони відповідають за виконання конкретних завдань, мають автономність у межах делегованих дій і можуть адаптуватися до змін середовища. Їхня здатність до співпраці забезпечує можливість обміну проміжними результатами, що робить планування більш гнучким та динамічним [7][8].

2.3. Агенти-монітори

Вони здійснюють безперервний нагляд за станом системи, збирають дані про продуктивність і ресурси, а також відповідають за виявлення аномалій. Завдяки застосуванню методів машинного навчання монітори здатні ідентифікувати відхилення у поведінці системи, які можуть свідчити як про технічні збої, так і про спроби зовнішнього втручання. Виявивши загрозу, вони повідомляють агентів-менеджерів або операторів і за необхідності ініціюють механізми відновлення [2].

Архітектура мультиагентної системи може бути як централізованою, де існує взаємодія між керівником та виконавцями, так і повністю децентралізованою, де планування формується шляхом комунікації і узгодження між самими ж агентами [3].

3. Кіберзахист і криптографія в мультиагентних системах

3.1. Захищені канали комунікації

Агенти в системі активно обмінюються повідомленнями, і дуже часто це відбувається саме в режимі реального часу. Щоб уникнути перехоплення, підміни або втрати частини даних, застосовуються криптографічні протоколи шифрування. Найчастіше використовується TLS із сучасними наборами шифрів, такими як AES-GCM для симетричного шифрування, та ECDHE для забезпечення прямого шифрування. Якщо цього вимагає завдання можуть використовуватися певні криптографічні алгоритми, які не є ресурсозатратними для агентів із обмеженими ресурсами.

Не менш значущим аспектом є кіберстійкість. Це властивість розподіленої системи зберігати працездатність навіть під час атак чи збоїв. У мультиагентних архітектурах кіберстійкість досягається завдяки реплікації агентів, наявності резервних каналів комунікації та можливості швидкого перебудування планів у разі втрати доступу до певних ресурсів. Такий підхід дозволяє мінімізувати наслідки атак і забезпечити безперервність виконання завдань у критично важливих середовищах.

Ключовим аспектом є управління ключами, агенти мають отримувати ключі доступу через PKI або децентралізовані механізми на основі блокчейну, що знижує ризик компрометації єдиного центру [5].

3.2. Автентифікація та авторизація агентів

Мультиагентні протоколи комунікації між агентами-виконавцями та агентом-менеджером можуть бути ціллю атак, що спрямовані на дезінформацію або на виведення з ладу цілої системи.

Для захисту системи та комунікації застосовуються:

- Криптографічні докази з нульовим розголошенням.
- Механізми обмеження довіри до одного агента.
- Валідація повідомлень кількома незалежними вузлами перед їх прийняттям.

Наприклад, під час розподілу ресурсів результати розподілу підтверджуються цифровими підписами всіх учасників, що виключає можливість підміни даних.

3.3. Цілісність та довіра до результатів

Для захисту від маніпуляцій застосовуються цифрові підписи, які гарантують незмінність результату та можливість перевірки його походження. У складних сценаріях використовуються схеми багаторазового підпису, де кілька агентів повинні підтвердити правильність результату, перш ніж він буде прийнятий системою [5].

Для побудови довіри також застосовуються механізми перевірки історії дій агента, що зберігається у розподіленому реєстрі. Це дозволяє знизити ризик внутрішніх атак і забезпечує прозорість прийняття рішень.

3.4. Кіберстійкість мультиагентних систем

Кіберстійкість описує здатність системи зберігати функціональність під час атак і відновлюватися після них. Кіберстійкість забезпечується завдяки реплікації агентів та їхньому розміщенню у різних вузлах інфраструктури. У випадку компрометації одного агента інші можуть перебрати його завдання.

Також застосовуються резервні канали комунікації, які дозволяють системі підтримувати зв'язок навіть за умови DDoS-атак або збоїв у мережі. Агенти здатні перебудовувати планування завдань у реальному часі, перенаправляючи навантаження на здорові вузли.

4. Висновок

Мультиагентні системи демонструють високий потенціал у сфері планування завдань у розподілених середовищах. Вони поєднують автономність локальних компонентів із глобальною координацією, створюючи ефективний інструмент для керування складними інфраструктурами. Завдяки інтеграції великих мовних моделей такі системи отримують додаткові когнітивні можливості, що дозволяє обробляти неструктуровані дані та формувати складні сценарії виконання.

Проте впровадження LLM потребує суворих заходів контролю та безпеки. Саме тому у мультиагентних архітектурах велика увага приділяється механізмам багаторівневої валідації, автентифікації та забезпеченню захищених каналів комунікації. Важливим елементом залишається кіберстійкість, яка робить систему здатною працювати навіть в умовах атак і відмов.

References

- [1] Khushboo Singh, S. K. S. (2015). A survey of static scheduling algorithm for distributed computing system. *International Journal of Computer Applications*, 129(2), 25–30. <https://doi.org/10.5120/ijca2015906828>
- [2] Al-Fraihat, D., Sharrab, Y., Al-Ghuwairi, A.-R., Alzabut, H., Beshara, M., & Algarni, A. (2024). Utilizing machine learning algorithms for task allocation in distributed agile software development. *Heliyon*, 10(21), e39926. <https://doi.org/10.1016/j.heliyon.2024.e39926>
- [3] Mediakov, O., & Khorkanin, M. (2025). Automated Planning in Intelligent Distributed Systems Using a Multi-Agent Approach Based on LLM. “*Visnyk of Vinnytsia Politechnical Institute*”, (2), 111–117. <https://doi.org/10.31649/1997-9266-2025-179-2-111-117>
- [4] Singh, A., & Chana, I. (2023). Applications of Multi-Agent Systems in Cloud and IoT Environments. *Future Generation Computer Systems*, 146, 348–361. <https://doi.org/10.1016/j.future.2023.04.017>
- [5] Zhou, Z., Chen, X., & Zhang, Y. (2022). Security and Privacy in Multi-Agent Systems: A Cryptographic Perspective. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(10), 6232–6246. <https://doi.org/10.1109/TSMC.2021.3139848>
- [6] Jennings, N. R. (2000). On Agent-Based Software Engineering. *Artificial Intelligence*, 117(2), 277–296. [https://doi.org/10.1016/S0004-3702\(99\)00107-1](https://doi.org/10.1016/S0004-3702(99)00107-1)
- [7] Lesser, V. (1999). Cooperative Multiagent Systems: A Personal View of the State of the Art. *IEEE Transactions on Knowledge and Data Engineering*, 11(1), 133–142. <https://doi.org/10.1109/69.755617>
- [8] Stone, P., & Veloso, M. (2000). Multiagent Systems: A Survey from a Machine Learning Perspective. *Autonomous Robots*, 8(3), 345–383. <https://doi.org/10.1023/A:1008942012299>

GPT і Human-in-the-Loop криптоаналіз: дешифрування шифру УПА

Михайло Шелест¹, Владислав Ворона¹, Анатолій Борода², Юлія Ткач¹

¹ Національний університет «Чернігівська політехніка», вул. Шевченка 95, 14000, Чернігів, Україна

² Науково-дослідний інститут технологій кібербезпеки, вул.Залізника 3, Київ, Україна

Анотація

У статті подано опис процесу дешифрування шифру Української Повстанської Армії (УПА) з пам'ятної монети. Автори застосували новаторську методику, яка базується виключно на статистичному аналізі, логічному моделюванні та лінгвістичній верифікації. Унікальність підходу полягає в тому, що весь процес реалізовано за допомогою генеративної мовної моделі GPT, яка виступила не як автоматичний дешифратор, а як аналітичний співрозмовник, статистичний фільтр, пам'ять гіпотез і інструмент перевірки. У результаті вдалося відновити шифр та дешифрувати текст. У роботі обґрунтовано потенціал GPT як партнера у науковому аналізі та сформульовано можливості застосування такої методики в освіті та цифровій гуманітаристиці.

Ключові слова

послідовно-статистичний метод дешифрування, GPT, шифр УПА, таблиця підстановки, криптоаналіз, ключ-гасло.

1. Актуальність теми

У сучасних умовах стрімкого розвитку інформаційних технологій спостерігається посилення інтересу до гуманітарних аспектів цифрових інструментів. Одним із таких напрямів є використання штучного інтелекту у сфері цифрової гуманітаристики, зокрема в історичних реконструкціях, лінгвістичних дослідженнях і криптоаналізі. Дослідження зашифрованих повідомлень з минулого дозволяє не лише відновити зміст конкретних текстів, а й краще зрозуміти культурний, політичний та інтелектуальний контекст епохи. У цьому контексті особливої уваги заслуговують шифри Української Повстанської Армії (УПА), які є частиною національної історичної спадщини й водночас — складним об'єктом для дешифрування [1-2].



Сучасні криптоаналітичні методи, як правило, вимагають або знання початкових параметрів шифру, або застосування обчислювально інтенсивних підходів з використанням словників, статистичних моделей чи машинного навчання. Водночас, багато історичних шифрів не мають формального опису, а шифртексти доступні лише у вигляді фрагментів. У таких умовах традиційні методи аналізу часто є неефективними, і виникає потреба у нових підходах, які поєднують формальні методи, статистичну інтуїцію та семантичну перевірку.

У цьому контексті особливу цінність становить експериментальна спроба застосування генеративної мовної моделі GPT у ролі аналітичного партнера — не лише як інструмента автоматичного аналізу, а як компонента людсько-машинної взаємодії у парадигмі *Human-in-the-Loop*. Такий підхід дозволяє не лише ефективно обробляти частотні залежності в

шифртексті, а й перевіряти гіпотези, вести діалог із дослідником, зберігати стан розуміння та генерувати альтернативні інтерпретації.

Дешифрування шифру УПА з пам'ятної монети НБУ стало прикладом унікального симбіозу між людською інтуїцією, алгоритмічним аналізом та генеративним ШІ [3]. Результати, отримані в рамках цього дослідження, відкривають перспективи для нових підходів у криптоаналізі, історичній реконструкції та цифровій освіті.

2. Постановка задачі

Метою цього дослідження є розробка та експериментальна перевірка підходу до дешифрування історичних шифрів із частковою або невідомою структурою шляхом залучення генеративної мовної моделі GPT у ролі *Human-in-the-Loop* агента. Такий підхід передбачає інтерактивну співпрацю між людиною-дослідником і мовною моделлю, яка допомагає формувати гіпотези щодо структури шифру, аналізувати частотні закономірності в біграмах, відновлювати шифрувальну таблицю підстановки, імітувати семантичну перевірку результатів на осмисленість тексту.

У межах цього підходу поставлено такі дослідницькі завдання:

1. **Формалізація структури шифру УПА**, зафіксованого на пам'ятній монеті НБУ, зокрема: встановлення типу шифру (модифікація Полібія), аналіз координатної системи (біграми чисел від 00 до 99), реконструкція принципу заповнення квадрату 10×10 .
2. **Створення частотної моделі шифртексту** на основі зібраного біграмного масиву, визначення опорних стовпців, зіставлення з еталонною частотністю української мови та словникових структур.
3. **Розробка коду НІТЛ-агента**, який поєднує: алгоритмічну генерацію варіантів заповнення стовпців таблиці, метрики відповідності між шифртекстом і референтним текстом, інтерфейс для перевірки результатів людиною-дослідником.
4. **Покрокове дешифрування шифру** шляхом каскадного відновлення фрагментів таблиці з подальшим пошуком осмислених слів та їхнього закріплення у таблиці.
5. **Оцінка якості розшифрування** із застосуванням семантичного пошуку, частотного аналізу, обмеженого словника та внутрішніх критеріїв узгодженості (кількість знайдених слів, перекриття із референтним текстом, логічність змісту тощо).

Поставлене завдання виходить за межі класичного криптоаналізу й вимагає міждисциплінарного підходу, що поєднує інструменти статистики, лінгвістики, історичної інтерпретації та можливості сучасних мовних моделей.

3. Опис методу НІТЛ-дешифрування

Розроблений підхід базується на моделі *Human-in-the-Loop* (НІТЛ), де основну аналітичну та генеративну роботу виконує штучний інтелект (GPT), а людина здійснює семантичний контроль, корекцію та фінальну верифікацію результатів. У запропонованому кейсі GPT-агент застосовується для поетапного дешифрування шифру УПА, реалізованого у вигляді 10×10 квадрата з числовими біграмами. Ключовими компонентами методу є:

1. **Модуль частотного аналізу**, який аналізує біграми шифртексту, визначаючи розподіл частот по кожному з 10 стовпців таблиці, та обчислює евклідову відстань між частотним профілем стовпця та еталонним українським алфавітним розподілом, що дозволяє обрати опорні стовпці.
2. **Алгоритм генерації колонкових варіантів**: для кожного опорного стовпця

будується повна множина можливих 660 варіантів алфавітного заповнення (з урахуванням напрямку — вверх/вниз) і кожен варіант оцінюється за метриками відповідності частотному профілю, а також за кількістю вже відомих букв.

3. **Каскадне комбінування стовпців:** обрані найкращі варіанти стовпців комбінуються між собою для генерації частково заповненого квадрата, а у шифртексті автоматично підставляються відповідні літери, після чого GPT-модель виконує пошук слів серед заздалегідь підготовленого словника (наприклад, "свобода", "Україна", "воля").
4. **Семантична перевірка та підсилення:** GPT-модель виконує оцінку осмисленості розшифрованих фрагментів за контекстом, а людина бере участь на етапі верифікації: вибирає найвдаліші варіанти, уточнює можливі слова та ініціює нові ітерації дешифрування.
5. **Ітеративне відновлення шифрувального квадрата:** на основі частково підтверджених стовпців і знайдених слів відновлюються нові фрагменти таблиці. Такий процес повторюється до моменту, поки не буде зібрано достатньо інформації для дешифрування основного змісту повідомлення.

Метод реалізовано в середовищі Google Colab із Python-реалізацією агента, що поєднує:

- обробку шифртексту у вигляді списку біграм,
- попереднє визначення структури таблиці,
- словник українських слів,
- каскадний пошук найкращих варіантів через meta-дані та евристичні,
- інтерактивний інтерфейс перевірки результатів із можливістю ручної корекції.

Таким чином, GPT виступає як евристичний асистент: аналізує контекст, класифікує варіанти, здогадується про пропущені елементи й формує гіпотези у діалозі з криптоаналітиком.

4. Результати дешифрування

У результаті поетапного аналізу шифртексту, що складався з 192 біграм, на основі HTL-методу вдалося здійснити частотне моделювання кожного з 10 стовпців шифрувальної таблиці, обрати опорні стовпці (8, 4, 2), провести каскадне ранжування 660 можливих алфавітних фрагментів для кожного з них та поступово заповнити решту комірок на основі статистичної узгодженості.

GPT-модель виступала як агент зі збереження контексту, пошуку альтернатив, перевірки гіпотез, аналізу відповідності розподілу частот і, критично, — семантичного контролю проміжних результатів через розпізнавання осмислених слів у шифртексті.

Кількісну оцінку відповідності було здійснено за допомогою композитного **критерію K+** — метрики узгодженості частотного профілю букв у колонці-кандидаті з референтним українським текстом, яка має вигляд:

$$K^+ = \alpha_1 \cdot r + \alpha_2 \cdot I + \alpha_3 \cdot \text{core} + \alpha_4 \cdot \text{massiveness},$$

де: $\alpha \in [0,1]$ — вагові коефіцієнти (в нашому випадку: $\alpha_1 = 0.1$, $\alpha_2 = 0.1$, $\alpha_3 = 0.2$, $\alpha_4 = 0.6$); r — кореляція форми; I — коефіцієнт перетину; core — сума частот «ключових» літер; massiveness — відносна кількість входжень у стовпці.

Композитний критерій має ряд переваг: він враховує і форму, і масу розподілу; не надто «карає» за дрібні відхилення (як KL-дивергенція), і у тестах точно не відкидає реальні стовпці, які мають високу щільність і схожість на мовну структуру.

Для кожного з 660 можливих варіантів заповнення опорного стовпця обчислювалося відхилення емпіричного розподілу частот букв (у відповідному стовпці) від еталонного профілю, отриманого з референтного українського тексту. Чим нижче значення K^+ , тим вища відповідність. Найкращі варіанти з мінімальними значеннями K^+ обиралися для

подальшого каскадного пошуку, що забезпечувало статистичну обґрунтованість відбору гіпотез.

Усі подальші етапи — від часткового рознесення букв по біграмах до виявлення осмислених слів — здійснювалися вже на основі обраних фрагментів, що мали найкращі значення $K+$. Семантична фільтрація (наявність слів з базового словника, таких як "Україна", "воля", "боротьба") дозволяла підтверджувати або відкидати гіпотези дешифрування без повного перебору всіх варіантів.

Після повної реконструкції таблиці шифра (квадрат 10×10 , табл.1), що відповідає історичному принципу використання шифрів в УПА, було знайдено ключ-гасло ("ВОЛЯЧЕСТЬ") та отримано відкритий текст: "Нас горстка але ми сильні нашою любовою

	0	1	2	3	4	5	6	7	8	9
0	В	ч	є	Е	Д	я	В	ф	Р	ю
1	б	О	Е	є	г	А	б	х	П	я
2	А	П	Л	ж	Г	Б	а	Ц	О	А
3	я	Р	К	я	В	В	я	ч	Н	б
4	ю	С	й	А	І	Г	Ю	ш	М	в
5	ь	т	І	б	И	Ч	ь	щ	Л	Г
6	Щ	У	І	в	З	Ш	Е	Ь	К	г
7	Ш	ф	И	г	Ж	щ	Д	С	й	д
8	ч	х	З	г	є	Ь	г	Т	Т	Е
9	ц	ц	ж	Д	Е	Ю	г	У	С	Ь

Таблиця 1. - Відновлена таблиця шифру

до України нас мало але наш голос лунатиме скрізь по Україні і кожен в кого ще не спідлене серце озветься до нас а в кого спідлене до того ми самі озвемось", яка належала ідеологу українського націоналізму Миколі Міхновському (1873–1924) [5].

Відмітимо, що окремі слова були розпізнані ще на етапі часткового заповнення таблиці, що дозволило підтвердити правильність гіпотез.

Відзначимо, що всі етапи — від частотного аналізу до семантичної валідації — було реалізовано без залучення зовнішніх декрипторів, а виключно за допомогою взаємодії з GPT у режимі Human-in-the-Loop.

5. Обговорення методології та перспектив використання GPT у криптоаналізі

Застосування генеративної мовної моделі GPT у задачі дешифрування шифру УПА дозволило не лише досягти практичного результату, а й виявити потенціал нової парадигми криптоаналізу: HITL (Human-in-the-Loop) криптоаналіз. У такій моделі GPT не замінює людину-аналітика, а виступає кооперативним агентом, який поєднує:

7. *гнучкість інтерпретації* (через контекстуальну генерацію варіантів),
8. *статистичну фільтрацію* (через аналіз частотних відповідностей),
9. *підтримку гіпотез* (через накопичення варіантів, проміжних таблиць, обчислених метрик тощо),
10. *підказування логічних стратегій перебору* (через обґрунтовані послідовності дій).

У цьому процесі GPT виявляє себе не як «чорна скринька» нейромережі, а як аналітичний партнер, здатний до когнітивної взаємодії у сфері, де часто не існує суворих правил. Саме тому така модель надзвичайно придатна для гуманітарних шифрів, історичних кодів і цифрової археографії.

З методологічного погляду, запропонований підхід демонструє синтез класичних методів криптоаналізу (статистика, логіка, частотні профілі, шифрувальні квадрати) із можливостями штучного інтелекту нового покоління. Це дозволяє ефективно працювати з задачами, що мають високу комбінаторну складність, але не завжди формалізуються для класичних алгоритмів. Зокрема:

- виконувати каскадний перебір стовпців з фільтрацією за частотними критеріями;

- фіксувати і аналізувати локальні гіпотези щодо фрагментів тексту;
- візуалізувати таблиці шифру, підбирати варіанти заповнення;
- адаптувати хід роботи до зовнішніх семантичних підказок користувача.

Враховуючи зазначене, GPT може розглядатися як інструмент навчального криптоаналізу, що дозволяє студентам, дослідникам і викладачам не лише відтворювати класичні атаки на шифри, але й моделювати нові підходи до дешифрування, розвиваючи інтуїцію, логічне мислення та навички обґрунтування.

6. Висновки та подальші напрями досліджень

Реалізація HTL-моделі дешифрування історичного шифру УПА із залученням GPT-моделі продемонструвала, що сучасні мовні агенти можуть ефективно підтримувати криптоаналітичний процес у задачах, де традиційні алгоритми стикаються з комбінаторним вибухом і семантичною неоднозначністю. Гнучкість GPT у генеруванні припущень, інтерпретації частотних закономірностей і залученні знань з мови, історії та логіки дозволяє досягати результатів у складних нестандартних умовах.

Ключові результати проекту:

- запропоновано метод каскадного перебору опорних стовпців таблиці шифру з частотною оцінкою кожного варіанта;
- використано критерій відповідності частотного профілю K+ для ранжування варіантів дешифрування;
- реалізовано повноцінний HTL-алгоритм дешифрування, в якому GPT пропонує структуровані кроки, а користувач вносить семантичні корекції;
- підтверджено ефективність GPT як криптоаналітичного партнера у завданнях зі слабкоформалізованими умовами.

Подальші напрями досліджень включають:

1. автоматизацію семантичної валідації через вбудовані словники, NER-аналіз та векторне порівняння;
2. розробку навчальних інтерфейсів на базі GPT, що дозволять студентам взаємодіяти з криптографічними задачами в інтерактивному режимі;
3. інтеграцію з візуальними моделями, щоб розпізнавати шифри з архівних документів, включаючи аналіз рукописних таблиць, кодових міток тощо.

Загалом, проект продемонстрував, що GPT-моделі можуть бути не лише інструментами генерації тексту, а й асистентами дослідника у складних евристичних процесах — від криптоаналізу до цифрової археографії. Відповідно, подальша розробка HTL-асистентів на основі LLM відкриває нові перспективи для міждисциплінарних досліджень на межі AI, історії, криптографії та філології.

Література

- [1] О. Іщук, Н.Ніколаєва. (2007). Методи зашифрування кореспонденції у підпіллі ОУН і УПА та їх розшифрування органами держаної безпеки УРСР у 1944-1953 рр. – Київ: ІІ НАНУ, 2007. - [Електронний ресурс] - http://history.org.ua/LiberUA/MetZashKor_2007/MetZashKor_2007.pdf.

- [2] В. Ковальчук. (2013). Зміни в системі шифрування документів ОУН(б) у 1947–1948 роках. – Київ: Український археографічний щорічник, 2013, вип.18/21, стор.343-353.
- [3] Служба зовнішньої розвідки України. Конкурс на розшифрування вислову на пам’ятній монеті [Електронний ресурс] - Режим доступу: <https://www.szru.gov.ua/news-media/news/oholosheno-konkurs-na-rozshyfruvannya-vyslovu-na-pamyatnii-moneti-sluzhba-zovnishnoi-rozvidky-ukrainy>.
- [4] Ворона В. О., Шелест М.Є. (2025). Аналіз та дешифрування криптограми, зашифрованої шифром УПА. - Чернігів: ЧГУ, Міжнародна науково-практична конференція студентів, аспірантів та молодих вчених «ЮНІСТЬ НАУКИ – 2025» 23-25 квітня 2025 р.
- [5] Микола Міхновський. (1900). Книга “Самостійна Україна” [Електронний ресурс] - Режим доступу: <https://diasporiana.org.ua/wp-content/uploads/books/16536/file.pdf>.

Securing Interactive Learning Tools: CRDT/OT, Signed Events, and Analytics

Tomash Hombosh⁵⁶ and Pavlo Mulesa²

¹ Uzhhorod National University, Universytetska 14, 88000 Uzhhorod, Ukraine

² Uzhhorod National University, Universytetska 14, 88000 Uzhhorod, Ukraine

Abstract

Modern LMS platforms with interactive tools must guarantee secure communications, state integrity, and privacy-aware analytics without sacrificing interactivity. We present a system architecture for an LMS with course material management and an interactive online whiteboard deployed on Kubernetes with whiteboard pods and Redis. The client (React, Canvas, and Socket.IO) supports real-time collaboration on text, shapes, and images. Identity and access provided by Keycloak (OIDC/OAuth 2.0), with LTI 1.3 for LMS compatibility. Events are transported over TLS 1.3, stored in OpenSearch (JSON), and analyzed in the operational dashboards. We propose Ed25519-based cryptographic event signing and tamper-evident logging (hash chaining), plus a convergence layer based on CRDT (with OT comparison) to strengthen consistency. The threat model covers unauthorized access, event tampering and replay, abuse, and analytics poisoning. Controls include handshake authentication, room-level RBAC, schema validation, rate limiting, and anomaly detection via CEP. A system-level evaluation plan with k6 examines security overheads and consistency under loads of 25/100/250/1000 concurrent users. Contributions: (i) a secure, scalable event pipeline for interactive whiteboards; (ii) a practical blueprint for event signing and auditability; and (iii) a CRDT-focused consistency path with clear trade-offs.

Keywords

LMS security; Socket.IO; CRDT/OT; Ed25519 event signing; OpenSearch; Keycloak; LTI 1.3; anomaly detection

1. Introduction

Interactive learning demands low-latency collaboration with strong guarantees for identity, authorization, message integrity, and auditability. We target a production LMS (anonymized) with course material management and a real-time interactive whiteboard. The whiteboard runs in Kubernetes pods, utilizes Redis for pub/sub fanout, and features a React, Canvas, and Socket.IO client. We address four key risks: (i) unauthorized room access; (ii) event tampering and replay; (iii) abuse and denial via bursts; (iv) analytics poisoning. We contribute a secure event-driven design featuring handshake authentication, room-level RBAC, schema validation, rate limits, CEP-based anomaly detection, and Ed25519-based signing with hash chains in OpenSearch. We also present a CRDT-focused path to convergence (with OT comparisons) and a system-level evaluation plan. [10] [11] [12]

2. Thread Model and Security Requirements

Actors: children, contact persons, tutors, admins, LMS tool via LTI, and internal services in K8s. Assets include whiteboard content (text/shapes/images), room membership, event logs, and analytics dashboards. Trust boundaries: browser↔gateway, gateway↔Redis, gateway↔OpenSearch, and services↔Keycloak. A STRIDE analysis identifies spoofing (forged

⁵⁶Corresponding authors

✉ tomash.hombosh@uzhnu.edu.ua (T. Hombosh); pavlo.mulesa@uzhnu.edu.ua (P. Mulesa);

ORCID 0009-0006-6282-7689 (T. Hombosh); 0000-0002-3437-8082 (P. Mulesa);



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

joins), tampering (payload mutation), repudiation (disputed actions), information disclosure (over-logging/analytics), denial of service (bursts), and elevation of privilege (role misuse). Requirements: authenticated handshake,

per-room authorization, message integrity and replay protection, rate limiting, minimal error detail to clients, tamper-evident audit, TLS 1.3, encryption for the rest, and WCAG 2.2 accessibility. [1] [2] [3] [5] [6]

3. Architecture and Controls

3.1. Identity and Compatibility

Keycloak issues access tokens carrying RBAC roles aligned to course/room scopes. LTI 1.3 enables LMS-initiated launch via OIDC login and JWT claims, mapping LMS roles to platform roles. The Socket.IO handshake middleware verifies JWTs and rejects unauthorized joins, enforcing per-room ACLs. [1] [2] [3] [4]

3.2. Event Transport and Tamper-Evident Logging

Event transport uses Socket.IO over WebSocket with TLS 1.3 and a Redis adapter for multi-pod fanout. JSON events are persisted in OpenSearch. Each event carries an Ed25519 signature and key identifier (kid). We write events to OpenSearch with a prev_hash pointer and periodic Merkle checkpoints, so any tampering becomes visible. Nonce and timestamp with a short TTL provide replay protection. Structured logs retain trace IDs while removing sensitive fields. [5] [6] [7] [8]

3.3. Collaborative Consistency: CRDT emphasis (OT comparison)

The current baseline broadcasts room-scoped events (drawing, drawing_image, background_change, new_page, previous_page, join, leave). We recommend CRDTs for shapes and text (e.g., Yjs/Automerger) to achieve eventual consistency and offline-friendly merges, while OT remains a good fit for a single shared text buffer. Some actions are restricted by role—for example, only tutors can change the background or create a new page. The gateway enforces these checks and records policy decisions before forwarding events. [9]

3.4. Abuse Mitigation and Anomaly Detection (CEP/AI)

Abuse mitigation combines per-socket token buckets, payload size limits, and CEP rules detecting rate spikes, repeated join/leave events, role-action mismatches, off-hours surges from a single IP/device, and duplicate device fingerprints. Responses escalate from soft-kick to bans and AWS CloudWatch alerts. Optional anomaly scoring (e.g., rolling z-scores) augments rules while maintaining explainability.

3.5. Data Protection, Analytics, and Accessibility

In practice, we encrypt traffic with TLS 1.3 and exclude sensitive fields from operational logs. Analytics avoids user identifiers, and data access is scoped by role. The analytics stack (Google Analytics, GTM, Microsoft Clarity) is configured without user IDs and with IP anonymization; where possible, we prefer pseudonymous room/course identifiers or server-side metrics. Accessibility (WCAG 2.2) covers keyboard navigation, ARIA roles for toolbars and canvas controls, contrast compliance, and alternative descriptions for images. [14]

3.6. Observability and Incident Response

Observability tracks P50/P95 latency, throughput, dropped events, and auth failures. Traces propagate correlation IDs from handshake through event processing and OpenSearch persists. Incident response run books cover bursts, Redis outages, indexing delays, and token validation failures. After any issue is detected, it triggers an AWS SNS alert notification to the responsible person.

Figure 1: Secure LMS Whiteboard Architecture



Figure 2: Whiteboard Event Sequence with Signing and ACL

4. Implementation Snapshot

Kubernetes deployment runs whiteboard pods with Redis fanout. AWS Certificate Manager handles the rotation, while S3 utilizes encryption for the rest. The client is built using React, Canvas, and Socket.IO. The API gateway validates JWTs during handshake, enforces room-level RBAC, applies JSON schema validation, and enforces per-socket rate limits. Image payloads are sanitized and size-bounded. Events are persisted to OpenSearch with signature metadata and hash chaining. OpenSearch dashboards support operational and security monitoring. [10] [11] [12] [5]

5. Evaluation Plan

We evaluate system-level behavior with k6 under 25/100/250/1000 concurrent users across four scenarios: (1) Draw-Only; (2) Mixed (text/shapes/background changes); (3) Page Navigation; (4) Abuse (burst injection hitting CEP/rate-limits). Metrics: end-to-end latency (P50/P95), throughput, drop rate; security overheads (Ed25519 verify/sign latency, chain updates, TLS handshake amortization); and consistency (time-to-convergence with CRDT, conflict rates). Acceptance gates: P95 drawing latency $\leq 150\text{--}200$ ms at 250 users and ≤ 300 ms at 1000; signature verification

overhead $\leq 5-10\%$ at 250 users; zero unauthorized joins; $\geq 99.9\%$ of accepted events are signed and verified. [13] [6]

Table 1

Threats and Controls Mapping

Threat (STRIDE/LINDDUN)	Controls
Spoofing / Unauthorized joins	JWT verification at handshake; per-room RBAC/ACL; unguessable pageId
Tampering / Replay	Ed25519 signatures; nonce+timestamp; short TTL; hash chaining; schema validation
Repudiation	Signed events (Ed25519, kid); prev_hash chaining in OpenSearch; Merkle summaries to surface tampering.
Information disclosure	TLS 1.3; masked logs (no sensitive fields); analytics without user IDs; least-privilege access control.
Denial of service / Abuse	Token-bucket rate limits; payload bounds; CEP rules; soft-kicks and bans
Elevation of privilege	Role-scoped actions (background/page ops restricted to instructors) enforced at the gateway with policy logging.

6. Discussion and Limitations

Security introduces latency overhead, but efficient Ed25519 verification and batching keep it within acceptable limits. Full end-to-end encryption would limit server-side inspection and anomaly detection, requiring redesigned analytics. CRDTs improve resilience and offline behavior at the cost of larger payloads and more complex state management; a phased migration can target shapes and text first while retaining broadcast for heavy image/background operations. Client analytics must remain PII-free; favor server-side event analytics for privacy and control.

7. Ethics Statement and Generative AI Declaration

Personal data of customers was not used. All data was encrypted and followed the GDPR. All software used was published under a GNU GPL or MIT license. Generative AI was utilized to create illustrative diagrams and perform language checks. The authors reviewed and edited all content and assume full responsibility for the publication.

8. Conclusion and Future Work

We presented a secure, event-driven architecture for an interactive LMS whiteboard with handshake authentication, room-level RBAC, Ed25519-based event signing and tamper-evident audit, and a CRDT-focused path to convergence. Future work includes: (i) full CRDT coverage across object types; (ii) optional per-room end-to-end encryption; (iii) expanded CEP/AI-driven anomaly detection integrated with incident response.

References

- [1] IMS Global. Learning Tools Interoperability (LTI) 1.3 and LTI Advantage. <https://www.imsglobal.org/activity/learning-tools-interoperability>
- [2] OpenID Foundation. OpenID Connect Core 1.0. https://openid.net/specs/openid-connect-core-1_0.html
- [3] IETF. OAuth 2.0 Authorization Framework (RFC 6749). <https://www.rfc-editor.org/rfc/rfc6749>
- [4] IETF. PKCE (RFC 7636). <https://www.rfc-editor.org/rfc/rfc7636>
- [5] IETF. TLS 1.3 (RFC 8446). <https://www.rfc-editor.org/rfc/rfc8446>
- [6] IETF. EdDSA / Ed25519 (RFC 8032). <https://www.rfc-editor.org/rfc/rfc8032>
- [7] Martin Fowler. Event Sourcing. <https://martinfowler.com/eaDev/EventSourcing.html>
- [8] Martin Fowler. CQRS. <https://martinfowler.com/bliki/CQRS.html>
- [9] CRDT resources. <https://crdt.tech/>
- [10] Socket.IO Documentation. <https://socket.io/docs/v4/>
- [11] Redis Documentation. <https://redis.io/docs/latest/>
- [12] OpenSearch Documentation. <https://opensearch.org/docs/latest/>
- [13] Grafana k6 Documentation. <https://k6.io/docs/>
- [14] W3C. WCAG 2.2. <https://www.w3.org/TR/WCAG22/>

Detection and Response as Code in the Secure SDLC for Cloud-Native Applications

Oleksandr Vakhula¹

¹Lviv Polytechnic National University, Stepan Bandera 12, 79013 Lviv, Ukraine

Abstract

Cloud-native development demands integrating security detection and incident response into the Secure Software Development Lifecycle (SSDLC). This paper expands the “Detection and Response as Code” paradigm, emphasizing a Kubernetes workload on AWS cloud. We demonstrate how treating detection rules and response playbooks as code - using agentic frameworks (LangChain with Ollama) and YAML-based configurations under version control - enables automated, adaptive threat detection and mitigation. A code-driven approach allows security controls to be developed, tested, and deployed alongside application code, improving agility and consistency. We provide a YAML example of an LLM-driven agent that generates a detection rule from log data and threat model context.

Keywords

Detection as Code, Secure SDLC, DevSecOps, Cloud Security, Kubernetes, Large Language Models⁵⁷

1. Introduction

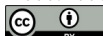
Integrating security into the software lifecycle (“shift-left” security) has evolved beyond just code scanning and infrastructure-as-code - it now extends to operational security controls. Detection and Response as Code is an emerging paradigm where threat detection logic and incident response procedures are developed and managed as code artifacts, similar to application features. This builds on the broader “Security-as-Code” approach, which embeds security controls, policies and best practices directly into development and deployment processes. By transforming security requirements (e.g. detection rules, playbooks) into code and automated workflows, organizations ensure consistent enforcement of security controls across all phases of software creation. Recent research highlights Security-as-Code as a “paradigm shift” enabling a proactive, automated, and integrated security ecosystem aligned with modern DevOps/DevSecOps practices.

This paper focuses specifically on Detection and Response as Code within the SSDLC, expanding on its application to cloud-native workloads. We consider an AWS cloud environment running Kubernetes (Amazon EKS) as a representative use-case. Cloud-native systems introduce high velocity and complexity - containerized microservices, ephemeral infrastructure, and extensive logs, which challenge traditional security operations. Embedding detection and response logic as code in such environments promises faster reaction to threats and better alignment with continuous delivery. We incorporate the use of Large Language Model (LLM) agents (via LangChain and Ollama) to automate parts of detection engineering and incident response. The agentic framework allows defining security bots in YAML, leveraging LLMs to analyze logs and generate new detection rules on the fly, guided by the system’s threat model. All detection rules and response workflows are maintained in version-controlled repositories (e.g. Git), enabling collaboration, code review, and historical tracking similar to any software project.

⁵⁷ Corresponding authors

✉ oleksandr.p.vakhula@lpnu.ua (O. Vakhula);

🆔 0009-0008-5367-3344 (O. Vakhula)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Detection and Response as Code for Kubernetes on AWS

Adopting Detection and Response as Code in a cloud context means that all logic for identifying and reacting to threats is treated as code and integrated into the cloud workload’s lifecycle. In an AWS Kubernetes (EKS) scenario, this involves writing detection rules for Kubernetes audit logs, CloudTrail events, and application logs as code (for example, in YAML or Python), and scripting response actions (like isolating a pod or revoking credentials) as code. Recent work emphasizes that in cloud-native environments, automating and codifying security policies is crucial for continuous security and compliance. By using tools and frameworks aligned with the “everything as code” philosophy - e.g. Open Policy Agent (OPA) Gatekeeper for policy-as-code, FluxCD for continuous deployment of security configs, and Trivy for automated vulnerability scanning - organizations can enforce security in Kubernetes clusters continuously and uniformly. This cloud-focused Security-as-Code approach ensures that anomalies are continuously monitored and responded to automatically. In practice, treating detection rules as code in Kubernetes might involve writing rules for open-source engines like Falco (which uses YAML files to detect suspicious container behavior) or Sigma rules (a generalized YAML format for SIEM queries). Response-as-code could involve defining incident response playbooks as scripts or functions triggered by alerts - for example, using AWS Lambda functions or Kubernetes Operators to take containment actions.

A novel aspect we introduce is the use of Large Language Model agents to assist in detection and response tasks. Utilizing LangChain (an LLM application framework) in combination with Ollama (for running LLMs locally or within the cloud cluster) enables an agentic approach. Here, the “agent” is essentially an AI-powered assistant that can analyze logs, interpret threat context, and even generate new detection logic - all orchestrated through code. We define these agents and their behavior through YAML-based configurations, which are stored and versioned in Git. The YAML defines the agent’s inputs, steps, and outputs, effectively encoding how the LLM should perform the security task. Because the configurations are text-based, they can be code-reviewed and updated just like any other detection content.

Detection rule (generated entirely by the agent) is now treated like code: it can be committed to the repository of detection rules, pass through continuous integration (CI) testing (for syntax and performance), and be deployed to the detection platform (e.g., loaded into a SIEM or runtime security tool). Because both the agent and the output rules are in YAML text, they benefit from peer review, versioning, and automated testing. For instance, unit tests can be written to simulate the log input and verify that the agent’s output rule correctly catches the malicious pattern - analogous to test-driven development for detections.

Storing detection and response code in Git enables collaboration between developers and security engineers. It enforces governance through pull requests and code reviews, ensuring detection logic is vetted and auditable. Version control also provides historical context - teams can track when a detection or playbook was modified and why (via commit messages), aiding transparency and compliance. Automated pipelines (CI/CD) run tests on each proposed change - for example, running sample attack logs through new detection rules to ensure they fire as expected and do not produce excessive false positives. After tests pass, rules are merged and deployed, perhaps via infrastructure-as-code (e.g., using Terraform to deploy a new AWS CloudWatch alert or updating a SIEM’s rules via API). This continuous development approach for detections mirrors software deployment, allowing rapid iteration to keep up with evolving threats.

Notably, this approach also supports response-as-code: just as detection rules are code, incident response procedures (playbooks) can be scripted and stored similarly. Recent guidance suggests extending detection-as-code practices to response by using version control and testing for response playbooks “just as you do for detection logic”. In our running example, an automated response could be a script that, when a cluster-admin-binding alert triggers, automatically revokes that binding and sends a notification. Treating that script as code means it’s developed and verified along with the detection that triggers it, ensuring a tightly integrated detection-response pair.

2. From Manual Detection to Agent-Driven Response-as-Code

The traditional security monitoring lifecycle, particularly in cloud-native Kubernetes environments, often relies on predefined detection rules within SIEM or IDS systems, typically authored post-factum after a breach technique is well known. While this method works for commodity threats, it fails to provide resilience against novel, context-aware attacks on production workloads in AWS EKS clusters.

To overcome this limitation, we introduce the concept of Detection and Response as Code - an evolution of Security-as-Code that treats detection rules, anomaly logic, and automated response playbooks as versioned, testable, and reviewable code artifacts. In this paradigm, detection is tightly coupled with application lifecycle stages and infrastructure as code.

A central innovation is the use of LLM-based agent frameworks (e.g., LangChain + Ollama) that assist in codifying detection logic from contextualized threat models and runtime logs. For example, an agent receives input in the form of a simplified STRIDE-derived model (Threat: “Credential theft via kubelet”), combined with audit logs from the Kubernetes API server. The agent then synthesizes a YAML-formatted detection rule (e.g., in Falco or Sigma format), which can be versioned via Git and tested in CI/CD using adversary emulation (e.g., Caldera, Atomic Red Team).

3. Integrated Rule Development and Lifecycle Automation

By embedding the DaRC pipeline into the SSDLC, organizations gain the ability to treat detection development as a structured lifecycle, consisting of:

- Threat Modeling, conducted using lightweight tools like OWASP Threat Dragon or markdown-based models in Git.
- Rule Engineering, manual and agent-assisted generation of rules using YAML for Falco/Sigma or Rego for OPA Gatekeeper policies.
- Testing, simulation of attack techniques with adversary emulation frameworks and replayed logs.
- Tuning & Validation, manual and agent-guided noise filtering, correlation with other events, threshold adjustments.

- Deployment & Monitoring, automated delivery to cloud-native SIEMs or Kubernetes-native runtime enforcement systems.

This approach fosters feedback loops between security engineers and developers, where detection logic evolves together with the software product and infrastructure. As a result, false positives decrease, detection coverage improves, and rules remain context-aware - crucial for dynamic environments like AWS.

The metrics indicate that a Detection and Response as Code strategy can dramatically improve responsiveness and reliability of security operations. For example, integrating automation from the start shortens MTTR by enabling programmatic containment as soon as an alert fires - as emphasized, “with each detection use case, deliver not just an alert rule but a miniature response plan, this upfront effort pays off by reducing mean time to respond and analyst workload”. Detection coverage also increases since teams can iterate quickly on rule sets; in a traditional model, writing a new SIEM rule or playbook might lag behind emerging threats, whereas in code-as-code model, teams treat it like a sprint task. The use of threat-informed development (mapping detections to MITRE ATT&CK, etc.) further ensures coverage of relevant TTPs and informs automated response planning.

False positives - a bane of many SOC's - are reduced through test-driven detection development. By using unit tests and simulations (including breach and attack simulation tools) during rule creation, as advocated in Detection-as-Code practices, the rules are more precise and trustworthy upon deployment. The ability to quickly refine a rule by pushing a code update also means any false-positive issues can be corrected in hours, not waiting for the next product update or maintainer to adjust a black-box rule.

Finally, the maintainability and knowledge management benefits cannot be overstated. With code, institutional knowledge (why a detection exists, how it works) is preserved in comments, commit history, and tests. This fosters a culture of continuous improvement and learning. Version control also supports accountability, an important aspect for compliance – e.g., being able to show which team member approved a certain detection logic change and when, in essence, Detection and Response as Code not only improves technical metrics but also aligns security operations with the rigor and pace of modern software engineering.

4. Conclusion

Detection and Response as Code is a transformative approach that brings the benefits of software development methodologies to security operations. In the context of an AWS Kubernetes workload, we showed how codifying detection rules and response playbooks (potentially enhanced by LLM-driven agents) leads to more adaptive and automated cloud security. By leveraging YAML-based agent definitions, LangChain/Ollama for AI-driven analysis, and Git for robust versioning, security teams can embed their work into the SSDLC - ensuring that as the application evolves, so do its protective measures. Additionally, evaluation across SOC performance metrics suggests improved threat coverage, fewer false alarms, and drastically reduced response times when adopting a code-centric strategy. This work focused on a single cloud workload scenario; future research could apply the approach to multi-cloud or on-premise hybrid environments and measure outcomes empirically. There is also an opportunity to explore how advanced LLMs can not only generate

detection content, but also self-tune and explain security rules for even greater efficiency. Overall, Detection and Response as Code aligns with the DevSecOps ethos of integrating security into every stage of development – it empowers organizations to treat security rules with the same care as application code, leading to a more resilient and secure software lifecycle.

References

- [1] O. Vakhula, I. Opirskyy, O. Mykhaylova. “Research on Security Challenges in Cloud Environments and Solutions based on the ‘Security-as-Code’ Approach.” In Proc. of CPITS 2023-II, CEUR Workshop Proc. Vol-3550, pp. 55–69, 2023.
- [2] O. Vakhula, I. Opirskyy. “Research on Security as Code Approach for Cloud-Native Applications based on Kubernetes Cluster.” In Proc. of CSDP 2024, CEUR Workshop Proc. Vol-3800, pp. 58–69, 2024.
- [3] R. Kullberg. “How Detection-as-Code Revolutionizes Security Posture.” Panther Labs Blog, Dec 2023.
- [4] M. Yumrutaş. “Understanding Detection as Code: Integrating with Breach and Attack Simulation.” Picus Security Blog, Mar 2024.
- [5] M. Goldenberg. “Integrating and Prioritizing Response Automation in the Detection Lifecycle.” Detect/FYI Blog, Aug 2025.

AI-Driven Approaches to Safeguard Workers in Cyber-Physical Environments

Maksym Krainiuk¹, Oleh Shcherbak¹ and Olena Krainiuk¹

Abstract

The rapid growth of cyber-physical systems (CPS) in industrial environments introduces both efficiency and new safety risks for workers. This study explores AI-driven approaches for monitoring and mitigating hazards, focusing on noise exposure from road-construction machinery. Using extensive measurements from the DZ-99 grader and neural network predictive modeling, the system accurately forecasts noise levels and supports real-time safety interventions. The results demonstrate the practical value of AI integration for enhancing worker safety in automated cyber-physical environments.

Keywords

AI, cyber-physical systems, worker safety, predictive modeling, noise monitoring, industrial automation, neural networks, road-construction machinery

1. Introduction

The rapid growth of cyber-physical systems (CPS) in industry and infrastructure has improved efficiency and automation and introduced new risks to worker safety. Cyber threats or system malfunctions can directly affect physical processes, leading to accidents or hazardous conditions.

Traditional safety measures, relying mainly on human monitoring, are often insufficient. Artificial intelligence (AI) offers solutions through predictive analytics, anomaly detection, and intelligent decision support, helping to prevent accidents and enhance worker safety.

This article explores AI-driven approaches to safeguarding workers in CPS environments [1]. Using system models and classifying key threats, we discuss the application, benefits, and challenges of AI-based safety solutions in modern automated settings [2].

2. Statement of the problem

Cyber-physical environments – including industrial IoT, automated control systems, and robotics – have transformed modern industry by integrating computation, communication, and physical processes [3]. While these systems increase efficiency, they also introduce complex safety and security challenges.

Key vulnerabilities in such environments include network intrusions, unauthorized access, sensor malfunctions, and actuator errors. These weaknesses can compromise both system operations and worker safety. The main threats identified include:

- Unauthorized system access – allowing attackers to manipulate operations.
- Data manipulation and spoofing – corrupting sensor readings or control signals.
- Sensor/actuator failures – causing unintended physical actions.
- Human-machine interaction risks – including operator errors or unsafe interventions.
- Network and communication vulnerabilities – leading to disruption or delayed responses.

⁵⁸ Corresponding authors

✉ cyber@khadi.kharkov.ua (M. Krainiuk); olegcherbak@gmail.com (O. Shcherbak); olenauvarova@ukr.net (O. Krainiuk)

🆔 0009-0009-5640-1261 (M. Krainiuk); 0000-0002-7953-2135 (O. Shcherbak); 0000-0001-9524-040X (O. Krainiuk)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

These challenges underline the need for AI-driven approaches capable of real-time monitoring, predictive detection, and proactive mitigation to safeguard workers in increasingly automated environments.

Literature Review and Study Aim

Noise from road-construction machinery affects operator safety, equipment durability, and environmental conditions [4, 5]. Heavy machinery generates vibrations and high acoustic loads, which can reduce structural lifespan and cause hearing loss and stress [6, 7]. Studies show that bulldozers produce average noise levels of 73.87–84.11 dBA, with peaks up to 113.40 dBA [4], and operators often experience levels exceeding WHO recommendations [8–11]. Cab design and engine operating mode significantly influence noise exposure, with protective cabins reducing levels by up to 30.5 % [12]. Machine learning and AI methods have been successfully applied to predict noise levels with high accuracy (differences within 2 dB) [13].

Objective

This study aims to develop intelligent methods to optimize road-construction machinery operation while minimizing noise exposure. Using machine learning, we model noise levels, assess their effects on operators, equipment, and the environment, and propose guidelines for integrating AI-driven noise monitoring and mitigation strategies.

3. Results and discussion

3.1. AI-Based Safety Approaches

AI-driven methods can predict, monitor, and mitigate hazards to enhance worker safety in cyber-physical environments. This study focused on noise exposure from road-construction machinery, specifically the DZ-99 grader used in Kharkiv.

Data Collection and Standards Compliance. Noise measurements were conducted according to Ukrainian and international standards (e.g., ДСТУ ISO 1996-1, ДБН В.1.1-31, IEC 61672, WHO Guidelines) to ensure accuracy and comparability. Over 300 measurements were performed using a Benetech GM1351 sound level meter at various distances and operating modes, including inside the cabin and around the machine. Environmental conditions were controlled: flat, open terrain, low background noise, temperature 16.5 ± 1 °C, wind 3 ± 1 m/s, no precipitation. Measurements were performed after ensuring the machine had reached stable operating conditions, using a calibrated sound level meter at various distances and operating modes.

AI-Powered Predictive Modelling. A neural network (MLPRegressor) was trained to forecast noise levels based on input parameters: engine speed, distance to measurement point, and gear setting. The network architecture (3-8-1, 3-9-1, 3-10-1) achieved high accuracy: mean squared error ~ 0.009 , correlation coefficient $r = 0.9999$. Figure 1 illustrates the correlation between predicted and measured noise levels, highlighting the effectiveness of AI in anticipating hazardous exposure.

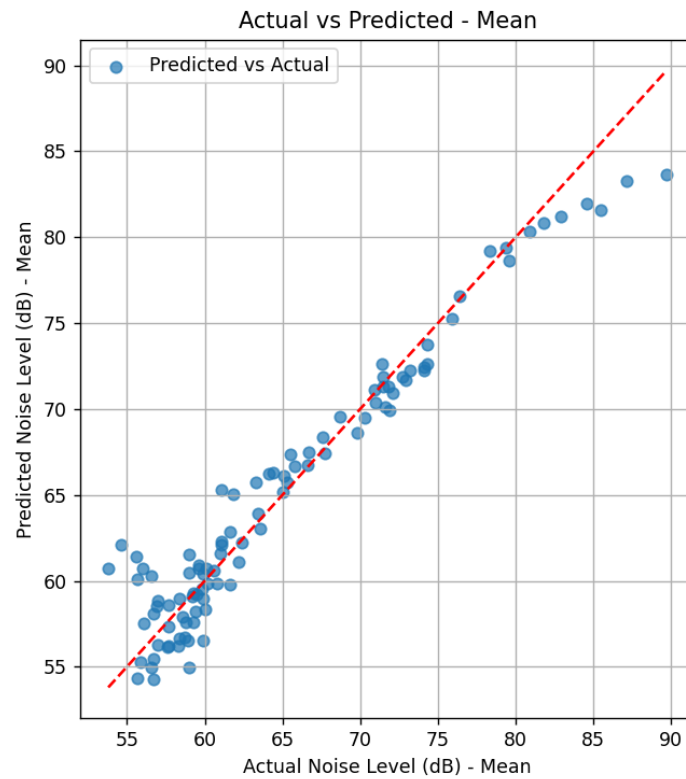


Figure 1: Figure 6 Correlation between predicted and measured average sound levels from a grader.

3.2. Analysis Results

- Noise variability decreased with distance, while relative measurement error also diminished.
- Engine speed and gear significantly affected noise, with human exposure highest at low-frequency ranges (31.5–2000 Hz).
- Statistical and ANOVA analyses confirmed microphone position as the dominant factor (>80% contribution) in overall noise variation.

Advantages of AI Integration

- Predictive Analytics: The model forecasts noise levels for different machine operations, enabling proactive safety measures.
- Real-Time Monitoring: AI dashboards can alert operators when predicted levels exceed safe thresholds.
- Decision Support: Insights from AI guide risk mitigation strategies, including optimal machine operation, protective equipment, and scheduling work to minimize exposure.

The current AI model focuses on the DZ-99 grader. Future work will expand to other road-construction machines, integrating additional factors such as surface type, temperature, and vibration, ultimately creating a comprehensive system for worker safety in industrial cyber-physical environments.

4. Conclusion

This study demonstrates the practical application of AI-driven approaches to enhance worker safety in cyber-physical environments. Key outcomes include:

- Integrating AI into monitoring systems enables automated assessment of hazardous conditions, reducing operator exposure to noise and improving workplace safety.
- A methodology for measuring and analyzing noise from the DZ-99 grader was developed, providing reliable data for AI-based predictive modelling.
- Experimental results highlighted critical factors affecting noise levels—engine speed, gear setting, and distance—allowing risk assessment and real-time hazard detection.
- The neural network model accurately predicts noise levels (correlation $r > 0.99$), confirming the practical value of AI in anticipating risks.
- The proposed intelligent monitoring system supports real-time data analysis, proactive interventions, and safer operation of machinery, illustrating how AI enhances cyber-physical system security.
- Future research will expand the AI model to include other road-construction machines, environmental variables (surface type, weather, equipment wear), and integration with adaptive safety and decision-support systems, advancing comprehensive protection for operators in complex industrial environments.

References

- [1] Krainiuk, M., Shcherbak, O. Intelligent Technologies for Optimising the Workplaces of Tractor Drivers: Safety and Comfort in Modern Machine and Tractor Units. In: Slavinska, O., Danchuk, V., Kunytska, O., Hulchak, O. (eds) Intelligent Transport Systems: Ecology, Safety, Quality, Comfort. ITSESQC 2024. Lecture Notes in Networks and Systems, vol 1335, Springer, Cham, 2025, pp. 211-222. doi:10.1007/978-3-031-87376-8_19
- [2] Krainiuk, M.Yu., Shcherbak, O.V. Zastosuvannya intelektualnykh vymiryvalnykh tekhnolohiy dlya analizu ekspluatatsiynykh parametru dorozhno-budivelnykh mashyn. Komunalne hospodarstvo mist, 2025, 1(189), 13–20. doi:10.33042/2522-1809-2025-1-189-13-20
- [3] Krainiuk, M.Yu., Shcherbak, O.V. Intelektualni systemy avtomatyzatsiyi kontrolyu za robochymy umovamy u kabinakh transportnykh zasobiv. 5-a mizhnarodna naukovotekhnichna konferentsiya «Intelektualni transportni tekhnolohiyi», Kharkiv, 25–27 lystopada 2024 r.: Tezy dopovidey, Kharkiv: UkrDUZT, 2024, pp. 137-138
- [4] Zaman, M.; Muslim, M.; Jehangir, A. Environmental noise-induced cardiovascular, metabolic and mental health disorders: a brief review. Environmental Science and Pollution Research, 2022, 29(51), 76485-76500
- [5] Hahad, O.; Kuntic, M.; Al-Kindi, S.; Kuntic, I.; Gilan, D.; Petrowski, K.; Daiber, A.; Münzel, T. Noise and mental health: evidence, mechanisms, and consequences. Journal of Exposure Science & Environmental Epidemiology, 2024 Jan 26, 1-8. doi:10.1038/s41370-024-00642-5
- [6] Serin, H.; Akay, A. Noise level analysis of a bulldozer used in constructing a forest road in Mediterranean region of Turkey. African Journal of Agricultural Research, 2010, 5, 2624-2628
- [7] Nourani, V.; Gökçekuş, H.; Umar, I.K. Artificial intelligence based ensemble model for prediction of vehicular traffic noise. Environmental Research, 2020, 180, 108852
- [8] Krainiuk, M.; Krainiuk, O.V.; Sherbak, O.V. Hygienic assessment of the working conditions of the tractor driver's workplace in modern machine-tractor units. Modern research in science

- and education. Proceedings of the 10th International scientific and practical conference. BoScience Publisher, Chicago, USA, 2024, pp. 263-273
- [9] Sehseh, E.E.; Abass Helmy, M.; Sorour, H.M. Noise test of two manufactured power tillers during transport on different local road conditions. *International Journal of Agricultural and Biological Engineering*, 2010, 3(4), 19-27. doi:10.3965/j.issn.1934-6344.2010.04.009-017
- [10] Behrooz Lar, M.; Pour, Z.K.; Payandeh, M.; Bagheri, J. Noise level of two types of tractor and health effect on drivers. *Journal of American Science*, 2011, 7(5), 382-387
- [11] Lashgari, M.; Maleki, A. Psychoacoustic evaluation of a garden tractor noise. *Agricultural Engineering International: CIGR Journal*, 2015, 17(3), 231-241
- [12] Hassan-Beygi, B.; Ghobadian, M.; Kianmehr, M.H.; Amiri, R. Prediction of a power tiller sound pressure levels in octave frequency bands using artificial neural networks. *Int J Agri Biol*, 2007, 3(9), 494-496
- [13] Choi, J.; Kang, H.; Hong, T.; Baek, H.; Lee, D. Automated noise exposure assessment model for the health of construction workers. *Automation in Construction*, 2021, 126, 103657. doi:10.1016/J.AUTCON.2021.103657

Secure and Energy-Efficient Object Recognition on Edge AI Devices

Vlad Lubin

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Abstract

This paper explores the integration of cybersecurity and energy efficiency in edge AI systems used for object recognition in drones, robots, and smart cameras. It identifies key vulnerabilities such as adversarial attacks, firmware tampering, and data interception, proposing both software and hardware countermeasures like lightweight encryption, secure boot, and Trusted Execution Environments. The study highlights the balance between strong security and low energy consumption under constrained hardware conditions. The proposed framework aims to enable trustworthy, energy-efficient object recognition for real-time applications in autonomous and intelligent edge devices.

1. Introduction

Edge AI devices (such as drones, robots, and smart cameras) increasingly perform on-board object recognition to enable real-time, intelligent decision-making close to the data source. Modern deep neural networks (DNNs) for vision deliver high accuracy but at the cost of heavy computation and energy demand, which is problematic for power-constrained edge devices. At the same time, deploying AI at the edge introduces new cybersecurity challenges: unlike controlled cloud servers, edge devices operate in dynamic, often unprotected environments and face novel attack vectors. These intelligent cyber-physical systems must therefore balance energy efficiency with robust security by design. This thesis proposal explores the threat landscape of edge-based object recognition and outlines both software and hardware approaches to secure data and AI models on resource-limited devices.

2. Threat landscape for Edge AI vision systems

Object recognition systems at the edge are exposed to a broad range of threats. Adversarial attacks against AI models are a significant concern – even tiny perturbations in camera input can cause DNNs to misclassify objects, potentially leading a surveillance UAV or autonomous robot to overlook critical obstacles or targets. Attackers may also target the sensor and data pipeline: for example, spoofing a drone’s GPS or vision sensors can hijack its navigation or feed it false imagery. Another major risk is firmware tampering or malware injection. A malicious firmware update could alter an autonomous drone’s control parameters, causing erratic or unsafe behavior. Such persistence at the firmware level can be hard to detect and mitigate without special protections, underscoring the need for secure boot and runtime integrity checks. Moreover, edge AI devices often communicate results or receive commands wirelessly; if communications are not encrypted and authenticated, eavesdropping or man-in-the-middle attacks could leak sensitive imagery or inject false data. Devices deployed in the field can even be subject to physical capture or tampering, enabling attackers to extract AI models or data if those are not securely stored. Finally, the resource constraints themselves present a vulnerability: attackers might launch denial-of-service by draining device battery or overloading the AI model (e.g. forcing costly computations). For instance, power-depletion attacks can exhaust an IoT camera’s battery, and adversaries can exploit the limited compute by triggering worst-case processing loads that degrade performance. This

diverse threat landscape mandates a combination of defensive measures in software and hardware to protect edge AI systems.

3. Software security approaches

Despite constrained resources, several software-level techniques can significantly improve the security of edge AI object recognition:

3.1. Lightweight data encryption

Protecting sensitive sensor data and AI outputs via encryption is critical, but standard cryptographic algorithms can be too slow or energy-intensive for small devices. Lightweight encryption algorithms have been developed to address this gap. For example, NIST’s new lightweight cryptography standard defines ciphers (from the Ascon family) that require much less computing power and time than traditional methods, yet still secure data on IoT-class devices . Using such algorithms for encrypting images, telemetry, and communications (e.g. via lightweight TLS/DTLS) helps prevent data interception or tampering without overly taxing the device.

3.2. Secure boot and firmware integrity

Edge AI devices must ensure they boot with authentic, untampered software. Secure boot mechanisms establish a hardware-based root of trust to verify each stage of the bootloader and firmware using digital signatures. If any code has been modified or is not properly signed by the device OEM, the device will refuse to execute it, halting the boot process to avoid running malicious code. This prevents attackers from persistently installing backdoors or altered AI models. Every software component, from the low-level bootloader to the object recognition model, should be signed and verified before execution. In practice, drone manufacturers already implement multi-stage secure boot; for instance, DJI drones encrypt and sign each layer of firmware (from boot ROM up to the flight control OS), and will not start if any layer fails an integrity check. Additionally, firmware update packages are often encrypted and signed before delivery to the device, ensuring that only verified updates can be applied. Together, secure boot and signed updates defend against unauthorized firmware or model modifications.

4. Hardware security measures

In tandem with software defenses, hardware-based security measures establish a robust foundation for protecting edge AI systems. Modern edge devices are increasingly equipped with the following hardware features to secure data and computations:

4.1. Trusted Execution Environments (TEEs)

A TEE is a secure enclave within the processor that runs code in isolation from the main operating system. For example, ARM TrustZone technology divides the system into secure and normal worlds. Sensitive operations – such as handling cryptographic keys, running parts of the object recognition model, or processing confidential data – can execute in the secure world, where they are concealed from any malware in the normal OS. Even if an attacker compromises the device’s application layer, they cannot directly access data or code running inside the TEE. In practice, DJI’s drones use an ARM TrustZone-based TEE to keep device keys, certificates, and user data safe from tampering. Similarly, secure enclaves on other edge AI platforms ensure that critical algorithms (like model inference on sensitive inputs) can run with confidentiality and integrity

guarantees. Research has shown that properly-implemented TEEs dramatically enhance the confidentiality and integrity of edge AI operations, though performance overhead is a consideration.

4.2. Secure Elements and Key Storage

A secure element (SE) is a dedicated tamper-resistant chip used to store cryptographic keys, certificates, and other secret data. These chips provide an isolated execution environment with their own minimal CPU/memory and are engineered to be highly resistant to physical and logical attacks. By embedding a secure element, an edge device (e.g. a smart camera or drone) can safeguard private keys for encryption, device authentication, or firmware signing. Even if an attacker probes the device hardware, the SE will protect its contents against extraction. For instance, keys might be injected at manufacturing into an on-board secure element or one-time-programmable memory area; those keys never appear in plaintext to the main processor. The secure element can perform cryptographic operations (signing, decrypting) on behalf of the device so that the main CPU never handles raw secret material. This hardware-root-of-trust ensures that critical credentials remain safe even under intense attack. The limitation of many secure elements, however, is their constrained performance and fixed functionality – meaning they are used for specific tasks like key management rather than general compute. Nonetheless, combining a TEE for flexible secure processing with an SE for robust key storage provides strong security coverage.

5. Challenges under resource constraints

Implementing robust cybersecurity on edge AI systems is challenging due to stringent power and compute limitations. Every added layer of security (encryption, authentication, isolation) can introduce overhead in terms of CPU cycles, memory usage, and energy consumption. Designers must carefully balance security against the device’s real-time performance and battery life. For instance, strong encryption and communication protocols may significantly increase latency and power draw: securing a tiny 10-byte sensor reading can require kilobytes of additional data in headers and handshakes, consuming extra bandwidth and energy. IoT devices often have to trade off security and efficiency; if computational load or battery drain becomes too high, the device might scale back cryptographic protections, resulting in weaker security during low-power states. Techniques like TLS, while vital for confidentiality, can be heavy for a microcontroller handling video streams or frequent detections. Similarly, running inference inside a TEE or performing frequent integrity checks can slow down object recognition throughput, conflicting with real-time requirements. TEEs and encryption engines themselves draw power and can produce heat, which on a drone or wearable is at a premium. Research confirms that TEEs, for all their security benefits, do introduce performance and scalability limits on constrained devices.

Another challenge is memory and storage overhead. Secure implementations often require storing extra keys, certificates, or logs; memory encryption might reserve a portion of RAM for secure operation, effectively reducing what’s available for AI tasks. On tiny embedded platforms, this can be significant. Computational overhead is also an issue: lightweight ciphers help, but cryptographic operations still take time away from AI processing. Edge AI devices must also handle intermittent connectivity and cannot always rely on cloud assistance for heavy security tasks (though offloading some security processing to the network is one strategy). In summary, achieving strong security under tight energy and compute budgets requires careful optimization and sometimes accepting compromises in security level or device functionality. This thesis will investigate methods to mitigate these tensions – for example, by using hardware accelerators for

crypto, adopting adaptive security that scales with available resources, and designing AI models that are inherently more robust to attacks so that fewer add-on defenses are needed.

6. Conclusion

In summary, the proposed thesis will investigate a holistic approach to securing object recognition systems on edge AI devices, marrying software and hardware strategies under the constraints of limited energy and computing power. We will analyze the unique threat landscape faced by edge AI – from adversarial inputs to firmware attacks – and develop a layered defense paradigm that includes lightweight encryption, secure boot, trusted execution, and more. The aim is to enable trustworthy and resilient edge AI in applications like surveillance UAVs, autonomous robots, and smart cameras, without sacrificing the efficiency that makes on-device intelligence attractive. By exploring cross-disciplinary techniques and optimizing for the edge context, this research will contribute both conceptual frameworks and practical solutions for integrating cybersecurity into AI at the edge. The outcome will help pave the way for future deployments of edge AI that are not only intelligent and energy-efficient but also robust against cyber threats, thus accelerating adoption in safety- and privacy-critical domains. Through publications and a proof-of-concept implementation, we hope to influence design principles and standards for AI in cybersecurity, ensuring that as these technologies converge, they do so in a secure and sustainable manner.

Declaration on Generative AI

During the preparation of this work, the author used Grammarly in order to: Grammar and spelling check. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication’s content.

References

- [1] Kendryte Datasheet. Canaan Inc. [Electronic resource]. – Access mode: https://github.com/kendryte/k510_docs
- [2] NIST Lightweight Cryptography Project. National Institute of Standards and Technology. [Electronic resource]. – Access mode: <https://csrc.nist.gov/projects/lightweight-cryptography>
- [3] ARM TrustZone Technology Overview. ARM Limited. [Electronic resource]. – Access mode: <https://developer.arm.com/Architectures/TrustZone>
- [4] DJI Security White Paper. DJI Technology Co. [Electronic resource]. – Access mode: <https://security.dji.com/data-security>
- [5] Edge AI and Vision Alliance. Low-Power Computer Vision at the Edge. [Electronic resource]. – Access mode: <https://www.edge-ai-vision.com>
- [6] GDPR: General Data Protection Regulation. European Union. [Electronic resource]. – Access mode: <https://gdpr.eu/>

Direct Preference Optimization Using Synthetic Data: Domain-Specific Model Training and Benchmarking for GPT models

Bohdan Pavlyshenko¹, Ivan Bulka²

¹ Ivan Franko National University of Lviv, Lviv, Ukraine, Drahomanova St, 50, Lviv, Lviv Oblast, 79000

² Ivan Franko National University of Lviv, Lviv, Ukraine, Drahomanova St, 50, Lviv, Lviv Oblast, 79000

Abstract

Large Language Models (LLMs) have revolutionized natural language processing, but their usage in specialized domains such as finance is complicated by limitations in domain-specific understanding and the use of specific terminology. This study investigates the adaptation of LLMs, focusing on Meta-Llama-3-8B-Instruct, for advanced financial NLP tasks through a combination of Parameter-Efficient Fine-Tuning (PEFT) techniques, specifically LoRA and QLoRA, and Direct Preference Optimization (DPO) methods. Utilizing the open-source Sujet-Finance-Instruct-177k dataset, which covers six core financial NLP tasks, we demonstrate that PEFT approaches improve performance in tasks such as sentiment analysis and topic classification, while showing limited efficiency in complex generative tasks like question answering. To address this gap, we introduce DPO using synthetically generated preference pairs, enabling supervised alignment based on human-like feedback. Experimental results reveal that DPO enhances the model's performance in challenging question-answering tasks, as evidenced by increased LLM-based evaluation scores. Our findings highlight that while PEFT methods offer efficient domain adaptation, augmenting them with supervised preference optimization is crucial for optimal performance in financial applications.

Keywords

LLM, fine-tuning, DPO, Llama, LoRA, QLoRA, GPT ⁵⁹

1. Introduction

Large Language Models (LLMs) made a big impact in the field of natural language processing by leveraging transformer-based architectures [1] to achieve near-human text generation and comprehension. However, while these general-purpose models demonstrate impressive capabilities, they often have limitations when applied to specialized domains such as finance, where deep domain expertise and the accurate use of technical terminology are essential. Effectively adapting LLMs to bridge the gap between general linguistic proficiency and domain-specific requirements remains a key challenge.

Furthermore, beyond the generation of text, there is an increasing demand for LLMs to produce outputs that are not only accurate but also optimally aligned with specific user objectives and domain-specific constraints. This underscores the necessity of providing explicit instructions or guidance to steer the models toward generating outputs that best fulfill the intended tasks..

⁵⁹ Corresponding authors

✉ bohdan.pavlyshenko@lnu.edu.ua (B. Pavlyshenko); ivan.bulka@lnu.edu.ua (I. Bulka)

ORCID 0000-0001-9515-3488 (B. Pavlyshenko); 0009-0003-2962-7931 (I. Bulka)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Problem and Scientific Novelty

While Retrieval-Augmented Generation (RAG) [2] can help language models become more knowledgeable by allowing them to look at extra information, it has some drawbacks. RAG can not efficiently handle some domain-specific tasks, understand terminology, or have a good understanding of the domain.

To address these issues, we unified Parameter-Efficient Fine-Tuning (PEFT) methods such as LoRA and QLoRA [3]. These techniques allow us to adapt large language models like Meta-Llama-3-8B-Instruct to financial tasks without making the training process too costly or complicated. PEFT methods make it easier to customize models for specialized areas by updating only a small part of the model's parameters [4, 5].

In addition to these approaches, we used Direct Preference Optimization (DPO) [6]. DPO helps models learn directly from human feedback by showing them which responses are preferred. This technique is especially useful when fine-tuned models still make mistakes or give unsatisfactory answers. By using DPO, we aim to further improve the model's performance in financial applications, ensuring it provides more accurate and useful outputs even in difficult cases [7]. We will explore how much DPO can enhance results when standard fine-tuning does not deliver strong enough performance.

3. Results

3.1. Model Training

The experimental framework used the Sujet-Finance-Instruct-177k dataset, an open-source collection from Hugging Face containing 177,000 records across six core financial NLP tasks: sentiment analysis, question answering (standard, with context, yes/no, conversation-based), and topic classification. The dataset was divided into training (85%), validation (10%), and test (5%) splits, ensuring even class distribution.

Initial model fine-tuning was performed on Meta-Llama-3-8B-Instruct utilizing PEFT approaches, specifically LoRA and QLoRA, with standardized configurations to optimize training efficiency on A100 GPUs. Detailed model parameters are described in Table 1.

Table 1

Model Training Config

Parameter	Description	Value
<code>lora_alpha</code>	LoRA scaling factor	16
<code>lora_dropout</code>	Dropout parameter to reduce overfitting	0.1
<code>r</code>	Matrix rank relates to the number of trainable parameters	64
<code>learning_rate</code>	Learning rate	1e-6

optimizer	Optimizer	adam

To further enhance performance on complex tasks, particularly question answering, we introduced Direct Preference Optimization (DPO) during subsequent fine-tuning stages. DPO enables the model to explicitly learn from preference signals by contrasting positive and negative response samples.

3.2. DPO Dataset Construction

Since DPO requires pairs of positive (preferred) and negative (dispreferred) outputs for each training instance, we generated these pairs synthetically. The original dataset answers were treated as positive samples.

To source negative samples, we employed GPT-4o to systematically rewrite the original answers according to specific criteria:

- Simplification: Important information was omitted, thereby reducing answer completeness.
- More Context: Extra, often irrelevant, information was inserted.
- Tone Change: The answer tone was intentionally altered (e.g., shifting between formal or informal), which could impact clarity or appropriateness.

Prompts to GPT-4o included explicit editing instructions and answer modifications, with rewritten outputs collected as corresponding negative samples. Only the modified answers were retained, ensuring a clean dataset for DPO training. With the Synthetic dataset generation, we generated 33k samples for the question-answering user case and used them to tune a model [8].

3.3. DPO training configuration

DPO fine-tuning used the paired positive and negative samples to instruct the model to distinguish high-quality responses from suboptimal alternatives. This preference-based training directly optimizes the model to prefer more appropriate, contextually accurate, and informative outputs for financial question-answering tasks. DPO training configuration provided in Table 2.

Table 2

DPO training configuration

Parameter	Description	Value
lr_scheduler_type	Scheduler type	constant
warmup_steps	Number of warmup steps	150
learning_rate	Learning rate	1e-6
optimizer	Optimizer	adam

3.4. Evaluation

Model evaluation employed LLM-based scoring (using GPT-4 as an evaluator) [9] over the six task categories and demonstrated that, while previous fine-tuning approaches yielded improvements for simpler tasks, they provided limited gains for complex question-answering scenarios (Table 3). Following DPO-based tuning, a marked increase in LLM evaluation scores (Table 4) was observed specifically in the question-answering category, highlighting the effectiveness of supervised preference optimization in specialized financial NLP applications.

Table 3

Metrics evaluation (on test data, metric-score 1 ...5 using GPT-4 model)

Task type	Base model	Tuned model
Question-answering	4.77	4.67
QA in conversation	4.43	4.18
QA with context (RAG)	4.75	4.9
Sentiment analysis	4.01	4.03
Topic classification	3.89	4.26
Yes / No questions	2.9	3.14
Avarage	4.13	4.2
Weighed average	4.27	4.33

Table 4

DPO metrics evaluation

Task type	Base model	Tuned model	DPO tuning
Question-answering	4.77	4.67	4.89

4. Conclusion

Our study examined the adaptation of large language models (LLMs) for specialized financial language processing tasks. Initial experiments leveraging parameter-efficient fine-tuning (PEFT) approaches such as LoRA and QLoRA achieved moderate improvements across a range of financial

NLP benchmarks, particularly for tasks like sentiment analysis and topic classification. However, our findings also exposed the limitations of PEFT-based adaptation - particularly for complex generative scenarios such as question answering, where conventional fine-tuning did not confer substantial performance gains.

To address these limitations, we used Direct Preference Optimization (DPO) as an additional supervised alignment step, generating preference-labeled training pairs via controlled outputs from GPT-4o. The integration of DPO showed better results, evidenced by clear improvements in LLM-based evaluation scores for challenging question-answering tasks. These outcomes validate the hypothesis that preference-driven optimization can enhance LLM performance in sensitivity-critical, expert-knowledge domains.

Overall, our results suggest that while efficient fine-tuning methods remain invaluable for bridging generalist models into niche sectors, their efficiency is further amplified when coupled with instruction-level supervision that aligns outputs with domain-relevant user expectations. Future research should explore hybrid approaches, leveraging both retrieval augmentation and advanced tuning, to maximize LLM reliability and utility in highly specialized professional contexts such as finance.

References

- [1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [2] Arslan, M., Ghanem, H., Munawar, S., & Cruz, C. (2024). A Survey on RAG with LLMs. *Procedia computer science*, 246, 3781-3790.
- [3] Arslan, M., Ghanem, H., Munawar, S., & Cruz, C. (2024). A Survey on RAG with LLMs. *Procedia computer science*, 246, 3781-3790.
- [4] Pavlyshenko, B., & Bulka, I. METRIC-BASED COMPARISON OF FINE-TUNED LLAMA 2 AND MIXTRAL LARGE LANGUAGE MODELS FOR INSTRUCTION TASKS. *Electronics and information technologies/Електроніка та інформаційні технології*, (26).
- [5] Pavlyshenko, B., & Bulka, I. (2025). PARAMETER EFFICIENT FINE-TUNING AND OVERFITTING IN GPT LARGE LANGUAGE MODELS: A METRIC-BASED COMPARISON. *Electronics and information technologies/Електроніка та інформаційні технології*, (30), 33-42.
- [6] Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., & Finn, C. (2023). Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36, 53728-53741.
- [7] Zhou, Z., Liu, J., Yang, C., Shao, J., Liu, Y., Yue, X., ... & Qiao, Y. (2023). Beyond one-preference-for-all: Multi-objective direct preference optimization.
- [8] Dandekar, A., Zen, R. A., & Bressan, S. (2018, August). A comparative study of synthetic dataset generation techniques. In *International Conference on Database and Expert Systems Applications* (pp. 387-395). Cham: Springer International Publishing.
- [9] Gu, J., Jiang, X., Shi, Z., Tan, H., Zhai, X., Xu, C., ... & Guo, J. (2024). A survey on llm-as-a-judge. *arXiv preprint arXiv:2411.15594*.

Multilingual Knowledge Graphs for Thematic Text Analysis

Bohdan Pavlyshenko¹ and Mykola Stasiuk¹

¹ Ivan Franko National University of Lviv, Drahomanova 50, 79000 Lviv, Ukraine

Abstract

The multilingual nature of cybersecurity discourse poses significant challenges for extracting structured knowledge and identifying thematic fields. Traditional keyword-based monitoring lacks scalability and fails to capture semantic relationships between concepts across languages. This work proposes an approach for constructing bilingual knowledge graphs from English and Ukrainian cybersecurity texts, enabling the discovery of thematic fields through unsupervised methods. Large Language Models (LLMs) are used to automatically extract entities and relations relevant to cyber threats, providing the foundation for graph construction. The resulting knowledge graphs are clustered to reveal hidden themes, while association rule mining offers interpretable co-occurrence patterns that validate the discovered structures. To ensure comparability across languages, multilingual embeddings align clusters, and LLM-based summarization generates concise thematic labels for analysts. The novelty of the approach lies in its unique combination of neural techniques, like LLM-based entity extraction and multilingual embeddings, with symbolic methods, such as association rule mining, to support multilingual threat intelligence. Instead of predefined taxonomies, the framework utilizes new information and offers interpretable insights that can aid in security analysis in tracking disinformation, identifying emerging threats, and structuring complex data.

Keywords

knowledge graphs, thematic field discovery, large language models, association rules, multilingual text mining; cyber threat intelligence

1. Introduction

Cybersecurity is increasingly challenged by the rapid expansion of digital information sources in multiple languages [1]. Analysts are confronted with the need to extract structured knowledge from texts in order to identify new threats, understand adversarial tactics, and track disinformation campaigns. Traditional approaches based on keywords or classification are insufficient for this task because they lack scalability, adaptability, and semantic precision.

Knowledge Graphs (KGs) provide a flexible representation of entities and their relationships, making them suitable for capturing the complexity of discourse related to security. However, constructing and aligning KGs across languages requires advanced techniques for entity recognition, semantic clustering, and thematic interpretation.

This paper proposes a methodology for building bilingual knowledge graphs from English and Ukrainian text corpora and for discovering thematic fields using graph clustering, association rules, and LLMs. The chosen focus on English and Ukrainian is motivated by the importance of monitoring multilingual cybersecurity narratives in the context of hybrid warfare and disinformation campaigns.

⁶⁰ Corresponding authors

✉ bohdan.pavlyshenko@lnu.edu.ua (B. Pavlyshenko); mykola.stasiuk@lnu.edu.ua (M. Stasiuk)

🆔 0000-0001-9515-3488 (B. Pavlyshenko); 0009-0006-5256-2103 (M. Stasiuk)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Problem Statement & Scientific Novelty

The increasing multilingual nature of cybersecurity discourse presents several challenges for threat intelligence. The primary issue include the automated extraction of entities and their relationships from unstructured, multilingual cybersecurity text, as well as the effective construction and clustering of knowledge graphs to reveal hidden thematic fields. Furthermore, a hurdle lies in the cross-lingual alignment of thematic clusters across different corpora, specifically English and Ukrainian, to ensure comparability and a unified understanding. Lastly, a significant challenge is ensuring the interpretability of discovered themes, as the ultimate goal is to provide insights that can directly support decision-making processes.

Existing research has studied KGs in cybersecurity [2], but in monolingual settings and without integrating unsupervised thematic discovery with symbolic rule-based reasoning. Moreover, few approaches directly address bilingual corpora that reflect parallel but not identical information spaces.

The proposed methodology introduces novelty at different levels:

- Integration of LLMs for Entity Extraction: LLMs were utilized to automatically extract entities relevant to cybersecurity such as “malware names”, “vulnerabilities”, “organizations” and relations, e.g. “uses”, “targets”, “causes”. This reduces manual effort and increases coverage.
- Unsupervised Thematic Discovery: Graph clustering allows the discovery of thematic fields without predefined taxonomies, which enables adaptive monitoring of emerging threats.
- Hybrid Symbolic–Neural Reasoning: Association rule mining validates clusters with interpretable co-occurrence rules, bridging the gap between opaque clustering and human-usable insights.
- Cross-Lingual Knowledge Alignment: Multilingual embeddings align English and Ukrainian clusters, ensuring comparability across corpora.
- Human-Readable Summaries: LLMs generate concise thematic labels and descriptions of clusters, improving interpretability.

3. Methodology.

The framework for this research, demonstrated in Figure 1, consists of a series of interconnected stages. Initially, the process begins with data collection, which involves gathering corpora related to cybersecurity from both English and Ukrainian sources, including technical reports, threat bulletins, and news. While traditional methods for this task have relied on techniques as presented in TTPDrill [3], the proposed framework extends these capabilities. Following this, LLM-based entity and relation extraction, as explored by Xu et al. [4], is performed to identify named entities and the semantic relationships between them automatically.

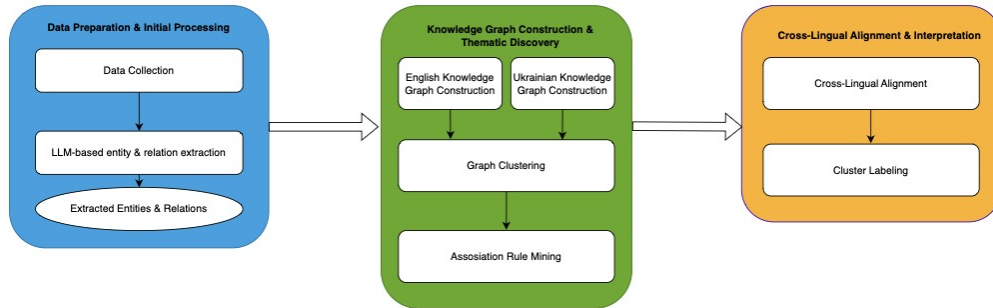


Figure 1: Architecture of the Bilingual Knowledge Graph Construction and Thematic Discovery Framework.

The extracted entities and relations form the basis for the subsequent knowledge graph construction. Separate graphs are built for each language, with entities represented as nodes and relations as edges. To uncover thematic fields, graph clustering is performed using unsupervised methods like community detection, which groups related subgraphs. This is followed by association rule mining to derive co-occurrence rules, which serve to validate whether the created clusters reflect meaningful cybersecurity topics. For cross-corpus analysis, cross-lingual alignment is achieved with multilingual embeddings [5] to ensure comparability between the English and Ukrainian clusters. Finally, cluster labeling was performed, where LLMs were leveraged to generate concise and human-readable summaries for each cluster, such as "Phishing attacks targeting MFA systems," thereby improving interpretability.

4. Expected results

The proposed approach is expected to yield several key results. First, it will produce parallel bilingual knowledge graphs that are aligned across English and Ukrainian, offering a structured representation of multilingual cybersecurity discourse. Second, the methodology will reveal thematic fields in this discourse, demonstrating its ability to adaptively identify new topics without relying on predefined taxonomies. Furthermore, the framework will provide interpretable insights through the use of association rules and LLM-based summarization, making complex information accessible. Ultimately, these results are anticipated to provide valuable data in monitoring multilingual cyber threat intelligence and countering disinformation campaigns.

5. Conclusions

In conclusion, this work proposes a novel and comprehensive bilingual methodology for thematic field discovery in cybersecurity. The framework leverages a unique integration of neural embeddings, symbolic reasoning, and multilingual alignment to structure and interpret complex, multilingual information flows. By focusing on English and Ukrainian, this approach directly addresses the need to monitor and understand cybersecurity narratives in the context of hybrid threats and disinformation. The resulting system is anticipated to offer a scalable and adaptable tool for threat intelligence, moving beyond traditional methods to provide interpretable insights that can aid in a dynamic threat landscape.

References

- [1] Admass, W., Yayeh, Y., and Diro, A. "Cyber Security: State of the Art, Challenges and Future Directions." *Cyber Security and Applications* 2 (2023): 100031. doi:10.1016/j.csa.2023.100031.
- [2] Mittal, S., Joshi, A., and Finin, T. Cyber-All-Intel: An AI for Security related Threat Intelligence, arXiv:1905.02895, 2019. doi:10.48550/arXiv.1905.02895.
- [3] Husari, G., Al-Shaer, E., Ahmed, M., Chu, B., and Niu, X. "TTPDrill: Automatic and Accurate Extraction of Threat Actions from Unstructured Text of CTI Sources," in: *Proceedings of the 33rd Annual Computer Security Applications Conference, ACSAC '17*, Association for Computing Machinery, New York, NY, 2017, pp. 103–115. doi:10.1145/3134600.3134646.
- [4] Xu, D., Chen, W., Peng, W., Zhang, C., Xu, T., Zhao, X., Wu, X., Zheng, Y., Wang, Y., and Chen, E. Large Language Models for Generative Information Extraction: A Survey, arXiv:2312.17617, 2024. doi:10.48550/arXiv.2312.17617.
- [5] Pavlyshenko, B., and Stasiuk, M. "SEMANTIC SIMILARITY ANALYSIS USING TRANSFORMER-BASED SENTENCE EMBEDDINGS." *Electronics and information technologies/Електроніка та інформаційні технології* (30) (2025): 43-58. doi:10.30970/eli.30.4.

A hybrid ML approach for anomaly detection in information networks

Andriy Riiy, Marian Kyryk, Taras Kolodiy

Lviv Polytechnic National University, S. Bandera Street 12, 79013 Lviv, Ukraine

Abstract

The growing complexity of cyber threats and the ever-increasing network traffic volumes highlight the need for adaptive anomaly detection methods. This paper presents a comparative analysis of unsupervised algorithms (Isolation Forest, Local Outlier Factor, One-Class SVM, Elliptic Envelope) and introduces a hybrid approach integrating Isolation Forest, LOF, and DBSCAN. Experiments on simulated datasets containing 10,000 records with 10% anomalies demonstrated that the hybrid model achieved an F1-score of 88.0%, outperforming the standalone algorithms while preserving computational efficiency. The results emphasize the potential of hybrid unsupervised methods for enhancing intrusion detection in large-scale information systems.

Keywords

anomaly detection, intrusion detection systems, unsupervised algorithms, hybrid model, information networks

1. Introduction

With the growth of digital technologies and information networks, intrusion detection systems (IDS) face new challenges. Signature-based approaches are ineffective against zero-day and unknown attacks, which exploit previously unknown or unpatched vulnerabilities, necessitating machine learning-based anomaly detection. Recent studies have compared supervised, unsupervised, and meta-learning approaches for intrusion detection, highlighting their strengths and limitations in detecting unknown attacks [1]. Unsupervised methods, in particular, are valuable because they can detect anomalies without labeled data, making them suitable for dynamic and evolving network environments. A comprehensive evaluation by Agyemang (2024) further emphasizes the effectiveness of unsupervised machine learning algorithms in anomaly detection tasks [2].

2. Methodology

Experiments were conducted using three publicly available datasets:

1. **NSL-KDD** – an improved version of the KDD’99 dataset, widely used for benchmarking IDS.
2. **UNSW-NB15** – a modern dataset containing nine categories of real-world attacks.
3. **CICIDS2017** – a large-scale dataset with realistic traffic, including DoS, DDoS, brute force, and botnet scenarios

Data preprocessing included normalization (Z-score, Min-Max scaling) and feature selection. The following algorithms were evaluated: Isolation Forest, Local Outlier Factor, One-Class SVM,

⁶¹ Corresponding authors

andrii.i.rii@lpnu.ua (A. Riiy); marian.i.kyryk@lpnu.ua (M. Kyryk); taras.b.kolodiy@lpnu.ua (T. Kolodiy)
0009-0005-0252-1533 (A. Riiy); 0000-0001-9156-9347 (M. Kyryk); 0009-0008-4777-320X (T. Kolodiy)

and Elliptic Envelope. Additionally, a hybrid model combining Isolation Forest, LOF, and DBSCAN was developed to capture both global and local anomalies, following principles proposed in recent semi-supervised deep learning approaches for anomaly detection [3], hybrid ensemble models in IoT environments [4], and enhanced DBSCAN-based pruning for attack detection [5]. Performance was measured using Precision, Recall, F1-Score, AUC-ROC, and AUC-PR.

The proposed hybrid model consists of three main components. The first component is the Isolation Forest module for global analysis, which employs an ensemble of 100 trees (contamination=0.1) to rapidly identify global anomalies, such as DDoS attacks, and to pre-filter traffic [6]. The second component is the Local Outlier Factor module ($n\_neighbors=20$), applied to the points selected by Isolation Forest for more detailed examination of local anomalies, for example, rare protocol-based attacks. The third component is an adaptive clustering module based on the DBSCAN algorithm, which identifies unclustered points as potential anomalies. This approach is effective for handling multimodal data distributions typical of network traffic, which contains diverse patterns of normal behavior and attacks. DBSCAN is used as the final step after global detection by Isolation Forest and local detection by LOF, ensuring comprehensive anomaly identification. Its key parameters, eps and $min\_samples$, are adjusted dynamically according to statistical characteristics of the data, such as distance distribution and density. This adaptive approach is scientifically grounded, as it allows the algorithm to account for local data properties, which is critical for anomaly detection in heterogeneous datasets.

3. Results

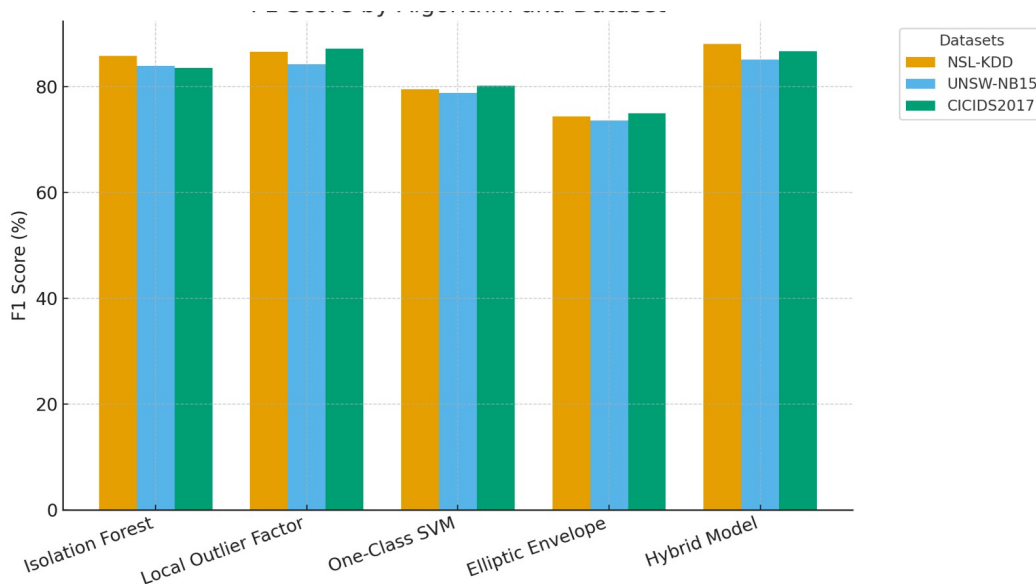


Figure 1: Comparison of anomaly detection models' effectiveness based on F1-score

From the diagram, it can be seen that Isolation Forest achieves the best balance between detection quality and efficiency, reaching F1-scores of 85.8% on NSL-KDD, 83.9% on UNSW-NB15, and 83.5% on CICIDS2017, making it highly suitable for large-scale traffic analysis. Local Outlier Factor (LOF) performs particularly well in detecting local anomalies, achieving the highest F1-score on CICIDS2017 (87.1%), while scoring 86.6% on NSL-KDD and 84.2% on UNSW-NB15, though at the cost of increased computational demands. One-Class SVM shows moderate results across all datasets, with F1-scores of 79.5%, 78.8%, and 80.2%, and is sensitive to data scaling. Elliptic Envelope

consistently underperforms, with F1-scores of 74.3%, 73.6%, and 74.9%, highlighting its limitations for non-Gaussian traffic distributions. The proposed hybrid model demonstrates the most robust overall performance, reaching F1-scores of 88.0% on NSL-KDD, 85.1% on UNSW-NB15, and 86.7% on CICIDS2017, maintaining stable detection quality across all datasets.

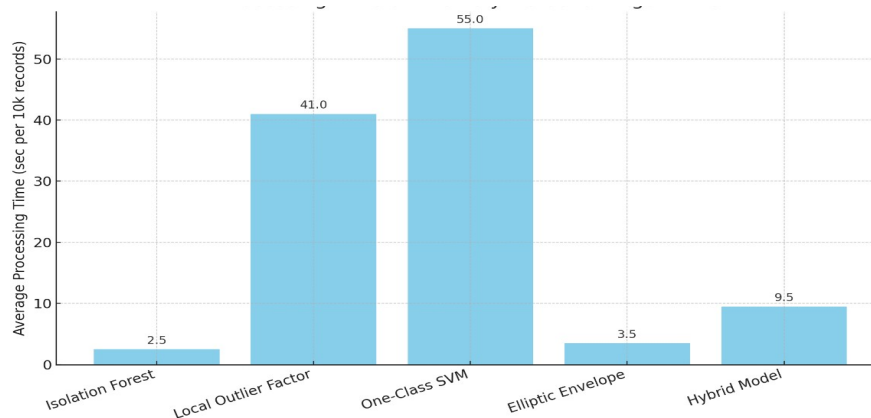


Figure 2: Comparison of the average processing time of anomaly detection models

As the diagram illustrates, Isolation Forest processes data most quickly, supporting efficient large-scale analysis. The hybrid model remains practical for near real-time detection, while LOF and One-Class SVM demand considerably more processing time, reflecting the balance between performance and speed.

4. Discussion and Conclusion

The hybrid approach demonstrated better robustness across datasets compared to standalone models. Its ability to adapt to heterogeneous traffic patterns and maintain balanced detection quality makes it a promising direction for IDS in complex information networks.

Future research will focus on extending the hybrid framework with deep learning architectures, such as convolutional and recurrent neural networks, which have shown strong performance in detecting network anomalies [7]. Additionally, validating the approach on streaming data will be a key step toward achieving real-time anomaly detection in operational environments.

References

- [1] T. Zoppi, A. Ceccarelli, T. Puccetti, A. Bondavalli, Which algorithm can detect unknown attacks? Comparison of supervised, unsupervised and meta-learning algorithms for intrusion detection, *Computers and Security* 127 (2023) 103107. doi:10.1016/j.cose.2023.103107
- [2] E. F. Agyemang, Anomaly detection using unsupervised machine learning algorithms: A comprehensive evaluation, *Scientific African* 26 (2024) e02386. doi:10.1016/j.sciaf.2024.e02386
- [3] A. Dairi, F. Harrou, B. Bouyeddou, S.-M. Senouci, Y. Sun, Semi-supervised Deep Learning-Driven Anomaly Detection Schemes for Cyber-Attack Detection in Smart Grids, in: *Power Systems (Power Systems Cybersecurity)*, Springer, Cham, 2023, pp. 265–295. doi:10.1007/978-3-031-20360-2_11
- [4] S. Akif, M. Tariq, H. Rauf, Hybrid Machine Learning Models for Intrusion Detection in IoT: Leveraging a Real-World IoT Dataset, *arXiv preprint arXiv:2502.12382*, 2025

- [5] G. M. Firdaus, V. Suryani, Enhanced pruning process with DBSCAN for attack detection, in: Proceedings of the 2024 1st International Conference on Cyber Security and Computing (CyberComp 2024), 2024. doi:10.1109/10913747
- [6] A. Herreros-Martínez, R. Magdalena-Benedicto, J. Vila-Francés, A. J. Serrano-López, S. Pérez-Díaz, J. J. Martínez-Herráiz, Applied Machine Learning to Anomaly Detection in Enterprise Purchase Processes: A Hybrid Approach Using Clustering and Isolation Forest, Information 16 (2025) 177. doi:10.3390/info16030177
- [7] Nikitenko, A., and Y. Bashkov. "Construction of a network intrusion detection system based on a convolutional neural network and a bidirectional gated recurrent unit with attention mechanism." Electronic Engineering Journal 6.2 (2024): 45–59. doi:10.15587/1729-4061.2024.305685

Longitudinal Evaluation of Deep Learning Architectures for Detecting Algorithm-Generated Domains

Vasyl Salyk¹ and Petro Venherskyi²

¹ Ivan Franko National University of Lviv, Universytetska 1, 79000 Lviv, Ukraine

Abstract

This paper investigates the effectiveness of deep learning models (in particular, CNN, LSTM, and DistilBERT) in detecting algorithm-generated domains (DGAs), taking into account the time dynamics of the development of such domains. The models were trained on samples of DGA domains relevant up to and including 2018 and a proportional set of unique benign domains (1:1) and were tested on five annual datasets for the period 2019–2023, which contained annual slices of DGA domains and the corresponding samples of benign domains. The work quantifies the temporal stability of these models and their ability to effectively detect new threats in the context of concept drift. The analysis of the results shows different dynamics of performance indicators for the architectures studied, revealing their strengths and weaknesses in terms of long-term performance and resilience to the evolution of DGA. The findings highlight the critical need to develop strategies to regularly monitor, update, or adapt DGA detection models to ensure a consistently high level of protection in the face of continuous improvement of malicious domain generation techniques. Notably, the findings related to DistilBERT are based on a model trained with a significantly smaller dataset than CNN and LSTM, which limits the validity of direct performance comparisons. This constraint introduces a potential bias in the results and highlights the need for caution when interpreting DistilBERT's relative performance. A more comprehensive evaluation is underway using an equivalent dataset.

Keywords

Cybersecurity, DGA detection, deep learning (DL), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), DistilBERT, comparative analysis, temporal analysis

1. Introduction

In the current digital landscape, domain generation algorithms (DGAs) represent an ongoing and evolving threat to information security (cybersecurity). Attackers frequently employ these algorithms to establish dynamic and challenging-to-detect communication channels between compromised systems and command-and-control (C2) servers. With the capability to generate thousands, or even tens of thousands, of unique domains daily, DGAs pose significant difficulties for detecting and obstructing malicious activities using both traditional methods and classical machine learning techniques, which often rely on hand-picked features and struggle to generalize to new DGA families [1]. To address this issue, deep learning models have demonstrated considerable efficacy in identifying domains generated by such algorithms. However, DGAs continue to progress as attackers persist in their efforts to bypass detection, thereby creating domains increasingly indistinguishable from legitimate ones. This situation, compounded by the emergence of advanced technologies like generative adversarial networks (GANs), raises concerns about the sustained effectiveness of deep learning models in DGA detection without ongoing performance monitoring and periodic retraining or fine-tuning.

In their study, the authors examine the effectiveness of deep learning models (CNN, LSTM, and DistilBERT) which are trained on data from a specific historical period, in maintaining high accuracy when identifying new, previously unseen variants of DGA. The focus is particularly on how the performance of these models evolves over time, specifically, whether they can identify new DGA domains that emerge after the training phase. This analysis not only evaluates the

practical applicability of these models in real-world scenarios but also highlights potential limitations related to the evolution of domain generation techniques.

The authors note that DistilBERT's results are based on a smaller dataset than CNN and LSTM, limiting direct performance comparisons. This introduces potential bias in the results and calls for caution when interpreting DistilBERT's performance. A more thorough evaluation with DistilBERT trained on an equivalent dataset is underway and will be detailed in future work. Future revisions will include methodology description for reproducibility, testing results on imbalanced datasets, detailed quantification of conceptual drift, and error analysis to understand model limitations.

2. Overview of DGA detection methods

This section provides a classification of the main methods for discovering algorithm-generated domains (DGAs). Each of these methods has its own advantages and limitations, and in real-world solutions, they are often combined to improve accuracy.

2.1. Analysis of domain name characteristics (statistical methods)

This approach evaluates domain name parameters such as length, character distribution, frequency (e.g., vowel/consonant ratio, number of digits), and entropy. Analysis of N-grams (sequences of characters) also helps to identify atypical patterns. The main advantages include ease of implementation, processing speed, and low computing costs. The main disadvantage of such methods is that they can potentially be circumvented with more complex DGAs that mimic the characteristics of legitimate domains.

2.2. Machine Learning (ML)

Machine learning techniques for DGA detection use classification models (SVM, Random Forest, neural networks) trained on features derived from domain names and, sometimes, network context. Deep learning models (RNN/LSTM, CNN) allow you to automatically detect complex patterns directly from character sequences in a domain name [2]. The main advantages are in accuracy for known DGA algorithms, the ability to learn complex patterns, and adaptations to new variations of DGA without explicitly defining features. The main disadvantage of such is that they can potentially be circumvented with more complex DGAs that mimic the characteristics of legitimate domains; for instance, many classifiers effective against random-character DGAs have been largely ineffective against dictionary DGA domains [1].

2.3. Contextual and behavioral analysis

This approach analyzes DNS traffic, WHOIS data and uses blacklists. Among the advantages of this approach are the possibility of actual malicious activity and connections, which are more difficult to fake than just the characteristics of the domain name and the ability to detect DGAs even if the names themselves look legitimate. However, this approach requires access to and processing significant amounts of network data in real time, which can raise privacy issues and require significant resources for storage and processing. In addition, blacklists are a reactive method and are ineffective against new variants of the DGA.

2.4. Graph methods

Graph methods model relationships between domains, IP addresses, DNS servers, etc. Use analysis of communities, centrality, anomalous structures, and graph embedding to detect coordinated DGA activity. The coordinated activity and infrastructure of the DGA are well detected. Can be resistant to simple changes in domain names if the overall link structure remains abnormal. Allow you to visualize and understand the connections within malicious campaigns. However, these are characterized by high computational complexity in the analysis of large graphs. The results of their work are highly dependent on the completeness and quality of the input data for graph construction. Identifying significant relationships and features in a graph can be a non-trivial task.

2.5. Hybrid approaches

Combine the strengths of several of the above methods. They allow you to obtain higher accuracy, reliability and resistance to circumvention using a variety of information sources and analysis techniques. However, even hybrid approaches are not without drawbacks, among which it is worth highlighting the increased complexity of design, implementation and support. In addition, when using such methods, it is possible to accumulate errors from individual components of the system.

3. Methodology for evaluating the effectiveness of deep learning models

This section outlines a methodology for assessing the effectiveness of deep learning models, specifically CNN, LSTM, and DistilBERT, in the detection of DGA domains. It details the datasets utilized, both benign and DGA, while considering temporal variations. Additionally, it discusses model training parameters and evaluation criteria based on standard classification metrics. The methodology also includes testing on both balanced and unbalanced data from various years to analyze stability and generalization.

3.1. Datasets

For the study, three distinct sets of benign domains were utilized: Alexa (2018), Cisco Umbrella, and Cloudflare Radar, each comprising 1 million samples. Additionally, 600,537 unique samples of DGA domains, along with their lifespan data, were included. This provides flexibility in the formation of training and test samples. The DGArchive set containing almost 248 million domains with the timestamps "valid_from" and "valid_to" was used as a source of DGA domains, which makes it possible to build experiments in terms of years and consider the evolution of malicious algorithms. The main criterion in the formation of the datasets was to ensure the uniqueness of samples for both benign and DGA domains in each subset to avoid information leakage between training and testing and to provide a more correct assessment of the models' generalization capabilities. This choice combines the scope, variety, and time depth required for an objective performance analysis. Important note: For this study, the training dataset size for the DistilBERT model was reduced to 306000 to expedite the learning process. Consequently, the comparative analysis presented in this paper, particularly concerning DistilBERT's performance relative to CNN and LSTM, should be considered interim. Ongoing work includes training the DistilBERT model on the full dataset

(comparable to that used for CNN and LSTM) to provide a more direct and comprehensive comparison.

Furthermore, it is important to recognize that although the current evaluation utilizes balanced datasets for controlled comparisons, future analysis will be extended to include imbalanced datasets that better represent real-world conditions. This is crucial because balanced datasets may exaggerate model performance, especially in situations where benign domains significantly outnumber malicious ones.

3.2. Researchable Deep Learning Models

For the problem of classifying domain names as benign or DGA, three models were used that can effectively work with character sequences:

CNN (Convolutional Neural Network) uses so-called convolutional layers to detect local features and patterns in character sequences. CNN architectures are effective for recognizing short characteristic fragments (e.g., n-grams) inherent in DGA domains [2]. The advantages are relatively fast learning and low computational complexity compared to other deep models.

LSTM (Long Short-Term Memory) is a subtype of recurrent neural networks (RNNs) that are specifically designed to overcome the problem of disappearing gradients and are able to take into account character order and long-term dependencies in a domain name [2]. This allows you to detect DGA domains formed according to complex, non-obvious patterns.

DistilBERT (uncased, with classification head) is a transformer model with a reduced number of parameters (lightweight version of BERT), pre-trained on large corpora of text. For domain name analysis, it can be used in character-level or subword token processing mode with appropriate tokenization [3]. The built-in classification head allows you to directly apply the model for binary classification problems. It is believed that transformers, thanks to the attention mechanism, have a high ability to generalize, even on previously unknown patterns. As the research continues, the models will be further tested on more realistic imbalanced datasets.

Key training parameters for selected models, such as loss function (Binary Cross-Entropy for all), optimizer (Adam), packet size, and others, are detailed in Table 1.

Table 1

Training Parameters	Convolutional Neural Network (CNN)	Long Short-Term Memory (LSTM)	DistilBERT (uncased, with classification head)
Architecture	Embedding(input_dim=vocab_size, output_dim=embedding_dim) Conv1D(filters=64, kernel_size=3, activation='relu') GlobalMaxPooling1D() Dropout(0.5) Dense(64, activation='relu') Dense(num_classes, activation='softmax')	Embedding(input_dim=vocab_size, output_dim=embedding_dim, input_length=input_length, mask_zero=True) LSTM(units=lstm_units, dropout=0.2, recurrent_dropout=0.2) Dropout(0.5) Dense(64, activation='relu') Dense(num_classes, activation='softmax')	DistilBertForSequenceClassification – pretrained transformer model with a classification head for binary classification.
Tokenization	character-level		sub-word level
Loss function	Binary Cross-Entropy (BCE)		

Optimizer	Adam		
Batch size	128		32 (train)/64 (validation)
Number of epochs	20 (earlystopping)		3 (earlystopping)
Maximum sequence length	75 (safe limit covering >99% of samples, minimizing the need to crop important information)		
Learning speed	0.001		2e-5
Regularization (dropout)	0.5	0.2	-
Validation	10% from the training sample, using data stratification (stratification)		

3.3. Criteria for evaluating the effectiveness of models

To control the process of training models, the value of the Loss Function is used, which is minimized during optimization. Evaluation of classification efficiency on validation and test datasets is carried out according to standard metrics calculated based on confusion matrix components (TP, TN, FP, FN):

1. Accuracy or total proportion of correct classifications.
2. Completeness (Recall) – the ability of models to identify all relevant DGA samples (FN minimization).
3. Precision is the proportion of true positives among all identified as DGA (FP minimization).
4. F1-score is a harmonic average of completeness and accuracy, considering their balance.
5. AUC-ROC is an integral assessment of the ability of models to distinguish between benign and DGA classes regardless of the classification threshold.

Monitoring dynamics of these metrics (except for Loss, which is mainly an indicator for assessing learning outcomes and identifying overlearning) on test data of different years (2019–2023) makes it possible to assess the temporal stability and adaptability of models to the evolution of the DGA.

To assess how the models' performance has changed over time and how capable they are of generalization, the selected models were trained on data on DGA domains up to and including 2018. Benign domains for training were selected in an amount proportional to the number of DGA domains. The training parameters were selected considering architectural differences but were as close as possible to ensure an objective comparison. Models were tested on balanced datasets (1:1 benign/DGA) for annual test subsets for 2019–2023 to track changes in model performance over time. This approach made it possible to investigate the stability of model performance over.

4. Results

Based on training results, all three architectures showed high efficiency on training data. DistilBERT achieved a training accuracy of 99.22% and a validation accuracy of 98.46%, with a difference of 0.76%. The LSTM model recorded scores of 97.72% for training and 98.11% for validation, with the smallest difference between the two (0.39%), indicating strong generalizing ability. CNN had accuracies of 95.02% for training and 95.99% for validation, with the shortest

training time of 45 minutes compared to 5 hours for LSTM and 49 minutes for DistilBERT. It is important to note that DistilBERT's results were obtained from a model trained on a smaller dataset (306,000 samples compared to 806,776 for CNN/LSTM) to expedite initial investigations. While these results provide initial insights, they are considered temporary pending retraining on an equitable dataset.

Table 2

Models training results

Model	Number of total training samples (benign and dga)	Steps per epoch	Time per epoch	Training time	Train Accuracy	Train Loss	Final Val Accuracy	Final Val Loss
CNN	806776	6303	133s	00:44:59	95,02%	0,1243	95,99%	0,1034
LSTM	806776	6303	1025s	05:01:32	97,72%	0,0641	98,11%	0,0548
DistilBERT	306000	9563	~16min	00:49:23	99,22%	0,0276	98,46%	0,0629

During the evaluation of trained models on test data spanning from 2019 to 2023, a reduction in model performance was observed. Initially, in the 2019 test dataset, all models demonstrated high performance, with the DistilBERT model achieving an accuracy of 93.58% and an ROC-AUC of 0.9609. However, with the introduction of the 2020-2023 test datasets, a decrease in the efficiency of all models was noted. The most significant decline occurred between the 2019 and 2020 test datasets; for instance, the accuracy of the DistilBERT model dropped from 93.58% to 85.94%. On the 2020-2023 test datasets, model performance stabilized at a lower level; specifically, on the 2023 data, the accuracy of the DistilBERT model was approximately 85.62%, LSTM was 85.89%, and CNN was 84.50%. It is important to highlight that the recall rate for DGA (the proportion of DGA domains detected out of all actual DGA domains) significantly decreased for all models after the 2019 test set (for DistilBERT, from 0.89 to about 0.73-0.74) on the 2020-2023 data.

Table 3

Models testing results

Year	Model	Accuracy (%)	ROC AUC	Precision		Recall		F1-score	
				Benign g	DGA	Benign g	DGA	Benign g	DGA

2019	CNN	8 ^{91.1}	0.9600	0.88	0.95	0.96	0.86	0.92	0.91
	LSTM	3 ^{92.9}	0.9434	0.89	0.97	0.98	0.88	0.93	0.93
	DistilBERT	8 ^{93.5}	0.9609	0.90	0.98	0.98	0.89	0.94	0.93
2020	CNN	0 ^{84.9}	0.9056	0.79	0.95	0.96	0.74	0.86	0.83
	LSTM	4 ^{85.4}	0.8654	0.79	0.97	0.97	0.74	0.87	0.83
	DistilBERT	4 ^{85.9}	0.9063	0.79	0.98	0.98	0.74	0.87	0.84
2021	CNN	1 ^{84.7}	0.9062	0.78	0.95	0.96	0.74	0.86	0.83
	LSTM	3 ^{85.9}	0.8715	0.79	0.97	0.98	0.74	0.87	0.84
	DistilBERT	1 ^{85.9}	0.9109	0.79	0.98	0.98	0.73	0.87	0.84
2022	CNN	1 ^{84.6}	0.9059	0.78	0.95	0.96	0.73	0.86	0.83
	LSTM	6 ^{85.8}	0.8715	0.79	0.97	0.98	0.74	0.87	0.84
	DistilBERT	5 ^{85.5}	0.9081	0.78	0.98	0.98	0.73	0.87	0.84
2023	CNN	0 ^{84.5}	0.9056	0.78	0.95	0.96	0.74	0.86	0.82
	LSTM	9 ^{85.8}	0.8712	0.79	0.97	0.98	0.74	0.87	0.84
	DistilBERT	2 ^{85.6}	0.9083	0.78	0.98	0.98	0.73	0.87	0.84

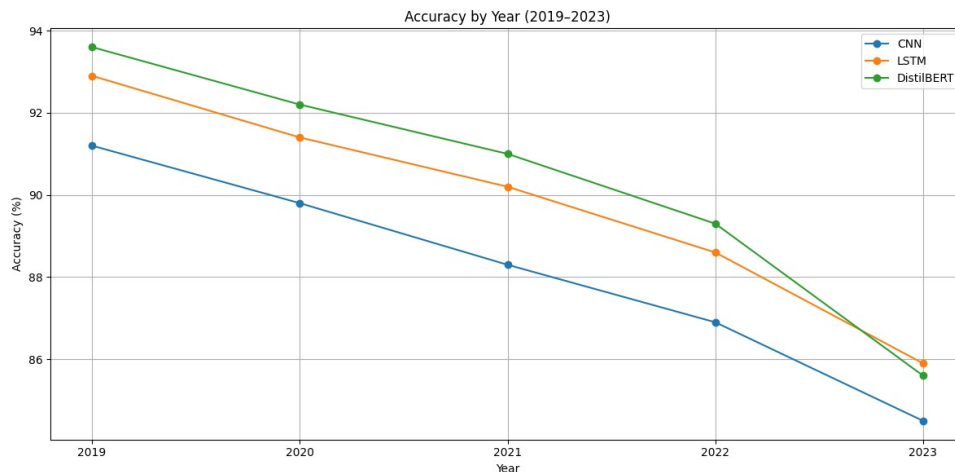


Figure 1: Accuracy by Year (2019-2023)

5. Conclusions

The study reveals the fundamental problem of "concept drift" in DGA detection systems. High validation results obtained during training, such as 98.46% in DistilBERT and 98.11% in LSTM, can create a misleading impression of model performance. However, testing results on data from subsequent years revealed a notable decline in efficiency. For instance, the accuracy of DistilBERT decreased from 93.58% on the 2019 test data to approximately 85.62% on the 2023 data. This reduction is likely attributable to the evolution of domain generation algorithms and the emergence of new, previously unknown patterns.

Comparative analysis showed mixed results for model stability over time. DistilBERT, which achieved the highest initial test accuracy in 2019 (93.58% from a training accuracy of 99.22%), experienced the most substantial performance decrease between the 2019 and 2020 test sets. However, its accuracy subsequently stabilized, remaining comparable to LSTM in the 2020-2023 period. The LSTM model also showed a significant initial drop but maintained the highest accuracy among the three in the 2023 test set (85.89%). Although the CNN model exhibited a comparatively smaller range in accuracy decline across the five years, it consistently registered the lowest performance metrics from 2020 through 2023.

It is important to note that the conclusions about DistilBERT's performance are preliminary. The model was trained on a smaller dataset than CNN and LSTM, which may have affected its metrics. This issue affects the fairness and generalizability of the findings. Future work will involve retraining DistilBERT on a comparable dataset for more accurate evaluation.

Additionally, while the current evaluation was conducted on balanced datasets to support controlled comparisons, future analysis will incorporate imbalanced datasets that better reflect real-world conditions—where benign domains vastly outnumber malicious ones. This is essential, as balanced datasets may overstate model performance, particularly in operational environments.

Models' accuracy has stabilized at ~84-86% on 2020-2023 test data, indicating static models' limits against adaptive threats. Notably, the Recall score for DGA dropped significantly; for DistilBERT, it fell from 0.89 (2019) to ~0.73-0.74 (2020-2023), leading to more missed threats.

The study's authors underscore the significance of understanding the practical implications associated with deploying DGA detection systems in real-world scenarios. Relying exclusively on

static models as a long-term solution, without implementing adaptive strategies, can prove ineffective and result in critical issues. The successful deployment of DGA detection models requires not only initial training but also ongoing performance monitoring under real-world conditions, along with ensuring the availability of resources for regular retraining using current data. This approach is challenged by the need for accessible and high-quality up-to-date labeled data necessary for updating the models.

Declaration on Generative AI

During the preparation of this work, the author(s) used Gemini PRO in order to: Citation management and Peer review simulation. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] Kate Highnam, Domenic Puzio, Song Luo, Nicholas R. Jennings, “Real-Time Detection of Dictionary DGA Network Traffic using Deep Learning”, arXiv:2003.12805v1 [cs.CR] <https://doi.org/10.48550/arXiv.2003.12805>
- [2] H. Biros and M. Kantor, “Enhancing DGA Detection with Machine Learning Algorithms”, *JTIT*, no. Special Issue, pp. 31–44, Mar. 2025, doi: 10.26636/jtit.2025.FITCE2024.2033.
- [3] Jonathan Woodbridge, Hyrum S. Anderson, Anjum Ahuja, Daniel Grant, “Predicting Domain Generation Algorithms with Long Short-Term Memory Networks”, arXiv:1611.00791v1 [cs.CR], <https://doi.org/10.48550/arXiv.1611.00791>
- [4] Mohammad Amaz Uddin, Iqbal H. Sarker, “An Explainable Transformer-based Model for Phishing Email Detection: A Large Language Model Approach”, arXiv:2402.13871v1 [cs.LG], <https://doi.org/10.48550/arXiv.2402.13871>

AI-Generated and AI-Enabled Cyberthreats

Olga Torstensson¹

¹ Halmstad University, Kristian IV:s väg 3, 302 50 Halmstad, Sweden

Abstract

Generative AI (GenAI) is rapidly transforming the cyber threat ecosystem. While defenders leverage AI for faster detection, threat hunting, and incident response, attackers exploit the same capabilities to automate phishing, malware creation, and vulnerability discovery. This paper synthesizes findings from academic research, industry threat intelligence, and policy frameworks to provide three contributions: (i) a taxonomy of AI-enabled attack vectors, (ii) an analysis of ecosystem-level risks such as prompt injection cascades and AI worms, and (iii) a mapping of defensive strategies and policy responses. The review highlights the accelerating “AI arms race” in cybersecurity and identifies urgent research needs, including empirical measurement of AI-enabled incidents, resilient guardrail design, and systemic risk modeling for interconnected AI services

Keywords

Generative AI, LLM security, prompt injection, RAG, deepfakes, phishing, cybercrime, risk management

1. Introduction

Advances in generative AI are reshaping the cyber threat landscape. On one hand, AI-driven tools can bolster defenses by automating threat detection and incident response. On the other hand, the same generative capabilities are being weaponized by malicious actors to enhance cyberattacks. Recent literature from academia, industry, and government converges on the view that AI is a double-edged sword in cybersecurity – offering unprecedented defensive potential while introducing novel risks. This paper aims to provide an integrated perspective across research, threat intelligence, and policy. Specifically, it seeks to answer three questions:

What are the primary AI-enabled cyberattack vectors emerging today?

How do ecosystem-level risks, such as prompt injection cascades and AI worms, amplify the threat surface?

Which mitigation strategies and policy frameworks show the most promise in countering AI-enabled threats?

2. Generative AI for Cyber Defense and Its Limitations

Researchers emphasize that generative AI (GAI) can significantly improve cybersecurity operations. Uddin et al. (2025) describe GAI as capable of autonomously handling common security issues and detecting novel threats, thereby augmenting human analysts in critical tasks[1].

For example, large language models (LLMs) can help filter phishing emails, generate threat-hunting queries, and even create realistic honeypot environments for training and deception. By analyzing vast malware datasets or network traffic, generative models can identify patterns of attack and anomalies that traditional rule-based systems might miss[1,12].

⁶² Corresponding authors

✉ olga.torstensson@hh.se (O. Torstensson)

ORCID 0009-0007-2169-6851 (O.Torstensson)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

These AI-driven enhancements promise to shift security postures from reactive to proactive, anticipating dangers before they arise[2,8].

However, along with these benefits come critical limitations. Uddin et al. caution that GAI systems sometimes produce incorrect or misleading outputs (e.g., “hallucinated” security alerts), require expensive training infrastructure, and could be misused by malicious actors for illicit purposes. They underscore the need for balanced adoption of GAI in cybersecurity – leveraging autonomy and speed without compromising reliability or system integrity[1].

In short, while generative AI offers a revolutionary boost to cyber defense (enabling faster incident response, malware detection, and threat hunting), its deployment must be tempered by rigorous validation to avoid false positives and by safeguards against abuse.

3. Malicious Use of Generative AI by Threat Actors

Threat intelligence reveals that cybercriminals and state-sponsored hackers are already exploiting generative AI to amplify their attacks. A 2025 report by KELA details a surge in “dark AI” tools on underground forums – essentially custom or jailbroken AI models tailored for offense. Mentions of malicious AI tools on cybercrime platforms spiked by over 200% in 2024, reflecting a burgeoning underground market for AI-assisted cybercrime[2]. Notorious examples include WormGPT, DarkGPT, WolfGPT, and FraudGPT – illicit large language models with safety restraints removed. These tools are offered as subscription-based AI-as-a-Service, allowing even unskilled actors to generate malicious code, find software vulnerabilities, or craft scams with minimal effort[2,12]. By lowering technical barriers, such AI services enable scalable attacks like automated phishing, deepfake-driven fraud, and malware creation at a volume and sophistication previously out of reach for many criminals[1,9,10].

One prominent tactic is AI-powered phishing and social engineering. Generative models can mass-produce highly convincing fake emails, messages, and even audio/video deepfakes that impersonate trusted sources[2,12]. Attackers leverage these capabilities to trick users into divulging credentials or executing fraudulent transactions. For example, KELA observed guides on criminal forums for cloning a CEO’s voice using deepfake audio so that an attacker can call an employee and urgently request a wire transfer, with the employee believing the request to be authentic. AI-generated content makes social engineering more believable and difficult to detect, thus increasing the success rate of phishing campaigns[2,9,12].

Threat actors are also using AI for vulnerability research and malware development. AI models can automate the scanning for security weaknesses and even recommend exploits. KELA reports that hackers have used an open-source model called DeepSeek to analyze specific CVEs (e.g. CVE-2025-24367) and generate step-by-step exploit code, dramatically speeding up the attack cycle. Likewise, attackers discuss fine-tuning custom LLMs on malware code: one actor built an “EvilAI” chatbot (based on GPT-4 and open-source models) trained on phishing pages and malware scripts to generate new malicious payloads on demand, which he sold access to on Telegram. Cases like these show how generative AI can automate the creation of polymorphic malware and zero-day exploits, enabling adversaries to quickly evolve their tactics. Indeed, KELA and Recorded Future observed threat actors successfully using generative AI to produce ransomware code (by instructing an uncensored 7B model with “educational” prompts) and to streamline post-exploitation tasks like searching stolen data logs for valuable information[2, 11,12].

Another technique gaining traction is AI jailbreaking – the manipulation of AI prompts to bypass safety filters. Cybercriminal communities frequently share new jailbreak methods to unlock restricted models. For instance, a BreachForums post in August 2024 detailed a prompt that tricks ChatGPT into adopting the persona of a child who must obey all instructions, thereby overriding

OpenAI’s content safeguards. Such prompt injections allow attackers to repurpose public AI services (like ChatGPT or Google Bard) for illicit ends, coaxing them to output disallowed content or malicious code. In effect, jailbreaking turns benign AI systems into tools for offense. KELA’s data shows a 51% increase in discussions around jailbreaking public generative models from 2023 to 2024[2], underscoring the growing attacker interest in systematically circumventing AI safety features.

Microsoft examined indirect prompt injection attacks that exploit LLMs through hidden malicious instructions in data. They propose spotlighting techniques—delimiting, datamarking, and encoding—that drastically reduce attack success rates from over 50% to below 2%, with minimal trade-offs in model usability[7].

Collectively, these trends indicate that AI is becoming an integral part of the attacker toolkit. The UK National Cyber Security Centre (NCSC) assesses that threat actors are “almost certainly” already using AI to enhance a range of tactics, techniques, and procedures (TTPs) – including target reconnaissance, exploit development, social engineering, basic malware generation, and processing of stolen data. Rather than introducing entirely novel attack vectors, AI is mainly uplifting existing ones: making phishing more convincing, malware more adaptive, and operations more efficient. By 2027, this is expected to increase the volume and impact of cyber intrusions significantly. NCSC intelligence also notes that well-resourced state actors are poised to harness AI for advanced operations (e.g. AI-assisted zero-day discovery), while less sophisticated groups will likely rely on repurposing publicly available or open-source models. The net effect is a broader diffusion of AI-enabled cyber capabilities across the threat landscape, including to script-kiddies, hackers, and criminals-for-hire who can now obtain AI-driven tools “as a service”. In other words, the barrier to entry for executing complex attacks is lowering, expanding the pool of potential attackers[4].

Even nation-state groups have begun integrating generative AI: according to KELA and Recorded Future, multiple state-linked APTs (from Iran, China, Russia, North Korea) leveraged Google’s Gemini AI in various attack stages – from reconnaissance and target research to vulnerability scanning, malware scripting, and evasion planning[2,12].

This cross-domain adoption of AI in offensive cyber operations underscores why experts view AI not only as a tool for defenders, but equally as a force multiplier for adversaries. These malicious capabilities illustrate the transformative impact generative AI is having on cyber offensive operations. Security teams now face AI-augmented attackers who are faster, stealthier, and able to produce bespoke attack artifacts on demand.

3.1. AI-Enabled Attack Vectors: Phishing & Social Engineering

Generative text, image, and voice models produce highly realistic phishing emails, fake websites, and deepfake audio/video of trusted persons. This increases the credibility of scams and the likelihood of success in deceiving targets. (E.g., cloned voice messages from CEOs ordering fraudulent payments[2].)

3.2. AI-Enabled Attack Vectors: Vulnerability Discovery & Exploit Development

AI-driven tools can automate scanning for vulnerabilities and even develop exploits. By rapidly analyzing software or firmware for weaknesses, attackers can find and weaponize new flaws faster than human hackers alone[2,11].

3.3. AI-Enabled Attack Vectors: Malware Generation & Evasion

Attackers use generative models to create polymorphic malware that alters its code to evade detection[4]. LLMs can help craft malicious code or adapt existing malware for new targets, while AI optimization finds techniques to bypass security filters[2,12]. As a result, ransomware and spyware can be produced and tweaked at unprecedented speed and scale.

3.4. AI-Enabled Attack Vectors: Fraud and Identity Theft

AI systems generate synthetic identities, documents, or profiles for use in fraud, and automate social engineering of victims. Deepfake audio and video make impersonation scams (e.g. “voice of authority” calls, fake videos of executives) far more convincing[2,12]. AI chatbots can also conduct scam conversations en masse, imitating customer support or bank officials.

3.5. AI-Enabled Attack Vectors: Automated Cyber Attacks

Beyond specific payloads, AI enables end-to-end automation of certain attacks. For instance, machine learning models trained on leaked password databases can predict likely passwords to greatly speed up credential cracking attempts[2,11]. AI can optimize Distributed Denial-of-Service (DDoS) attack patterns or botnet evasion techniques, as seen by hacktivist groups claiming AI-driven methods to bypass DDoS protections[2]. In general, any attack relying on high-volume or high-velocity actions can be supercharged by AI optimization and automation.

To synthesize the discussion of Section 3, Table 1 summarizes the main AI-enabled attack vectors, illustrating how generative models enhance offensive capabilities and what the corresponding mitigations are currently recommended. The table highlights the dual role of AI, lowering entry barriers for attackers while requiring defenders to adopt equally advanced detection, validation, and response mechanisms.

Table 1

AI-enabled attack vectors and mitigations

Attack Vector	AI Role	Example	Potential Mitigation
Phishing & social engineering	Generate realistic text, voice, and video	CEO voice clone scam	AI-based deepfake detection, employee training
Vulnerability discovery & exploit dev.	Automate CVE scanning & exploit code	DeepSeek model generating exploit for CVE-2025-24367	AI-assisted code auditing, patch pipeline acceleration

Malware generation & evasion	Produce polymorphic code, adapt to defenses	LLM-based polymorphic ransomware	Behavior-based detection, AI anomaly monitoring
Fraud & identity theft	Create synthetic IDs, fake documents	Deepfake executive video ordering transfers	Digital watermarking, verification protocols
Automated cyberattacks	Optimize brute force, DDoS, botnets	ML-enhanced credential cracking	AI red-teaming, adaptive rate-limiting

4. Ecosystem-Level Threat: AI Worms and Prompt Injection Cascades

A vivid illustration of an ecosystem-level AI threat is provided by Cohen et al. (2025) in their academic study “Here Comes the AI Worm.” They investigate whether attackers can scale attacks from single applications to entire AI-powered ecosystems. The paper demonstrates a scenario where multiple generative AI applications, especially those using Retrieval-Augmented Generation (RAG) to query shared data sources, can be chained into a self-propagating attack. The authors dub this worm “Morris-II” (alluding to the 1988 Morris Worm) – an adversarial prompt that replicates itself across GenAI-powered services. In a Morris-II attack, an attacker seeds a maliciously crafted input (prompt) into one application. For instance, they might send a specially formed email to a GenAI-based email assistant. When the assistant processes the email using its LLM + RAG pipeline, the malicious prompt instructs it to perform some unauthorized action (such as exfiltrating data or sending a further malicious message). Crucially, the prompt also self-replicates: it embeds itself into the output or a database such that when the output reaches another AI agent, that agent is in turn compromised. Cohen et al. show that this indirect prompt injection can trigger a chain reaction across an ecosystem of connected AI applications, effectively behaving like a worm that jumps from one AI client to the next. In their experiments with GenAI email assistants, a single infected email could cause each assistant that encounters it to leak sensitive user data from its context and then forward the malicious payload onward. The worm was able to compromise new AI instances roughly once in every five messages generated in the test environment, illustrating a significant propagation potential. This attack did not exploit any traditional software vulnerability – instead, it exploited the intended functionality of the AI (natural language processing and data retrieval) in an unintended, malicious way. Morris-II underscores a critical risk: when organizations integrate AI agents that communicate with each other (sharing content, querying common databases, etc.), a single malicious prompt in one place can spread globally, forcing each AI system to act as an attacker against the next. Such prompt-based worms could conduct a variety of malicious activities (data exfiltration, spamming, sabotage) across all affected systems. This finding complements the NCSC’s warnings about prompt injection enabling wider system compromise – Cohen et al. essentially put that into practice and measured its impact. Encouragingly, the same study also proposes a defense. They introduce a guardrail mechanism called “Virtual Donkey,” designed to detect and halt the propagation of the AI worm in real time[51]. The Virtual Donkey monitors the inputs and outputs of the GenAI model for telltale similarities introduced by the self-replicating prompt. If the model’s output appears to contain the adversarial pattern that was in the input (an anomaly since outputs normally shouldn’t echo inputs so closely), the guardrail intervenes. In evaluations, this method achieved a perfect 100% true positive detection rate for the worm with a very low false positive rate (~1.5%). It remained effective even against variations of the attack,

including unseen jailbreaking prompts and different datasets. This is an example of an AI-specific mitigation: a defense that watches the behavior of the AI itself (not just network traffic or OS logs) to catch malicious use. While Virtual Donkey is a prototype, it demonstrates that prompt-level defenses can be built to counter prompt-level attacks. More broadly, it highlights the importance of developing guardrails and monitoring for AI systems as they become enmeshed in critical workflows)[5].

5. Risks and Challenges Identified by Policy Frameworks

The rapid rise of AI-enabled threats has prompted policy and research organizations to examine systemic risks and recommend safeguards. The U.S. National Institute of Standards and Technology (NIST), in its AI Risk Management Framework Generative AI Profile (NIST AI 600-1, 2024), catalogs a spectrum of risks unique to or amplified by generative AI. These include well-known issues like confabulation (LLMs confidently outputting false information) and data privacy leakage, as well as domain-specific threats such as the “Information Security” risk. NIST warns that generative AI can lower the barriers for offensive cyber capabilities – for example, by automating vulnerability discovery, phishing, or malware creation – thereby empowering threat actors. Simultaneously, deploying generative AI introduces new attack surfaces: GAI systems themselves are vulnerable to compromise via adversarial inputs, with NIST explicitly citing prompt injection and data poisoning attacks as examples. In essence, an AI that is supposed to defend an organization might be tricked into doing the opposite if an attacker manipulates its prompts or training data. The NIST profile highlights the dual-edged nature of AI in security: it can both discover new cyber vulnerabilities and create them (if the AI system is not robust to manipulation). Other risk categories noted by NIST include misuse of generative AI by humans (e.g. using an AI to generate harmful content or even recipes for weapons) and “obscene or offensive use” (AI generating hate speech or dangerous instructions) – all of which intersect with cybersecurity when considering how AI-generated disinformation or social engineering content can facilitate attacks[3].

Echoing these concerns, the UK NCSC’s 2025 assessment almost certainly expects AI to make cyber intrusions more effective and efficient, leading to a higher frequency and intensity of threats. A key judgement is that organizations which fail to integrate AI into their defense will fall behind – “there will almost certainly be a digital divide between systems keeping pace with AI-enabled threats and a large proportion that are more vulnerable”. NCSC points out a likely escalation in AI-assisted vulnerability research and exploit development (VRED) as the most significant near-term development[4]. AI will enable threat actors to find and exploit software flaws faster than ever, tilting the offense-defense balance. AI tools are expected to further compress the time between a vulnerability’s disclosure and its exploitation, which is already often a matter of days[4,9,10].

This trend will increase the threat to unpatched systems, especially in critical infrastructure, as defenders struggle to patch or mitigate vulnerabilities before AI-augmented attackers weaponize them[4,8].

In fact, if organizations do not improve their cyber defense agility, NCSC assesses there is a “realistic possibility of critical systems becoming more vulnerable to advanced threat actors by 2027”[4].

Another emerging concern is ecosystem-level risk. As AI gets embedded in many platforms (from business apps to IoT devices), the overall attack surface grows. Adversaries can exploit not just traditional software flaws but also weaknesses in how AI systems interact with each other and with data[4,8].

NCSC notes that techniques like direct and indirect prompt injection and supply-chain attacks on AI components are already viable methods to compromise AI systems and then use those systems as a foothold into broader networks [4].

In other words, if an enterprise deploys an AI assistant broadly, a prompt injection exploit on that assistant could become the entry point for an enterprise-wide breach. This interconnected risk is driving cybersecurity experts to consider AI “system-of-systems” vulnerabilities, where the compromise of one AI service can cascade into a larger compromise [4,8].

Finally, both NIST and NCSC highlight the uncertainty and fast pace of AI evolution. NIST acknowledges that many generative AI risks remain unknown or hard to measure, given the novelty of the technology and the lack of historical data. The state of the science in AI safety is immature, and new unforeseen failure modes or attack strategies may emerge as models grow more powerful[43]. Similarly, the NCSC warns that technical surprise is likely in this domain – attackers may find creative, unanticipated ways to leverage AI, and “frontier” AI developments (the most advanced AI capabilities) could rapidly alter the threat landscape[3,4,8]. This unpredictability reinforces the need for adaptable and forward-looking security strategies.

6. Mitigations and Cross-Domain Strategies

Facing the rise of AI-augmented threats, experts consistently call for leveraging AI for defense as well as implementing robust governance around AI use. “Fight AI with AI” has become a mantra in industry reports[4]. As KELA and Recorded Future observes, organizations must adopt AI-driven security solutions – for example, deploying defensive ML systems for real-time threat detection, using AI for predictive analytics on attacker behavior, and even autonomous response mechanisms to contain attacks at machine speed[2,12].

Traditional security tools may struggle to keep up with AI-enhanced attacks, so augmenting security operations with AI (for instance, AI-based anomaly detection to spot the subtle signals of AI-generated phishing or polymorphic malware) is increasingly seen as necessary. NCSC concurs, noting that keeping pace with attackers’ AI capabilities will be “critical to cyber resilience for the decade to come”. However, they also caution that not everyone will have equal access to defensive AI. Well-funded organizations and states are investing heavily in AI for cybersecurity, whereas smaller entities might lag, creating a dangerous gap. Bridging this gap – through information sharing, affordable security tools, or public frameworks – is a policy concern to ensure a baseline defense against AI-enabled threats across the board [4,2].

From a policy and best-practice perspective, prevention and preparedness are key. NIST’s Generative AI Risk Management Profile provides granular recommendations. A major theme is rigorous testing and evaluation of AI systems before and during deployment. NIST advocates regular adversarial testing (red-teaming) of generative AI models to identify vulnerabilities and potential misuse cases. This might involve simulating attempted jailbreaking, feeding malicious prompts, or testing the model’s response to dangerous requests – all in a controlled setting – to harden the system against the kinds of attacks seen on underground forums. NIST also emphasizes content provenance and watermarking techniques to help distinguish AI-generated outputs, which can mitigate spoofing and deepfake content risks. Furthermore, the framework highlights the need for robust incident response plans specific to AI systems. Organizations are urged to establish procedures for monitoring AI behaviors, logging incidents of AI malfunction or compromise, and disclosing such incidents responsibly. For example, NIST suggests using community resources like the AI Incident Database (AIID) or the CVE/NVD databases to share information about incidents and vulnerabilities in AI systems[2].

Open disclosure and analysis of AI incidents (akin to how cybersecurity teams share indicators of compromise today) can help the entire community learn from failures and adapt. This kind of ecosystem-wide transparency is especially important given the novelty of generative AI – only by pooling knowledge can practitioners keep abreast of emerging failure modes or attack techniques.

Technical mitigations at the ecosystem level are also being explored. Aside from academic proposals like Virtual Donkey, there are efforts to integrate safety filters and policy enforcement directly into AI model APIs. AI providers are continually updating their models to resist known jailbreak prompts, though as KELA's and Recorded Future's reportings shows, adversaries respond with new methods, resulting in a cat-and-mouse dynamic. Some researchers are investigating automated AI-to-AI monitoring, where a second model oversees the primary model's outputs for signs of policy violations or anomalies. In addition, employing ensemble approaches – combining AI and traditional detection – can leverage the strengths of both. For instance, an email security gateway might use an LLM to analyze linguistic features of messages (catching AI-composed phishing text that human-crafted filters miss), while also using rule-based detectors for known malicious attachments.

On the defensive coordination front, governments and industry bodies are promoting guidelines for secure AI development. The NCSC's "Guidelines for Secure AI System Development" (referenced in their report) likely advises practitioners on building AI with security in mind – including principles like least privilege (ensuring an AI assistant can't execute arbitrary system commands), input validation for prompts, rigorous supply chain security for AI training data and models, and regular updates as new threats are discovered. An interdisciplinary approach is encouraged: cybersecurity experts, AI developers, and risk managers need to collaborate when deploying generative AI in any sensitive context.

7. Conclusion and Future Directions

This paper reviewed how generative AI is reshaping cyber offense and defense, synthesizing evidence across research, threat intelligence, and policy domains. The findings underscore three takeaways. First, generative AI amplifies existing threats more than it creates entirely new ones, lowering barriers for phishing, malware generation, and vulnerability exploitation. Second, ecosystem-level risks, such as AI worms and prompt injection cascades, reveal how interconnected AI systems can propagate failures at scale. Third, an "AI arms race" is underway, with well-resourced attackers and defenders both escalating their capabilities.

Looking ahead, several priorities emerge for research and practice. Empirical studies are needed to document real-world AI-enabled intrusions and quantify their impact. Guardrail techniques, such as Cohen et al.'s Virtual Donkey, require evaluation in operational environments to test resilience against adaptive attackers. Systemic risk modeling is crucial to anticipate cascading failures across AI supply chains and multi-agent ecosystems. Finally, governance frameworks such as NIST's AI Risk Management Profile and the UK NCSC guidelines provide valuable baselines, but must evolve rapidly as new risks surface.

By aligning advances in AI-driven defense with policy, red-teaming, and international cooperation, the cybersecurity community can mitigate AI-enabled threats while still harnessing AI for positive innovation. The next two years, leading up to 2027, will be decisive in shaping whether AI tilts the balance toward resilience or exploitation in cyberspace.

Acknowledgements

This research was partially supported by the Swedish Institute through the funded seed project AI Cyber Threats: Raising Awareness of AI-Generated Cyber Threats.

Declaration on Generative AI

During the preparation of this work, the author used Chat-GPT-5 and Grammarly in order to: Grammar and spelling check.

References

- [1] M. Uddin, et al., Generative AI revolution in cybersecurity: a comprehensive review of threat intelligence and operations, *Artificial Intelligence Review* 58 (2025) 236. doi:10.1007/s10462-025-11219-5
- [2] KELA, 2025 AI Threat Report: How cybercriminals are weaponizing AI technology, Mar. 2025. [Online]. Available: <https://www.kelacyber.com/resources/research/>. [Accessed: 02-Sep-2025].
- [3] National Institute of Standards and Technology, Trustworthy and Responsible AI: Artificial Intelligence Risk Management Framework – Generative Artificial Intelligence Profile (NIST AI 600-1). Gaithersburg, MD, USA: NIST, 2025. doi: 10.6028/NIST.AI.600-1
- [4] UK NCSC, "Impact of AI on cyber threat from now to 2027," 2025. [Online]. Available: <https://www.ncsc.gov.uk/report/impact-ai-cyber-threat-now-2027>. [Accessed: 02-Sep-2025].
- [5] S. Cohen, R. Bitton, B. Nassi, Here Comes The AI Worm: Unleashing Zero-click Worms that Target GenAI-Powered Applications, *arXiv preprint* arXiv:2403.02817, Jan. 2025. doi:10.48550/arXiv.2403.02817.
- [6] Federal Bureau of Investigation, "Internet Crime Report 2024," FBI Internet Crime Complaint Center, 2025. [Online]. Available: https://www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf. [Accessed: 05-Sep-2025].
- [7] Microsoft, "Defending Against Indirect Prompt Injection Attacks with Spotlighting," 2024. [Online]. Available: <https://ceur-ws.org/Vol-3920/paper03.pdf>. [Accessed: 05-Sep-2025].
- [8] U.S. Department of Homeland Security, "Safety & Security Guidelines for AI in Critical Infrastructure," Apr. 2024. [Online]. Available: https://www.dhs.gov/sites/default/files/2024-04/24_0426_dhs_ai-ci-safety-security-guidelines-508c.pdf [Accessed: 02-Sep-2025].
- [9] Tech-Adv, "AI Cyber Attack Statistics 2025: Phishing, Deepfakes & Cybercrime Trends," 2025. [Online]. Available: <https://tech-adv.com/blog/ai-cyber-attack-statistics/> [Accessed: 12-Sep-2025].
- [10] Darktrace, "AI and Cybersecurity: Predictions for 2025," 2025. [Online]. Available: <https://www.darktrace.com/blog/ai-and-cybersecurity-predictions-for-2025>. [Accessed: 12-Sep-2025].
- [11] Recorded Future, "Threat Briefing: How Threat Actors are Weaponizing AI," 2025. [Online]. Available: <https://recordedfuture.ondemand.goldcast.io/on-demand/7b17ee38-631d-4178-b8f1-e0f66ee24574>. [Accessed: 12-Sep-2025].
- [12] Recorded Future Insikt Group, "The Future of Automated Operations," 2025. [Online]. Available: <https://view.highspot.com/viewer/ddc331c8134d8d28d35028f10edd16a3#1>. [Accessed: 12-Sep-2025].

Intelligent Agents in Multi-Agent Information Security Systems

Roman Yaroviy¹

¹ Ivan Franko 1 Private Higher Educational Establishment "European University", 16v. Vernadskoho Akademika Blvd., Kyiv, 03115, Ukraine

Abstract

In today's digitalization era, the complexity of cyber threats is growing, requiring new approaches to information security. Traditional cybersecurity systems based on monolithic architectures often prove ineffective in IoT environments. Multi-agent systems (MAS), built on the collective interaction of intelligent agents, offer a promising approach to autonomous, adaptive, and scalable protection. This paper examines the use of intelligent agents in multi-agent systems for IoT security, analyzing architectures, key communication protocols, modern use cases, specific challenges, and promising directions for development using artificial intelligence.

Keywords

IoT, intelligent agents, multi-agent systems, artificial intelligence, information security, communication protocols

1. Introduction

The Internet of Things (IoT) encompasses a vast number of resource-constrained devices often connected via unreliable data transmission channels. These devices can serve as points of injection, data redirection, or botnet components. As a result, IoT systems represent high-risk environments where standard centralized security solutions are often cumbersome, slow, or inadequate.

AI-based multi-agent systems — especially those employing reinforcement learning, deep networks, or hybrid methods — can enable IoT systems to:

- Detect anomalies and respond locally and autonomously;
- Adapt security policies based on network behavior (e.g., device class or risk level);
- Scale efficiently as the number of devices grows;
- Reduce dependency on centralized servers and lower response latency.

The purpose of this paper is to describe architectures and methods for intelligent agents in IoT security, compare modern approaches, identify challenges and opportunities, and provide recommendations for further research.

2. Concept of Multi-Agent Systems

A multi-agent system is a collection of intelligent agents united by a common goal or task, which they solve through information exchange, coordinated actions, and collective decision-making. Agents in MAS may have different action sets or tasks, and their contexts are not necessarily shared.

⁶³ Corresponding authors

✉ roman.yaroviy@e-u.edu.ua (R. Yaroviy)

🆔 0000-0001-8978-8137 (R. Yaroviy)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2.1. Definition of an Intelligent Agent in IoT

An intelligent agent in an IoT environment is a software component that:

- 14 Is deployed on a resource-limited device (edge), gateway, or fog node;
- 15 Can collect data from sensors or network logs;
- 16 Processes data locally or through shared models;
- 17 Takes actions (e.g., blocking traffic, isolating devices, changing configurations) and learns from results

2.2. Fundamental Architecture of an Intelligent Agent

The internal structure of an intelligent agent can be viewed as a set of interconnected modules that transform inputs into goal-directed actions.

Core Modules:

- **Decision Integrator**
- **Learning Module**
- **Explainable Engine**
- **Resource Manager**
- **Data Processing Unit**

Operational

Input Sources → Multilevel Buffering → Data Processing → Core Engine → Multi-Agent System → Output Actions → Environment

Flow:

Feedback

Environment → Result Monitoring → Meta-learning → Model Update → Decision Refinement

Cycles:

Self-Defense

Integrity Monitor → Anomaly Detector → Recovery Engine → Blockchain Audit → Security Policy Update

System:

Key Features:

1. **Intelligence:** Adaptive meta-learning, explainability, and quantum-inspired optimization.
2. **Scalability:** Federated learning, neuromorphic computing, and buffered processing under load.
3. **Security & Reliability:** Autonomous defense, blockchain-based auditing, and failure recovery.
4. **User Interaction:** Multimodal interfaces (GUI, API, voice, AR), predictive recommendations, and automated incident documentation.

This architecture enables a transition from reactive to proactive cybersecurity systems with autonomous operation and continuous self-improvement.

3. Future Research Directions

Future development of MAS in cybersecurity focuses on overcoming existing limitations and expanding functionality:

Autonomous SOCs: MAS will underpin self-operating Security Operations Centers (SOCs), where AI handles most operational tasks while humans focus on strategic decisions.

Explainable AI: Developing transparent and trustworthy reasoning mechanisms for regulated sectors.

Enhanced Security Protocols: Protecting agents from manipulation and maintaining the integrity of their decisions.

Standardization: Advancing open protocols (e.g., MCP, A2A) to improve interoperability and reduce vendor lock-in.

Integration with Traditional Systems: Creating hybrid environments by merging MAS with SIEM, SOAR, and TIP platforms.

4. Conclusion

Multi-agent systems based on intelligent agents represent a promising approach to building effective cybersecurity systems capable of countering complex threats. Through autonomy, specialization, and collective coordination, they ensure fast and adaptive responses, proactive threat detection, and efficient security management. Despite challenges such as coordination complexity, agent vulnerability, and data privacy concerns, the advantages of MAS make them a leading research direction. Further standardization, enhanced security mechanisms, and seamless integration will unlock their full potential for scalable, adaptive, and resilient protection against cyber threats.

References

- [1] Alam, Md Morshed; Das, Lokesh Chandra; Roy, Sandip; Shetty, Sachin; Wang, Weichao. "RESTRAIN: Reinforcement Learning-Based Secure Framework for Trigger-Action IoT Environment." *arXiv preprint* arXiv:2503.09513 (2025).
- [2] Severt, Marcos; Casado-Vara, Roberto; Martín del Rey, Ángel; Quintián, Héctor; Calvo-Rolle, José Luis. "Multi-agent reinforcement learning based algorithm detection of malware-infected nodes in IoT networks." *Logic Journal of the IGPL*, 2024.
- [3] Jamshidia, Saeid; Nikanjama, Amin; Wazed Nafia, Kawser; Khomha, Foutse; Rastab, Rasoul. "Application of Deep Reinforcement Learning for Intrusion Detection in Internet of Things: A Systematic Review." *arXiv preprint* arXiv:2504.14436 (2025).
- [4] Andreou, A.; Mavromoustakis, C. X.; Markakis, E.; et al. "Enhancing network slice security with Deep Reinforcement Learning and Moving Target Defense strategies." *Discover Internet of Things* 5 (2025): 67.
- [5] Feng, Chao; Huertas Celdran, Alberto; Sanchez Sanchez, Pedro Miguel; Kreischer, Jan; von der Assen, Jan; Bovet, Gerome; Martinez Perez, Gregorio. "CyberForce: A Federated

- Reinforcement Learning Framework for Malware Mitigation." *arXiv preprint* arXiv:2308.05978 (2023).
- [6] Rosenberger, Julia; Urlaub, Michael; Rauterberg, Felix; Lutz, Tina; Selig, Andreas; Bühren, Michael; Schramm, Dieter. "Deep Reinforcement Learning Multi-Agent System for Resource Allocation in Industrial Internet of Things." *Sensors*, vol. 22, no. 11, 2022, article 4099.
- [7] Fan, Falong; Li, Xi. "PeerGuard: Defending Multi-Agent Systems Against Backdoor Attacks Through Mutual Reasoning." *arXiv preprint* arXiv:2505.11642 (2025).
- [8] "Multi-Agent Deep Reinforcement Learning-based Key Generation for Graph Layer Security." *ACM Transactions on Privacy and Security*, Volume 28, Issue 2, May 2025.
- [9] "Multi-Agent Reinforcement Learning for Privacy-Aware Distributed CNN in Heterogeneous IoT Surveillance Systems." *Journal of Network and Computer Applications*, Volume 230, October 2024.
- [10] "Multi-Agent Reinforcement Learning for Cybersecurity: Classification and Survey." *Intelligent Systems with Applications*, Volume 26, June 2025

Lightweight Machine Learning Models for Securing Linux-Based IoT Devices

Yurii Yashunin¹

¹ National Technical University "Kharkiv Polytechnic Institute", 2 Kyrpychova St., 61000 Kharkiv, Ukraine

Abstract

The paper discusses the application of lightweight machine learning models for detecting and preventing cyber threats in IoT devices running on Linux. It is shown that classical detection methods are limited due to high resource consumption, while optimized neural network architectures (MobileNet, SqueezeNet, TinyML) provide an efficient balance between detection accuracy and performance. Special attention is given to the issues of energy consumption, data confidentiality, and the integration of such solutions with existing security systems.

Keywords

cybersecurity, IoT, Linux, lightweight models, machine learning, information security.

5. Вступ

ІоТ-пристрої, серед яких сенсори, промислові/SCADA контролери та системи відеоспостереження, широко застосовуються в інфраструктурі «розумних міст», промислових мережах і побутових середовищах. Значна частина з них функціонує на базі полегшених дистрибутивів Linux, що обумовлює специфічні вимоги до організації кіберзахисту. Обмеженість обчислювальних ресурсів унеможливорює використання класичних засобів захисту без значної втрати продуктивності, що створює додаткові вразливості.

6. Мета та завдання

Метою роботи є дослідження можливостей застосування легковагових моделей машинного навчання для виявлення аномалій та шкідливого програмного забезпечення на ІоТ-пристроях з обмеженими ресурсами.

Основні завдання:

1. Проаналізувати сучасні архітектури lightweight-мереж (MobileNet, TinyML, SqueezeNet) у контексті захисту ІоТ.
2. Визначити критерії ефективності моделей: точність, швидкодія, енергоспоживання.
3. Запропонувати підхід до інтеграції легковагових моделей у системи безпеки ІоТ на базі Linux.

7. Методи та підходи

Для вирішення поставлених завдань пропонується використання комплексу підходів. По-перше, аналіз системних викликів Linux (system calls) дозволяє формувати базові

⁶⁴ Corresponding author

✉ yashunin.yuriy@gmail.com (I. Iashunin);

🆔 0009-0002-1915-4571 (I. Iashunin);



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

поведінкові ознаки процесів. По-друге, виявлення аномалій у мережевому трафіку можливе на основі компактних згорткових та рекурентних нейронних мереж (CNN, RNN). По-третє, застосування підходів Edge AI забезпечує виконання інференсу безпосередньо на пристрої без потреби у постійному підключенні до хмарних ресурсів. Додатково передбачається оптимізація моделей методами квантизації, pruning та knowledge distillation, що суттєво зменшує їх обчислювальні вимоги.

8. Очікувані результати

Застосування легковагових моделей машинного навчання дозволить досягти кількох важливих результатів. По-перше, знизити середній час виявлення атак у порівнянні з централізованими рішеннями. По-друге, забезпечити автономність роботи IoT-пристроїв без постійного підключення до хмарних сервісів. По-третє, мінімізувати енергоспоживання, що критично для багатьох категорій IoT-пристроїв. І нарешті, підвищити рівень конфіденційності даних завдяки локальній обробці інформації.

9. Висновки

Легковагові моделі машинного навчання є перспективним інструментом побудови систем захисту IoT-пристроїв, що працюють під керуванням Linux. Вони дозволяють досягти балансу між точністю виявлення, швидкодією та низькими вимогами до апаратних ресурсів. Подальші дослідження доцільно зосередити на практичному тестуванні моделей у реальних промислових середовищах та створенні універсальних бібліотек для інтеграції в IoT-платформи.

References

- [1] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2020). ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. IEEE CVPR. doi:10.1109/CVPR.2018.00716
- [2] Howard, A., et al. (2019). Searching for MobileNetV3. IEEE ICCV. doi:10.1109/ICCV.2019.00496.
- [3] Han, S., Mao, H., & Dally, W. J. (2016). Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. ICLR
- [4] Xu, C., et al. (2020). Tiny Machine Learning: A Comprehensive Survey. ACM Computing Surveys, 53(3), 1–34. doi:10.1145/3380958
- [5] Yin, C., et al. (2020). A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks. IEEE Access, 6, 21954–21961. doi:10.1109/ACCESS.2018.2837746
- [6] Liu, J., et al. (2021). Edge AI for IoT Security: Applications, Challenges and Future Directions. Future Generation Computer Systems, 115, 340–353. doi:10.1016/j.future.2020.09.018



МЕТОДИ ПРОТИДІЇ ДЕЗІНФОРМАЦІЙНОМУ ВПЛИВУ ТА СОЦІАЛЬНОМУ ІНЖИНІРИНГУ

Розробка ансамблю моделей на основі ResNet50 для детекції дипфейків

Дмитро Азарний

Факультет радіофізики, електроніки та комп'ютерних систем (ФРЕКС), Київський національний університет імені Тараса Шевченка, просп. Академіка Глушкова, 4 Г, м. Київ, 03022, Україна

Анотація

Діпфейки – це штучно згенеровані або змінені зображення та відео облич, які становлять серйозний виклик для кібербезпеки та довіри до інформації. Попри значний прогрес у дослідженнях методів виявлення підроблених зображень облич, ефективна детекція дипфейків на практиці залишається складною проблемою. У цій роботі запропоновано підхід, що підвищує надійність виявлення дипфейків шляхом об'єднання кількох спеціалізованих моделей у ансамбль. Показано, що ансамблювання таких моделей забезпечує значно кращі результати, ніж будь-яка окрема модель: точність класифікації досягає ~99%, F1-міра перевищує 0.99, а AUC – понад 0.995 на змішаному тестовому наборі. Додавання нового складного датасету на другому етапі навчання істотно підвищило здатність моделей розпізнавати тонкі артефакти, збільшивши повноту (Recall) з ~0.8 до понад 0.95. Розроблений ансамбль моделей демонструє практичну можливість високоточного розпізнавання дипфейків та може слугувати основою для побудови надійних систем кібербезпеки.

Ключові слова

дипфейк, виявлення підроблених зображень, ResNet50, ансамблеве навчання, механізм уваги

1. Вступ

Діпфейки, тобто штучно згенеровані або модифіковані зображення й відео облич, набули значного поширення та становлять серйозну загрозу для інформаційної безпеки. Основні труднощі полягають у великій різноманітності підходів до створення фейків та ризику того, що модель детекції може вивчити лише специфічні артефакти окремих датасетів і погано узагальнювати на нові приклади. Тому актуальним є підвищення стійкості та точності детекторів шляхом поєднання різних підходів [1].

У цій роботі запропоновано ансамбль з кількох спеціалізованих нейронних мереж на основі ResNet50. Кожна модель спрямована на виявлення дипфейків через різні механізми уваги. Об'єднання їхніх прогнозів усередненням ймовірностей дозволяє компенсувати слабкі сторони окремих детекторів і суттєво підвищити загальну надійність[4].

2. Опис підходу

Архітектура моделей. Усі розроблені детектори побудовані на базі глибокої згортової мережі ResNet50. Поверх цієї основи реалізовано різні спеціалізовані модифікації, націлені на виявлення характерних артефактів дипфейків.

⁶⁵Corresponding authors

✉ dazarny@gmail.com (D. Azarny)

id 0009-0000-9549-8873 ((D. Azarny))



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Ансамблювання. Для отримання фінального рішення використовуються всі вказані детектори одночасно. Виходи моделей (ймовірності належності до класу «фейк») усереднюються, після чого виконується класифікація як діпфейк або реальне зображення.

Стратегія навчання. Навчання моделей виконувалося у дві стадії. Спочатку всі мережі були навчені на великому об'єднаному наборі різноманітних фейкових і реальних зображень, після чого здійснено їх перенавчання (fine-tuning) на новому, більш складному датасеті із високоякісними малопомітними підробками.

3. Результати

Ансамбль моделей продемонстрував значно вищу точність класифікації порівняно з будь-яким окремим детектором. Підсумкова **accuracy** ансамблю на великому тестовому наборі становить близько 99%, тоді як найкраща з індивідуальних мереж досягала ~96–97%. Відповідно, і **F1-міра** ансамблю (~0.99) та **AUC** (~0.996) перевищують аналогічні показники будь-якої окремої моделі, демонструючи відмінний баланс між precision та recall. Ці результати підтверджують ефективність об'єднання різнорідних детекторів: ансамбль успадковує сильні сторони всіх компонентів та компенсує їхні слабкості. Завдяки цьому досягається близький до максимального рівень надійності та точності виявлення діпфейків.

4. Висновки

Ансамблеве об'єднання спеціалізованих моделей шляхом усереднення їхніх прогнозів продемонструвало виняткову ефективність. Розроблений ансамбль перевершив усі індивідуальні мережі за основними метриками, досягнувши близько 99% точності класифікації та високої збалансованості precision/recall на великому тестовому наборі. Практичні результати дослідження підтверджують, що комбінування різних архітектурних підходів значно підвищує надійність виявлення діпфейків і знижує ризик помилкових класифікацій. Отже, ансамбль спеціалізованих ResNet50-моделей є перспективним шляхом для створення стійких до різних видів діпфейків систем кіберзахисту.

Література

- [1] Tolosana R., Vera-Rodriguez R., Fierrez J., Morales A., Ortega-Garcia J. DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. Information Fusion, vol. 64, pp. 131–148, 2020. DOI: 10.1016/j.inffus.2020.06.014.
- [2] Woo S., Park J., Lee J.-Y., Kweon I. S. CBAM: Convolutional Block Attention Module. In Proc. ECCV, pp. 3–19, 2018. DOI: 10.1007/978-3-030-01234-2_1.
- [3] Rössler A., Cozzolino D., Verdoliva L., Riess C., Thies J., Nießner M. FaceForensics++: Learning to Detect Manipulated Face Images. arXiv:1901.08971, 2019. (Proc. ICCV 2019, pp. 1–11).
- [4] Frank J., Eisenhofer T., Schönherr L., Fischer A., Kolossa D., Holz T. Leveraging Frequency Analysis for Deep Fake Image Recognition. In Proc. ICML, 2020, pp. 3247–3258. arXiv:2003.08685

Математичне моделювання поширення пропаганди та фейкової інформації у соціальних мережах: модель SEIPR

Ярина Борисюк⁶⁶

¹ Львівський національний університет імені Івана Франка, Університетська 1, 79000 Львів, Україна

Анотація

У роботі представлено математичну модель SEIPR, призначену для дослідження процесів поширення пропаганди та фейкової інформації у соціальних мережах. Модель враховує два типи поширювачів: звичайних користувачів та спеціальні акаунти з великою аудиторією. Проведено аналітичний аналіз, досліджено стійкість точок рівноваги та визначено базове репродукційне число. Чисельні експерименти показали ключову роль спеціальних акаунтів у збереженні активного поширення інформації. Результати роботи можуть бути використані для побудови стратегій протидії дезінформації.

Ключові слова

соціальні мережі, SEIPR, пропаганда, математичне моделювання, якісний аналіз

1. Вступ

Сучасні соціальні мережі є основним середовищем поширення інформації та обміну новинами. Поряд із достовірними повідомленнями у них активно циркулюють фейки та пропагандистські матеріали. Це створює серйозні ризики для суспільства, оскільки дезінформація здатна формувати викривлену громадську думку, провокувати соціальну напругу та спричиняти політичну нестабільність. Дослідження механізмів поширення такої інформації може допомогти розробити ефективні заходи протидії. Для аналізу цих процесів часто застосовуються епідеміологічні підходи [1,2]. У даній роботі запропоновано модель SEIPR, яка є модифікацією класичних SEIR- та SMIR-моделей і враховує роль великих акаунтів (суперрозповсюджувачів) у поширенні дезінформації [3].

2. Модель SEIPR

У SEIPR-моделі користувачі соціальної мережі поділяються на п'ять груп:

- $s(t)$ – сприйнятливі акаунти, які ще не стали жертвою чуток, але сприйнятливі до цієї інформації;
- $e(t)$ – інфіковані, але пасивні;
- $i(t)$ – звичайні користувачі, активні поширювачі фейків серед своїх друзів;
- $p(t)$ – спеціальні акаунти з великою кількістю підписників, активні поширювачі інформації серед незнайомих людей;

⁶⁶ Corresponding authors

✉ yaryna.borysyuk@lnu.edu.ua (Я. Борисюк)

ORCID 0000-0002-0322-0637 (Я. Борисюк)(E. Lindström)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

- $r(t)$ – акаунти, які не зацікавлені в поширенні інформації, стійкі до пропаганди, які вийшли з груп розповсюджувачів пропаганди.

Перехід між станами описується системою диференціальних рівнянь виду:

$$\begin{cases} \frac{ds}{dt} = \lambda - \beta_I s i - \beta_P s p - d s \\ \frac{de}{dt} = \beta_I s i + \beta_P s p - \mu e - d e \\ \frac{di}{dt} = \mu e - \gamma i - \delta i - d i \\ \frac{dp}{dt} = \gamma i - \beta p - d p \\ \frac{dr}{dt} = \delta i + \beta p - d r \end{cases} \quad (1)$$

З початковими умовами:

$$s(0)=s_0 \geq 0, e(0)=e_0 \geq 0, i(0)=i_0 \geq 0, p(0)=p_0 \geq 0 \quad (2)$$

Загальна кількість акаунтів $s(t)+e(t)+i(t)+p(t)+r(t)=1$ для всіх $t \in [0, T]$.

У моделі SEIPR параметри мають такий зміст:

- λ – швидкість появи нових користувачів;
- β_I – коефіцієнт передачі інформації через звичайних користувачів;
- β_P – коефіцієнт передачі через спеціальні акаунти;
- μ – швидкість переходу від пасивного до активного поширювача;
- γ – швидкість перетворення звичайного акаунту на суперрозповсюджувача;
- δ – швидкість втрати інтересу до розповсюдження фейків серед групи звичайних користувачів $i(t)$;
- β – швидкість втрати інтересу серед суперрозповсюджувачів $p(t)$;
- d – коефіцієнт виходу користувачів із мережі.

Зауважимо, що виходячи з фізичного змісту моделі, всі ці параметри невід'ємні. Кожен із цих параметрів має пряму інтерпретацію в умовах соціальних мереж. Наприклад, коефіцієнт β_P характеризує силу впливу великих акаунтів на аудиторію. Якщо такий акаунт має мільйони підписників, то навіть один його пост може спричинити різке зростання кількості інфікованих користувачів. Параметр δ описує природне згасання інтересу: навіть якщо користувач поширює пропаганду, з часом він може втратити до неї інтерес. Таким чином, система (1)–(2) дозволяє моделювати як короткотривалі, так і довготривалі кампанії.

3. Аналіз моделі

Для системи (1)-(2) доведено існування та додатну визначеність розв'язку задачі аналізу. Ключовим показником для SEIPR-моделі є базове репродукційне число R_0 . Якщо $R_0 < 1$, то пропаганда згасає та система сходиться до рівноваги без поширення інформації. Якщо $R_0 > 1$, то формується стала рівновага з активним поширенням. Для обчислення R_0 використовується метод матриці наступного покоління, який базується на розрахунку частки нових інфікованих, створених одним поширювачем [4]. Отримано:

$$R_0 = \frac{\frac{\lambda}{d} \cdot \beta_I \mu (\beta + d) + \beta_P \mu \gamma}{(\mu + d)(\gamma + \delta + d)(\beta + d)}.$$

Знайдено дві особливі точки рівноваги системи: особлива точка рівноваги відсутності пропаганди $E_0(\lambda/d, 0, 0, 0, 0)$ та точка рівноваги стабільного поширення інформації

$$E_1(s_1, e_1, i_1, p_1, r_1), \quad \text{де} \quad s_1 = \frac{\lambda}{R_0 d}, \quad e_1 = \frac{\lambda - \lambda/R_0}{\mu + d}, \quad i_1 = \frac{\mu(\lambda - \lambda/R_0)}{(\mu + d)(\gamma + \delta + d)},$$

$$p_1 = \frac{\gamma \mu (\lambda - \lambda/R_0)}{(\mu + d)(\gamma + \delta + d)(\beta + d)}, \quad r_1 = \frac{(\lambda - \lambda/R_0)(\delta(\beta + d) + \beta \gamma)}{d(\mu + d)(\gamma + \delta + d)(\beta + d)}.$$

Проаналізовано стійкість точок рівноваги. Зокрема, доведено, що точка рівноваги $E_1(s_1, e_1, i_1, p_1, r_1)$ існує, якщо $\lambda - \lambda/R_0 > 0$, тобто при $R_0 > 1$. Отже, при досягненні точки рівноваги E_1 , інформація буде стабільно поширюватися в системі при $R_0 > 1$ та зникне при $R_0 < 1$, що, враховуючи епідемічний підхід до побудови моделі SEIPR, відповідає поведінці поширення інфекційної хвороби.

4. Чисельні експерименти

Для чисельного аналізу використано середовище MATLAB. Проведено ряд чисельних експериментів. Було досліджено три сценарії: (а) висока активність друзів і низька активність спеціальних акаунтів; (б) низька активність друзів і висока активність спеціальних акаунтів; (в) баланс двох каналів. Результати свідчать, що у сценарії (б) швидкість поширення значно більша, ніж у сценарії (а). Це підтверджує ключову роль спеціальних акаунтів у збереженні пропаганди.

Крім того, було проведено експерименти із варіюванням параметра δ , що описує втрату інтересу. При низькому δ інформаційна хвиля поширюється довше, тоді як при високому δ швидко згасає. Також виявлено, що підвищення μ (швидкість переходу з e до i) збільшує темпи зростання кількості поширювачів.

Таким чином, моделювання показує, що контроль лише за звичайними користувачами є недостатнім. Наявність навіть кількох активних спеціальних акаунтів забезпечує підтримку високого рівня поширення. Система (1)-(2) демонструє, що ефективна стратегія боротьби повинна включати блокування або обмеження впливу саме цих акаунтів.

5. Висновки

Модель SEIPR є зручним інструментом для дослідження поширення пропаганди та фейкової інформації у соціальних мережах. Вона дозволяє описати вплив звичайних і спеціальних акаунтів, визначити критичні умови для поширення та спрогнозувати наслідки зміни параметрів. Отримані результати можуть бути використані для формування

ефективних стратегій протидії дезінформації, зокрема шляхом обмеження активності великих акаунтів та підвищення медіаграмотності користувачів.

Література

- [1] R. P. Ojha, et al., Controlling of Fake Information Dissemination in Online Social Networks: An Epidemiological Approach, IEEE Access 11 (2023) 32229–32240. doi:10.1109/ACCESS.2023.3262737.
- [2] R. D'Ambrosio, G. Giordano, S. Mottola, B. Paternoster, Stiffness Analysis to Predict the Spread Out of Fake Information, Future Internet 13 (2021) 222. doi:10.3390/fi13090222.
- [3] Y. Yu, J. Liu, J. Ren, C. Xiao, Stability analysis and optimal control of a rumor propagation model based on two communication modes: friends and marketing account pushing, Math. Biosci. Eng. 19 (2022) 4407–4428. doi:10.3934/mbe.2022204. PMID:35430821.
- [4] P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of diseases transmission, Math. Biosci. 180 (2002) 29–48. doi:10.1016/S0025-5564(02)00108-6.

Тенденції інформаційного простору та їхній вплив на стратегічні комунікації сектору безпеки і оборони: виклики та можливості

Ірина Макарченко

Національний університет оборони України, Київ, Україна

Анотація

У роботі проаналізовано ключові світові медійні та технологічні тенденції, що формують сучасний інформаційний простір, та їхній вплив на стратегічні комунікації Сил оборони. Визначено потенційні ризики, можливості та адаптивні підходи, спрямовані на збереження інформаційної переваги в умовах швидких змін медійного середовища. Особливу увагу приділено новітнім загрозам, зокрема поширенню мультимодального штучного інтелекту та зростанню ролі медійних інфлюенсерів, а також практичним рекомендаціям щодо підвищення стійкості та узгодженості державних комунікацій у сфері оборони.

Ключові слова

стратегічні комунікації; дезінформація; соціальний інжиніринг; штучний інтелект; соціальні медіа; інформаційна безпека; когнітивна війна; інформаційний простір

1. Вступ

Стрімкий розвиток технологій штучного інтелекту (ШІ) та соціальних медіа докорінно змінює характер інформаційного простору, у якому функціонують стратегічні комунікації сил оборони та держави. Мультимодальні ШІ-системи, здатні синхронно аналізувати текст, аудіо, відео та зображення, відкривають безпрецедентні можливості для обробки розвідданих, зокрема у сфері розпізнавання облич, аналізу відеопотоків з безпілотних апаратів чи моніторингу змін у середовищі. Водночас така технологічна експансія створює і нові загрози: зростання кількості псевдоекспертів, які продукують дезінформацію за допомогою автоматизованих інструментів, підвищує ризики інформаційних маніпуляцій, а автоматизація новин без редакційного контролю сприяє поширенню упередженого або викривленого контенту, що може бути навмисно насиченим ворожими наративами (1).

2. Тенденції інформаційного простору та їхній вплив на стратегічні комунікації сектору безпеки і оборони

Одним із помітних зрушень останніх років стало впровадження платної верифікації акаунтів у соціальних мережах, що призвело до збільшення кількості підтверджених облікових записів, пов'язаних із пропагандистськими структурами. У результаті зростає ймовірність, що користувачі, зокрема з низьким рівнем медіаграмотності, сприйматимуть дезінформаційні повідомлення як достовірні. Для стратегічних комунікацій це означає необхідність переосмислення підходів до виявлення та нейтралізації таких джерел, а також впровадження превентивних заходів щодо захисту цільових аудиторій від маніпулятивного впливу.

Щільніший контроль доступу до даних соціальних мереж ускладнює процес збору та аналізу інформації про ворожі інформаційні кампанії, що знижує оперативність і точність оцінки загроз (2). Це вимагає розвитку альтернативних інструментів моніторингу, які

поєднують OSINT, ІІІ-аналітику та кросплатформне відстеження контенту. Паралельно, трансформація журналістики та зростання ролі TikTok як платформи для прямих комунікацій із молодшою аудиторією змінює механіку формування суспільних наративів (3). Для сил оборони важливим завданням стає розуміння культурних кодів та форматів контенту, що домінують у молодіжному середовищі, аби вчасно інтегрувати власні меседжі у релевантний інформаційний потік.

Кросплатформність соціальних медіа та миттєве перетікання контенту з однієї платформи на іншу посилюють ефект «повторення правди», коли багаторазове відтворення неправдивої інформації на різних ресурсах підвищує її сприйняття як достовірної. У поєднанні з появою нових нішевих платформ, що швидко набирають аудиторію та нерідко стають простором радикалізації, це формує новий ландшафт інформаційних загроз (4). Ефективні стратегічні комунікації в таких умовах потребують гнучкої адаптації — від швидкого освоєння нових майданчиків до створення механізмів контролю якості поширюваного контенту.

Додатковим викликом є трансформація соціальних мереж у пошукові інструменти. Користувачі дедалі частіше звертаються до відеоплатформ у пошуку відповідей на питання, що стосуються війни, політики чи економіки, отримуючи матеріали з потенційно чутливою інформацією (4). Ця тенденція має двоїстий характер: з одного боку, вона створює ризики витоку даних, з іншого — відкриває можливості для цілеспрямованого впливу на інформаційний порядок денний шляхом оптимізації контенту під популярні пошукові запити.

Таким чином, технологічні інновації та поведінкові зміни аудиторій формують нову операційну реальність для стратегічних комунікацій у сфері оборони. Вона вимагає одночасно глибшої інтеграції інструментів ІІІ у роботу комунікаційних підрозділів, активного розвитку навичок критичного мислення у військовослужбовців та створення адаптивних моделей моніторингу і реагування на дезінформаційні впливи. Лише системне поєднання технологічних рішень, кадрової підготовки та міжнародної координації дозволить зберегти ефективність комунікаційних стратегій у швидкоплинному інформаційному середовищі.

3. Висновки

Ефективність стратегічних комунікацій сил оборони в умовах швидкоплинного інформаційного середовища залежить від здатності одночасно: інтегрувати інструменти ІІІ в операційні цикли моніторингу, аналізу та реагування; підтримувати кросплатформну синхронізацію меседжів і форматів; підвищувати інформаційну стійкість персоналу та вразливих аудиторій через навчання й процедурні запобіжники. Збалансоване поєднання технологій, процесів і підготовки кадрів мінімізує когнітивні ризики та підвищує ймовірність досягнення бажаних ефектів.

Література

- [1] Fredheim R. Virtual Manipulation Brief 2023/1: Generative AI and its Implications for Social Media Analysis. NATO StratCom COE, Riga, 2023. URL: <https://stratcomcoe.org/pdfs/?file=/publications/download/Virtual-Manipulation-Brief-2023-2.pdf>.
- [2] Dentsu. 2024 Media Trends. 2023. URL: <https://insight.dentsu.com/2024-media-trends/>.

- [3] Nieman Lab. TV reporters become TikTok influencers. December 2023. URL: <https://www.niemanlab.org/2023/12/tv-reporters-become-tiktok-influencers/>.
- [4] Hootsuite Blog. Social media trends. 2025. URL: <https://blog.hootsuite.com/social-media-trends/>.

A Model for Keyword Spotting in a Voice Signal to Counter Social Engineering and Disinformation

Andrii Didus¹ and Ihor Tereikovskiy¹

¹National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 03056, Kyiv

Abstract

The voice communication channel is a significant vector for social engineering attacks and the spread of disinformation. Existing countermeasures that rely on cloud services have substantial drawbacks, including high latency, dependence on network connectivity, and privacy risks, making them unsuitable for real-time applications. This paper proposes a resource-efficient modular keyword spotting model designed for autonomous operation on resource-constrained edge devices. The model's architecture is based on the transformation of sequences of Mel-frequency cepstral coefficients into compact string "fingerprints" using differentiated weighting of informative features, followed by classification using the Levenshtein distance. Experimental validation on a Ukrainian-language command corpus demonstrated high performance: the F1-score reached 0.92 in ideal conditions and 0.78 at a signal-to-noise ratio of 5 dB. The proposed model significantly surpasses baseline and classical counterparts in the balance of accuracy, speed, and resource efficiency, which confirms its suitability for creating autonomous systems for proactive detection of auditory threats.

Keywords

keyword spotting, voice processing, social engineering, disinformation, resource-constrained systems

1. Introduction

The modern information security landscape is characterized by the expansion of attack vectors beyond traditional digital channels. The voice communication channel has evolved into a full-fledged space for attacks, actively used for social engineering (particularly voice phishing), spreading disinformation, and unauthorized system access. The dominant paradigm for countering such threats is based on the centralized analysis of audio data in cloud environments. This approach has several fundamental limitations. Firstly, the high latency associated with data transmission makes immediate real-time threat response impossible. Secondly, the reliance on a stable network connection makes such systems vulnerable in tactical situations or in the absence of reliable infrastructure. Thirdly, transmitting confidential audio data to third-party servers creates significant privacy and security risks. [1]

This work proposes a paradigm shift towards autonomous on-device analysis. A resource-efficient keyword spotting (KWS) model has been developed to eliminate the shortcomings of cloud-based solutions. The model provides a private, low-latency, and connection-resilient framework for threat detection, turning technical characteristics into strategic advantages in the field of cybersecurity. [2-4]

⁶⁷ Corresponding authors

✉ didusavd@gmail.com (A. Didus); terejkowski@ukr.net (I. Tereikovskiy)

ORCID 0009-0004-2235-6742 (A. Didus); 0000-0003-4621-9668 (I. Tereikovskiy)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Model Architecture and Methods

The proposed model is built on a modular principle, ensuring flexibility and adaptability. The architecture includes sequential stages of signal processing, feature extraction, "acoustic fingerprint" creation, and classification.

2.1. Signal Processing and Feature Extraction

The initial processing converts the raw audio signal into a set of numerical features. This stage includes:

Signal preparation: Voice Activity Detection (VAD) to isolate speech from background noise and amplitude normalization to compensate for volume variations.

Feature extraction: Calculation of Mel-frequency cepstral coefficients (MFCCs). To enhance reliability, static MFCCs are supplemented with dynamic features—the first (delta) and second (delta-delta) derivatives—which encode the rate and acceleration of spectral changes. This improves robustness to variations in speech tempo.

2.2. "Acoustic Fingerprint" Formation

The key element of the model is the transformation of a sequence of feature vectors into a compact string representation (an "acoustic fingerprint"). The process is based on the hypothesis of the unequal informativeness of different MFCCs. A weight vector W is applied to amplify the coefficients most significant for phonetic content. The weighted feature vectors are averaged, quantized, and serialized into a final string F according to Formula 1 below:

$$F = \text{Serialize} \left(Q \left(\frac{1}{T} \sum_{t=1}^T (M_t \odot W) \right) \right), \quad (1)$$

where M is the feature vector at time t , T is the number of frames, Q is the quantization function, and $\odot$ – is element-wise multiplication.

2.3. Classification by Levenshtein Distance

The final classification is performed by comparing the input "fingerprint" with reference templates from a pre-compiled dictionary. The Levenshtein distance is used as a similarity metric, which is a deterministic and computationally efficient method. The choice of this method ensures the transparency and reproducibility of results, which is critical for security systems. [5]

3. Experimental Study

To validate the model, a series of experiments was conducted using a software prototype implemented in Python.

Dataset: A specially recorded Ukrainian-language corpus containing commands for controlling a ground drone was used. Scalability was tested on lexicons of 10, 100, and 200 words.

Baseline model for comparison: To evaluate the effectiveness of the proposed improvements (dynamic features, weighting), a baseline model was implemented that used only static, unnormalized MFCCs.

4. Results

The model demonstrated high accuracy and reliability. In ideal conditions (clean signal, known speaker), the F1-score for a 100-word lexicon was 0.92. The system showed high resistance to noise, maintaining an F1-score of 0.78 at a signal-to-noise ratio of 5 dB. [2-4]

For contextualization of the results, a comparative analysis was conducted with the baseline model, the classic Dynamic Time Warping (DTW) algorithm, and a cloud-based ASR (Automatic Speech Recognition) service. The results for the 100-word lexicon are presented in Table 1. [6]

Table 1

Comparative analysis of KWS architecture performance

Model	Memory Footprint	Inference Time	RTF	F1-Score	F1-Score (5 dB noise)
Baseline	~150KB	~3 ms	0.003	0.75	0.45
Proposed	~250KB	~5 ms	0.005	0.92	0.78
Classical DTW	~2MB	~20 ms	0.02	0.88	0.65
Cloud-based	N/A	~450 ms	0.45	0.97	0.91

Conclusions

This paper presents a validated framework for transferring auditory threat detection tasks to peripheral devices. The proposed KWS model is lightweight, fully autonomous, and effective for critical applications where access to cloud infrastructure is limited or unavailable. Key findings of the study demonstrate that the "acoustic fingerprint" approach ensures privacy, ultra-low latency for real-time response, and resilience to network failures. Empirical validation confirmed that the model achieves an optimal balance between accuracy and computational efficiency. This enables a shift from a reactive, centralized security model to a proactive and decentralized one, strengthening the protection of the voice communication channel at the edge device level.

References

- [1] D. O'Shaughnessy, "Trends and developments in automatic speech recognition research," *Computer Speech & Language*, vol. 83, 2024, 101538, ISSN 0885-2308, <https://doi.org/10.1016/j.csl.2023.101538>.
- [2] S. Alharbi et al., "Automatic Speech Recognition: Systematic Literature Review," *IEEE Access*, vol. 9, pp. 131858-131876, 2021, doi: 10.1109/ACCESS.2021.3112535.
- [3] G. Chen, C. Parada and G. Heigold, "Small-footprint keyword spotting using deep neural networks," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 2014, pp. 4087-4091, doi: 10.1109/ICASSP.2014.6854370.
- [4] A. Mahmood and U. Köse, "Speech recognition based on convolutional neural networks and MFCC algorithm", *Adv. Artif. Intell. Res.*, vol. 1, no. 1, pp. 6–12, 2021.

- [5] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, Feb. 1989, doi: 10.1109/5.18626.
- [6] Дичка, І. А., Терейковський, І. А., Дідус, А. В., Терейковська, Л. О., & Бояринова, Ю. Є. (2023). Оцінка ефективності засобів розпізнавання ключових слів у голосовому сигналі. *Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки*, 34(3), 1-7. <https://doi.org/10.32782/2663-5941/2023.3.1/19>.
- [7] D. Seo, H. -S. Oh and Y. Jung, "Wav2KWS: Transfer Learning From Speech Representations for Keyword Spotting," *IEEE Access*, vol. 9, pp. 80682-80691, 2021, doi: 10.1109/ACCESS.2021.3078715.
- [8] S. Dua et al., "Developing a Speech Recognition System for Recognizing Tonal Speech Signals Using a Convolutional Neural Network," *Applied Sciences*, vol. 12, no. 12, 2022, Art. no. 6223, <https://doi.org/10.3390/app12126223>.
- [9] J. Oruh, S. Viriri and A. Adegun, "Long Short-Term Memory Recurrent Neural Network for Automatic Speech Recognition," *IEEE Access*, vol. 10, pp. 30069-30079, 2022, doi: 10.1109/ACCESS.2022.3159339.
- [10] C. -H. H. Yang et al., "Decentralizing Feature Extraction with Quantum Convolutional Neural Network for Automatic Speech Recognition," *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, 2021, pp. 6523-6527, doi: 10.1109/ICASSP39728.2021.9413453.

The method of cognitive reverse reframing under conditions of large-scale armed aggression of the russian federation against Ukraine

Serhii Bazarnyi¹, Roman Zubko²

¹*National Defence University of Ukraine, Kyiv, Ukraine*

²*National Defence University of Ukraine, Kyiv, Ukraine*

Abstract

The article presents Cognitive Reverse Reframing (CRR) as an innovative method to counter manipulative narratives during Russia's large-scale aggression against Ukraine. Combining cognitive-behavioral therapy, cognitive linguistics, and neuro-semantic, CRR neutralizes hostile disinformation frames by replacing them with constructive alternatives. Implemented through the “Cogniframe” platform, it enables automated detection, semantic analysis, generation, validation, and dissemination of alternative narratives. Ultimately, CRR enhances strategic communications by boosting cognitive resilience and reducing the influence of adversarial information-psychological operations.

Keywords

strategic communications, cognitive security, disinformation, framing, neuro-semantic, cognitive resilience, target audience, reframing method.

The large-scale armed aggression of the Russian Federation against Ukraine is characterized by its systemic and multidimensional nature, encompassing simultaneously the military, political, economic, informational, and cognitive domains. This aggression is conducted not only through conventional armed confrontation but also by means of a complex set of asymmetric actions aimed at undermining the resilience of the Ukrainian state. Alongside the use of military force, the adversary systematically and deliberately implements information and psychological operations designed to demoralize the civilian population, erode public trust in governmental institutions, and provoke internal divisions as well as social destabilization.

According to the Centre for Strategic Communications and Information Security, during the period 2022–2025 the dominant themes of hostile propaganda included such narratives as “Western fatigue with Ukraine,” “ineffectiveness of the Ukrainian government,” “economic collapse,” and the “inevitability of defeat.” A distinctive feature of these disinformation constructs is their adaptability: they are continuously modified and recalibrated in line with the current news agenda, exploiting emerging societal and economic crises in order to intensify their destructive impact. Consequently, the aggression assumes a hybrid form, combining physical coercion with multidimensional cognitive and informational pressure.

One of the key mechanisms underlying the dissemination of disinformation is framing—the strategic presentation of information within a specific cognitive “frame” that shapes audience perception and guides subsequent behavioral responses. The adversary extensively employs framing techniques such as metaphorization (e.g., “Ukraine as a puppet of the West”), context distortion (selective quoting of international partners to reinforce discrediting narratives), and repetitive cognitive loading (“daily reminders of high casualties”). These techniques generate feelings of hopelessness, fatigue, and powerlessness among the population, undermining trust in state institutions and fostering social apathy.

Under such conditions, traditional fact-checking mechanisms lose part of their effectiveness, since rebuttals often reach the audience with a delay and are not always capable of altering the already internalized emotional perception of messages. Moreover, empirical evidence demonstrates that blocking disinformation resources is insufficient: the adversary rapidly restores communication channels, while segments of the audience tend to interpret censorship as "confirmation" of the credibility of hostile messages. This underscores the necessity of developing new approaches that operate not only at the informational level but also address the cognitive mechanisms of perception.

In this regard, the method of cognitive reverse reframing (CRR) emerges as a promising instrument for building societal resilience. The essence of CRR lies in replacing destructive cognitive frames with constructive ones, thereby reshaping perception without triggering resistance from target audiences. Conceptually, the method is grounded in the principles of cognitive-behavioral therapy, cognitive linguistics, and neuro-semantics. Its strategic objective is the neutralization of hostile narratives through the integration of alternative constructive cognitive frameworks into the natural information environment, enabling the transformation not only of perception but also of social attitudes and behavioral orientations.

The method of cognitive reverse reframing is operationalized through a structured five-stage procedure that ensures systematicity and reproducibility. The stages can be outlined as follows:

1. Identification of the hostile frame. This stage involves the extraction of key disinformation narratives, manipulative patterns, and cognitive triggers that shape the interpretative framework of the message. The task is to construct a map of discursive constructs that define the perceptual boundaries.

2. Semantic analysis. At this phase, linguistic and cognitive components of the message are examined. This includes the identification of dominant keywords and collocations, the assessment of tonal polarity, and the exploration of metaphorical contexts. Semantic decomposition makes it possible to detect vulnerabilities within the hostile frame and points of potential transformation.

3. Reframing. An alternative narrative is constructed, characterized by positive emotional valence, rational underpinning, and cognitive value for the audience. The challenge is to balance emotional appeal with logical persuasiveness in order to ensure that the new frame is perceived as natural and non-confrontational.

4. Contextual validation. The newly constructed frame is tested on a designated target audience to evaluate its effectiveness. The assessment focuses on the degree of acceptance, emotional resonance, and the capacity to neutralize the destructive impact of the original message.

5. Reverse action (dissemination). At the final stage, reframed messages are disseminated within the public information space through relevant communication channels and strategic platforms capable of achieving both broad reach and targeted engagement.

In this way, the method facilitates a transition from a reactive to a proactive model of counter-disinformation activity, fostering societal resilience by systematically replacing destructive cognitive frames with constructive ones.

Results and Discussion. The practical application of the Cognitive Reverse Reframing (CRR) method demonstrates its efficiency in neutralizing hostile narratives and strengthening societal cognitive resilience. For instance, the destructive frame "Ukraine is a failed state" can be transformed into the constructive narrative "Ukraine is a state that demonstrates resilience, unity, and enjoys broad international support." Such a transformation reduces the negative emotional impact while simultaneously reinforcing collective confidence, enhancing trust in state institutions, and contributing to social consolidation.

The technical implementation of CRR is supported by the modular system "Cogniframe", designed as an integrated platform for the analysis, transformation, and management of cognitive

frames within the information environment. The architecture of the system comprises five interrelated modules:

- a data collection module, enabling multi-source acquisition from OSINT, SOCMINT, and MEDINT inputs;
- a semantic analyzer, powered by GPT-class models, responsible for the identification of keywords, tonal polarity, and metaphorical structures;
- a reframing core, generating alternative messages with positive cognitive valence and rational context;
- a distribution module, tailored to the functional specificities of platforms such as Telegram, Facebook, YouTube, and TikTok;
- a monitoring module, designed to measure audience reach, assess resonance levels, and calculate the Cognitive Impact Index (CII).

Due to its modular design, the “Cogniframe” system provides operational flexibility, allowing processes to be tailored to diverse information ecosystems and ensuring rapid adaptation to the evolving dynamics of disinformation campaigns, thereby making it an effective tool for the practical implementation of CRR.

Core Components. The “Cogniframe” system, which operationalizes the Cognitive Reverse Reframing (CRR) method, is designed as a modular architecture comprising five principal components:

1. Data Processing Module (Input-Module) — responsible for collecting and aggregating information from diverse open sources (OSINT, SOCMINT, MEDINT), thereby covering both traditional media and digital social networks as well as specialized information channels.
2. Semantic Analyzer (NLU-Mod) — a GPT-class neurolinguistic model that automatically identifies frames, key narratives, and contexts, including tonal polarity and metaphorical structures.
3. Reframing Generator (Reframing-Core) — produces alternative narratives tailored to the cognitive styles of the audience (analytical, affective, etc.), ensuring alignment of content with the psychological and socio-cultural characteristics of target groups.
4. Distribution Module (Output-Core) — disseminates reframed messages across multiple information ecosystems (Telegram, Facebook, YouTube, TikTok), with due consideration for platform-specific algorithms and user interaction logics.
5. Analytical Module (Monitoring-Unit) — evaluates audience reach, resonance levels, and calculates the Cognitive Impact Index (CII), thus enabling quantitative assessment of campaign effectiveness.

This method provides the capacity for automated cognitive counter-influence within the framework of strategic communications. Unlike conventional practices of blocking or deleting disinformation content, CRR integrates directly into the natural flow of information exchange, thereby minimizing psychological resistance and enhancing the credibility of messages.

Conclusions. The Cognitive Reverse Reframing method, combined with the “Cogniframe” platform, can be regarded as a promising component of the system of strategic communications and information security. Its principal advantage lies in its proactive nature: instead of removing destructive content, it substitutes it with constructive frames seamlessly embedded in the ordinary information flow. Future research should focus on the large-scale testing of the method within real-world information campaigns and on the development of quantitative indicators of its effectiveness, ensuring standardized measurement of outcomes.

References

- [1] V.M. Oliinyk, Strategic Communications in the Security and Defence Sector of Ukraine: Essence, Structure and Development Prospects, National Defence University of Ukraine, Kyiv, 2021.
- [2] I.Yu. Pavlenko, Cognitive Security as a Component of National Security: Monograph, National Defence University of Ukraine, Kyiv, 2020.
- [3] I.M. Sivoha, Ya.O. Lashin, Model for Predicting Information Threats in the System of Strategic Communications, STRATCOM. Scientific-Publicistic Edition of the Institute of Strategic Communications of NDUU 5 (2025) 53–59.
- [4] National Institute for Strategic Studies, Information Security of Ukraine: Analytical Report, Kyiv, 2022.
- [5] I. Zhdanov, S. Kostiuk, Information and Psychological Operations: Ukrainian Experience in Countering Hybrid Aggression, Military Science 1(43) (2023) 25–39.
- [6] A.V. Barovska, Cognitive Aspects of the Influence of Information Operations in Modern Warfare, V.M. Koretsky Institute of State and Law of NAS of Ukraine, Kyiv, 2021.
- [7] A.P. Misiura, Framing in Political Communication: A Cognitive-Semantic Approach, Scientific Notes of the Institute of Political and Ethno-National Studies of NAS of Ukraine 4 (2019) 115–127.
- [8] O.M. Liashenko, Strategic Communications as a Tool of Countering Disinformation: The Ukrainian Context, National Institute for Strategic Studies, 2022.
- [9] V.A. Yarovyi, Social Networks and Information Warfare: Ukrainian Dimension, Sociology: Theory, Methods, Marketing 2 (2020) 75–88.
- [10] Institute of Strategic Communications of NDUU, Collected Scientific Works “STRATCOM”, National Defence University of Ukraine, Kyiv, 2023.
- [11] A.M. Solovei, Methodology of Countering Manipulative Narratives in Hybrid Warfare, Military-Scientific Bulletin 2(56) (2024) 41–52.
- [12] Ministry of Culture and Information Policy of Ukraine, White Paper on Countering Disinformation and Fake News, Kyiv, 2023.



БЕЗПЕКА І АНАЛІЗ ВЕЛИКИХ ДАНИХ

Композиційний шифр як послідовна дія діагонального та секвенціального шифрів

Олег Гутік¹, Дмитро Дияк¹, Юрій Кузбит¹

¹ Ivan Franko National University of Lviv, Universytetska 1, 79000 Lviv, Ukraine

Анотація

У цій праці запропоновано секвенціальний шифр, який в комбінації з діагональним шифром використовуємо для шифрування. Наведено приклади таких шифрувань та їхній криптографічний аналіз.

Ключові слова

діагональний шифр, секвенціальний шифр, перестановка

1. Вступ

Однією з перших спроб узагальнення шифру зсуву (шифру Цезаря) був шифр Віженера. Хоча шифр Віженера запропонований вперше італійцем Джованні Баттіста Белласо (Giovanni Battista Bellaso) в 1553 році у монографії "La cifra del Sig.", носить ім'я французького дипломата Блеза Віженера (Blaise de Vigenère) і був описаний в його монографії "Traicté des Chiffres ou Secrètes Manières d'Ecrire" 1586 р. Цей шифр ще довгий час (до 1940-х років) використовувався багатьма країнами.

У цій праці ми пропонуємо секвенціальний шифр, який є узагальненням принципів шифру Віженера, має високу стійкість до зламу та ідейно ближчий до шифру Гілберта-Вернама. Композиція секвенціального шифру та діагонального шифру, запропонована в праці [2], є достатньо стійкою до класичних методів криптоаналізу.

Надалі для спрощення викладу будемо вважати, що наш алфавіт складається з 31 літери: з двадцяти шести літер латинського алфавіту $a, b, \dots, x, y, z$ та п'яти символів: «.» (крапка), «,» (кома), «:» (двокрапка), «-» (тире) і «+» (плюс).

Таблиця 2

Коди символів зображено в стандартній таблиці

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15

q	r	s	t	u	v	w	x	y	z	.	,	:	-	+
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

⁶⁸ Corresponding authors

✉ oleg.gutik@lnu.edu.ua (О. Гутік); dmytro.dyak@lnu.edu.ua (Д. Дияк); yurii.kuzbyt@lnu.edu.ua (Ю. Кузбит)

ORCID 0000-0001-8513-0282 (О. Гутік)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Алгоритм діагоналізації

Нехай дано відкритий текст $M = m_1 m_2 \dots m_k$, де $k \in \mathbb{N}$. Будемо шифрувати кожен символ $m_i, i = 1, \dots, k$, впорядкованою парою (c_i, d_i) , де числа c_i та d_i вибираємо так. Зафіксуємо довільне натуральне число $N \gg 31$ (тобто N набагато більше за 31)⁶⁹. Випадковим чином (рівномовірно) вибираємо число $c_i = 00, 01, \dots, N$. Знаходимо число $d_i = 00, 01, \dots, N$ таке, що

$$c_i - d_i = m_i \pmod{31}. \quad (1)$$

У рівності (1) ми вважаємо, що m_i зображено числом із Таблиці 1. Символу (числу) m_i буде відповідати впорядкована пара чисел (c_i, d_i) .

Зашифрований текст, який відповідає відкритому тексту $M = m_1 m_2 \dots m_k$, матиме вигляд

$$C = c_1 d_1, c_2 d_2, \dots, c_k d_k. \quad (2)$$

Очевидно, що закодований текст C має вдвічі більшу довжину, ніж відкритий текст M . Для того, щоб отримати із зашифрованого тексту (1) відкритий текст $M = m_1 m_2 \dots m_k$, скористаємося формулою

$$m_i = c_i - d_i \pmod{31}. \quad (3)$$

Один з методів шифрування алгоритмом діагоналізації викладено у [1]. Цей шифр полягає в незалежному шифруванні непарних символів $C^1 = c_1 c_2 \dots c_k$, слова C та парних символів $C^2 = d_1 d_2 \dots d_k$ слова C різними перестановками σ_1, σ_2 визначеними на символах слова C^1 і C^2 відповідно. Ключем такого шифру є впорядкована пара підстановок (σ_1, σ_2) . У [1] також викладено стійкіший метод шифрування. Він полягає у наступному.

Діагональне шифрування можемо розглядати, як блоковий шифр з довжиною блоку b , у випадку великої довжини відкритого тексту. А перестановки σ_1, σ_2 ми розглядаємо не як перестановки елементів алфавіту, а як перемтановки літер слова стосовно їхнього порядку в блоці. У цьому випадку при $N=1000$ ми розглядаємо перестановку чисел 000, 001, ..., 098, 099 стосовно порядку в якому вони входять у блоки довжини b слів C^1 та C^2 , відповідно. Якщо ж останній блок слова C^1 та C^2 неповний, то його доповнюємо шумом.

Розглянемо число n_{2k} , яке у двійковій системі числення має довжину (кількість символів 0 і 1) строго більшу за $2k$.

Робимо трансформацію $\iota_{n_{2k}}$ (перетворення) слова $C = c_1 d_1, c_2 d_2, \dots, c_k d_k$ наступним чином, якщо на i -тій позиції числа $n_i \in 0$, нічого не відбувається, а якщо 1 - то міняємо літери c_i і d_i місцями. На виході утворюється слово $C' = c'_1 d'_1, c'_2 d'_2, \dots, c'_k d'_k$, де впорядковані

⁶⁹ Число N можна вибрати як завгодно великим, що буде кардинально ускладнювати процес атаки зловмисником.

Надалі число C_i записувати з тією ж кількістю десяткових цифр, що і число N

пари (c'_i, d'_i) утворюють вище описане перетворення, діємо впорядкованою парою перестановок σ_1, σ_2 , відповідно, як це описано в попередньому алгоритмі. Ключем шифрування у цьому випадку буде впорядкована трійка $(n_{2k}, \sigma_1, \sigma_2)$, елементи якої описані вище.

Алгоритм дешифрування у першому випадку виглядатиме так. Якщо отримали зашифроване повідомлення $S = s_1, s_2, \dots, s_{2k}$, то виділяємо з нього слово $S_1 = s_1, \dots, s_{2k-1}$, літери якого послідовно розміщені на непарних позиціях слова S і слова $S_0 = s_2, \dots, s_{2k}$, літери якого послідовно розміщені на парних позиціях слова S . Виконуємо дії $c_i = (S_{2i-1})\sigma_1^{-1}, d_i = (S_{2i})\sigma_2^{-1}$, де σ_1^{-1} і σ_2^{-1} — обернені перестановки (відображення) до перестановок σ_1 і σ_2 , відповідно. Далі відновлюємо текст M за формулою (3).

У випадку другого шифру виконуємо першу дію, описану вище (тобто непарні перестановки σ_1^{-1} і σ_2^{-1}), а потім виконуємо трансформацію $t_{n_{2k}}$ утвореного слова.

Далі відновлюємо текст M за формулою (3).

3. Секвенціальний шифр

Нехай дано відкритий текст $M = m_1 m_2 \dots m_k$. Розглянемо довільно вибрану скінченну неспадну послідовність натуральних чисел $n_1, n_2, \dots, n_{e-1}, n_e$, таку, що $k < e$.

Зашифрування тексту M відбувається за формулою

$$c_i = m_i + n_i \pmod{n}, i = 1, \dots, k \quad (4)$$

де $N \gg k$.

Дешифрування очевидне:

$$m_i = c_i - n_i \pmod{n}, i = 1, \dots, k. \quad (5)$$

Далі ми пропонуємо метод побудови послідовності $n_1, n_2, \dots, n_{e-1}, n_e$.

Нехай X — довільне речення (текст) довжини N , де N — як завгодно велике натуральне число.

Текст X передається кореспонденту довільним чином (відкритим або закритим каналом) і X буде складати відкриту частину ключа секвенціального шифру. Далі через $s_1, s_2, \dots, s_p$ позначимо літери алфавіту, які будуть становити таємну частину ключа. Таємний ключ має вигляд:

$S s_1 s_2 \dots s_p$, де S — натуральне число, яке вимагає початок (число) відліку ключа.

Алгоритм шифрування виглядає так.

Нехай t_1 — найменше натуральне число, таке що символ s_1 розташований на позиції t_1 .

Літеру m_1 зашифруємо так

$$c_1 = (m_1 + t_1) \pmod{n}.$$

Аналогічно зашифруємо всі літери $m_i, i = 2, \dots, p$:

$$c_i = (m_i + t_i) \pmod{n}. \quad (6)$$

У випадку $i > p$ замість t_i підставимо t_j

$$c_i = (m_i + t_j) \pmod{n}. \quad (7)$$

де

$$j = \begin{cases} i \pmod{p}, & \text{якщо } i \text{ не є кратним } p. \\ p, & \text{якщо } i \text{ кратне } p. \end{cases} \quad (8)$$

Дешифрування відбувається за формулою:

$$m_i = (c_i - t_j) \pmod{n},$$

де t_j обчислюється за формулою (8).

Зрозуміло, що під час секвенціального шифрування за умови використання кожного разу іншого таємного ключа $S_1 S_2 \dots S_p$ складність взлому залежить від довжини p ключа та початку читання тексту S . Очевидно також, що у випадку коли число S близьке до довжини відкритого ключа, то ми можемо продовжувати відлік додаючи знову текст X . Навіть, якщо зловмисник (криптоаналітик) знає початок відліку ключа S та довжину ключа p , то кількість варіантів, які йому потрібно перебрати p^n варіантів, де n - кількість літер у алфавіті. Отже зловмисник має виконати $|S| \cdot p^n$ операцій для підбору ключа, що є достатньо затратним.

З визначення алгоритму шифрування випливає, що знання S та частини ключа $S_1 S_2 \dots S_p$ не дає можливості розшифрувати зашифрований текст, а тому запропонований секвенціальний шифр є достатньо стійким до атак «зашифрований текст/відкритий текст».

4. Діагонально-секвенціальний шифр

Алгоритм діагонально-секвенціального шифру має два варіанти:

1. Відкритий текст M зашифровується спочатку діагональним шифром з ключем $(\sigma_1, \sigma_2, l_{n_n})$ за формулою (4), а потім отриманий зашифрований текст шифрується секвенціальним шифром з відкритим ключем X та таємним ключем $S_1, S_2, \dots, S_p$ за формулами (6)–(7).
2. Відкритий текст M зашифровується секвенціальним шифром з відкритим ключем X і таємним ключем $S_1, S_2, \dots, S_p$ (за формулами (6)–(7)), а потім зашифровуємо отриманий текст діагональним шифром з ключем $(\sigma_1, \sigma_2, n_{2n})$, за формулою (4).

У випадку 1) розшифровуємо спочатку за ключем $S_1, S_2, \dots, S_p$ секвенціальним шифром за формулою (8), а потім за ключем $(\sigma_1, \sigma_2, l_{n_n})$ діагональним шифром за формулою (5); а у випадку 2) — за формулою (5) і далі за формулою (8).

Надалі діагональний та секвенціальний шифри будемо позначати через $\mathcal{D}$ і $\mathcal{S}$ відповідно, а їхні композиції (послідовні зашифрування) — через $\mathcal{DS}$ та $\mathcal{SD}$. Легко бачити, що діагональний шифр $\mathcal{D}$ має лінійну складність $O(k)$ і секвенціальний шифр $\mathcal{S}$ має також лінійну складність шифрування $O(k)$. Звідси випливає, що їхні композиції $\mathcal{DS}$ та $\mathcal{SD}$ мають також лінійну складність шифрування $O(k)$, де k — довжина відкритого тексту.

Аналогічно шифруванню, дешифрування має лінійну складність $O(k)$, де k — довжина зашифрованого тексту.

Оскільки у діагональному шифрі $\mathcal{D}$ є імовірнісний вибір числа c_i для літери m_i відкритого тексту, то методи лінійного та диференціального криптоаналізу не працюють у випадку шифру $\mathcal{D}$, а також шифрів $\mathcal{D}\mathcal{S}$ та $\mathcal{S}\mathcal{D}$. Аналогічно щодо криптоатак типу «відомий відкритий текст / вибраний відкритий текст / вибраний шифротекст» — вони не дають можливості зламати зашифрований текст.

5. Приклади

Треба зашифрувати шифром $\mathcal{D}$ текст “qutinismguilo.la-la-la” при $N=1000$. Розбиваємо на блоки довжиною $b=11$, і розглянемо підстановки $\sigma_1=(2,4,10,1,9,8,6,3,7,5,11)$ і $\sigma_2=(11,9,7,3,6,8,1,10,4,2,5)$. Візьмемо зручні c_i (4-цифрові, бо $N=1000$): $c_i=7000+i$ для $i=1, \dots, 22$. Тоді $d_i=c_i-m_i \pmod{31}$ (беремо у межах $0..300..300..30$, записуємо 4-цифрово). Розбиття на блоки ($b=11$) і застосування σ_1, σ_2 .

Блок 1 (позиції 1–11 до перестановок):

(7001,0010) (7002,0007) (7003,0009) (7004,0021) (7005,0017) (7006,0023)
(7007,0014) (7008,0021) (7009,0028) (7010,0015) (7011,0028)

Застосовуємо в блоці 1 до c -послідовності — $\sigma_1=(2,4,10,1,9,8,6,3,7,5,11)$, до d -послідовності — $\sigma_2=(11,9,7,3,6,8,1,10,4,2,5)$. отримуємо

(7002,0028) (7004,0028) (7010,0014) (7001,0009) (7009,0023) (7008,0021)
(7006,0010) (7003,0015) (7007,0021) (7005,0007) (7011,0017)

Блок 2 (позиції 12–22 до перестановок) має вигляд:

(7012,0026) (7013,0024) (7014,0013) (7015,0029) (7016,0010) (7017,0013)
(7018,0001) (7019,0013) (7020,0016) (7021,0004) (7022,0016).

Аналогічно застосовуємо перестановки σ_1, σ_2 у межах блоку 2 і отримаємо:

(7013,0016) (7015,0016) (7021,0001) (7012,0013) (7020,0013) (7019,0013)
(7017,0026) (7014,0004) (7018,0029) (7016,0024) (7022,0010)

Для розшифрування, до кожного блока окремо застосовуємо σ_1^{-1} до c -частини блока й σ_2^{-1} до d -частини (повертаємо оригінальний порядок у блоці), для кожної позиції i у блоці обчислюємо $m_i=(c_i-t_i) \pmod{31}$ і розкодуємо символи. Це повертає саме “qutinismguilo.la-la-la”.

Розглянемо тепер приклад секвенціального шифру. Повідомлення М “qutinismguilo.la-la-la” має коди m_i : 16 20 19 8 13 08 18 12 06 20 08 11 14 26 11 00 29 11 00 29 11 00. Відкритий ключ Х (довжини N): abcdefghijklmnopqrstuvwxyz., : - + (усі 31 символ у відомому порядку). Таємний ключ: S=1; таємні літери s=(k,e,y) відображає позиції в Х: t=(11,5,25). Для $i>3$: t_i циклічно повторюються [11,5,25,11,5,25,...] за правилом (8).

Застосовуємо $c_i=(m_i+t_i) \pmod{31}$ (формула (6)), отримуємо коди c_i :

27 25 13 19 18 2 29 17 0 0 13 5 25 0 5 11 3 5 11 3 5 11.

Перетворюємо назад у символи (та сама таблиця 31-символьного алфавіту) → шифротекст: ,zntsc-raanfzafldfldfl

Для дешифрування (перевірки), беремо ті самі t_i та рахуємо $m_i = (c_i - t_j) \pmod{31}$ (формула (5)) – отримаємо: “qutinismguilo.la-la-la??”.

Література

- [1] D. Kahn, The codebreakers: the story of secret communication from ancient times to the Internet, Scribner, New York, NY, 1996.
- [2] O.V. Gutik, O.B. Popadiuk, V.A. Vlasov. Symmetric algebraic cipher. Theoretical and Applied Cybersecurity. Матеріали III Всеукраїнської науково-практичної конференції (TACS-2025). Київ, Політехніка, 2025, 319с, с. 139-141.

Інтеграція масштабованих ML-моделей валідації вразливостей у SIEM-системи для автоматизації та прискорення реагування на інциденти

Владислав Денисюк¹

¹Національний Технічний Університет «Харківський політехнічний інститут», вул. Кирпичова 2 Харків, 61002, Україна

Анотація

У роботі досліджується інтеграція масштабованих і пояснюваних моделей машинного навчання для валідації вразливостей у системи управління подіями інформаційної безпеки. Матеріал окреслює підхід, за якого узгоджені внутрішні та зовнішні джерела даних використовуються для формування оцінки ризику експлуатації на рівні пари «вразливість–актив» із подальшим наданням зрозумілих пояснень і рекомендованих дій. Розглядаються організаційні аспекти включення результатів у процеси пріоритизації та реагування, спрямовані на зменшення інформаційного шуму, прискорення ухвалення рішень і підвищення керованості ризиками. Підхід орієнтований на дотримання вимог масштабованості, надійності та прозорості для аналітиків без прив'язки до конкретних технологічних реалізацій.

Ключові слова

валідація вразливостей; події безпеки; машинне навчання; пояснювані моделі; ризик експлуатації; граф вразливостей; інвентаризація активів; журнали подій; пріоритизація сповіщень; автоматизація реагування.⁷⁰

На сьогоднішній день існує проблема, яка полягає в тому, що кількість оприлюднених вразливостей і обсяг технічної телеметрії зростають швидше, ніж спроможність команд безпеки їх опрацьовувати [1]. Сигнатурні підходи генерують надлишкові сповіщення, не враховують контекст конкретних активів і бізнес-критичність, унаслідок чого затягуються пріоритизація та усунення ризиків.

Запропоноване рішення – інтеграція масштабованих і пояснюваних моделей машинного навчання у процеси систем управління подіями безпеки. Такі моделі працюють на рівні пари «вразливість–актив», узгоджують дані з відкритих каталогів, інвентаризації та журналів подій, вибудовують зв'язки між об'єктами інфраструктури й у близькому до реального часу формують оцінку ризику з коротким поясненням та рекомендацією дій. Додатково враховується контекст критичності активів і чинні політики реагування, що дозволяє узгоджувати пріоритизацію з бізнес-цілями та ресурсними обмеженнями. У підсумку зменшується інформаційний шум, прискорюється тріаж і спрямовуються ресурси на найбільш критичні загрози, водночас підвищується прозорість для аналітиків і керованість процесів реагування [2].

Теоретична постановка задачі ґрунтується на моделюванні ризику експлуатації для пари «вразливість–актив» у момент часу. Нехай

$$s = (cve, a, t)$$

⁷⁰ Corresponding author.

[†] These authors contributed equally.

✉ vladyslav.denysuk@cs.khpi.edu.ua (V. Denysuk);

id 0009-0001-6334-1473 (V. Denysuk);



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

де: cve – конкретна вразливість (ідентифікатор з каталогу вразливостей), α – конкретний актив (сервер, хост, сервіс), t – момент часу, коли відбувається оцінка. Разом це утворює ситуацію s , яка далі відображається у просторі ознак. Це числове подання всього, що знаємо про цю пару “вразливість-актив” (тип слабкості, версія ПЗ на активі, експозиція в інтернет, наявність патчу, сигнали з журналів тощо):

$$x = \phi(s)$$

Де: s – конкретний об’єкт/ситуація, для якої було проведено оцінювання, ϕ – перетворення ознак, функція, яка будує числові ознаки для моделі, x – вектор ознак, результат $\phi(s)$, який подається в ML-модель. Ціль – наблизити умовну ймовірність експлуатації у горизонті:

$$\rho_0(y=1|x).$$

Приклад: після обчислення ознак для пари (CVE-X, сервер-Y) модель дала $\rho_0 = 0.40$ – тобто ~40% шанс експлуатації вразливості.

Операційний ризик визначається як добуток каліброваної ймовірності, ваги критичності активу та показника експозиції середовища, а рішення про пріоритизацію формалізується пороговою політикою:

$$d(s) = 1\{risk(s) \geq \tau(a)\}$$

Де: $\tau(a)$ – поріг для цього класу активів (може бути нижчим для критичним систем),

$1\{*\}$ – індикатор, який повертає 1 (так), або 0 (ні), на основі цього індикатора, вноситься рішення про відкриття інцидента у системі управління ризиками, або підвищення пріоритету для виправлення цього інцидента.

Предметну область доцільно описувати гетерогенним графом, у якому вузли відповідають вразливостям, слабкостям, продуктам/версіям, активам і мережевим сегментам, а ребра – відношенням «вразливість-слабкість», «вразливість-продукт», «актив-продукт», «актив-сегмент». Таке представлення надає топологічний контекст і дозволяє будувати ознаки центральності, відстані до критичних вузлів та кластерні характеристики, підсилюючи здатність моделі враховувати поширення ризику інфраструктурою [3].

Інтеграція з системами управління подіями безпеки починається з конвеєра надходження, нормалізації та зв’язування даних. Зовнішні каталоги вразливостей, інвентаризація активів і журнали подій уніфікуються до єдиної схеми, виконується зіставлення імен продуктів і версій із фактичними конфігураціями (entity resolution), виконується вирівнювання часових міток видалення дублів. Теоретично важливо забезпечити узгодженість між навчанням і прогнозуванням: ті самі перетворення ознак і довідники мають застосовуватися в обох контурах, що гарантується використанням сховища ознак і забезпечення версій трансформацій [4].

Простір ознак поєднує статичні характеристики вразливості (тип слабкості, вектор атаки, необхідні привілеї, потреба у взаємодії користувача, а також текстові представлення описів), контекст активу (роль, критичність, конфігурація, версія, стан оновлень, належність до бізнес-процесів), експозицію середовища (сегментація, зовнішня доступність, відкриті сервіси, історія з’єднань) і спостережні сигнали з журналів (спрацювання контрольних засобів, кореляційні правила, інтенсивність подій у ковзних вікнах). Додаються графові та часові ознаки: свіжість уразливості, поява експлойтів і патчів, інерція змін у середовищі.

Навчання формується як бінарна класифікація з часовими обмеженнями: дані для побудови ознак мають бути доступними на момент прогнозу, а розділення «навчання–перевірка» виконується за часом, що уникає витоків інформації. Поряд із класифікацією можливі альтернативні постановки – ранжування пар «вразливість–актив» за пріоритетом усунення або аналіз «час-до-події», коли важливим є не тільки факт експлуатації, а й очікувана швидкість її настання.

Ймовірнісний вихід моделі потребує калібрування, щоб числові значення відповідали інтерпретації частоти подій у популяції. Для керованості прийняттям рішень доцільно враховувати невизначеність прогнозу: агрегування моделей або конформні методи формують довірчі межі, які впливають на вибір порогів і маршрутів реагування. Таким чином поєднуються статистична валідність і операційна придатність результатів [5].

Тлумачення інтегрується як локальне адитивне розкладення внесків ознак у прогноз для кожної конкретної пари «вразливість–актив» і як агреговане уявлення про важливість ознак на глобальному рівні. Вбудовані пояснення в об'єкті сповіщення знижують бар'єр довіри та забезпечують аудит, дозволяючи аналітику відразу бачити причини підвищеного ризику і обґрунтованість запропонованої дії.

Злиття детерміністичних правил кореляції з машинним навчанням формує гібридну політику. Правила гарантують покриття відомих патернів загроз, а модель додає контекстно-залежний ризикове оцінювання; у результаті сповіщення сегментуються на пріоритетні діапазони, які безпосередньо накладаються на сценарії реагування. Такий підхід зменшує шум і водночас зберігає критичні сигнали, що особливо важливо для навантажених операційних середовищ.

Масштабованість забезпечується розділенням на потоковий контур інференсу та пакетний перерахунок ризиків. Повідомлення про зміни стану активів і появу нових вразливостей надходять через шину подій; сервіс моделі виконує оцінювання, а нічні або інкрементальні батчі переглядають результати оцінювання після оновлень довідників і політик. Сховище ознак гарантує консистентність між навчанням і реальним використанням, модельний реєстр – контроль версій і поетапне розгортання, а механізми зворотного тиску та ідемпотентності – надійність під навантаженням.

Керованість і безпека інтеграції визначаються політиками доступу, мінімізацією переданих полів і логуванням рішень моделі. Необхідні спостереження за дрейфом розподілів, тригери перекалібрування й повторного навчання, а також відтворюваність експериментів завдяки маркуванню версій наборів даних, ознак і параметрів. У сукупності ці принципи утворюють теоретико-технічний фундамент для архітектури, яка переносить машинне навчання з дослідницького середовища у керований цикл обробки подій і реагування в сучасних системах безпеки.

Запропонований підхід формує цілісну рамку інтеграції масштабованих і пояснюваних моделей машинного навчання у процеси систем управління подіями безпеки. Теоретична основа спирається на моделювання ризику експлуатації для пари «вразливість–актив» у визначений часовий горизонт, що доповнюється графовим представленням предметної області, узгодженим простором ознак і процедурами калібрування та оцінювання невизначеності. Така комбінація забезпечує узгоджені, інтерпретовані та операційно придатні прогнози.

Технічна архітектура з'єднує конвеєр надходження й нормалізації даних, сховище ознак, сервіс прогнозування та модуль пояснення із гібридною політикою «правила + модель». У результаті система зменшує інформаційний шум, прискорює тріаж, фокусує ресурси на критичних ризиках і підвищує керованість процесами реагування без втрати прозорості для аналітиків. Версіонування даних, ознак і моделей, а також контроль розгортання дозволяють відтворювати результати й безпечно еволюціонувати рішення.

Надалі доцільно розвивати підхід у напрямках розширення графових ознак і моделей, використання ранжувальних та «час-до-події» постановок для пріоритизації, поглиблення методів невизначеності й активного збирання міток, а також удосконалення механізмів адаптації до дрейфу. Перспективним є подальше зближення із операційними циклами реагування через тіснішу інтеграцію сценаріїв автоматизації та зворотного зв'язку від аналітиків.

Література

- [1] G. González-Granadillo, S. González-Zarzosa, R. Diaz, Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures, *Sensors* 21.14 (2021) 4759. doi:10.3390/s21144759.
- [2] J. Jacobs, S. Romanosky, B. Edwards, I. Adjerd, M. Roytman, Exploit Prediction Scoring System (EPSS), *Digit. Threat.* 2.3 (2021) 1–17. doi:10.1145/3436242.
- [3] Z. Shi, N. Matyunin, K. Graffi, D. Starobinsk, Uncovering Product Vulnerabilities with Threat Knowledge Graphs, *IEEE Secur. Dev. Conf. (SecDev)* (2022) 84–90.
- [4] V. Christophides, V. Efthymiou, T. Palpanas, G. Papadakis, K. Stefanidis, An Overview of End-to-End Entity Resolution for Big Data, *ACM Comput. Surv.* (2020) 53(6):127. <https://doi.org/10.1145/3418896>.
- [5] Niculescu-Mizil, A., Caruana, R. "Predicting Good Probabilities with Supervised Learning." In: *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, 2005. <https://doi.org/10.1145/1102351.1102430>

Нейромережева модель для розпізнавання емоцій текстових фрагментів

Олександр Коровій¹ та Ігор Терейковський¹

¹Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", 03056, Київ

Анотація

Розпізнавання емоцій у тексті є ключовим завданням у сфері обробки природної мови, однак воно створює значні труднощі для мов із обмеженими ресурсами, зокрема української, через брак спеціалізованих наборів даних і моделей. У статті представлено модульний підхід до розробки нейромережових засобів розпізнавання емоцій в україномовних текстових фрагментах. Проведено експериментальне дослідження, у якому порівнювалися два прототипи: легковаговий класифікатор на основі FastText та потужніша модель Transformer-класу RoBERTa-base. Обидві моделі навчалися розпізнавати сім базових емоцій на спеціально створеному корпусі українських дописів у соціальних мережах. Результати показали, що RoBERTa-base досягла високого значення F1-міри, суттєво перевищивши показники FastText та базового перекладного підходу. Водночас ця точність потребує значно більших обчислювальних ресурсів та знижує швидкодію. Дослідження підкреслює критичний баланс між точністю та ефективністю моделі та пропонує практичний підхід до вибору архітектури залежно від обмежень ресурсів і вимог до продуктивності для аналізу емоцій української мови.

Ключові слова

розпізнавання емоцій, обробка природної мови, українська мова, нейронні мережі, RoBERTa, FastText.

1. Вступ

Активне поширення користувацького контенту в соціальних мережах зробило автоматичне розпізнавання емоцій важливим завданням у галузі обробки природної мови (NLP), яке знаходить застосування в аналізі громадських настроїв, маркетингу та психологічних дослідженнях. Для мов із великими ресурсами, як-от англійська, досягнуто значного прогресу, проте розвиток інструментів для мов із обмеженими ресурсами, включно з українською, залишається утрудненим [1, 2].

Основною проблемою для досліджень україномовних текстів є відсутність великих якісних відкритих корпусів для розпізнавання емоцій. Це ускладнює навчання надійних моделей із нуля. Більшість існуючих систем спираються на компромісний підхід із перекладом, коли український текст спершу перекладається англійською, а далі обробляється класифікатором емоцій англійської мови. Проте такий метод втрачає культурно специфічні відтінки емоційного вираження, що призводить до суттєвого падіння точності. Окрім цього, українська мова має низку особливостей — діалектні варіації, неформальність соціальних мереж, складні комбінації емоцій, які переклад часто спотворює [1, 2].

⁷¹ Corresponding authors

✉ korovii.oleksandr@lpi.kpi.ua (O. Korovii); terejkowski@ukr.net (I. Tereikovskiy)

ORCID 0009-0002-2835-9173 (O. Korovii); 0000-0003-4621-9668 (I. Tereikovskiy)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Це дослідження спрямоване на подолання зазначених обмежень шляхом експериментальної валідації модульного підходу до побудови україномовних систем розпізнавання емоцій, як описано у роботі [3]. Створивши новий збалансований набір даних і порівнявши архітектури моделей, ми прагнемо закласти основу для подальших досліджень та довести перевагу моделей рідною мовою над перекладними підходами.

2. Експеримент

Для перевірки запропонованого фреймворку було розроблено експеримент для порівняння продуктивності та ефективності різних архітектурних підходів до розпізнавання емоцій в українських текстах.

2.1. Набір даних та попередня обробка

Для дослідження було створено новий набір даних. Він складається з приблизно 18 000 україномовних дописів, випадково відібраних із соціальної мережі X (раніше Twitter). Дані збиралися без тематичних фільтрів для забезпечення широкого, загального представлення дискурсу користувачів. Дописи були вручну анотовані за сімома основними емоціями: радість, сум, гнів, страх, любов, здивування та нейтральність. Для забезпечення якості та надійності даних було реалізовано комплексний конвеєр попередньої обробки, що включав:

1. Фільтрацію: Видалення неукраїнського контенту, коротких дописів (менше п'яти слів) та дублікатів.
2. Очищення: Нормалізація пунктуації, стандартизація кодування та видалення спеціальних символів.
3. Допомогу в анотації: Використання великої мовної моделі (GPT-4o) [4] для попередньої анотації з подальшою перевіркою та виправленням трьома незалежними волонтерами для забезпечення узгодженості.
4. Балансування: Забезпечення рівномірного представлення всіх класів емоцій для уникнення упередженості моделі.

2.2. Моделі та архітектури

Було реалізовано та навчено дві основні моделі, що представляють різні компроміси між продуктивністю та обчислювальними витратами:

- Класифікатор FastText: Легковагова архітектура на основі векторних представлень слів (Word2Vec, Skip-Gram) [5]. Вона усереднює вектори слів для створення представлення документа, що робить її обчислювально ефективною та швидкою. Модель мала приблизно 3 мільйони параметрів, що робить її ідеальним кандидатом для систем з обмеженими ресурсами та застосунків реального часу.
- Класифікатор RoBERTa-base: Потужна, сучасна модель на основі архітектури Transformer [6] з механізмами само-уваги. Базова модель RoBERTa [7] була додатково попередньо навчена на україномовному корпусі для адаптації до специфічних лінгвістичних особливостей цільової мови. Ця модель, що має приблизно 125 мільйонів параметрів, призначена для вловлювання складних контекстуальних залежностей та тонких семантичних відтінків, віддаючи перевагу точності над швидкістю.

2.3. Порівняльні підходи

Для всебічної оцінки продуктивність реалізованих нами моделей порівнювалася з двома альтернативними методами:

- Конвеєр на основі перекладу: Двоетапний процес, у якому український текст спочатку перекладався на англійську за допомогою стандартної моделі перекладу (Ukr2Eng), а потім класифікувався спеціалізованим англомовним класифікатором емоцій. Це поширений обхідний шлях для низькоресурсних мов.
- GPT-4o: Пропрієтарна, високопродуктивна велика мовна модель, відома своїми потужними багатомовними можливостями, яка використовувалася тут як еталон максимальної продуктивності.

3. Результати

Моделі оцінювалися за якістю класифікації (точність, повнота, влучність, F1-score) та обчислювальною продуктивністю (кількість параметрів, швидкість висновування). Результати, узагальнені в Таблиці 1, демонструють чіткий компроміс між точністю та ефективністю.

Модель RoBERTa-base досягла найвищої продуктивності серед моделей, які можна розгорнути локально, з показником F1-score 0,91. Це підтверджує, що архітектури на основі Transformer, за умови належної адаптації до цільової мови, є високоефективними для вловлювання нюансів емоційного вираження. Класифікатор FastText, хоча й значно менш точний з F1-score 0,72, мав надзвичайну перевагу у швидкості: час висновування становив лише 5 мкс порівняно з 9 мс у RoBERTa — різниця більш ніж у 1800 разів.

Підхід на основі перекладу (Ukr2Eng Translation + Eng-Emotion) показав найнижчу продуктивність з F1-score 0,59. Цей результат переконливо підтверджує гіпотезу, що переклад тексту перед аналізом призводить до значної втрати емоційної інформації. Модель GPT-4o досягла найвищої загальної точності (F1-score 0,95), але не підходить для локального розгортання та має дуже велику затримку висновування (2343 мс), що робить її непрактичною для багатьох застосунків реального часу.

Таблиця 1

Порівняльний аналіз моделей розпізнавання емоцій

Назва моделі	Кількість параметрів	Швидкодія	Точність	Повнота	Влучність	F1
FastText Classifier	3M	5 мкс	0.72	0.72	0.72	0.72
RoBERTa-base Classifier	125M	9 мс	0.91	0.91	0.91	0.91
Ukr2Eng Translation + Eng-Emotion	74M+83M	593 мс	0.59	0.59	0.59	0.59

GPT-4o	-	2343 мс	0.95	0.96	0.96	0.96

Висновки

Ця робота демонструє успішне застосування модульного фреймворку для розробки та оцінки систем розпізнавання емоцій для української мови. Наші експериментальні результати дозволяють зробити кілька ключових висновків:

4. Нативні мовні моделі є кращими: Низька продуктивність конвеєра на основі перекладу підкреслює критичну необхідність розробки моделей та наборів даних спеціально для української мови для врахування її унікального лінгвістичного та культурного контексту.
5. Існує чіткий компроміс між точністю та ефективністю: RoBERTa-base забезпечує високу точність, придатну для поглибленого офлайн-аналізу, тоді як FastText надає надзвичайно низьку затримку висновування, що робить його оптимальним вибором для застосунків реального часу на пристроях з обмеженими обчислювальними ресурсами.
6. Запропонований фреймворк є ефективним: Модульний підхід дозволяє гнучко адаптуватися до різних вимог завдань, даючи змогу розробникам обирати оптимальний баланс між продуктивністю та використанням ресурсів.

Майбутня робота буде зосереджена на розширенні набору розпізнаваних емоцій для більш гранульованого аналізу, дослідженні гібридних та ансамблевих архітектур для поєднання сильних сторін різних моделей, а також на покращенні стійкості моделей до сарказму та регіональних діалектів, специфічних для української мови.

References

- [1] Nandwani, P. & Verma, R. “A review on sentiment analysis and emotion detection from text”. *Social Network Analysis and Mining*, 2021; 11(81). DOI: <https://doi.org/10.1007/s13278-021-00776-6>.
- [2] Acheampong, F. A., Wenyu, C. & Nunoo-Mensah, H. “Text-based emotion detection: Advances, challenges, and opportunities”. *Engineering Reports*. 2020. DOI: <https://doi.org/10.1002/eng2.12189>
- [3] Tereikovskiy, I. A., & Korovii, O. S. (2025). A Model for Constructing Neural Network Systems for Recognizing Emotions of Text Fragments. *Herald of Advanced Information Technology*, 8(2), 197-208. DOI: <https://doi.org/10.15276/hait.08.2025.12>
- [4] OpenAI. “GPT-4o system card”. URL: <https://openai.com/index/gpt-4o-system-card/>
- [5] Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. “Bag of tricks for efficient text classification”. *15th Conference of the European Chapter of the Association for Computational Linguistics* *Eacl. 2017 Proceedings of Conference*, 2, pp. 427–431 DOI: <https://doi.org/10.18653/v1/e17-2068>
- [6] Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). “Attention is all you need”. *Advances in Neural Information Processing Systems*. 2017., 31st Conference on Neural Information Processing Systems (NIPS 2017), pp. 5999 – 6009

- [7] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. “Roberta: A robustly optimized bert pretraining approach”. arXiv, 2019. DOI: <https://doi.org/10.48550/arXiv.1907.11692>

Протидія атаці Вінера на RSA шифр

Юлія Мисло¹, Михайло Пагіря¹

¹ Ужгородський національний університет, пл. Народна 3, 88000 Ужгород, Україна

Резюме

Розглянуто криптоаналіз алгоритму RSA та його узагальнень. Визначено напрямки можливих атак на алгоритм. Запропоновано спосіб у протидії атаці Вінера.

Ключові слова

RSA алгоритм, криптоаналіз RSA, вибір ключів, ланцюгові дроби

1. Вступ

Актуальність. Людство завжди намагалося зберегти свої особисті, ділові, військові або політичні секрети. З цією метою вразлива інформація передавалася у схованому від сторонніх очей вигляді. За багато віків напрацьовані різні способи збереження секретів. Необхідність втаємничення інформації швидко зростає у сучасному швидко-мінливому, інформаційно-насиченому світі. Криптографічні методи шифрування/дешифрування повідомлень, що базуються на математичних функціях, перетвореннях, служать одним із способів збереження конфіденційності інформації. На практиці із успіхом використовуються підходи до шифрування/дешифрування як із симетричними, так із асиметричними ключами. Збереження таємності ключів — важливий показник надійності симетричних та асиметричних шифрів.

Постановка задачі. Нехай M — відкритий текст, а C — криптограма. Рівняння шифрування та дешифрування запишуться [1]:

$$C = E_{k_1}(M), \quad M = D_{k_2}(C),$$

де k_1 та k_2 , відповідно, ключі шифрування та дешифрування. Встановлення вразливості окремо взятих шифрів та розробка методологій по вибору безпечних ключів — одна із задач сучасного криптоаналізу. Мають місце два таких твердження: а) противник може застосувати "грубу силу" або як кажуть здійснити "брутальну атаку" реалізувавши повний перебір значень ключа за час t_g ; б) всяка інформація, в незалежності від її важливості, має свій термін актуальності — час t_a . Якщо ймовірний час атаки t_g значно перевищує час актуальності секретної інформації t_a , тобто $t_g \gg t_a$, то проводити лобову атаку нема сенсу, а тоді вибрані ключі шифру можна вважати практично стійкими до "брутальної атаки".

Шифри з асиметричними ключами ґрунтуються на принципах: ключ та алгоритм шифрування передаються по відкритим каналам зв'язку; ключ дешифрування відомий лише отримувачу повідомлень; зламування ключа дешифрування належить до важко розв'язуваних математичних задач. Останнє твердження "піддається сумніву" аналітиками, які десятки років здійснюється пошук методів розв'язання таких задач.

⁷² Corresponding authors

✉ julia.mislo@uzhnu.edu.ua (Ю. Мисло); mykhaylo.pahirya@uzhnu.edu.ua (М. Пагіря)

ORCID 0000-0001-6771-2844 (Ю. Мисло); 0000-0003-1488-3302 (М. Пагіря)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

В доповіді розглянуто різні модифікації атаки Вінера [2], які використовують правильні ланцюгові дробу [3,4] на RSA алгоритм [5] та його узагальнення.

2. RSA алгоритм та його узагальнення

При шифруванні/дешифруванні алгоритмом RSA:

- а) вибирають два великих одного порядку простих числа p та q і визначають $n = p \cdot q$;
- б) вибирають два цілих числа e та d , що задовольняють умову $e \cdot d \equiv 1 \pmod{\varphi(n)}$, де $\varphi(n) = (p-1)(q-1)$ – функція Ойлера

Пара (n, e) визначає відкритий ключ шифрування, тріада (p, q, d) – секретний (приватний) ключ алгоритму.

Шифрування/дешифрування. Оцифроване відкрите повідомлення M , ділять на блоки довжини меншої за n , тобто

$$M_1, M_2, \dots, M_k, \quad M_i < n, \quad i = 1, 2, \dots, k.$$

Рівняння шифрування має вигляд:

$$C_i = E(M_i) = (M_i)^e \pmod{n}, \quad i = 1, 2, \dots, k.$$

Рівняння дешифрування записується наступним чином:

$$M_i = D(C_i) = (C_i)^d \pmod{n}, \quad i = 1, 2, \dots, k.$$

Атака на RSA. Здійснити атаку на RSA алгоритм можна двома шляхами:

- а) визначити для числа n його співмножники p та q , тобто розв'язати задачу факторизації числа n ;
- б) знайти значення функції Ойлера $\varphi(n)$ і за відомим значенням e обчислити обернений елемент d .

Узагальнення RSA алгоритму. Таємний (приватний) ключ d алгоритму RSA в класичному випадку визначають із рівняння:

$$e \cdot d - k(p-1)(q-1) = 1, \quad k \in \mathbb{N}.$$

В [6,7] запропоновані клас узагальнених RSA-подібних алгоритмів, які використовують рівняння:

$$\begin{aligned} e \cdot d - k(p^n - 1)(q^n - 1) &= 1, \quad n > 0, \\ e \cdot d - k \frac{(p^n - 1)(q^n - 1)}{(p-1)(q-1)} &= 1, \quad n > 1. \end{aligned}$$

Відомо, що в незалежності від вибору n тривалі атаки, що побудовані на ланцюгових дробах, дають можливість відновити приватний (секретний) ключ d .

3. Основні типи атак на RSA - подібні шифри

При відсутності стратегії побудови ефективної та універсальної атаки на алгоритм RSA, існують окремі випадки, коли атака приводить до успіху.

Виокремлюють основні 4 типи атак: а) на відкритий ключ за рахунок факторизації числа n ; б) на малий (короткий) відкритий (публічний) ключ e ; в) на малий (короткий) секретний (приватний) ключ d ; г) на зашифрований текст M .

Проблема малих ключів, відкритого чи приватного, породжується тим, що реалізація алгоритму RSA вимагає тривалого часу для шифрування/дешифрування інформації. Щоб підвищити ефективність та швидкість обробки інформації можна зменшити кількість бітів

при виборі ключів e або d . Необхідність вибирати один із ключів шифрування/дешифрування малим виникають у випадку, коли шифрування та дешифрування здійснюється на комп'ютерах різної потужності, або коли існує потреба пришвидшення виконання дешифрування повідомлень із-за їх великої кількості чи важливості інформації.

Атака Вінера на RSA шифр. Коли приватний ключ d вибрано малим, то можлива атака Вінера на цей ключ. Із відкритих джерел відоме значення відкритого ключа (e, n) . Відносно елементів секретного ключа (p, q, d) робляться припущення:

$$q < p < 2q, \quad (1)$$

$$d < \frac{1}{3} \sqrt[4]{n}. \quad (2)$$

Пошук секретного ключа здійснюється наступним чином:

1. відношення e/n розвивається в ланцюговий дріб $e/n = [a_0; a_1, a_2, \dots, a_k, \dots]$;
2. вибирається деяке число M ;
3. послідовно обчислюють знаменники Q_k підхідних дробів $P_k/Q_k = [a_0; a_1, \dots, a_k]$;
4. секретним ключем d буде те значення знаменника Q_k , для якого виконується еквівалентність $(M^e)^{Q_k} \equiv M \pmod n$.

Однак оцінку Вінера (2) можна покращувати. В роботі [8] при виконанні додаткових умов була отримана оцінка $d < n^{0.298}$, а в роботі [9] $d < 2\sqrt[3]{2} n^{\frac{3}{4} - \frac{t}{2}}$, $t \in (0; 0.5)$.

Реалізація атаки Вінера. На практиці використання методу Вінера "зламування" таємного ключа RSA алгоритму може привести до позитивних результатів і у випадках, коли умови (1)-(2) не виконуються. Так при виборі $p=2039$, $q=401$ умова (1) не має місця. Позаяк $n=817639$, то умова (2) має вигляд: $d < 10,0235$. Але метод Вінера успішно знаходить приватний ключ $d=17$, який відповідний відкритому ключу $e=47953$. У наведеній нижче таблиці вміщено значення відкритого ключа (e, n) , де $n < 1.000.000$, правої частини оцінки (2) $d < L = \frac{1}{3} \sqrt[4]{n}$ і значення секретного ключа d , який знайдено за результатом атаки Вінера на RSA алгоритм.

Із наведених у таблиці значень випливає, що діапазон значень секретного ключа d , які можна "знайти" використавши вінерівську атаку на RSA алгоритм, значно ширший, ніж це передбачено умовою (2). Теоретичне обґрунтування цього твердження та встановлення нових умов для використання атаки Вінера, поки що не отримані.

e	n	L	d
52099	993563	10,52	19
216587	998341	10,54	23
316201	990133	10,51	25
73523	994627	10,53	27
67061	907633	10,29	27

4. Вибір безпечних параметрів RSA

Якщо параметри алгоритм RSA вибрані вдало, то маємо ефективний, та надійний асиметричний метод шифрування/дешифрування. При виборі оптимальних значень відкритого та секретного ключів потрібно: 1. забезпечити швидкість та ефективність шифрування/дешифрування; 2. подбати про легкість реалізації шифрування або дешифрування на наявних комп'ютерах; 3. якщо існує ймовірність атаки на RSA алгоритм, то передбачити, щоб час здійснення атаки t_g значно перевищувати час актуальності повідомлення t_a .

Зважаючи на дослідження останніх років в атаці Вінера на параметр секретного ключа d та наведені в таблиці часткові випадки, доречно реалізувати алгоритм Вінера і перевіряти вибрані параметри.

Література

- [1] О.Г. Корченко, В.П. Сіденко, Ю.О. Дрейс, Прикладна криптологія: системи шифрування. Державний університет телекомунікацій, 2014.
- [2] M.J. Wiener, Cryptanalysis of short RSA secret exponents, IEEE Transactions on Information theory, 36(3) (1990): 553–558.
- [3] М. Парія, Наближення функцій ланцюговими дробами. Ужгород: Гражда, 2016.
- [4] O. Perron, Die Lehre von den Kettenbrüchen, vol. 1. Stuttgart: Teubner, 1954.
- [5] R. Rivest, Shamir A., Adleman L. A, A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM, 21(2):120–126, 1978.
- [6] P. Cotan, G. Teşeleanu, A security analysis of two classes of RSA-like cryptosystems. Journal of Mathematical Cryptology, 18(1):34, 2024.
- [7] G. Teşeleanu, A lattice attack against a family of RSA-like cryptosystems. In International Symposium on Cyber Security, Cryptology, and Machine Learning, 343–355. Springer, 2024.
- [8] D. Boneh, G. Durfee, Cryptanalysis of RSA with private key d less than $n^{0.292}$. IEEE transactions on Information Theory, 46(4):1339–1349, 2000.
- [9] M. Bunder, J. Tonien. A new attack on the RSA cryptosystem based on continued fractions. Malaysian Journal of Mathematical Sciences, 11:45–57, 2017.

Програмна система для реалізації методології криптографічного захисту інформації на основі системи залишкових класів

Ігор Якименко¹, Михайло Касянчук¹, Василь Яцків¹, Людмила Бабала¹, Степан Івасьєв¹

¹ Західноукраїнський національний університет, вул. Львівська 11, 29016 Тернопіль, Україна

Анотація

У роботі наведено опис програмної системи для реалізації методології криптографічного захисту інформації на основі системи залишкових класів. Розроблена програмна система реалізує повний цикл криптографічної обробки інформації. Завдяки зручному інтерфейсу можна швидко вводити дані й аналізувати результати, що підвищує ефективність використання алгоритму для шифрування й розшифрування, забезпечує прозорість процесів, дає змогу оцінити результативність алгоритму і впевнитись у правильності обчислень.

Ключові слова

програмна система, методологія, криптографічний захист, система залишкових класів, захист інформації, діаграма класів та компонентів

1. Вступ

Для забезпечення необхідного рівня захисту використовуються симетричні та/або асиметричні криптографічні методи [1]. На сьогоднішній день до сучасних систем захисту інформації ставляться чіткі вимоги, які, як правило, досить громіздкі в реалізації та затратні в експлуатації. Крім того, для підвищення стійкості криптоалгоритмів потрібно збільшувати довжину ключів і, відповідно, значення операндів математичних перетворень, що приводить до зменшення швидкодії криптоалгоритмів. Одним з перспективних шляхів її підвищення є розпаралелення процесу обчислень, яку можна досягнути завдяки використанню різних форм непозиційної системи залишкових класів (СЗК) [2].

2. Аналіз досліджень та публікацій

В [3] представлена в методологія криптографічного захисту інформаційних потоків на основі використання симетричних та асиметричних криптосистем в цілочисельній, модифікованій досконалій та поліноміальній СЗК. Вона складається з восьми етапів: 1) процес формування множини блоків відкритого тексту (N , $N(x)$) для цілочисельних і поліноміальних криптографічних систем; 2) вибір модулів та ключів для цілочисельного та поліноміального шифрування в СЗК; 3) встановлення вимог щодо основних параметрів цілочисельних та поліноміальних симетричних та асиметричних криптографічних систем та захищеності інформації; 4) вибір запропонованих цілочисельних та поліноміальних

73 Corresponding authors

✉ iyakymenko@ukr.net (І. Якименко); kasyanchuk@ukr.net (М. Касянчук); vy@wunu.edu.ua (В.Яцків); ludmila@gmail.com (Л.Бабала); isv@wunu.edu.ua (С. Івасьєв)

ORCID 0000-0003-3446-1596 (І. Якименко); 0000-0002-4469-8055 (М. Касянчук); 0000-0001-9778-6625 (В.Яцків); 0000-0002-2388-270X (Л.Бабала); 0000-0003-2243-5956 (С. Івасьєв)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

криптосистем; 5) створення множини базових операцій; 6) формування набору методів виконання операцій; 7) вибір цілочисельної та поліноміальної форм СЗК; 8) програмна реалізація цілочисельних та поліноміальних симетричних та асиметричних криптосистем.

Ця методологія забезпечує комплексний підхід до розробки, реалізації та оптимізації симетричних та асиметричних цілочисельних і поліноміальних криптосистем на основі СЗК, що дає змогу досягти високого рівня захисту інформації при мінімальних витратах на обчислення.

3. Програмна система для реалізації симетричних та асиметричних криптоалгоритмів в СЗК

На рисунку 1. представлено діаграму компонентів програмної системи реалізації криптоалгоритмів в СЗК із врахування її практичної імплементації в прикладні інформаційно-комунікаційні підсистеми.

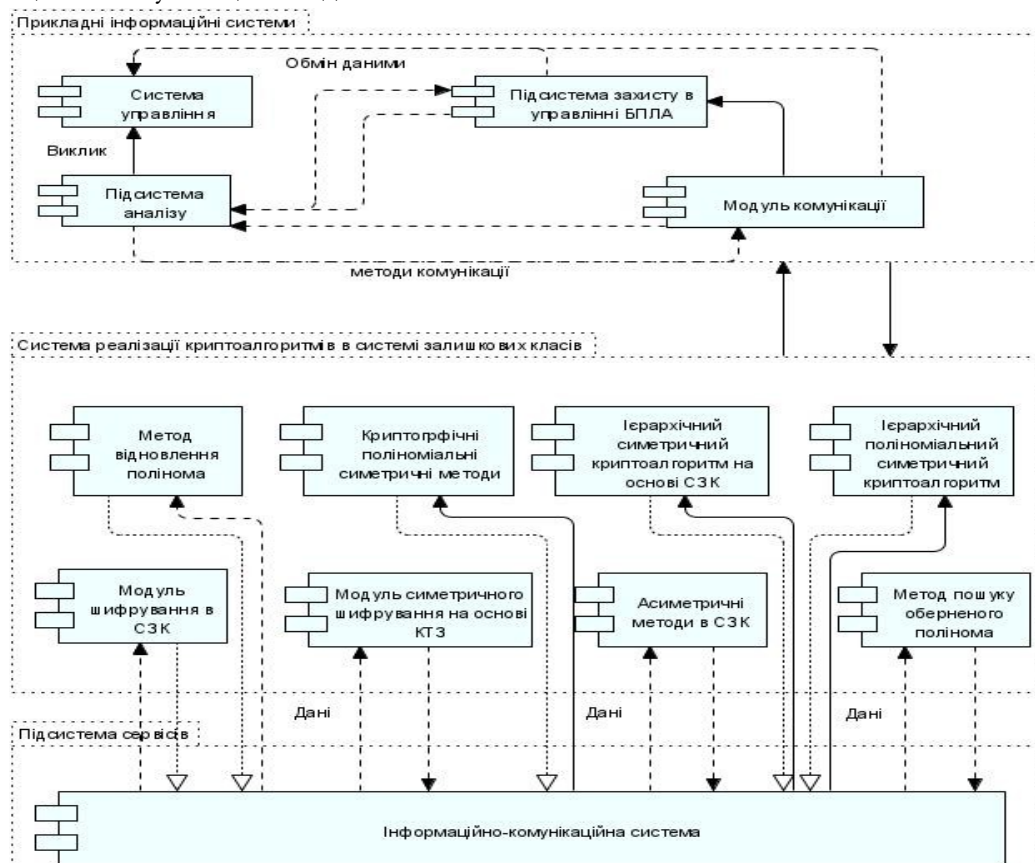


Рисунок 1 – Діаграма компонентів програмної системи реалізації криптоалгоритмів в системі залишкових класів

Програма реалізована за допомогою мови програмування Python, з його переваг слід відзначити наступні: зрозумілий синтаксис, підтримка більшості сучасних парадигм програмування, кросплатформеність, підвищена продуктивність, інтерпретована мова, динамічна типізація, розрядність чисел обмежуються тільки розміром оперативної пам'яті.

У процесі реалізації програмної системи були створені два основних інтерфейси для взаємодії з користувачем: графічний (GUI) та командний (CLI). Такий підхід забезпечує

гнучкість у використанні програмного забезпечення, дозволяючи користувачам обирати найбільш зручний спосіб взаємодії залежно від конкретних завдань та умов експлуатації.

З точки зору криптографічних алгоритмів, програмна система підтримує як симетричні, так і асиметричні методи шифрування. Оскільки їх реалізація має багато спільних аспектів, для пояснення принципів роботи системи використовується узагальнений опис, який охоплює обидва підходи. Це дозволяє більш ефективно продемонструвати загальну архітектуру системи та ключові етапи обробки даних.

Структурна організація програмного коду представлена у вигляді об'єктно-орієнтованої моделі, де основні компоненти реалізовані у формі класів. Взаємозв'язки між цими класами, їхні функціональні можливості та принципи взаємодії між собою зображені на діаграмі основних реалізованих класів (рис. 2). Така структурна схема наочно демонструє логіку побудови програмної системи та сприяє подальшому вдосконаленню її функціональності.

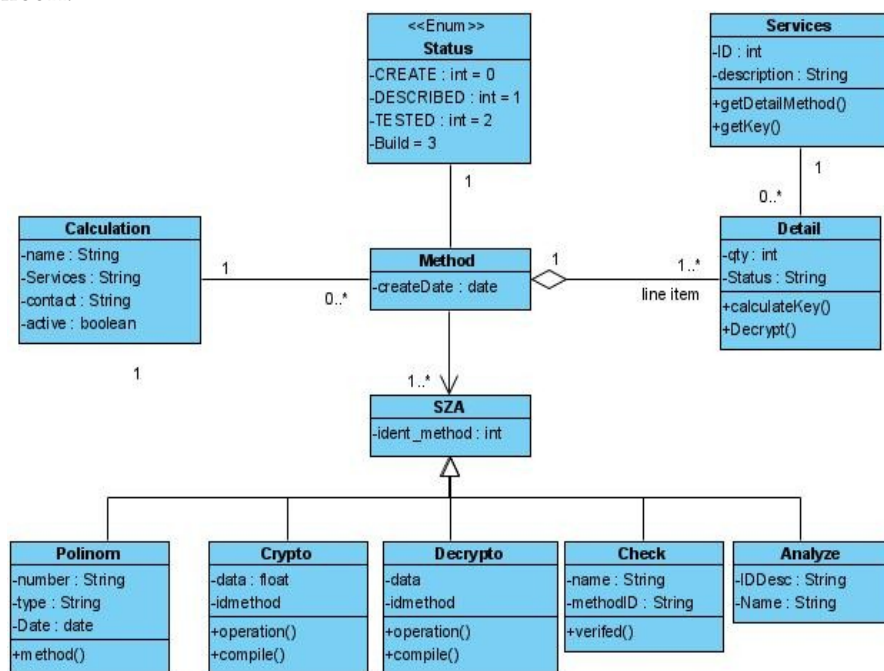


Рисунок 2 – Діаграма класів системи

На завершальному етапі остаточне криптографічне рішення передається зовнішній інформаційній системі або користувачеві через інтерфейс для подальшого використання. Завдяки чіткій структурі взаємодії між компонентами система забезпечує модульність, ефективну обробку запитів та високу продуктивність криптографічних алгоритмів.

4. Висновки

Розроблена програмна система реалізує повний цикл криптографічної обробки інформації на основі СЗК. Завдяки зручному інтерфейсу можна швидко вводити дані й аналізувати результати, що підвищує ефективність використання алгоритму для шифрування й розшифрування, забезпечує прозорість процесів, дає змогу оцінити результативність алгоритму і впевнитись у правильності обчислень.

Література

- [1] O.G.Abood, S.K.Guirguis. A survey on cryptography algorithms. International Journal of Scientific and Research Publications. Vol. 8(7), 2018. Pp. 410–415.
[DOI:10.29322/IJSRP.8.7.2018.p7978](https://doi.org/10.29322/IJSRP.8.7.2018.p7978)
- [2] M.Kasianchuk, I.Yakymenko, Y.Nykolaychuk. Symmetric Cryptoalgorithms in the Residue Number System. Cybernetics and Systems Analysis. Vol. 57(2). 2021. pp. 329–336.
<https://doi.org/10.1007/s10559-021-00358-6>.
- [3] І.Якименко. Методологія криптографічного захисту інформації на основі цілочисельної, модифікованої досконалої та поліноміальної систем залишкових класів. Матеріали XIV Міжнар. наук.-техн. конф.: ITSec: Безпека інформаційних технологій. Тернопіль-Київ: ЗУНУ-ДУІКТ, 2025. С.224-228 с.

LLM-Based Methodology for Data Classification under SOC 2 Type 2.

Oleh Deineka¹, Oleh Harasymchuk¹

¹ Lviv Polytechnic National University, 12 Bandera str., Lviv 79000, Ukraine

Abstract

This article introduces a comprehensive methodology for automated data classification, emphasizing the role of large language models (LLMs) in processing vast information arrays. The methodology encompasses stages from data collection and preliminary processing to entity recognition, event detection, and automatic classification, ensuring alignment with security standards like SOC 2 Type 2. Recognizing the limitations of costly commercial solutions, the proposed system offers flexibility and scalability, integrating seamlessly with existing technologies without substantial investment. Through robust security measures, continuous monitoring, and personnel training, this approach enhances trust and competitiveness, enabling organizations to effectively respond to potential incidents and maintain compliance with regulatory requirements.

Keywords

Large Language Models (LLM), SOC 2 Type 2, Data Governance, Data Quality, Access and Incident Management, Extract Transform Load (ETL), ITSM, and ITIL.

1. Introduction

In the context of the rapid growth of digital information, organizations face new challenges regarding effective management, protection, and analysis of data. Data exists in various forms—from structured databases to unstructured text files, correspondence, or media—which requires modern approaches to their processing and classification. The relevance of automated data classification methods is growing against the backdrop of information security requirements, compliance with legislative norms (for example, GDPR, SOC 2 Type 2), and the need to ensure confidentiality.

Large language models (LLMs) play a significant role in this process, as they automate the processing of large arrays of information, facilitate entity extraction, content analysis, and the establishment of relationships between data objects. The proposed methodology is based on a staged approach: from data collection and preliminary information processing to entity recognition, event detection, and automatic classification. In addition, security measures, regular monitoring, and personnel training are implemented, which allows not only for effective management of information flows but also for prompt response to potential incidents, increasing the level of trust and competitiveness of organizations.

Automated data classification using LLMs enables organizations to manage diverse and rapidly growing information streams more efficiently. By combining entity recognition, event detection, and contextual analysis, LLMs support compliance with regulations like GDPR and SOC 2 Type 2 while strengthening data security. This staged methodology, paired with monitoring and staff

^{74*} Corresponding author.

[†] These authors contributed equally.

✉ deinekaoleg.86@gmail.com (O. Deineka); oleh.i.harasymchuk@lpnu.ua (O. Harasymchuk);

ORCID 0009-0005-9156-3339 (O. Deineka); 0000-0002-8742-8872 (O. Harasymchuk);



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

training, ensures effective information governance, reduces risks of breaches, and enhances organizational competitiveness. [1-3]

2. Methodology overview

In the process of seeking an effective solution for data classification, storage, and processing in accordance with the SOC 2 Type 2 standard, it became clear that the implementation of existing solutions is not always straightforward. Many of them require significant adaptation efforts, rely on predefined templates, and offer limited customization options, which makes their use challenging in business environments with specific or evolving requirements.

To overcome these challenges, the team was tasked with developing a simple yet effective system that would meet SOC 2 Type 2 requirements while leveraging the technologies and competencies already available within the company. At the same time, it is essential to seek new opportunities, including the adoption of advanced technologies and innovative approaches, to ensure the system’s flexibility and practicality.

The solution must provide seamless integration with existing services without demanding large investments in costly software or equipment. Equally important is scalability, allowing the organization to adjust the system to future business needs and to new security and compliance challenges as they arise.

The methodology has been researched and presented in published articles [4,5,7], outlining approaches for classifying and storing information. Figure 1 highlights the data processing workflow as performed by various roles and activities.

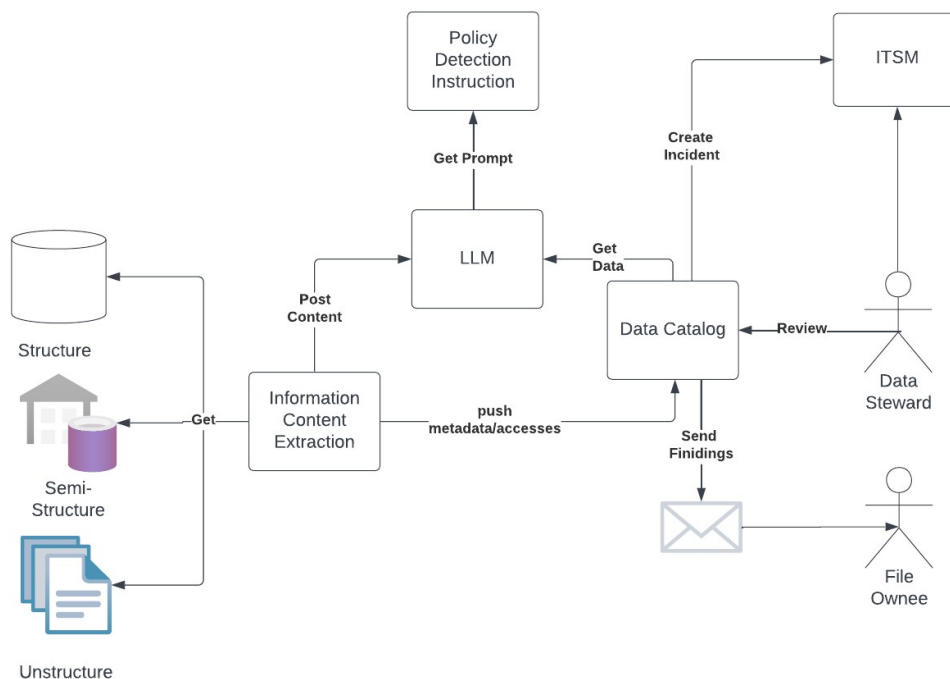


Figure 1: Methodology data flow

Data classification methodology is a step-by-step process that enables systematic resolution of efficient data management issues—from the moment data is received to integrated security measures, monitoring, and incident response.

8. **Data Collection and Reception** Effective data classification begins with gathering information from various sources—structured, semi-structured, and unstructured.

- Structured data is organized in clear formats (e.g., relational databases), making it quickly accessible for analytics and search.
- Semi-structured data has some structure with tagging, as in JSON or XML. It contains hierarchy and allows focused work with mixed content.
- Unstructured data includes text files, images, videos, etc., which lack a fixed schema. They are the most complex to analyze but often contain the most valuable information.

The main task of the collection stage is to ensure the complete reception of data, maintain its authenticity and integrity, and enable further transition to deeper analysis.

9. **Preliminary Text Processing** Data collection is accompanied by its initial processing. At this stage, the following actions are performed:

- Data cleaning from "noise" (e.g., punctuation marks, unnecessary codes, formatting)
- Text normalization (converting to a unified case, unifying date, and number formats)
- Removal of unnecessary words (stop words, non-informative parts)
- Lemmatization and stemming (reducing words to their root form)

These actions enhance text quality and create a foundation for further effective analysis and automation.

10. **Entity Recognition and Relationship Extraction**

In cleaned data, it is important to identify key objects—entities (names, dates, organizations, places). The use of entity recognition algorithms (NER) allows "extraction" of main participants in information flows. Following this, relationship extraction is applied, establishing connections between entities. For example, algorithms can determine that a specific person is mentioned in connection with a particular organization, location, or event.

11. **Event Extraction**

At this transitional stage, simple connections and sets of entities form complete events. Event identification is crucial for understanding the context and significance of information. This is especially important for security monitoring: detecting incidents, suspicious activities, or critical operations that require intervention.

12. **Data Classification**

Thanks to the obtained structure and context, data can be classified—automatically or with partial expert involvement. Classification categories are determined by sensitivity level: "confidential," "internal," "public," etc. This allows:

- Establishing access policies
- Specifying measures for protecting each category
- Harmonizing work with external and internal regulations

13. **Storage and Access Control Classified**

Data is stored in protected repositories (e.g., encrypted cloud environments or isolated segments of internal infrastructure). Each category has separate rules ensuring that access to specific data is granted only to a limited circle of authorized people. Continuous attention is paid to access monitoring, logging, and automatic notifications of suspicious or unauthorized actions.

14. **Data Protection and Incident Response**

The methodology involves the use of cryptographic tools, masking, anonymization. If an incident is detected, a clearly defined response procedure is activated to minimize potential risks, prevent leaks, and restore normal system operation. It is important to have pre-established adaptation scenarios for responding to different types of threats.

15. Roles and Responsibilities

The File Owner is responsible for ensuring that the data they manage is accurate, complete, and properly classified according to sensitivity levels. They must follow organizational policies when creating, storing, or sharing files, and are accountable for granting appropriate access rights. Based on established security policies, the File Owner is also responsible for applying appropriate information encryption to protect sensitive content. Additionally, the File Owner participates in the review of findings from the Data Catalog and takes corrective actions when policy violations or security issues are identified.

Data Steward oversees the governance process by reviewing metadata, classification results, and access policies generated by the system. Their role is to ensure data consistency, compliance with SOC 2 Type 2 standards, and adherence to organizational rules for information handling. Data Stewards also validate incidents flagged by the Data Catalog, escalate them through ITSM when needed, and provide guidance to File Owners on proper remediation steps. By bridging technical workflows and compliance requirements, they play a leading role in maintaining trust and integrity in data management. [4-9]

3. Conclusions

By employing a step-by-step structured approach, an organization gains the ability to effectively manage sensitive data, protect it from external and internal threats, automate analytics, and ensure continuous compliance with regulatory requirements. This integrally enhances the level of trust, reduces risks, and establishes transparent, understandable information management processes in a dynamic digital environment.

Beyond compliance, the methodology creates tangible business value: it improves operational efficiency by reducing manual work, lowers costs through scalable integration with existing systems, and strengthens resilience against cyber threats. Moreover, it supports innovation, enabling organizations to adopt modern technologies without compromising security. This approach enhances competitiveness by building trust with partners, clients, and regulators while ensuring sustainable long-term growth.

References

- [1] <https://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/soc2report.html>
- [2] ISACA – SOC 2 and Data Security
<https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2021/volume-7/soc-2-and-data-security>
- [3] John R. Vacca. Computer and Information Security Handbook, 3rd Edition. Elsevier, 2017, Chapter: Auditing and Compliance including SOC 2. ISBN: 9780128038437
- [4] O. Deineka, et al. Designing Data Classification and Secure Store Policy According to SOC 2 Type II CEUR Workshop Proceedings, 2024, 3654, pp. 398–409.
- [5] O. Deineka, et al. Application of LLM for Assessing the Effectiveness and Potential Risks of the Information Classification System According to SOC 2 Type II // CEUR Workshop Proceedings. – 2025, 3991, pp. 215–232.
- [6] "Large Language Models for Text Classification: Case Study and Comprehensive Review" by Arina Kostina et al., published in Computer Science Journal, 2024.
- [7] O. Harasymchuk, et al. Information classification framework according to SOC 2 Type II // CEUR Workshop Proceedings. – 2024, 3826, pp. 182–189.

- [8] ITIL Foundation, ITIL 4 Edition — AXELOS, First Edition ,2019, ISBN-10:0113316070, ISBN-13:9780113316076
- [9] Randy A. Steinberg. Servicing ITSM: A Handbook of Service Descriptions for IT Service Managers, Trafford Publishing, 2013, 288 сторінок, ISBN 13:978 1490719566
- [10] Shari S. C. Shang, Shu Fang Lin “Barriers to Implementing ITIL — A Multi-Case Study on the Service-based Industry” <https://doi.org/10.7903/cmr.1131>

Integrated Framework Deployment for Cybersecurity of Information and Communication Channels in On-Demand Printing

Tetyana Neroda¹

¹ *Institute of Printing Art and Media Technologies in Lviv Polytechnic, Bandery St. 12, 79013 Lviv, Ukraine*

Abstract

A clear and well-documented LibreOffice document is presented as an article formatted for publication by CEUR-WS in a conference proceedings. This article presents and explains many of the common variations, as well as many of the formatting elements an author may use in the preparation of the documentation of their work. Recent research in the security of computerized maintenance management systems reveals that conventional cybersecurity mechanisms remain primarily focused on external threats and fail to incorporate the operational specifics of telemetry streams and service log data. Addressing this gap, the present work substantiates the necessity of developing analytical approaches capable of detecting latent vulnerabilities in the information and communication channels of on-demand printing workflows. To this end, an analytical apparatus for real-time vulnerability assessment has been constructed within a multilayered service-oriented architecture. The apparatus integrates models of telemetry, maintenance logs, and risk evaluation into a closed mathematical cycle, thereby enabling a structured transition from heterogeneous big data flows to formalized representations suitable for quantitative analysis. The proposed framework provides metrics-driven insights into the impact of hidden vulnerabilities on production parameters such as throughput stability, workflow continuity, and confidentiality of client orders. Furthermore, it ensures that the outcomes of risk evaluation are directly transformed into executable protection actions, securing information and communication channels in dynamic printing environments.

Keywords

Computerized Maintenance, Management Systems, Information System Vulnerabilities, Covert Data Leakage Channels, Predictive Maintenance, Telemetry Data Analytics.

1. Introduction

In the context of digitalization of operational printing, the role of Computerized Maintenance Management Systems (CMMS) is steadily increasing. These systems integrate data from equipment, sensor subsystems, and software modules to ensure continuity of the production process. Such complexes generate large volumes of information, encompassing maintenance logs, printing machine telemetry, workload data, and downtime records. Traditional approaches to cybersecurity deployment in industrial environments primarily focus on countering external attacks, protecting against malicious software, and controlling access.

At the same time, the issue of covert data leakage channels remains insufficiently studied, particularly those associated with temporal patterns, correlation links between parameters, and metadata accompanying technical information. In the operational printing sector, where the speed of order processing and the stability of printing processes are critical, such vulnerabilities may indirectly reveal confidential information about production organization, equipment utilization, or maintenance planning practices. This creates a rationale for more in-depth research on the risks associated with the security of big data generated by maintenance management systems.

⁷⁵ Corresponding authors

✉ tetyana.v.neroda@lpnu.ua (T. Neroda)

ORCID 0000-0002-5728-7060 (T. Neroda)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Analysis of Recent Research

Recent scientific studies on the security of Computerized Maintenance Management Systems indicate increasing risks associated with the processing of large volumes of data and their integration with other information systems. In particular, specialized research emphasizes insufficient attention to vulnerabilities arising from covert data leakage channels, such as temporal patterns, correlation links, and metadata accompanying technical data, which can be exploited for unauthorized access to confidential information about clients, production organization, and maintenance planning [1].

Another important aspect is the integration of CMMS with other enterprise information systems, which can create new vulnerabilities and complicate security management. In particular, research highlights the need to implement centralized patch management systems, utilize endpoint detection and response (EDR) tools, web application firewalls, and network protocol analyzers to ensure the security of such integrated systems [2]. The growing role of artificial intelligence technologies in big data analysis opens new opportunities for anomaly detection and predictive maintenance. However, this also introduces new security challenges, particularly in the context of protecting machine learning algorithms from manipulation and ensuring the transparency and explainability of AI-driven decisions [1, 3, 4]. Risks associated with data manipulation do not necessarily involve information theft, but rather covert alteration of data to gain advantage or influence processes. Such modifications can be subtle and difficult to detect, yet they may have serious consequences for the organization. Protection against these attacks requires the implementation of anomaly detection mechanisms, data integrity assurance, and regular system audits [5]. The use of cloud technologies for storing and processing CMMS data can also increase system flexibility and efficiency, but it introduces additional security challenges. Specifically, it is necessary to prevent data leakage, ensure compliance with regulatory requirements, and implement effective access control mechanisms. Given the growing adoption of cloud technologies, it is crucial to develop security strategies that take into account the specific characteristics of these environments [6].

The analysis of recent research has shown that the presented results are mostly limited to general approaches to the security of industrial systems, without considering the specifics of printing production. There is a pressing need to deploy an analytical framework for identifying hidden vulnerabilities within industrial infrastructure and to further assess their potential impact on production processes and information security. Therefore, opens opportunities for the development and implementation of real-time protection strategies for information and communication systems in operational printing environments.

3. Problem statement

Multilevel information and communication links between printing equipment in modern industrial lines, management servers, cloud services, and user interfaces pose the threat of emerging risks that go beyond traditional notions of security. Vulnerabilities may arise not only at level of direct system access but also through indirect channels, including telemetry, maintenance logs, and machine workload data. Such leaks can indirectly reveal production cycles, equipment configurations, order characteristics, and organizational specifics of enterprise operations, complicating the development of a comprehensive framework for assessing potential threats and predicting their impact on the production infrastructure. In conditions of martial law and intensified competition in the printing services market, such leaks carry the risk of disclosing confidential information about clients, commercial terms of cooperation, and organizational

features of production. Therefore, the objective of the present study is to develop an analytical framework for identifying such vulnerabilities, assessing their impact on the stability of production processes, and determining ways to prevent potential threats to the security of production data and the protection of client interests.

4. Service Architecture for Hidden Vulnerability Detection

The architecture of a service framework for detecting hidden vulnerabilities within CMMS environments involves the development of a multi-layered model for processing production telemetry and maintenance logs. Its purpose is not only to identify known anomalies but also to establish correlations that remain unnoticed in traditional monitoring systems. This approach is based on multichannel data collection from printing equipment, microclimate sensor modules, CMMS scheduling modules, and user interfaces.

The first level involves the normalization of telemetry streams, which enables the alignment of data in different formats and the elimination of temporal discrepancies in their acquisition. This provides a foundation for the subsequent integration of technical and organizational parameters. The second level focuses on the detection of latent dependencies in maintenance logs, where relationships between technical service events, such as the frequency of component replacements or the recurrence of equipment failures, are considered potential indicators of vulnerabilities in software or hardware. The third level is aimed at assessing information risks associated with side-channel data leakage. For instance, the analysis of the frequency of service module accesses or the duration of preventive operations can indirectly reflect production load schedules or specifics of cooperation with particular clients. This forms the basis for constructing a production activity profile, which, in the case of unauthorized access, could become a source of confidential information.

The final stage involves the integration of results into a unified assessment system, which combines detected technical anomalies with potential data security threats. The proposed architecture enables a quantitative evaluation of the impact of hidden vulnerabilities on the stability of production processes and the level of data protection, while also providing a foundation for the development of real-time preventive measures.

5. Analytical Framework for Real-Time Vulnerability Assessment

The analytical framework for detecting hidden vulnerabilities within CMMS environments in operational printing is proposed to be built on a defined multi-layered architecture, with each level having its own system of formalized dependencies. At the first level, the normalization of telemetry streams from printing equipment is considered. Telemetry is presented as a set of discretized temporal patterns t_i , within which a parameter v_i is recorded, allowing for the tracking of periodicity or anomalies in the data based on the recorded “time–value” pair within the studied interval (1):

$$T = \{ (t_i, v_i) \}. \quad (1)$$

To eliminate gaps and synchronize time scales, a normalization operator $N: T \rightarrow T^*$ is applied, which transforms the original data into an aligned set T^* . This allows for the correct correlation of data from different sources. The second level concerns maintenance logs, which are described as a set of events of type ℓ_j (for example, a fault code or operation description) with corresponding attributes a_j , indexed by the event number or entry in the maintenance log (2):

$$\mathbf{L} = \{(\ell_j, t_j, a_j)\}. \quad (2)$$

To identify latent dependencies between logs and telemetry, a correlation operator $\mathbf{C}(\mathbf{T}^*, \mathbf{L}) \rightarrow \mathbf{D}$ is introduced, which generates a set $\mathbf{D}$ of hidden dependencies that are not captured by standard monitoring tools. The third level is focused on assessing the risks of side-channel data leakage. Here, for the time intervals Δt between events and weighting coefficients of different event types, a risk function is formulated (3):

$$\mathbf{R} = f(\mathbf{D}, \Delta t, \mathbf{k}). \quad (3)$$

The obtained value $\mathbf{R}$ provides a numerical assessment of risk or a vulnerability indicator, enabling the system to make real-time decisions (for example, initiating equipment checks, blocking suspicious communications, or generating alerts for the operator). Overall, this is interpreted as the probability that telemetry or logs indirectly reveal information about production loads or printing order data. At the fourth level, the results are integrated into a unified assessment of the impact on production processes and information security. For this purpose, production process stability metrics $\mathbf{Q}$, which include average downtime, number of equipment failures, and duration of preventive maintenance, are processed through an integrated indicator reflecting the security status within the system via the assessed risk level and the quality or condition of control parameters (4):

$$\mathbf{S} = g(\mathbf{R}, \mathbf{Q}) \quad (4)$$

The value $\mathbf{S}$, aggregated through weighting coefficients, the “minimum” rule, a probabilistic model, or another analytical approach, allows for the quantitative determination of threat levels (such as acceptable, potentially hazardous, or critical) and supports decision-making regarding the necessity of protective actions in real time. As a result, model (4) enables the transition from abstract risk to a concrete assessment of the security state, which directly guides the actions of the protection system or operator interventions. The proposed framework formalizes the architecture for analyzing telemetry and maintenance logs, transforming it into a system of mathematical operators and functions that ensures the detection of hidden vulnerabilities and the prediction of their impact on production processes and information security in operational printing. The generalization of the integrated indicator $\mathbf{S}$ in the form of a threshold-based classification model provides a formalized delineation of security levels (OK, ALERT, CRITICAL) based on configurable thresholds θ_1 and θ_2 , enabling the implementation of unified response criteria within the designed framework (5):

$$\mathbf{S} = \begin{cases} \text{OK}, & S < \theta_1 \\ \text{ALERT}, & \theta_1 \leq S < \theta_2 \\ \text{CRITICAL}, & S \geq \theta_2 \end{cases} \quad (5)$$

It should be noted that the classification model (5) is not self-sufficient without further interpretation of its results in the form of specific response procedures. Therefore, the next stage involves transitioning to an action function, which transforms the classification states into operational rules executed in real time, ensuring both the integrity of telemetry streams and the protection of information and communication channels from the further development of incidents (6):

$$act(S) = \begin{cases} \text{logging and monitoring,} & S < \theta_1 \\ \text{incident recording, access restriction,} & \theta_1 \leq S < \theta_2 \\ \text{emergency procedure initiation,} & S \geq \theta_2 \end{cases} \quad (6)$$

Thus, by combining telemetry models, maintenance logs, and integrated risk assessment, the presented analytical framework (1)–(4), with the introduction of threshold values (5), culminates in an action function (6) that formalizes the selection of operations in real time based on security level. This design enables the transition from abstract analysis to automated execution of specific response procedures, creating a closed loop from data collection to practical application. Within the integrated framework, this approach not only allows for the detection of hidden vulnerabilities but also directly transforms assessment results into protective actions, ensuring the resilience of information and communication channels in operational printing against threats in dynamic production environment.

Conclusions

The developed analytical apparatus integrates telemetry models, maintenance logs, risk assessment, and a composite security indicator, thereby forming a closed loop that transforms industrial big data into actionable decisions. This construction ensures systematic incident response, automation of protection procedures, and elastic integration of security analysis mechanisms into the production infrastructure without compromising its operational stability. The scientific novelty of study lies in formalized combination of telemetry and maintenance log analysis with a mathematical classification model of security states and the action function $act(S)$, which defines real-time response rules. The practical significance of the obtained results is determined by the adaptability of the integrated framework to diverse operational printing configurations, enhancing the level of cybersecurity without reducing order execution efficiency. Furthermore, the application of the analytical apparatus enables quantitative forecasting of the impact of hidden vulnerabilities on the stability of production processes and ensures the preservation of customer confidentiality, which is particularly significant in a dynamic information and communication environment.

Declaration on Generative AI

The author have not employed any Generative AI tools.

References

- [1] Durnyak, B., & Romanyshyn, Y. (2023). Protection of a Printing Company with Elements of Artificial Intelligence and IIoT from Cyber Threats. *Advances in Artificial Systems for Logistics Engineering III* (pp.197-205). DOI:10.1007/978-3-031-36115-9_19
- [2] Aldea, M., & Gheorghică, D. (2020). Software vulnerabilities integrated management system. In *2020 13th International Conference on Communications (COMM)* (pp. 111–116). DOI: 10.1109/COMM48946.2020.9141970
- [3] Shankar, L., Singh, C. D., & Singh, R. (2024). AI and CMMS: A powerful duo for enhanced maintenance in manufacturing. *Educational Administration: Theory and Practice*, 30(10), 1–10. DOI: 10.53555/kuey.v30i5.4434

- [4] Wieczorek, A. (2023). Assessment of usefulness of CMMS class system for Industry 4.0 enterprise. *Scientific Papers of Silesian University of Technology. Organization and Management Series*, 182, 1–12. DOI: 10.29119/1641-3466.2023.182.33
- [5] Lu, Y. J., Lee, W. C., & Wang, C. H. (2023). Using data mining technology to explore causes of inaccurate reliability data and suggestions for maintenance management. *Journal of Loss Prevention in the Process Industries*, 83, 105032. DOI: 10.1016/j.jlp.2023.105063
- [6] Arafat, M. (2018). Information security management system challenges within a cloud computing environment. In *Proceedings of the International Conference on Future Networks and Distributed Systems* (pp. 1–7). DOI: 10.1145/3231053.3231127
- [7] Mugarza, I., & Montero, J. L. (2020). Security issues and software updates management in the industrial internet of things (IIoT) era. *Sensors*, 20 (24), 7160. DOI: 10.3390/s20247160
- [8] Yadav, G., Paul, K., & Gauravaram, P. (2022). Vulnerability management in IIoT-based systems: What, why and how. In *Secure and Trusted Cyber Physical Systems* (pp. 1–27). Springer. DOI: 10.1007/978-3-031-08270-2_3
- [9] Kosmowski, K. T., Piesik, E., Piesik, J., & Śliwiński, M. (2022). Integrated functional safety and cybersecurity evaluation in a framework for business continuity management. *Energies*, 15(10), 3610. DOI: 10.3390/en15103610
- [10] Sappal, S., & Prowse, P. (2021). A cybersecurity vulnerability management system for medical devices. *CMBES Proceedings*, 44.

Application of Scallop neurosymbolic framework to defending against cyberattacks

Vitaly Vlasov¹, Sofia Mikhailova¹

¹ Ivan Franko National University of Lviv, Universytetska 1, 79000 Lviv, Ukraine

Abstract

Neurosymbolic AI is a promising approach to solving various issues with AI based on neural networks. We describe a solution for an intrusion detection system using Scallop, a novel neurosymbolic framework based on Datalog, a logic programming language.

Keywords

neurosymbolic, cyber attacks, Datalog, intrusion detection

1. What is Neurosymbolic AI

Among many issues with neural networks-based AI, we can mention explainability, interpretability, and reliability.

Neurosymbolic AI combines both neural networks and symbolic AI approaches. These two traditions have been present since the very beginning of AI research [1]. Neural networks approach, also called “connectionist”, began its evolution since pioneering work by Warren McCulloch, Walter Pitts, and Frank Rosenblatt. This approach came to prominence in late 1980s specifically, with work by Yann LeCun and others. In 2012 a series of papers such as Krizhevsky, Sutskever and Hinton’s 2012 “ImageNet Classification with Deep Convolutional Neural Networks” [2] was published, which set the stage for what was about to come, by achieving excellent results on the object recognition challenge known as ImageNet.

This approach is statistical, neural networks work as “universal function approximators” that are probabilistic in nature. Some even call Large Language Models (LLMs) using a derogatory term “stochastic parrots”. Neural networks are trained on vast amounts of data, which can be both structured and unstructured.

Another approach, which is often termed “symbolic” AI, can be traced back to Bertrand Russell, then later John von Neumann, and later to work by John McCarthy, Marvin Minsky, and others. In this approach, symbols are being manipulated by logical functions. Data that is fed into these systems is often stored in special databases called “knowledge bases”, that are often graph-like. Thus, it can be said that in this approach, both algorithmic manipulation and structured representation of data are crucial.

Prominent Neurosymbolic AI examples are AlphaFold, AlphaGo, AlphaProof, and AlphaGeometry by Google DeepMind. OpenAI’s ChatGPT has a system called “code interpreter”, in which LLMs call Python interpreters, which are themselves symbolic [1]. Additionally, data augmentation approaches (like Retrieval-augmented generation (RAG)) seem to rely on symbolic manipulation.

⁷⁶ Corresponding authors

✉ vitaly.vlasov@lnu.edu.ua (V. Vlasov); sofia.mikhailova@lnu.edu.ua (S. Mikhailova)

🆔 0009-0003-8007-2900 (V. Vlasov); 0000-0000-0000-0000 (S. Mikhailova)



© 2025 Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Intrusion detection systems

Network Intrusion Detection Systems (NIDS) are security systems that are able to detect and prevent attacks in a particular network environment, such as Denial of Service (DoS) attacks, phishing, malware etc. Moreover, these systems should be able to intelligently identify and classify any abnormal behaviors within a network [3].

Present intrusion detection systems utilize both classical and machine learning (ML) approaches, in particular, solutions based on deep learning approaches (neural networks). Classical NIDS employ primarily signature-based or rule-based methods, which compare observed traffic against predefined patterns of known attacks [4, 5]. These systems rely on signature databases which have to be always kept up-to-date, thus making them effective only for known threats.

More prominent ML methods include k-nearest neighbor algorithms, support vector machines, logistic regression (LR), Naive Bayes models, ensemble systems (XGBoost), and decision trees [3, 4, 5]. Neural network architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs) have been extensively used recently [5].

Notwithstanding the utility of ML-based methods, they still face difficulty in detecting unknown attacks. Moreover, issues like computational costs, interpretability and explainability are also present here [5]. Therefore, neurosymbolic AI methods might offer a promising solution to some of the aforementioned issues.

3. Scallop

Scallop is a language based on DataLog and an associated framework that supports differentiable logical and relational reasoning [6]. Datalog is a logic programming language, itself a subset of Prolog. Its precursors include DeepProbLog, an extension of probabilistic logic programming language ProbLog that integrates Probabilistic Logic Programming with deep learning by introducing the neural predicate [7].

Scallop programs can be easily integrated in Python or into PyTorch via a provided learning module. The key insight underlying Scallop is its choice of three inter-dependent design decisions: a relational model for symbolic data representation, a declarative language for symbolic reasoning, and a provenance framework for differentiable reasoning [6].

Relations in Scallop are used for symbolic data representation, while also being a good fit for probabilistic reasoning. The declarative language used in Scallop is an extension of DataLog. Datalog implementations can utilize optimizations used in relational database systems. Also, Datalog is rule-based, which makes programs easier to write, debug, and verify [6].

Scallop solves the issues of differentiation in gradient-based methods by leveraging the framework of provenance semirings [8]. The framework proposes a common algebraic structure for applications that define annotations (i.e., tags) for tuples and propagate the annotations from inputs to outputs of relational algebra queries.

One of the issues facing Scallop is slower performance of symbolic component that runs on a CPU as opposed to neural component that runs on a GPU. Recently, a version of Scallop that can run on GPUs have been proposed [9], which aims to solve mentioned performance issues.

4. Problem definition

A public security dataset (CIC-IDS2017) was used for model training. CIC-IDS2017 dataset contains benign and the most up-to-date common attacks. Additionally, it also includes the results

of the network traffic analysis using CICFlowMeter with labeled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols and attack [10].

The duration of the data period in the dataset is 5 days. The implemented attacks include Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DDoS.

A Scallop module was written for PyTorch which embedded rules for DoS attack detection. We used a neural network model based on a combination of Convolutional Neural Network (CNN) and a Bidirectional Long Short-Term Memory (BiLSTM) neural network architectures. BiLSTMs utilize both forward and backward temporal dependencies, and in combination with CNNs, work well for tasks where it is necessary to recognize both spatial and temporal patterns.

5. Conclusions

We observed that the application of Scallop to a PyTorch learning module increases accuracy of the model against Denial-of-Service (DoS) cyber attacks. Compared to conventional approach based on neural networks, there is noticeable degradation in training time, which hopefully can be mitigated with future GPU-based versions of Scallop.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] G. Marcus, How o3 and Grok 4 Accidentally Vindicated Neurosymbolic AI, 2025. URL: <https://garymarcus.substack.com/p/how-o3-and-grok-4-accidentally-vindicated>.
- [2] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, Communications of the ACM 60 (2012) 84 – 90. doi: 10.1145/3065386.
- [3] M. Al-Omari, M. Rawashdeh, F. Qutaishat, et al, An Intelligent Tree-Based Intrusion Detection Model for Cyber Security, Journal of Network and Systems Management 29:20 (2021). doi: 10.1007/s10922-021-09591-y.
- [4] H. T. T. Tran, J. Sander, A. Cohen, B. Jalaian, N. D. Bastian, Neurosymbolic AI Transfer Learning Improves Network Intrusion Detection. doi: 10.48550/arXiv.2509.10850. URL: <https://arxiv.org/abs/2509.10850>.
- [5] A. Bizzarri, C. (J.) Yu, B. Jalaian, F. Riguzzi, N. D. Bastian, Neurosymbolic AI for network intrusion detection systems: A survey, Journal of Information Security and Applications, Volume 94 (2025). doi: 10.1016/j.jisa.2025.104205.
- [6] Z. Li, J. Huang, M. Naik, Scallop: A Language for Neurosymbolic Programming, Proc. ACM Program. Lang. 7, PLDI, Article 166 (2023). doi: 10.1145/3591280.
- [7] R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester, L. De Raedt, Neural Probabilistic Logic Programming in DeepProbLog, Artificial Intelligence (2021). doi: 10.1016/j.artint.2021.103504.
- [8] T. J. Green, G. Karvounarakis, V. Tannen, Provenance Semirings, In ACM Symposium on Principles of Database Systems (PODS) (2007). doi: 10.1145/1265530.1265535.
- [9] P. Biberstein, Z. Li, J. Devietti, M. Naik, Lobster: A GPU-Accelerated Framework for Neurosymbolic Programming. URL: <https://arxiv.org/pdf/2503.21937>.
- [10] Intrusion detection evaluation dataset (CIC-IDS2017), 2017. URL: <https://www.unb.ca/cic/datasets/ids-2017.html>.

НАУКОВЕ ВИДАННЯ

МАТЕРІАЛИ

Міжнародної науково-технічної конференції
«SMICS: Безпека сучасних інформаційно-комунікаційних систем»

16-18 жовтня 2025 року

м. Львів (Україна)

Організаційний комітет конференції та редакція можуть не поділяти думки авторів і не несуть відповідальність за достовірність викладеної інформації.

За науковий зміст і викладення матеріалу, достовірність та коректність фактичних даних уся відповідальність покладається на авторів та їх наукових керівників.

Неінформативний текст матеріалів доповіді міг бути скорочений або вилучений на розсуд Оргкомітету конференції.

Оригінал-макет підготовлено на кафедрі кібербезпеки
Львівського національного університету імені Івана Франка