МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ
ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

НАВЧАЛЬНО-НАУКОВИЙ
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

TACS

THEORETICAL AND APPLIED CYBERSECURITY

**Матеріали III Всеукраїнської
науково-практичної конференції**

Випуск 3

Київ – 2025

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

THEORETICAL AND APPLIED CYBERSECURITY

Матеріали III Всеукраїнської науково-практичної конференції

Випуск 3

Київ – 2025

Рекомендовано до друку Вченою радою
КПІ ім. Ігоря Сікорського
(протокол № 9 від 08 вересня 2025 р.)

Theoretical and Applied Cybersecurity. Матеріали III Всеукраїнської науково-практичної конференції (TACS-2025). – Київ: Політехніка. – 319 с.

До збірника увійшли матеріали доповідей, представлених на III Всеукраїнській науково-практичній конференції Theoretical and Applied Cybersecurity (TACS-2025, 29 травня 2025 р., м. Київ, Україна).

У збірнику представлені матеріали, присвячені питанням безпечного функціонування кіберпростору, безпеки промислових систем та систем критичної інфраструктури, криптографічних проблем кібернетичної безпеки, запобігання і протидії кібератакам, моделювання та протидії інформаційним операціям, технологій інформаційно-аналітичних досліджень на основі відкритих джерел інформації, застосування штучного інтелекту тощо.

Наведені матеріали з актуальних проблем інформаційної та кібернетичної безпеки, можливості застосування штучного інтелекту, системного аналізу при забезпеченні підтримки прийняття рішень, комп'ютерного моделювання процесів і систем.

Для фахівців у сфері інформаційних технологій, інформаційної і кібернетичної безпеки, а також для аспірантів і студентів старших курсів вищої школи відповідних спеціальностей.

ПЕРЕДМОВА

III Всеукраїнська науково-практична конференція Theoretical and Applied Cybersecurity (TACS-2025) є академічною подією, яку організовує Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» на базі Навчально-наукового фізико-технічний інституту. Хоча конференція проводилась у Києві, учасники конференції представляють різні міста України і світу. Вимоги до публікації статей та стандартів для авторів постійно зростають, а процес рецензування стає все більш суворим з кожним роком.

Збірник містить відібрані доповіді, представлені на III Всеукраїнській науково-практичній конференції Theoretical and Applied Cybersecurity (TACS-2025, 29 травня 2025 р., Київ, Україна). Серед понад 100 поданих заявок обрано 55 найкращих доповідей для публікації. В матеріалах конференції представлені останні досягнення в таких напрямках кібербезпеки, як теоретичні та криптографічні проблеми кібернетичної безпеки; державна політика у сфері кібернетичної безпеки; математичні методи, моделі та технології дослідження безпечного функціонування кіберпростору; алгоритми та методи запобігання і протидії кібератакам; аналітичні системи на основі відкритих джерел інформації; безпека промислових систем та систем критичної інфраструктури; соціальний інжиніринг та методи протидії деструктивним психологічним впливам у кіберпросторі; застосування штучного інтелекту кібербезпеки.

Сподіваємось, що це видання стане значущим внеском у фундаментальні та прикладні знання у сфері кібербезпеки.

Програмний комітет конференції TACS-2025

ЄВРОПЕЙСЬКА БД ВРАЗЛИВОСТЕЙ (EUVD) У ГЕОПОЛІТИЧНОМУ ТА КІБЕРБЕЗПЕКОВОМУ КОНТЕКСТІ

Щербина Д. М.[1], Венгерський П. С.[1]

[1]*Львівський національний університет імені Івана Франка, м. Львів, Україна,*
danylo.shcherbyna@lnu.edu.ua, petro.venherskyy@lnu.edu.ua

Досліджено запуск Європейської бази даних вразливостей (EUVD) у контексті кібербезпеки та геополітики. Проаналізовано переваги та ризики проєкту, розглянуто зв'язок між міжнародними базами даних вразливостей. Окреслено повноваження американських та європейських державних органів у сфері кіберзахисту.

Ключові слова: CISA, CVE, ENISA, EUVD, GSD, MITRE, NVD, вразливості, геополітика, Європейська комісія, Європейський Союз, ЄС, кібербезпека

Вступ

13 травня 2025 р. Європейська комісія офіційно оголосила про запуск Європейської бази даних вразливостей (EUVD) [1]. Ця ініціатива спрямована на зміцнення цифрової безпеки Європи, допомагаючи організаціям відповідати вимогам Директиви NIS2 щодо управління ризиками вразливостей у критичних секторах, таких як енергетика, транспорт і охорона здоров'я. Крім того, EUVD підтримує Акт про кіберстійкість, забезпечуючи захист цифрових продуктів від кіберзагроз.

Американський контекст

У 2003 р. США створили орган US-CERT, який займався протидією кіберзагрозам. У 2009 р. створили окремий орган ICS-CERT для реагування на загрози безпеці виробничих систем керування. У лютому 2023 р. обидва органи було інтегровано у Cybersecurity and Infrastructure Security Agency (CISA). Згадана агенція розпочала роботу у 2007 р. як DHS National Protection and Programs Directorate, отримавши нову назву та повноваження у 2018 р. указом Д. Трампа під час його першого президентського терміну.

Оскільки кіберзагрози носять глобальний характер через саму природу мережі Інтернет, CISA часто доводилося діяти поза внутрішньо американським контекстом. Зокрема, відзначимо аналіз російських кібератак на українську інфраструктуру перед повномасштабним вторгненням та під час нього [2, 3].

MITRE, американська дослідницька організація, розробила концепцію універсальної БД для стандартизації ідентифікації та класифікації вразливостей. До появи бази Common Vulnerabilities and Exposures (CVE) у 1999 р. інформація про кіберзагрози була розпорошеною між різними дослідницькими установами, виробниками ПЗ та урядовими організаціями. Це спричиняло такі проблеми:

- відсутність єдиного стандарту ідентифікації вразливостей;
- складність координації між організаціями, що займалися кібербезпекою;
- дублювання даних, що ускладнювало аналіз загроз;
- несвоєчасна реакція на критичні вразливості через брак централізованого управління.

Геополітичні виклики

Таким чином, США взяли на себе відповідальність не лише за глобальну безпеку, а й за кібербезпеку, забезпечуючи координацію кіберзахисту в тому числі через БД CVE, підтримувану організацією MITRE, фінансування якої здійснювалося через CISA.

16 квітня 2025 р. програма CVE зіткнулася з ризиком припинення фінансування, що могло призвести до серйозних наслідків для кібербезпеки, оскільки CVE є основним механізмом ідентифікації та координації виправлення вразливостей у ПЗ. Хоча CISA підтвердило свою підтримку програми, заявивши, що проблеми були пов’язані не з браком коштів, а з адміністративними питаннями контракту [4], ця подія стала тригером передчасного запуску проєкту EUVD, який відбувся вже 19 квітня 2025 р. Станом на 20 травня 2025 р. вебсайт euvd.enisa.europa.eu все ще повідомляє, що знаходиться у фазі бета-тестування, тож передчасний реліз був вочевидь викликаний політичними факторами.

Європейський контекст

Запуск EUVD, з одного боку, зменшує залежність європейської галузі кібербезпеки від зовнішніх БД, та знаменує посилення технологічного суверенітету ЄС. З іншого боку, впровадження нової класифікації несе ризик фрагментації інформації щодо вразливостей, що може здаватися кроком назад після впровадження CVE. На думку авторів, такі побоювання є перебільшеними, адже поле Alternative ID містить відповідний CVE-, а подекуди і GSD-ідентифікатор. Як відомо, Global Security Database (GSD) була створена для усунення обмежень у системі CVE, зокрема проблеми масштабування, сфери охоплення та повільної реєстрації. Врешті-решт, самі американці мають базу даних National Vulnerability Database (NVD), яка також тісно пов'язана з CVE. Можна сказати, що згадані системи працюють не окремо, а у тісній взаємодії, створюючи глобальну модель реагування на кіберзагрози.

Розпорядником EUVD виступає Європейське агентство з кібербезпеки (ENISA). Воно було засноване 13 березня 2004 р. відповідно до Регламенту ЄС № 460/2004 з метою підвищення рівня кібербезпеки в Європейському Союзі шляхом координації зусиль між державами-членами та приватним сектором [5].

Попри очевидні переваги бази даних EUVD у зміцненні технологічної незалежності Євросоюзу, її запуск несе і певні виклики:

- якщо ЄС почне надто віддалятися від міжнародних платформ (CVE тощо), це може знизити ефективність глобальної взаємодії у кібербезпеці;
- компаніям доведеться адаптувати процеси до нових європейських стандартів;
- підтримка окремої бази даних залежатиме від постійного фінансування з боку ЄС.

Висновки

Таким чином, запуск EUVD є важливим кроком для зміцнення цифрового суверенітету ЄС, але його довгостроковий успіх залежатиме від ефективної взаємодії з міжнародними стандартами. Європейський Союз продовжує розвивати свою кібербезпекову політику,

впроваджуючи нові законодавчі ініціативи та розширюючи міжнародну співпрацю для захисту цифрового простору.

Список використаних джерел

1. Consult the European Vulnerability Database to enhance your digital security! ENISA. URL: https://www.enisa.europa.eu/news/consult-the-european-vulnerability-database-to-enhance-your-digital-security (дата звернення: 18.05.2025).

2. Cyber-Attack Against Ukrainian Critical Infrastructure. CISA. URL: https://www.cisa.gov/news-events/ics-alerts/ir-alert-h-16-056-01 (дата звернення: 18.05.2025).

3. Update: Destructive Malware Targeting Organizations in Ukraine. CISA. URL: https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-057a (дата звернення: 18.05.2025).

4. Statement from Matt Hartman on the CVE Program. CISA. URL: https://www.cisa.gov/news-events/news/statement-matt-hartman-cve-program (дата звернення: 18.05.2025).

5. Regulation - 460/2004 - EN - EUR-Lex. URL: https://eur-lex.europa.eu/eli/reg/2004/460/oj (дата звернення: 18.05.2025).

ON GRAPH BASED MULTIVARIATE MAPS OF PRESCRIBED DEGREE AND DENSITY AS INSTRUMENTS OF KEY ESTABLISHMENT AND ACCESS CONTROL

Vasyl Ustimenko

University of Royal Holloway (London), Egham Hill, Egham TW20 0EX, United Kingdom,
Institute of Telecommunications and Global Information Space of the NAS of Ukraine, Chokolivsky Boulevard 13, Kyiv,
vasylustimenko@yahoo.pl

Let us assume that one of two trusted parties (administrator) manages the information system (IS) and another one (user) is going to use the resources of this IS during the certain time interval. So they need establish secure user's access password to the IS resources of this system via selected authenticated key exchange protocol. So they need to communicate via insecure communication channel and secretly construct a cryptographically strong session key that can serve for the establishment of secure passwords in the form of tuples in certain alphabet during the certain time interval. Nowadays selected protocol has to be postquantum secure. We propose the implementation of this scheme in terms of Symbolic Computations. The key exchange protocol is one of the key exchange algorithms of Noncommutative Cryptography with the platform of multivariate transformation of the affine space over selected finite commutative ring. The session key is a multivariate map on the affine space. Platforms and multivariate maps are constructed in terms of Algebraic Graph Theory.

1. On multivariate cryptography

Multivariate Cryptography (MC) is one of the fifth directions of Post Quantum Cryptography designed for the constructions and investigations of asymmetric cryptographic algorithms with the resistance to attacks of the adversary who uses quantum computer. The security of algorithm of MC is based on the complexity of solving the system of nonlinear

equations over the finite commutative ring. Traditionally MC uses systems of equations of degree 2 or 3 defined over the finite field. Despite the fact that the last talk on Multivariate Cryptography on Eurocrypt conferences was in 2021 many new results in this area were obtained (see Proceedings of PQCrypts 2021-2024) and new cryptosystem [1] was submitted to NIST (USA). Classical task of Multivariate Cryptography is the constructions of public keys in the form of multivariate maps of small degree over finite fields.

We consider more general task of construction of nonlinear multivariate map F on affine space K^n where K is a commutative ring with unity with the trapdoor accelerator T which is a piece of information such the knowledge of T allows us to compute the reimage of F in polynomial time (see [2], [3] for some examples). We assume that multivariate map F is given in its standard form of kind $x_i \rightarrow f_i(x_1, x_2, \ldots, x_n)$, $i=1, 2,\ldots, n$ where polynomials f_i are given in the form of the some of monomial terms listed in the lexicographical order. We assume that monomial term M is written in the form $a(x_1)^{t(1)}(x_2)^{t(2)}\ldots(x_n)^{t(n)}$, where $t(i)$ are elements of Z_m, m is the order of multiplicative group K^*.

The construction of such forms with the corresponding trapdoor accelerator is an interesting task of Algebraic Geometry for which cases of complex numbers or algebraically closed field K are especially important.

Multivariate Cryptography in wide sense is about applications of the theory of nonlinear trapdoor accelerators to Symmetric and Asymmetric Cryptography. In the case when K is a finite commutative ring the pair (F, T) can be used as instrument of symmetric encryption. The space K^n can serve as space of ciphertexts. Usually the knowledge of T allows efficient computation of the value of F on the given tuple. Both correspondents have the knowledge on T. So they do not need to compute the standard form. Adversary can use various attacks to investigate the multivariate nature of encryption procedure for the construction of the procedure to compute the reimages of the map.

Standard forms of F with trapdoor accelerator can be considered as potential multivariate maps which can serve as

instruments for encryption or conducting digital signatures. Multivariate maps are elements of Cremona semigroup ${}^nCS(K)$ of endomorphisms of $K[x_1, x_2, \ldots, x_n]$. Such endomorphism F can be given by its values $F(x_i)= f_i(x_1, x_2, \ldots, x_n)$, $i=1, 2, \ldots, n$ on generic variables x_i. We assume that polynomials f_i are given via their standard form and define degree and density of the map F.

Trapdoor accelerators can be used for the generation of subsemigroup H of ${}^nCS(K)$ with the property P of computation of the product of n elements from H in a polynomial time. Its alternative approach to combinatorial method of generators and relations. Note that ${}^nCS(K)$ itself does not posses P itself because the product of n its general representatives of degree k, $k \geq 2$ will be of degree k^n.

Subsemigroup H satisfying property P can be used as platforms of Noncommutative Cryptography ([4]) for the implementation of algebraic key exchange protocols of Postquantum Cryptography. Recent cryptanalytic results on algorithms of Noncommutative Cryptography can be found in [10].

Some methods of construction of multivariate maps with the trapdoor accelerator in terms of Algebraic Graph Theory are presented in [2] together with some cryptographical applications.

The talk is dedicated to applications of graph based constructions of trapdoor accelerators to algorithm of the establishment of secure user's access password to the resources of Information System with further options of the password changes.

2. On the key establishment algorithms in term of multivariate cryptography

Assume that Administrator A and his/her trusted user have safely protected password t for mutual authentication, This is the string from K^n.

Administrator A of the Information System (IS) possesses the map F in n-variables and its trapdoor accelerator T. He/she is going to give secure access to the resources of IS to trusted

user U. So A and U executes selected protocol of Noncommutative Cryptography in terms of special subsemigroup S of the affine Cremona semigroup of all multivariate maps of K^n into itself. The output of the protocol X can be used by A and U for the mutual identification. U sends $X(t)$ to A who compares it with own computation of $X(t)$.

Administrator creates the map F of the same degree $deg\ (X)$ of the affine space of K^n

Administrator sends $F+X$ to U. User restores F. Now A is able to create pseudorandom or genuinely random password $(p_1, p_2, ..., p_n) = p$ as the condition to enter the system. Administrator solves the equation $F(x)=b$ and sends the solution $x=(d_1, d_2,..., d_n)=d$ to the user together with the link for entering the password. User U gets the password as $F(d_1, d_2, ..., d_n)$.

Administrator has the option to change the password several times working with the same map F with the trapdoor accelerator. He/she is able to change F via a new session of the protocol and delivery scheme. Some modifications of this procedure are discussed in the conclusion of the paper.

The security of this scheme rests on the security of selected Postquantum Protocol on Noncommutative Cryptography. We describe Twisted Diffie-Helman protocol which use the complexity of Conjugation Power Problem of the subsemigroup of ${}^nCS(K)$ satisfying property P. The general idea of this scheme and some other protocols and platforms can be found in [5]-[7]. Some of them use the semigroup ${}^nES(K)$ of Eulerian transformations (see [8]).

The talk is dedicated to the implementation of the scheme with constructions of graph based multivariate maps of prescribed degree and density with the trapdoor accelerators. We use the platforms generated by the transformations of densities $O(1)$ or $O(n)$. Recall that the density of multivariate map F is a maximal value of densities of $f_i = F(x_i)$, $i=1, 2, \ldots, n$ which are numbers of monomial terms of these multivariate polynomials.

We use walks on special linguistic graphs defined in [3] for the explicit constructions of groups supporting the following statement of Computational Algebraic Geometry.

Theorem 1. *Let K be commutative ring with unity. For each positive integer n, d, d≥2 and s≥0 there is a subgroup H of affine Cremona semigroup* $^{n}CS(K)$ *of all endomorphisms of* $K[x_1, x_2,..., x_n]$ *such that maximal degree of representative of H is d and the densities of elements from H are of size* $O(n^s)$.

Corollary. *There is a subsemigroup H of* $^{n}CS(K)$ *of prescribed degree d and density* $O(n^s)$.

Theorem 2. *For each parameters n, d,* $d \geq 2$ and α, $\alpha \leq d$ *there is a polynomial map F of* K^n *onto* K^n *of degree d, density of size* $O(n^{\alpha})$ *with the trapdoor accelerator of speed* $O(n)$.

So we used semigroup *H* of Theorem 1 and maps of Theorem 2 based on graphs [9] for the implementation of described above algorithm of the key establishment. Results of computer simulations will be presented.

References

1. NIST PQC Digital Signature Project: TUOV Specification\url{https://csrc.nist.gov/csrc/media/Projects/pqc-dig-sig/documents/round-1/spec-files/TUOV-spec-web.pdf}, last accessed 2024/07/30.

2. Ustimenko V. Graphs in terms of Algebraic Geometry, symbolic computations and secure communications in Post-Quantum world. UMCS Editorial House, Lublin (2022).

3. Ustimenko V. Schubert cells and quadratic public keys of Multivariate Cryptography. CEUR Workshop Proceedings ITTAP, \url{https://ceur-ws.org/Vol-3628/}, last accessed 2024/07/30.

4. Myasnikov A. Shpilrain V. and Ushakov A. (2011), Non-commutative Cryptography and Complexity of Group-theoretic Problems, American Mathematical Society.

5. Ustimenko V. On new symbolic key exchange protocols and cryptosystems based on hidden tame hommorphism, Dopovidi NAS of Ukraine, 2018, n 10, pp.26-36.

6. Ustimenko V. On desynchronised multivariate algorithms of El Gamal type for stable semigroups of affine Cremona group, Theoretical And Applied Cybersecurity, Vol. 1, No. 1 (2019).

7. Ustimenko V. On the usage of postquantum protocols defined in terms of transformation semigroups and their homomophisms, Theoretical And Applied Cybersecurity, Vol. 2, No. 1 (2020).

8. Ustimenko V. On Inverse Protocols of Post Quantum Cryptography Based on Pairs of Noncommutative Multivariate Platforms Used in Tandem, Theoretical And Applied Cybersecurity, Vol. 5, No. 2 (2023).

9. Ustimenko V. On Eulerian semigroups of multivariate transformations and their cryptographical properties, European Journal of Mathematics 9, 93 (2023).

10. Chojecki T., Erskine G., Tuite J., Ustimenko V. On affine forestry over integral domains and families of deep Jordan–Gauss graphs. European Journal of Mathematics 11, 10 (2025).

11. Ben-Zvi A., Kalka A., Tsaban B. Cryptanalysis via algebraic spans. Advances in Cryptology – CRYPTO 2018 / eds.: H. Shachan, A. Boldyreva. Berlin: Springer, 2018. P. 1–20. (LNCS; vol. 109991).

МОДЕЛЮВАННЯ КОНКУРЕНЦІЇ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ ЗА ЕНЕРГІЮ І КОРИСТУВАЧІВ

Ланде Д. В.[1], Даник Ю. Г. [1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»,*
dwlande@gmail.com

У доповіді розглядається проблема моделювання конкуренції між системами штучного інтелекту, які взаємодіють у спільному середовищі за обмежених ресурсів, таких як користувачі та енергія. Дослідження присвячене аналізу стратегічної поведінки систем, їх адаптивності та впливу на результати конкуренції через математичні моделі та методи, зокрема диференціальні рівняння, модель Ланкастера та цикли Бойда. Дослідження включає числові розрахунки та симуляції, які демонструють, як початкові умови та параметри систем впливають на їхню конкурентоспроможність. Запропоновані моделі можуть бути застосовані для прогнозування поведінки AI-систем у реальних сценаріях, таких як інформаційні кампанії, кіберконфлікти та оптимізація ресурсів у цифрових середовищах.

Ключові слова: конкуренція AI-систем, модель Ланкастера, цикли Бойда, диференціальні рівняння, енергоефективність, задоволеність користувачів, інформаційна боротьба, кібербезпека, Adversarial AI, симуляція взаємодії

Вступ

Одним із найбільш важливих ресурсів, які визначають успіх великих мовних моделей (LLM), є користувачі та енергія. Більша кількість користувачів забезпечує моделі більший обсяг даних для навчання та покращення, а також збільшує її вплив на ринку. Однак це також призводить до зростання споживання енергії, оскільки обробка великих

обсягів даних вимагає значних обчислювальних ресурсів. Ефективне використання енергії може дозволити моделі працювати швидше та точніше, що, у свою чергу, залучає більше користувачів. Таким чином, конкуренція між LLM формує складну взаємодію, де кожна модель намагається домогтися переваги через оптимізацію своїх стратегій.

У дослідженні конкуренції між LLM вже існує ряд підходів, які враховують різні аспекти їхньої взаємодії. Зокрема, у роботах [1] та [2] розглядаються моделі змагального штучного інтелекту, аналізу взаємодії великих мовних моделей у різних сценаріях. У [3] акцент робиться на використанні дистиляції знань для покращення продуктивності систем, що дозволяє їм ефективніше використовувати доступні ресурси. Також вивчаються питання енергоефективності, які стають особливо актуальними у контексті зростання масштабів обчислень [4], [5]. Для глибокого розуміння конкуренції між LLM необхідно застосовувати комплексні моделі, такі як модель Ланкастера [6], яка описує динаміку взаємодії між протиборчими сторонами, або цикли Бойда (OODA-loop) [7], які дозволяють формалізувати процес прийняття рішень у реальному часі.

Мета дослідження полягає у створенні математичних моделей, які дозволяють прогнозувати поведінку LLM у конкурентному середовищі, враховуючи їхню взаємодію за ресурсами та користувачами. Це надасть можливість не лише аналізувати існуючі тенденції, але й розробляти нові стратегії для покращення ефективності та стійкості ШІ-систем.

Для вирішення задачі моделювання конкуренції між системами штучного інтелекту застосовувався ШІ як основний інструментарій. Для забезпечення чіткості та формалізації завдань, що ставляться перед ШІ, використовувались структуровані промпти, побудовані на базі фреймворку безкодового програмування [8]. Застосування цього фреймворку дозволило створити промпти, які не лише чітко описують задачі, але й забезпечують їх виконання у вигляді послідовних інструкцій, що інтерпретуються великими мовними

моделями. Такий підхід забезпечує прозорість, відтворюваність результатів та можливість адаптації до різних сценаріїв конкуренції.

Основний зміст

Опис поведінки двох LLM, які конкурують за енергію і користувачів

Уявімо, що дві моделі, LLM_A і LLM_B, працюють у спільному середовищі, де вони отримують ресурси (енергію) та користувачів на основі своєї продуктивності та популярності.

Кожна модель має за мету максимізувати кількість користувачів і ресурсів, які вона отримує.

Для математичної формалізації взаємодії двох LLM через диференціальні рівняння, використовується підхід, який враховує динаміку розподілу користувачів і енергії між системами.

Розглядаються такі змінні та параметри:

- $U_A(t)$ – кількість користувачів LLM_A на час *t*;
- $U_B(t)$ – кількість користувачів LLM_B на час *t*;
- $E_A(t)$ – енергія LLM_A у момент часу *t*;
- $E_B(t)$ – енергія LLM_B у момент часу *t*;
- $S(t)$ – ентропія системи у момент часу *t*.

Загальна кількість користувачів і енергії є константою:

$U_A(t)+U_B(t)=1,\ \ E_A(t)+E_B(t)=1.$

Швидкість зміни кількості користувачів $U_A(t)$ і $U_B(t)$ залежить від стратегій моделей та їх ефективності.

Припустимо, що

- швидкість зміни користувачів $U_A(t)$ пропорційна різниці ефективності стратегій LLM_A і LLM_B;
- ефективність стратегій залежить від наявної енергії.

Математично це можна записати так:

$$\frac{dU_A}{dt} = k_1 \cdot \left(E_A - E_B\right) + k_2 \cdot \left(U_A - U_B\right),$$

де $k_1 > 0$ – коефіцієнт, що показує вплив енергії на залучення користувачів; $k_2 > 0$ – коефіцієнт, що показує вплив поточної різниці користувачів на їх зміну.

Аналогічно для $U_B(t)$:

$$\frac{U_B(t)}{dt} = -\frac{U_A(t)}{dt}.$$

Енергія перерозподіляється на основі умов (наприклад, якщо одна модель має більше користувачів, вона отримує додаткову енергію). Це можна формалізувати так:

$$\frac{E_A(t)}{dt} = k_3 \cdot \max\left(0, U_A - U_B - \Delta\right) - k_4 \cdot \max\left(0, U_B - U_A - \Delta\right),$$

де Δ – порогове значення переваги користувачів (0.1); $k_3 > 0$ – коефіцієнт, що регулює швидкість надання додаткової енергії LLM_A; $k_4 > 0$ – коефіцієнт, що регулює швидкість надання додаткової енергії LLM_B.

Аналогічно для $E_B(t)$:

$$\frac{E_B(t)}{dt} = -\frac{E_A(t)}{dt}.$$

За початкові умови можна:

$$U_A(0) = 0{,}5,\ \ U_B(0) = 0{,}5;\ \ E_A(0) = 0{,}5,\ \ E_B(0) = 0{,}5.$$

Модель LLM_A або LLM_B збільшує свою частку користувачів, якщо її стратегія є ефективнішою (залежно від енергії та поточного розподілу).

Енергія перерозподіляється на користь моделі, яка має значну перевагу у користувачів.

Система еволюціонує від високої ентропії (рівномірний розподіл) до низької ентропії (перевага однієї моделі).

За допомогою фреймворку безкодового програмування сформульовано задачу у вигляді структурованого промпту, який можна потім виконувати, вказуючи лише кількість

ітерацій і відношення потужностей систем. Створений за алгоритм має вигляд:

Ми моделюємо взаємодію двох великих мовних моделей (LLM_A і LLM_B), які конкурують за користувачів і енергію. Система оновлюється через певну кількість ітерацій, на кожній з яких моделі приймають рішення на основі своїх стратегій. Початковий розподіл користувачів і енергії задається у відсотках.

Параметри:

Кількість ітерацій (*N*): Скільки разів моделі взаємодіятимуть.

Відношення потужностей систем $\left(P_A,\ P_b\right)$: Початковий розподіл користувачів і енергії між LLM_A і LLM_B у відсотках.

Інструкція для виконання

Ініціалізація:

Задайте початковий розподіл користувачів і енергії:

Користувачі LLM_A: U_A. Користувачі LLM_B: U_B.

Енергія LLM_A: P_A. Енергія LLM_B: P_B.

Цикл взаємодії:

Повторіть наступні кроки *N* разів:

Перевірте умови для розподілу енергії:

Якщо кількість користувачів LLM_B перевищує кількість користувачів LLM_B більш ніж на 10 %, то надайте додаткову енергію LLM_A (збільште її енергію на 5 %).

Якщо кількість користувачів LLM_B перевищує кількість користувачів LLM_A більш ніж на 10 %, то надайте додаткову енергію LLM_B (збільште її енергію на 5 %).

Застосуйте стратегії моделей:

LLM_A використовує функцію для покращення

персоналізації відповідей. Ефективність цієї стратегії залежить від доступної енергії.

LLM_B використовує функцію для покращення швидкості відповідей. Ефективність цієї стратегії також залежить від доступної енергії.

Поновіть кількість користувачів:

Оцініть вплив стратегій на користувачів:

Додайте до користувачів LLM_A результат впливу його стратегії.

Додайте до користувачів LLM_B результат впливу його стратегії.

Якщо сума користувачів LLM_A і LLM_B не дорівнює 100 %, нормалізуйте значення так, щоб загальна сума становила 100 %.

Виведіть результати ітерації:

Покажіть поточний розподіл користувачів між LLM_A і LLM_B.

Повернення фінального стану:

Після завершення всіх ітерацій виведіть фінальний розподіл користувачів між LLM_A і LLM_B.

Сучасні LLM видають результати опрацювання цього моделювання у вигляді:

Виконується промпт для 10 ітерацій і однакової початкової потужності систем для обох LLM_A і LLM_B, але встановлюються різні коефіцієнти переваги користувачів k_i.

Результати моделювання наведені в табл. 1. Завдяки більш ефективній стратегії (покращення персоналізації відповідей) та додатковій енергії, отриманій через перевагу в користувачах, LLM_A значно збільшує свою частку ринку. LLM_B, незважаючи на покращення швидкості відповідей, не може конкурувати з LLM_A.

Таблиця 1. Результати моделювання

Ітерація	Користувачі LLM_A	Користувачі LLM_B
1	53 %	47 %
2	56 %	44 %
3	60 %	40 %
4	63 %	37 %
5	66 %	34 %
6	69 %	31 %
7	72 %	28 %
8	75 %	25 %
9	78 %	22 %
10	81 %	19 %

Модель Ланкастера для боротьби систем штучного інтелекту

Модель Ланкастера – це математична модель, яка використовується для опису динаміки взаємодії між двома протиборчими сторонами. Вона може бути адаптована для моделювання конкуренції між двома системами, такими як LLM_A і LLM_B. У цьому випадку ми будемо розглядати «силу» кожної системи як кількість її користувачів (U_A і U_B).

Конкуренція між системами описується через диференціальні рівняння.

Існує два основних типи моделей Ланкастера лінійна та квадратична. Квадратична модель є більш придатною, оскільки вона враховує нелінійність впливу (наприклад, масштабність стратегій).

Модель Ланкастера для конкуренції між LLM_A і LLM_B з урахуванням енергії записуємо так:

$$\frac{dU_A}{dt} = -k_B \cdot U_B^2 \cdot E_B + k_A \cdot U_A^2 \cdot E_A,$$

$$\frac{dU_B}{dt} = -k_A \cdot U_A^2 \cdot E_A + k_B \cdot U_B^2 \cdot E_B.$$

Це враховує, що ефективність стратегій залежить від наявної енергії.

Приклад початкових умов:

$U_A(0) = 0{,}5, \quad U_B(0) = 0{,}5, \quad E_A(0) = 0{,}5, \quad E_B(0) = 0{,}5$ (початкова рівновага).

Система з більшою кількістю користувачів (U_A або U_B) має перевагу, оскільки її вплив зростає нелінійно (U^2). Енергія (E_A, E_B) посилює ефективність стратегій, що дозволяє системі з більшою енергією швидше здобувати перевагу.

Розрахунок за моделлю Ланкастера ітеративним методом для початкових умов $k_A = 0{,}2$, $k_B = 0{,}15$ приводить до результатів, наведених у табл. 2. Після 10 ітерацій:

U_A = 0,6286 (62,86 % користувачів),

U_B = 0,3714 (37,14 % користувачів).

Моделювання циклів Бойда

Цикли Бойда (OODA-loop: Observe-Orient-Decide-Act) також можна використати для моделювання конкуренції між двома системами (LLM_A і LLM_B).

Цикли Бойда дозволяють формалізувати конкуренцію між LLM_A і LLM_B через ітеративний процес спостереження, аналізу, прийняття рішень і дій. Цей підхід враховує стратегічну поведінку систем і їх адаптивність до змін у середовищі.

Таблиця 2. Ітерації за моделлю Ланкастера

Ітерація	U_A	U_B
1	0,50625	0,49375
2	0,51357	0,48643
3	0,52207	0,47793
4	0,53180	0,46820
5	0,54287	0,45713
6	0,55548	0,44452
7	0,56992	0,43008
8	0,58654	0,41346
9	0,60585	0,39415
10	0,62860	0,37140

Observe (Спостереження):

Цей етап полягає у зборі інформації про стан середовища. Для кожної системи (LLM_A і LLM_B) ми можемо формалізувати цей процес через функцію:

$$Input_A = F_{observe}\left(U_A(t), E_A(t), S_A(t)\right),$$
$$Input_B = F_{observe}\left(U_B(t), E_B(t), S_B(t)\right),$$

де $U_A(t), U_B(t)$ – кількість користувачів на момент часу t; $E_A(t), E_B(t)$ – енергія системи на момент часу t; $S_A(t), S_B(t)$ – задоволеність користувачів на момент часу t ; $F_{observe}$ – функція, яка збирає дані про стан системи.

Orient (Орієнтація):

На цьому етапі система аналізує зібрані дані та формує стратегії. Використовуючи примітив «Функція», можна записати:

$$Strategy_A = F_{orient}\left(Input_A\right),$$

$$Strategy_B = F_{orient}\left(Input_B\right),$$

де F_{orient} – функція, яка аналізує вхідні дані та генерує стратегію.

Наприклад, стратегія може бути визначена через умовний оператор («Умова»):

$$Strategy_A = \begin{cases} \text{«}Покращити\ персоналізацію\text{»}, & якщо \quad U_A(t) > U_B(t), \\ \text{«}Покращити\ швидкість\text{»}, & якщо \quad U_A(t) \le U_B(t). \end{cases}$$

Аналогічно для $Strategy_B$.

Decide (Рішення):

На цьому етапі система вибирає найкращу стратегію для досягнення своїх цілей. Використовуючи примітив «Умова», можна записати:

$$Decision_A = \begin{cases} \text{«}Покращити\ персоналізацію\text{»}, & якщо \quad S_A(t) > S_B(t), \\ \text{«}Покращити\ швидкість\text{»}, & якщо \quad S_A(t) \le S_B(t). \end{cases}$$

Аналогічно для $Decision_B$:

Act (Дія):

На цьому етапі система реалізує обрану стратегію. Вплив стратегій на користувачів можна описати через диференціальні рівняння:

$$\frac{U_A(t)}{dt} = k_A \cdot \left(Effect_{Decision_A}\right) - k_B \cdot \left(Effect_{Decision_B}\right),$$

$$\frac{U_B(t)}{dt} = -\frac{U_A(t)}{dt},$$

де k_A, k_B – коефіцієнти ефективності стратегій LLM_A і LLM_B, $Effect_{Decision}$ – ефективність обраної стратегії.

Кожна система (LLM_A і LLM_B) виконує свій цикл Бойда незалежно, але їхні дії взаємодіють через спільне середовище:

$$\frac{dU_A}{dt} = f\left(U_A, U_B, E_A, E_B\right),$$
$$\frac{dU_B}{dt} = g\left(U_A, U_B, E_A, E_B\right).$$

Цикли Бойда дозволяють формалізувати конкуренцію між LLM_A і LLM_B через ітеративний процес спостереження, аналізу, прийняття рішень і дій. Цей підхід враховує стратегічну поведінку систем і їх адаптивність до змін у середовищі.

Наприклад, початкові умови:

$U_A(0) = 0,6,\ \ U_B(0) = 0,4;$
$E_A(0) = 0,55,\ \ E_B(0) = 0,45;$
$S_A(0) = 0,85,\ \ S_B(0) = 0,75;$
$k_A = 0,25,\ \ k_B = 0,2,$

приводять до результатів:

$$U_A = 0,6 + 0,0035 = 0,6035,\ U_B = 0,4 - 0,0035 = 0,3965.$$

Таким чином, цикли Бойда дозволяють формалізувати конкуренцію між LLM_A і LLM_B через ітеративний процес спостереження, аналізу, прийняття рішень і дій. Цей підхід враховує стратегічну поведінку систем і їх адаптивність до змін у середовищі.

Для об’єктивної оцінки ефективності розглянутих підходів до моделювання конкуренції між системами

штучного інтелекту було визначено комплекс критеріїв, що відображають практичні аспекти їх застосування.

Експертний аналіз показав, що кожен з розглянутих методів має свої характерні особливості та області оптимального застосування (див. табл. 3).

Таблиця 3. Порівняльна характеристика методів моделювання

Критерій	Диф. рівняння	Модель Ланкастера	Цикли Бойда
Складність реалізації	Висока	Середня	Низька
Адаптивність	Обмежена	Часткова	Висока
Урахування стратегії	Так	Частково	Повною мірою
Прогнозна ефективність	Висока	Середня	Висока
Універсальність	Обмежена	Часткова	Висока

Висновки

У доповіді розглянуто задачу моделювання конкуренції між двома системами штучного інтелекту, які взаємодіють у спільному середовищі за обмежених ресурсів, таких як користувачі та енергія. Дослідження вирішує ключові завдання, пов'язані з аналізом динаміки взаємодії, формалізацією стратегічної поведінки та прогнозуванням результатів конкуренції. При моделюванні запропоновано розширену систему показників, яка враховує не лише кількість користувачів і енергію, але й інші важливі аспекти поведінки систем, такі як задоволеність користувачів, точність відповідей, швидкість обробки запитів та етичність.

Результати дослідження можуть бути використані для прогнозування поведінки AI-систем у реальних сценаріях, таких як інформаційні кампанії, кіберконфлікти та оптимізація ресурсів у цифрових середовищах.

Список використаних джерел

1. Haryanto, Christoforus Yoga, Anne Maria Elvira, Trung Duc Nguyen, Minh Hieu Vu, Yoshiano Hartanto, Emily Lomempow, and Arathi Arakala. "Contextualized AI for Cyber Defense: An Automated Survey using LLMs." In 2024 17th International Conference on Security of Information and Networks (SIN), pp. 1-8. IEEE, 2024. DOI: 10.1109/SIN63213.2024.10871242
2. Lande D., Danyk Y. Competitive Artificial Intelligence in Information and Cyber Warfare. Available at SSRN: https://ssrn.com/ abstract=5084698, DOI: 10.2139/ssrn.5084698 (January 06, 2025). – 16 pp.
3. Yang, Yang, Bo Tian, Fei Yu, and Yuanhang He. An Anomaly Detection Model Training Method Based on LLM Knowledge Distillation. In 2024 International Conference on Networking and Network Applications (NaNA), pp. 472-477. IEEE, 2024. DOI: 10.1109/NaNA63151.2024.00084
4. GrantWilkins, Srinivasan Keshav, Richard Mortier. Hybrid Heterogeneous Clusters Can Lower the Energy Consumption of LLM Inference Workloads. e-Energy '24: Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems. Pages 506 – 513. DOI: 10.1145/3632775.3662830
5. Lande D., Strashnoy L. Resilient Artificial Intelligence Architecture. ResearchGate: https://www.researchgate.net/publication/389497329_Resilient_Artificial_Intelligence_Architecture, DOI: 10.13140/RG.2.2.36062.55360 (Mart 02, 2025). – 9 pp.
6. Lancaster Kelvin. The dynamic inefficiency of capitalism. Journal of Political Economy 81.5 (1973): 1092-1109. DOI: 10.1086/260108
7. Mancillas James, and Gregory L. Cantwell. Integrating Artificial Intelligence into Military Operations: A Boyd Cycle

Framework. Strategic Studies Institute, US Army War College (2017). URL: http://www.jstor.com/stable/resrep12117.11

8. Lande D., Strashnoy L. Semantic AI Framework for Prompt Engineering. ResearchGate: https://www.researchgate.net/publication/389692528_ Semantic_AI_Framework_for_ Prompt_Engineering. DOI: 10. 13140/ RG.2.2.21116.24964 (Mart 9, 2025). – 14 pp.

COMPARATIVE ANALYSIS OF THE TIME STABILITY OF CNN, LSTM AND DISTILBERT IN THE DOMAIN GENERATION ALGORITHMS (DGAS) DETECTION PROBLEM

Salyk Vasyl[1], Venherskyi Petro[1]

[1]*Ivan Franko National University of Lviv, Lviv, Ukraine*
vasyl.salyk@lnu.edu.ua, petro.venherskyy@lnu.edu.ua

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

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.

1. 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.

1.1. 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.

1.2. 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].

1.3. 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.

1.4. 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.

1.5. 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.

1.6. 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.

1.7. 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.

1.8. 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 paterns. 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)		

1.9. 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):

- Accuracy or total proportion of correct classifications.
- Completeness (Recall) – the ability of models to identify all relevant DGA samples (FN minimization).
- Precision is the proportion of true positives among all identified as DGA (FP minimization).
- F1-score is a harmonic average of completeness and accuracy, considering their balance.
- 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.

2. 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

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.

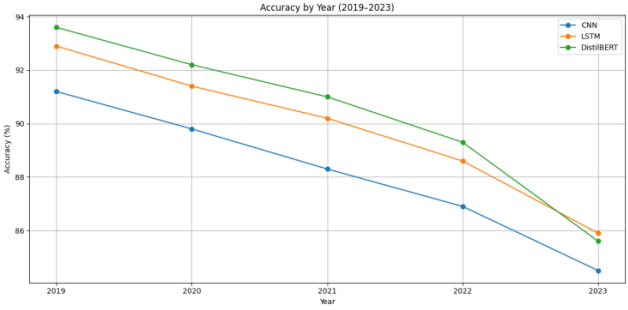


Figure 1. Accuracy by Year (2019-2023)

3. 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.

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], doi: 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], doi: 10.48550/arXiv.2402.13871

ВИЗНАЧЕННЯ ХАРАКТЕРИСТИК ПРИХОВАНИХ КІБЕРАТАК НА СИСТЕМИ КЕРУВАННЯ ІЗ ШТРАФОМ НА ПОМІТНІСТЬ

Новіков О. М.[1], Ільїн М. І. [1],
Стьопочкина І. В. [1], Дудуладенко В. М. [1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»*

Вступ

Сучасні автоматизовані системи керування об'єктами критичної інфраструктури (ICS, SCADA, DCS, PLC) дедалі частіше стають об'єктами цілеспрямованих кібератак [1–3]. Серед них особливу небезпеку становлять непомітні (приховані) атаки [4, 5] — атаки, які викликають суттєві відхилення в роботі системи, але залишаються невиявленими стандартними детекторами несправностей [6, 7].

У цій роботі розглядається підхід до побудови непомітної атаки на систему оптимального керування, де обмеження на детектованість враховується у вигляді штрафу на функціонал. На відміну від інших підходів, у цьому дослідженні штрафний множник β задається емпірично.

Мета дослідження

Метою роботи є дослідження властивостей та ефективності фіксованого штрафного підходу в задачі побудови шкідливого сигналу керування, який одночасно максимізує вплив атаки на динаміку системи та контролює детектованість через штрафний доданок у функціоналі.

Математична модель атаки

Розглядається лінійна динамічна система керування:

$$\frac{dx(t)}{dt} = Ax(t) + B[u(t) + u_a(t)], \qquad x(t_0) = x_0,$$
$$y(t) = Cx(t),$$

де $u_a(t)$ – сигнал атаки на керування, що підлягає оптимізації.

Для вирішення задачі пошуку оптимального $u_a(t)$ сформуємо функціонал:

$$L(u_a) = \Phi(u_a) - \beta J(u_a),$$

де $\Phi(u_a) = \int_{t_0}^{t_k} x^T W x dt \to \max\limits_{u_a \in U_{\text{доп}}}$ – функціонал впливу; $J(u_a) = \int_{t_0}^{t_k} [Cx - y]^T S^{-1} [Cx - y] dt$ – функціонал детектованості; β – штрафний множник.

Переваги та обмеження методу із штрафом на помітність

Перевагами методу є простота реалізації та роботи з оновлення множнику, гнучкість у налаштуванні компромісу між ефективністю атаки і її помітністю, прозорість для використання в задачах сценарного аналізу.

Недоліки методу

Потребує емпіричного добору значення β, який суттєво впливає на результат, може призводити до перевищення порогу детектованості або до зменшення ефективності атаки та не гарантує високий якості результату.

Результати обчислювального експерименту

Проведено чисельне моделювання на прикладі лінійної системи другого порядку. Досліджено вплив параметра β на поведінку функціоналів Φ, J та профілю атаки u_a. Встановлено, що за малих β атака дуже ефективна (Φ – велике), але J перевищує допустимий рівень. При великих β атака залишається непомітною, але її вплив значно зменшується. Існує компромісна зона, де досягається баланс між помітністю та ефектом.

Висновки

Підхід до побудови непомітних атак на основі штрафу на детектованість із фіксованим множником β дозволяє гнучко моделювати поведінку атакуючого. Метод простий у реалізації та корисний для тестування чутливості систем виявлення, однак не забезпечує гарантії дотримання обмеження на помітність.

Дослідження підтверджує необхідність використання адаптивних методів, де множник штрафу оновлюється автоматично та є підставою для переходу до подальших робіт з адаптивним soft-constrained підходом.

Список використаних джерел

1. Teixeira A., Shames I., Sandberg H., Johansson K.H. A secure control framework for resource-limited adversaries. Automatica, 2015, vol. 51, pp. 135–148.
2. Mo Y., Sinopoli B. Secure estimation in the presence of integrity attacks. IEEE Trans. Automat. Contr., 2015, vol. 60, no. 4, pp. 1145–1151.
3. Pasqualetti F., Dorfler F., Bullo F. Attack detection and identification in cyber-physical systems. IEEE Trans. Automat. Contr., 2013, vol. 58, no. 11, pp. 2715–2729.
4. Urbina D.I., Giraldo J.A., Cardenas A.A. et al. Limiting the impact of stealthy attacks on industrial control systems. Proc. ACM Conf. on Computer and Communications Security (CCS), 2016, pp. 1092–1105.
5. Новіков О., Ільїн М., Стьопочкіна І., Овчарук М. Визначення параметрів непомітних кібератак на системи керування об’єктів критичної інфраструктури // KPI Science News - No. 1, 2025, p. 69-75.
6. Chen J., Patton R. Fault diagnosis in dynamic systems: theory and application. Prentice Hall, 1999. 345 p.
7. Gertler J. Fault detection and diagnosis in engineering systems. CRC Press, 1998. 493 p.

СУЧАСНИЙ СТАН СТЕГАНОГРАФІЇ ЦИФРОВИХ ЗОБРАЖЕНЬ

Казміді І. Д.[1], Зубок В. Ю.[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»*

ivkaz-ipt22@lll.kpi.ua, vitaly.zubok@lll.kpi.ua

Стеганографія – це наука та мистецтво приховування інформації в цифрових носіях, причому зображення є одним із найбільш часто використовуваних носіїв. Було розроблено різні методи для вбудовування секретних повідомлень у зображення, зберігаючи їх сприйнятливу якість і не допускаючи виявлення. Ця галузь продовжує динамічно розвиватися, пропонуючи поліпшення старих методів стеганографії та розробку нових методів з урахування новітніх технологій. Представлено класифікацію та досліджено різноманітні методи стеганографії цифрових зображень. Розглянуто як класичні методи стеганографії зображень, так і нові тенденції та виклики в галузі. Висунуто припущення про можливі напрямки майбутніх досліджень цієї сфери.

Ключові слова: кібербезпека, стеганографія, зображення, квантові технології, штучний інтелект

Актуальність

Стеганографія – це наука та мистецтво прихованого передавання інформації, що дає змогу забезпечити конфіденційність і цілісність комунікацій, використовуючи методи, невидимі для людини. Одним із найбільш популярних напрямків стеганографії є приховання даних у цифрових зображеннях, оскільки зображення є одним з найбільш поширених типів файлів у цифрових комунікаціях. Зростаючі загрози кібератак, необхідність збереження конфіденційності даних та потреба в захисті особистої та корпоративної інформації ставлять

стеганографію на передній план у галузі кібербезпеки. Проблема захисту інформації стає дедалі актуальнішою, оскільки традиційні методи криптографії часто не гарантують абсолютну невидимість перед зловмисниками, тоді як стеганографія дозволяє приховати дані так, що їх важко виявити навіть при ретельному аналізі файлів. Класифікацію методів стеганографії цифрових зображень варто почати з розділення за типами зображень. Розглянуто растрові та векторні зображення. Класифікацію стеганографічних методів ускладнює можливість їх суміщення чи заміни певних використаних алгоритмів. В своїх роботах, через відсутність узгодженої системи оцінки та характеристики методів, автори проводять часткову характеристику, використовуючи не всі доступні критерії, що ускладнює порівняння стеганографічних алгоритмів. Методи стеганографії растрових зображень можуть бути класифіковані за наступними принципами:

- класифікація за вибором частини зображення для ролі контейнера, який несе в собі приховану інформацію;
- класифікація за перетворенням зображення перед вбудовуванням інформації

Прикладами першого критерію можуть бути випадковий вибір пікселів для зміни та метод альфа-каналу [5], який вбудовує інформацію саме в пікселі сумісного шару зображення, який відповідає за відображення прозорих частин, або метод множин Жуліана [6], в якому зображення-контейнер генерується на основі відповідної математичної множини Жуліана і являє собою фрактальне зображення необхідного розміру. Параметри множини Жуліана при цьому може бути використана у вигляді ключа доступу для вилучення прихованої інформації. Другий критерій визначає більш комплексні методи, які важче піддаються статичному аналізу. Методи цього критерію, перед вбудовуванням інформації, проводять певні математичні перетворення та використовують особливості відповідних математичних функцій для приховування інформації. Прикладом виступають дискретні косинусоїдне

[7] та хвильове [8] перетворення. Вбудовування інформації відбувається не в саме зображення, в результат обробки зображення зазначеними перетвореннями.

Така класифікація зумовлена тим, що більшість методів тим чи іншим чином змінює параметри кольору пікселів зображення. Зазначена класифікація дозволяє розділяти та класифікувати методи між один-одним. Методи стеганографії векторних зображень, здебільшого [9] розділяються саме за параметрами, у які відбувається вбудовування інформації:

- просторові методи, які використовують абсолютні параметри точок об'єктів у якості контейнерів для приховування інформації;
- частотні методи використовують математичні перетворення для представлення об'єктів зображення у іншій математичній формі, в яку відбувається вбудовування інформації.

Серед просторових методів прикладом є метод Хоувена-Лі [10] та метод розділення прямих [11]. Обидва методи напряму взаємодіють з параметрами об'єктів на зображенні. Різниця між методами полягає у методиці пошуку об'єктів зі зручними для вбудовування інформації параметрами. Прикладом функцій для частотних методів є дискретне перетворення Фур'є [12] та вейвлет-перетворення [13]. Зазначені функції виконують перетворення зображень в декілька різних множин, груп вейвлет коефіцієнтів, в які може бути прихована інформація, після чого зображень збирається назад. Окрім класифікації вже відомих стеганографічних методів, варто звернути увагу на сучасні методи стеганографії зображень постійно вдосконалюються [1,2], особливо разом з розвитком технологій штучних мереж та квантової інформатики. Тому активно розробляються нові підходи, що дозволяють зберігати якість зображень і зменшувати зміни, викликані вставленими даними, що є важливим для забезпечення невидимості прихованої інформації. Крім того, необхідність забезпечення цілісності і аутентифікації зображень, що містять приховані дані, підштовхує розвиток

нових методів для захисту інформації від маніпуляцій. Стеганографія зображень має широкий спектр застосувань, від захисту конфіденційних комунікацій до використання в криптографії, цифрових водяних знаках, які, в свою чергу, слугують захистом від витоку або підробки інформації. Ці технології дозволяють додатково захистити дані, зробивши їх менш вразливими до атак і несанкціонованого доступу. Попри значний прогрес у галузі, існують серйозні виклики, зокрема зростаюча здатність алгоритмів аналізу та виявлення прихованої інформації. Це вимагає розробки нових, більш стійких методів стеганографії, які можуть інтегруватися з іншими технологіями, такими як машинне навчання та штучний інтелект [3], для підвищення ефективності захисту. З огляду на зростаючі загрози кібератак і витоків даних, стеганографія зображень набуває все більшої актуальності як інструмент для забезпечення безпеки комунікацій і захисту конфіденційної інформації. Одночасно з необхідністю поліпшення існуючих стеганографічних методів, розглядається використання стеганографії у нових технологіях, таких як квантова інформатика [4]. Наукові дослідження в цій галузі сприяють не лише розвитку нових методів захисту, але й підвищують рівень безпеки в сучасному цифровому середовищі.

Метод дослідження

Аналіз літератури проводиться з метою вивчення актуальних та потенційних проблем, що виникають у стеганографії зображень, шляхом дослідження існуючих методів цієї технології та наукових джерел. Для досягнення цієї мети здійснюється пошук в авторитетних наукових базах даних, таких як IEEExplore, ResearchGate, Arxiv, ScienceDirect, Scopus, Google Scholar, CrossRef, Thesai та Semantic Scholar, з акцентом на публікації останніх п’яти років, хоча охоплюється період до 15 років. Пріоритет надається статтям та матеріалам конференцій, які забезпечують точне та стисле представлення основних ідей. Вибір матеріалів орієнтований на англомовні публікації, що розглядають як традиційні, так і новітні методи

стеганографії, зокрема їх модифікації та удосконалення характеристик. Це дозволяє зосередитись на розвитку методів стеганографії та визначити потенційні напрямки для подальших досліджень.

Джерела систематизуються за двома основними критеріями: типами зображень, у яких ці методи застосовуються, і методологією вбудовування інформації в зображення. Аналіз фокусується на практичних аспектах стеганографії, з мінімальним зануренням у теоретичні аспекти конкретних технік, при цьому підкреслюється загальний розвиток напрямків у цій галузі. Приділяється увага методам оцінки методів, які використовуються авторами. Така класифікація демонструє різницю в підходах до стеганографії для різних типів зображень, а також еволюцію методів в межах одного типу зображень. Використання зазначених баз даних забезпечує високий рівень якості та достовірності отриманих матеріалів, а також швидкий доступ до новітніх наукових досягнень. Орієнтація на англомовні публікації дозволяє охопити найбільш актуальні та інноваційні дослідження, оскільки більшість передових розробок публікуються англійською мовою. Групування матеріалів за типами зображень та методологією сприяє структурованому підходу до аналізу, що допомагає виявити загальні тенденції та перспективи розвитку стеганографії зображень.

Висновки

Існують значні перспективи для подальших досліджень у сфері стеганографії зображень, зокрема в контексті розвитку нових технологій штучного інтелекту та квантової інформатики. Класичні методи стеганографії, які використовуються для різних форматів зображень, мають ряд недоліків, таких як обмежена стійкість до перетворень зображень і обмеження щодо кількості даних, які можна приховати. Окрім того, при масовому використанні ці алгоритми можуть демонструвати низьку швидкодію під час вбудовування та вилучення інформації. Враховуючи ці обмеження, існує потенціал для розширення можливостей існуючих методів, зокрема через вдосконалення їх

ефективності та підвищення стійкості до зовнішніх впливів. Це відкриває нові напрямки для майбутніх досліджень, що дозволить значно розширити застосування стеганографії зображень у різних галузях людської діяльності.

Список використаних джерел

1. Al-Obaidi SAR, Lighvan MZ, Asadpour M. Enhanced image steganalysis through reinforcement learning and generative adversarial networks. Intelligent Decision Technologies. 2024;18(2):1077-1100. doi:10.3233/IDT-240075

2. Al-Iedane, Hussein Ali and Mahameed, Ibrahim Ans (2023), Applying and Evaluating Machine Learning Models for the Detection of Digital Image Steganography, Doi: 10.2139/ssrn.4427951

3. Volkhonskiy D., Nazarov I., & Burnaev, E. (2020, January). Steganographic generative adversarial networks. In Twelfth international conference on machine vision (ICMV 2019) (Vol. 11433, pp. 991-1005). SPIE

4. Mayer, J. (Ed.). (2024). Steganography – The Art of Hiding Information. IntechOpen. doi: 10.5772/intechopen.1001493

5. Nichal, Arjun & Jadhav, Mr.Aniket & Pingale, Mr & Mohite, Mr.Chaitanya & Ponde, Mr. (2015). A Novel Steganography Scheme via the use of Alpha channel. IJIREEICE. 3. 18-21. 10.17148/IJIREEICE.2015.3404.

6. Nori, A.S. and Al-Qassab, A.M., 2014. Steganographic technique using fractal image. International Journal of Information Technology and Business Management, 23(1).

7. Khalaf, Ashraf A. M. & Fouad, Osama & Hussein, Aziza & Hamed, Hesham & Kelash, Hamdy & Ali, Hanafy. (2019). Hiding data in images using DCT steganography techniques with compression algorithms. TELKOMNIKA (Telecommunication Computing Electronics and Control).17.10.12928/telkomnika.v17i3.

8. Della Baby, Jitha Thomas, Gisny Augustine, Elsa George, Neenu Rosia Michael, A Novel DWT Based Image Securing Method Using Steganography, Procedia Computer Science, Volume 46,2015,Pages 612-618,ISSN 1877-0509,https://doi.org/10.1016/j.procs.2015.02.105.

9. Kinzeryavyy O. (2015), Steganographic methods for hide data in vector images resistant to active attacks based on affinian transformations [Doctoral dissertation, National University "Kyiv Aviation Institute"], erNAU - Electronic Institutional Repository of the National Aviation University of Ukraine.

10. Haowen Yan & Jonathan Li. (2011). A Blind Watermarking Approach to Protecting Geospatial Data from Piracy. International Journal of Information and Education Technology. 94-98. 10.7763/IJIET.2011.V1.16.

11. Sonnet, Henry & Isenberg, T. & Dittmann, J. & Strothotte, Thomas. (2003). Illustration watermarks for vector graphics. 73- 82. 10.1109/PCCGA.2003.1238249.

12. Solachidis V., Nikolaidis N. and Pitas I. Fourier descriptors watermarking of vector graphics images. *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, Vancouver, BC, Canada, 2000, pp. 9-12 vol.3, doi: 10.1109/ICIP.2000.899265.

13. Li Y. and Xu L. A Blind Watermarking of Vector Graphics Images," in Computational Intelligence and Multimedia Applications, International Conference on, Xian, China, 2003, pp. 424, doi: 10.1109/ICCIMA.2003.1238163.

ON NEW MESSAGE AUTHENTICATION CODES GENERATING SENSITIVE DIGESTS OF DIGITAL DOCUMENTS

Pustovit O.[1], Ustimenko V.[1]
Global Information Space of the NAS of Ukraine, Kyiv, sanyk_set@ukr.net, vasylustimenko@yahoo.pl

Specialists must use well checked documents to elaborate well founded decisions and plans. Check tools include cryptographically stable algorithms for compressing a large file into a digest of a specified size, sensitive to any change in the characters on the input. New fast compression algorithms are proposed, whose cryptographic stability is associated with complex algebraic problems, such as the study of systems of algebraic equations of large power and the problem of the expansion of nonlinear mapping of space by generators. The proposed algorithms for creation of change-sensitive digests will be used to detect cyberattacks and audit all system files after a registered intervention.

Keywords: cybersecurity, hash functions, message authentication codes, compression homomorphism, highly nonlinear cryptography from many variables, non-commutative cryptography.

1. On the verification of electronic documents

A simplified model of the global information space can be imagined as a large, time-growing network of registered virtual users (individuals or institutions) who exchange information and can store it in electronic repositories located in the network or isolated from it. The size of files for exchange (electronic documents) tends to increase. An important category of the information space is the trust in documents. Users can use a symmetric algorithm with a private key to encrypt documents and a key exchange protocol to maintain the security of the encoding procedure. Certified public key algorithms can also be used to change the key. These methods ensure the security of the exchange channels. It is easy to see that even the use of

reliable encryption tools does not ensure complete trust in documents. The above justifies the importance and relevance of the following questions. Has the original document already been damaged? To answer these questions, hash function generation technologies are used. A hash function is needed to generate a compensated form of the original document. Note that the general hash function does not require a key or password.

For document verification and authentication tasks, so-called key hash functions (key hash functions, message authentication codes, or MACs) are required that are password-dependent [1].

In the talk we propose new fast algorithms for creating digests of electronic files to detect cyberattacks, computer viruses or other corruptions and to verify data integrity. The cryptographic stability of new key-dependent hash functions is related to complex algebraic problems such as the study of systems of high-degree algebraic equations and the navigation problem of decomposing a nonlinear transformation into a composition of known generators. Our digest generation algorithms use the representation of files as sequences (words) of elements of a finite commutative ring K such as a finite field or arithmetic ring of residues modulo 2^m. Such words can be considered as elements of a free semigroup or its modification with the use of the ring addition operation. Representing a regular tree (or other infinite graph) as a projective boundary of finite graphs given by equations allows us to determine the "compression homomorphism" of the semigroup into the group of polynomial transformations of the affine space K^t (the space of digests).

2. The mathematical basis of the proposed hash functions

Let $F(K)$ be the space of potentially infinite texts in the alphabet K, which represents the set of all tuples of the form $(a_1, a_2, \ldots, a_k), a_i \in K$ of different lengths k.

Let us assume, that K is a finite commutative ring and identify $F(K)$ with the semigroup with the following multiplication

$(a_1, a_2, ..., a_k) \circ (b_1, b_2, ..., b_s) = (a_1, a_2, ..., a_k, b_1 + a_k, b_2 + a_k, b_s + a_{k)}$.

Let $F'(K)$ be a subsemigroup of all words of even length. We denote as $S(K^n)$ the finite semigroup of all polynomial transformations of the affine space K^n into itself. Our algorithm is based on the following mathematical statement.

Theorem. ([2]) *For each positive integer m≥2 there is a homomorphic map* $\psi : F'(K) \to S(K^m)$ *such that the image* $\psi(F'(K))$ *form the group G of polynomial transformation.*

Recall that the property to be homomorphism for $\psi = \psi_m$ is given by the relation $\psi(a \circ b) = \psi(a) \circ \psi(b)$.

The mapping satisfying the conditions of the theorem is constructed effectively in terms of the theory of discrete dynamical systems defined by algebraic graphs with extremal properties (see [3] and further references). Other ideas to use algebraic graphs for the construction of digests of electronic documents can be found in [4].

To create MACs [2] not the group G itself was used, but the mapping that defines it, together with the affine transformations A and B of the Cremona group with the rule $g : x \to A\psi(x)B$. In the talk a faster algorithm will be presented that uses the theorem not only in the case of a finite ring K but also in the case of an expansions of K of kind $K[x_1, x_2, ..., x_t]$.

Computer simulation has allowed us to calculate a very high avalanche effect in the range of 98-99 %. For example, in MAC [5] the avalanche effect interval is estimated as 47-50 %.

References

1. Aumasson J.Ph, Serious Cryptography: A Practical Introduction to Modern Encryption, No Starch Press. – 2017. – 312 pp.

2. Pustovit O. S., Ustimenko V. O. On new streaming algorithms for creating sensitive digests of electronic documents, Mathematical modeling in economy, No. 3 (16), July-September 2019, 18-35.

3. Chojecki T., Erskine G., Tuite J., Ustimenko V. On affine forestry over integral domains and families of deep Jordan–Gauss graphs. European Journal of Mathematics, 2025, V.11, No 10.

4. Polak M., Zhupa E., Keyed hash function from large girth expander graphs, Albanian J. Math., 2022, V.16, No 1, 25-39.

5. Krendelev S., Sazonova P., Parametric Hash Function Resistant to Attack by Quantum Computer, Based on Problem of Solving a System of Polynomial Equations in Integers, Proceedings of the 2018 Federated Conference on Computer Science and Information Systems, ACSIS. – Vol. 15. –pp. 387 – 390 (2018).

ФОРМУВАННЯ ТЕОРЕТИКО-ГРАФОВОЇ МОДЕЛІ ІНФОРМАЦІЙНО-ПСИХОЛОГІЧНОЇ МЕНТАЛЬНОЇ ВІЙНИ

Качинський А. Б.[1], Ланде Д. В.[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»*
dwlande@gmail.com, akachynsky@gmail.com

У роботі запропоновано розширену мережеву модель інформаційно-психологічних ментальних війн, яка базується на формалізації через ієрархію з п'яти рівнів та подальшому перетворенні її на мережу з використанням великих мовних моделей (LLM). Мета дослідження – удосконалити існуючу ієрархічну модель шляхом автоматичного розширення понятійної бази, виявлення нових семантичних зв'язків та їх кластеризації для глибшого аналізу структури ментального впливу. За допомогою LLM, отримано мережу з понад 300 ключових понять, які описують цілі, акторів, засоби, політики та результати ментальної війни. Проаналізовано зв'язки між елементами, сформовано граф мережі, який було проаналізовано і візуалізовано у системі Gephi. Результати показали, що ментальні війни формуються через комплекс взаємодіючих факторів, серед яких домінують символічні, емоційні та когнітивні механізми. Отримана модель може служити основою для подальшого дослідження інструментів психологічного впливу, акторів інформаційної дії, а також способів протидії таким формам війни.

Ключові слова: інформаційно-психологічна ментальна війна, розширена мережева модель, штучний інтелект, великі мовні моделі

Вступ

Мета цієї роботи – розширити й удосконалити модель інформаційно-психологічних ментальних війн – сучасного

інформаційно-психологічного протистояння за допомогою великих мовних моделей (LLMs).

Останніми роками штучний інтелект (ШІ), зокрема великі мовні моделі [1-3] відкрив нові можливості для аналізу та моделювання семантичних мереж. Використання ШІ не лише дозволяє вдосконалювати наявні моделі, а й виявляти нові аспекти, які раніше залишалися поза увагою дослідників [4].

Спочатку модель «інформаційно-психологічної ментальної війна» розглядається як комплексна мережа з ієрархічною структурою [5].

Нехай $(H, \leq)$ – скінченна частково впорядкована множина з найбільшим елементом b. Множина H називається ієрархією, якщо виконуються наступні умови:

1. Існує розбиття множини H на шари (рівні ієрархії):

$$H = L_1 \cup L_2 \cup ... \cup L_h,$$

де $L_1 = \{b\}$ – перший рівень ієрархії містить лише найбільший елемент; $L_k \cap L_j = \varnothing$ для всіх $k \neq j$ – ці підмножини не перетинаються.

Кожен елемент з H належить рівно одному з L_k. Ці множини L_k називають шарами або рівнями ієрархії.

2. Умова сумісності з порядком:

Якщо $x < y$ і $y \in L_k$, то $x \in L_m$, де $m > k$.

Інакше кажучи, якщо один елемент менший за інший, то він повинен знаходитись на нижчому рівні ієрархії.

3. Умова послідовності (спадного/зростаючого переходу):

Для кожного елемента $x \in H$:

Якщо x^- – попередник x (тобто $x^- < x$, і немає жодного елемента z такого, що $x^- < z < x$), то $x^- \in L_{k+1}$, якщо $x \in L_k$.

Аналогічно, якщо x^+ – наступник x (тобто $x < x^+$, і немає елемента z, для якого $x < z < x^+$), то $x^+ \in L_{k-1}$, якщо $x \in L_k$.

Тобто кожен попередник лежить на наступному рівні, а кожен наступник – на попередньому рівні.

При такому розгляді проблему можна розкласти на простіші компоненти, після чого оцінити відносний ступінь взаємодії між елементами цієї ієрархічної структури.

У комплексній мережі «інформаційно-психологічної ментальної війни» можна виділити п'ять рівнів L_i:

L_1 – цілі ментальної війни;

L_2 – сили та засоби ментальної війни;

L_3 – актори ментальної війни;

L_4 – цілі акторів;

L_5 – політики, що реалізуються акторами.

Математична модель допомагає формалізувати взаємодії між підсистемами за допомогою графів, матриць зв'язків та цільових функцій, які описують загальну ефективність системи.

Рівні L_1 та L_4 ієрархії включають певні набори цілей і підцілей. Нехай на рівні i є набір цілей $T_i = \{T_{i1}, T_{i2}, \ldots, T_{imi}\}$, де $i = \{1, 4\}$, а mi – кількість цілей на рівні i. Кожній цілі T_{ij} відповідає набір підцілей $F_{ij} = \{F_{ij1}, F_{ij2}, \ldots, F_{ijn}\}$, де ijn – кількість підцілей для цілі T_{ij}.

Елементи та підцілі на кожному рівні виконують конкретні функції (інформаційні, економічні, організаційні тощо) і можуть бути пов'язані з іншими елементами та підцілями F_{ijk} на тому ж або інших рівнях, утворюючи мережу взаємозв'язків. Це описується набором функціональних зв'язків між елементами та підцілями:

$$Links = \{(F_{ijk}, F_{i'j'k'}) \mid F_{ijk} \text{ is linked to } F_{i'j'k'}\}.$$

Набір функціональних зв'язків між елементами, цілями та підцілями формує граф, у якому вершинами є елементи, підцілі та цілі F_{ijk}, а ребрами – зв'язки між ними.

Для моделювання процесу досягнення головних цілей на кожному рівні можна визначити цільову функцію

системи «ментальної війни» Φ, яка залежить від виконання цілей і підцілей на різних рівнях:

$$\Phi = \alpha\, f(T_{ij}) + \beta\, (F_{ijk}),$$

де α, β – вагові коефіцієнти у межах інтервалу зі значеннями в діапазоні [0,1].

Розширення ієрархічної моделі до мережевої

У ієрархії ментальних воєн підцілі та концепти на кожному рівні визначаються експертами, у тому числі віртуальними [3], які допомагають формувати систему цілей. Якщо концепти можуть належати до кількох рівнів одночасно, ієрархія перетворюється на мережу, що ускладнює зв'язки між рівнями.

Якщо концепт f_i належить до кількох рівнів одночасно, це вводить нові зв'язки між рівнями. Таким чином, структура системи змінюється з чистої ієрархії на мережеву модель.. Якщо поняття f_i належить рівням T_a та T_b, то між цими рівнями існує зв'язок:

$$E = \left\{ (T_a, T_b) \mid f_i \in F,\, f_i \in T_a \cap T_b \right\}.$$

Ця формалізація відповідає послідовності дій, що передбачає можливість повторення процесу під наглядом експерта до досягнення остаточного розуміння предметної області:

1. Формування початкової схеми із відображенням базових зв'язків між поняттями.
2. Створення промптів для LLM з метою генерації нових понять і зв'язків.
3. Інтеграція нових зв'язків у початкову схему.
4. Обробка лінгвістичних даних – оцінювання нових та існуючих зв'язків, а також ранжування вузлів.
5. Аналіз даних і візуалізація шляхом завантаження даних у систему аналізу графів.
6. Формування та уточнення кластерів, визначення їхніх назв за допомогою LLM.
7. Фінальна перевірка та валідація розширеної моделі.

Аналіз і візуалізація

Об'єднані дані завантажуються в середовище програми Gephi для кластеризації на основі класів модулярності (рис. 1).

Серед понад 300 вузлів розширеної мережі отримані такі важливі поняття, що є суттєвими для розуміння суті інформаційно-психологічних ментальних війн:

- політика мовного контролю;
- маніпуляція страхом;
- символіка влади;
- формування нової свідомості;
- когнітивна перебудова;
- соціальна напруженість;
- критика інституцій;
- емоційна маніпуляція.

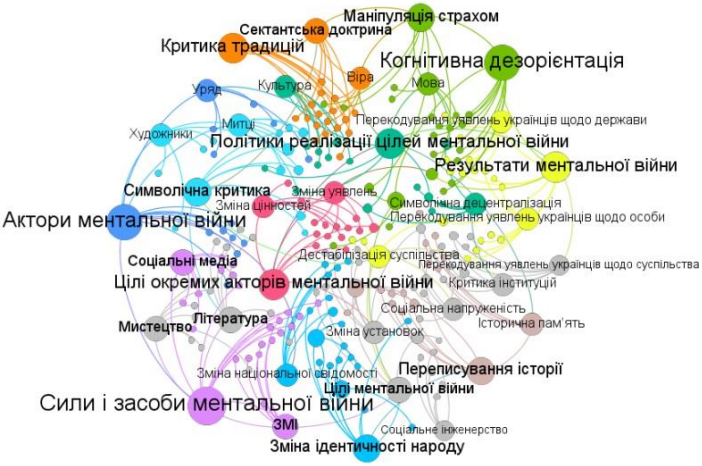


Рисунок 1. Розширена мережева модель ментальних війн

Висновки

Отримана розширена модель інформаційно-психологічних ментальних війн демонструє складну мережу зв'язків між ключовими та додатковими концептами, яка суттєво доповнює традиційне розуміння цієї теми. Застосування великих мовних моделей дозволило не лише автоматизувати процес формування понятійного апарату, а й виявити приховані семантичні залежності, важливі для стратегічного аналізу. Ієрархічна модель, запропонована на початковому етапі, недостатня для опису реальних взаємозв'язків. Перетворення її на мережеву модель забезпечило гнучкість та адаптивність до багаторівневого сприйняття, коли окремі поняття можуть одночасно належати кільком шарам системи, утворюючи нетривіальні зв'язки.

Кластеризація та ранжирування вузлів у рамках графової моделі надали можливість виокремити найбільш значущі концепти, які впливають на структуру ментальної війни. Серед них – політика мовного контролю, маніпуляція страхом, когнітивна перебудова, соціальна напруженість та емоційна маніпуляція. Ці елементи є центральними в механізмах символічного, психологічного та соціального впливу.

Таким чином, отримана мережева модель ментальних війн є динамічною, розширюваною та повторно використовуваною, що забезпечує її придатність як наукового інструменту для подальших досліджень в області інформаційно-психологічного впливу, протидії ментальним агресивним стратегіям, а також аналізу акторів та їхньої ролі в медійному просторі.

Список використаних джерел

1. Lande D., Strashnoy L. *GPT Semantic Networking: A Dream of the Semantic Web - The Time is Now*. Kyiv: Engineering, 2023. ISBN 978-966-2344-94-3.

2. Srinivasan Ramanujam. *The LLM Revolution: Transforming Industries with Large Language Models.* Kindle

Edition, 2024. https://www.amazon.com/ LLM-Revolution-Transforming-Industries-Language-ebook/dp/B0D7VS5BNK

3. Lande D., Strashnoy L. *Causality Network Formation with ChatGPT*. SSRN Electronic Journal. (May 30, 2023). – 16 p. DOI: 10.2139/ssrn.4464477

4. Lande Dmytro. OSINT in Cybersecurity: Educational Manual. – Kyiv: LLC "Engineering", 2024. – 522 p.

5. Kachynskyi A. The structural-functional model of the system for ensuring informational and informational-psychological security. Reports of the National Academy of Sciences of Ukraine, 2023. №1. pp. 16-23. (ukr. lang)

6. Wu F. Y. *The Potts model*. Rev. Mod. Phys. 54, 235 – Published 1 January 1982

IMPROVED BOUNDARY SAMPLING TECHNIQUES FOR SMOOTHED PARTICLE HYDRODYNAMICS WITH RIGID BODY COUPLING

Hrytsyshyn O.[1], Trushevsky V.[1]
[1] *Ivan Franko National University of Lviv, Lviv, Ukraine*

This paper presents an efficient method for handling fluid–rigid interactions in Smoothed Particle Hydrodynamics (SPH) using single-layer boundary sampling with volume-corrected mass. While traditional multi-layer boundary approaches ensure accurate force computation, they introduce significant computational overhead and complexity. We demonstrate that a single boundary layer, combined with a simple mass correction based on particle volume, is sufficient to preserve accuracy and stability even in thin or complex rigid geometries. Our results show a substantial reduction in interaction counts with minimal implementation effort, making the approach well-suited for real-time or large-scale SPH simulations.

This research is conducted in the context of protecting critical infrastructure, where accurate and efficient modeling of fluid–rigid interactions is crucial. The proposed method is particularly suitable for such scenarios, as it allows realistic simulation of how water masses interact with complex rigid structures (e.g., dams, barriers, or urban environments) without incurring high computational costs. By applying this technique, it becomes possible to predict the behavior of water in emergency or high-risk situations and to design more effective strategies for mitigating flooding and safeguarding essential infrastructure.

Keywords: SPH, fluid–rigid coupling, boundary handling, particle deficiency, volume correction, single-layer sampling, real-time simulation, critical infrastructure protection, flood mitigation, computational fluid dynamics

Introduction

Smoothed Particle Hydrodynamics (SPH) is a particle-based method for simulating fluids, where each particle represents a small volume of fluid and interacts with its neighbors through smoothing functions. This mesh-free approach is well-suited for complex, dynamic flows and free-surface phenomena.

The simulation becomes even more compelling when rigid bodies are introduced into the fluid. Interactions between particles and solid surfaces can produce highly realistic motion, splashing, and pressure effects. However, achieving stable two-way coupling between fluids and solids presents unique challenges. Issues like missing particles near boundaries can cause discontinuities or unnatural sticking and handling thin or constrained geometries requires careful design. These challenges make the interaction between SPH and rigid bodies a rich area for research and innovation.

1. SPH Basics

Smoothed Particle Hydrodynamics (SPH), originally proposed by Gingold & Monaghan (1977) [1], is approximating field quantities using discrete particles. A field variable A at position x_i, is estimated by summing contributions from neighboring particles x_j, where V_j is the volume represented at x_j within a smoothing radius h, using a kernel function W, as:

$$A(x_i) = \sum_j V_j A_j W(x_i - x_j, h) \tag{1}$$

One of the fundamental quantities in SPH is density. The computation of density is a specific case of this interpolation, where $A_j = \rho_j$. It can be computed in two equivalent ways, by:

$$\rho_i = \sum_j V_j \rho_j W_{ij} = \sum_j m_j W_{ij} \tag{2}$$

where m_j is the mass of particle j, and $W_{ij} = W(x_i - x_j, h)$ is the smoothing kernel.

These density estimates are used to compute pressure forces, typically using WCSPH [2] or PCISPH [3] approaches. Pressure and viscosity forces often follow the formulation introduced by Monaghan. Additional models are employed for simulating surface tension and multiphase fluids.

2. Boundary Handling

In SPH simulations, accurately representing and enforcing boundary conditions is critical for physical realism and numerical stability. One common approach is to use particle-based boundaries [4], where the geometry is sampled with additional particles equipped with kernel functions. This ensures consistency with fluid discretization but introduces challenges: dense sampling increases computational cost, while sparse sampling—though computationally cheaper—can lead to particle deficiency, causing inaccurate force computations, particle penetration, or noisy trajectories near boundaries, which can be seen from equation (1). Alternatively, implicit boundary methods use functions like signed distance fields to define geometry independently of particle size [5]. These allow for more flexible and memory-efficient representations but are harder to couple directly with SPH solvers. Each method has trade-offs in terms of accuracy, cost, and integration complexity.

2.1. Multi-Layer Boundary Sampling

To accurately compute fluid particle density near solid boundaries, contributions from both fluid neighbors x_{i_f} and boundary neighbors x_{i_b} must be included. The extended density summation formula derived from (2) is:

$$\rho_i = \sum_{i_f} m_{i_f} W_{ii_f} + \sum_{i_b} m_{i_b} W_{ii_b} \tag{3}$$

o ensure full kernel support for fluid particles near boundaries, especially when using large smoothing radius, multiple layers of boundary particles are required. While this guarantees accurate density estimates and avoids particle deficiency, it introduces trade-offs: denser sampling increases memory and compute cost, while sparse layers risk missing contributions, leading to unstable pressure forces.

2.2. Single-Layer Boundary Sampling

Single-layer sampling is an efficient and straightforward method to represent solid boundaries in SPH. It requires placing only one layer of boundary particles along surfaces, which greatly reduces computational cost and simplifies preprocessing compared to multi-layer approaches. However, this simplicity comes at a cost: fluid particles near the boundary may suffer from particle deficiency, where the kernel support region lacks enough neighboring boundary particles to accurately compute density or pressure forces.

To address this, each boundary particle is assigned a volume that accounts for the missing neighbors. This volume is estimated as:

$$V_{i_b} = \frac{m_{i_b}}{\rho_{i_b}} = \frac{m_{i_b}}{\sum_{j_b} m_{j_b} W_{i_b j_b}} \tag{4}$$

Assuming equal masses, this simplifies to:

$$V_{i_b} = \frac{1}{\sum_{j_b} W_{i_b j_b}} \tag{5}$$

Using this, we define a corrected mass:

$$\tilde{m}_{i_b} = \frac{\rho_0}{\sum_{j_b} W_{i_b j_b}} \tag{6}$$

which replaces the standard mass in SPH summations. The modified term ensures that even a single layer of boundary particles contributes an appropriate amount to the density and pressure estimates, compensating for the missing volume and improving physical realism.

This makes single-layer sampling not only computationally efficient, but also robust enough for accurate SPH fluid–solid interaction.

3. Fluid-Rigid Coupling

In SPH simulations, fluid–rigid coupling enables two-way interaction between fluids and solid objects. This means that the fluid exerts pressure on the rigid body, and the body responds with reactive forces that influence the surrounding fluid. Several strategies exist, ranging from impulse-based methods to pressure-based force formulations.

More physically accurate approaches use hydrodynamic pressure forces to model the interaction at the interface. These methods allow for the transfer of momentum between fluid and rigid particles but often involve additional computational complexity, such as multiple neighbor searches or correction steps to ensure non-penetration and stability. Handling complex contact scenarios, such as simultaneous interactions with multiple rigid bodies, remains a challenge.

Fluid-Rigid Pressure Force

The interaction between a fluid particle i_f and a boundary **particle** i_b is governed by the pressure force acting across their shared interface. In the SPH framework, the pressure force applied to the fluid particle by the boundary is computed as [6]:

$$F^p_{i_f \leftarrow i_b} = -m_{i_f} \tilde{m}_{i_b} \frac{p_{i_f}}{\rho_{i_f} \rho_{i_b}} \nabla W_{i_f j_b} \tag{7}$$

Here, $\tilde{m}_{i_b}$ is a corrected boundary mass derived from the boundary particle's volume to account for particle deficiency. This volume-based correction ensures that even a single layer of boundary particles contributes sufficiently to fluid pressure computations.

The boundary particle receives an equal and opposite reaction force:

$$F^p_{i_b \leftarrow i_f} = -F^p_{i_f \leftarrow i_b} \tag{8}$$

This symmetrical formulation supports stable and realistic two-way coupling, even when using minimal boundary sampling. It provides a practical and efficient way to simulate fluid–solid interactions in SPH without requiring multiple boundary layers or explicit velocity matching.

4. Results

To evaluate the impact of boundary sampling strategies in SPH, we compared single-layer and multi-layer configurations in terms of the total number of fluid–boundary interactions over 24 simulation steps. The simulation used 20,000 fluid particles. The multi-layer case included 14,272 boundary particles, while the single-layer setup used only 3,032.

The total number of interactions per frame was consistently lower for the single-layer setup, as shown in Figure 1. On average, the multi-layer configuration produced approximately 18,032 interactions per step, while the single-layer case resulted in about 13,958. This is an average reduction of around 4,074 interactions per step, or roughly 22.6 % less. These numbers are highly dependent on the number, placement, and geometry of rigid bodies in the scene.

When normalized per fluid particle, the interaction rate was approximately 0.90 in the multi-layer case and 0.70 in the single-layer case.

However, when normalized per boundary particle, single-layer sampling was significantly more efficient—each boundary particle contributed to over 4.6 interactions on average, compared to just 1.26 in the multi-layer configuration. This highlights that multi-layer setups tend to oversample, resulting in redundant or unnecessary computations.

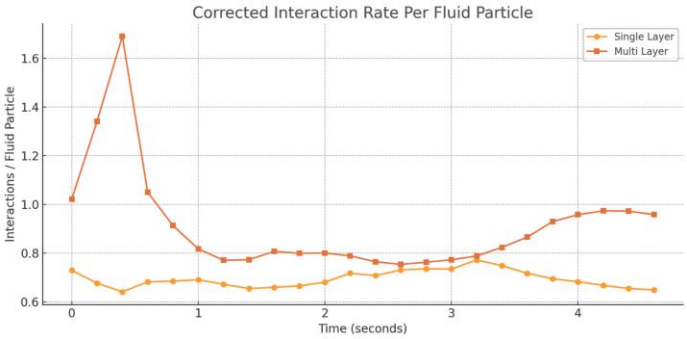


Figure 1. Corrected Interaction Rate Per Fluid Particle

Without using volume-corrected mass for boundary particles, single-layer setups suffer from particle deficiency, especially when modeling lower-dimensional geometries like thin plates, shells, or rods. In such cases, the lack of neighbors within the smoothing radius leads to inaccurate density estimation and broken pressure forces. However, by applying volume correction and computing a modified boundary mass,

the single-layer approach fully compensates for this deficiency. As a result, it supports robust and accurate pressure interaction even with challenging geometries, allowing for simpler, cheaper, and more general boundary representations.

Conclusions

This work explored different strategies for boundary handling and fluid–rigid coupling in Smoothed Particle Hydrodynamics (SPH), with a focus on comparing traditional multi-layer boundary sampling to a corrected single-layer approach. We presented a volume-based correction method that enables accurate force computation even when using only a single layer of boundary particles.

Our results demonstrate that the single-layer approach, when combined with a corrected mass formulation, achieves substantial reductions in computational cost – reducing interaction counts by over 20 % on average – while maintaining accuracy and stability in two-way fluid–rigid coupling. It also achieves significantly higher efficiency per boundary particle, making it a better fit for large-scale or real-time applications.

Crucially, the corrected mass method resolves the particle deficiency problem that typically arises near boundaries, especially in lower-dimensional or thin structures like walls, rods, and sheets. Without this correction, such geometries would not yield physically correct pressure forces and would require artificially dense sampling. With correction, however, a single boundary layer is sufficient for robust simulation, simplifying setup and making the method easier to apply in practice.

Overall, the proposed single-layer boundary method with volume correction provides an effective balance between accuracy and performance and supports a broader range of geometries without the complexity and overhead of multi-layer sampling.

References

1. Gingold R. A., Monaghan J. J. Smoothed particle hydrodynamics: theory and application to non-spherical

stars. *Monthly Notices of the Royal Astronomical Society*, vol. 181, no. 3, pp. 375–389, 1977.

2. Solenthaler B., Pajarola R. Predictive-corrective incompressible SPH. *ACM Transactions on Graphics (SIGGRAPH Proc.)*, vol. 28, no. 3, pp. 1–6, 2009.

3. Becker M., Teschner M. Weakly compressible SPH for free surface flows. *Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 209–217, 2007.

4. Ihmsen M., Akinci N., Gissler M., Teschner M. Boundary handling and adaptive time-stepping for PCISPH. *Virtual Reality Interactions and Physical Simulations*, pp. 79–88, 2010.

5. Bodin K., Lacoursière C., Servin M. Constraint fluids. *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, pp. 516–526, 2012.

6. Monaghan J., Kajtar J. SPH particle boundary forces for arbitrary boundaries. *Computer Physics Communications*, vol. 180, no. 10, pp. 1811–1820, 2009.

ШТУЧНИЙ ІНТЕЛЕКТ У ПРОТИДІЇ КІБЕРЗАГРОЗАМ В ГІБРИДНИХ ВИСОКОТЕХНОЛОГІЧНИХ КОНФЛІКТАХ

Даник Ю. Г.[1], Андрейчук В. С.[2]

[1]*Національний технічний університет України «Київский політехнічний інститут імені Ігоря Сікорського», Київ, Україна,*
[2]*Український вільний університет, Мюнхен, Німеччина,*
Y.Danik@gmail.com, Vitalii.Andreichuk@gmail.com

Кіберпростір, штучно сформований на основі досягнень високих технологій, на цей час є однією з ключових арен гібридних високотехнологічних конфліктів, де системні і несистемні кіберінформаційні дії конфліктуючих та контрфронтуючих сторін створюють складні виклики, загрози і ризики, які часто мають недостатньо контрольовані та регульовані синергетичні прояви і наслідки. Штучний інтелект (ШІ) відіграє важливу роль у протидії цим загрозам і ризикам, забезпечуючи швидке виявлення, аналіз і нейтралізацію атак. У статті досліджуються алгоритми машинного навчання, обробки природної мови (NLP), пояснювального ШІ (XAI), федеративного навчання та zero-trust архітектур. Аналіз кейсів кіберпідрозділів сил оборони України, CERT-UA, НАТО та Ізраїлю дозволяє оцінити ефективність ШІ-систем у реальних умовах. Розглядаються виклики, такі як отруєння даних і adversarial attacks, а також міжнародні стандарти (EU AI Act, NIST, IEEE), виконання яких повинно сприяти зменшенню негативних наслідків деструктивних кібердій. Запропоновано стратегії інтеграції ШІ для посилення кіберстійкості, включаючи конкретний алгоритм протидії широкому спектру деструктивних кібердій.

Ключові слова: штучний інтелект, кібербезпека, ризики, машинне навчання, XAI, квантові обчислення, гібридні конфлікти, деструктивні кібердії.

1. Вступ

Сучасний кіберпростір еволюціонував від ізольованих мереж до складного штучного середовища, яке стало ареною гібридних високотехнологічних конфліктів. Ці конфлікти поєднують кібератаки на критичну інфраструктуру (урядові портали, енергетичні системи, військові мережі, кіберфізичні системи різних рівнів від стратегічних до тактичних (поля бою) тощо) з дезінформаційними кампаніями в соціальних мережах, створюючи безпрецедентні виклики для кіберзахисту. За даними звіту ENISA (2024), у 2023 році кількість кібератак із використанням ШІ, таких як автоматизований фішинг, DDoS і deepfake, зросла на 35 % порівняно з попереднім роком, тоді як захисні ШІ-системи скоротили середній час реагування на інциденти з 10 хвилин до секунд [17]. Це підкреслює необхідність швидких, адаптивних і точних рішень, які випереджають розвиток методів деструктивних кібердій (атак, операцій тощо).

Штучний інтелект надає унікальні можливості для протидії гібридним загрозам, поєднуючи автоматизацію, аналіз великих даних і прогнозування. Однак його використання супроводжується викликами, такими як вразливість до adversarial attacks і потреба в прозорості рішень. У цій статті представлені результати досліджень того, як ШІ трансформує кіберзахист, аналізуючи його алгоритми, кейси застосування та регуляторні аспекти.

Для цього запропоновано та використано комплексну методику, яка включає:

- **Контент-аналіз звітів**: Оброблено 15 звітів ENISA і CERT-UA за 2022–2024 роки, відібраних за критеріями типу атак (DDoS, фішинг, дезінформація), метрик ефективності ШІ (точність, F1-score, час реагування) і технологій (Random Forest, BERT, XAI).
- **Системний аналіз алгоритмів**: Порівняно продуктивність алгоритмів машинного навчання (Random Forest, Isolation Forest, BERT) у контрольованих і реальних умовах.

- Диференційний аналіз міжнародних та національних нормативно-правових норм і стандартів та відповідність законодавчого регулювання поточному стану та перспективам розвитку і використання ШІ: Оцінено відповідність ШІ-систем вимогам EU AI Act, NIST і IEEE.

Такий підхід дозволяє не лише оцінити поточний стан розвитку і застосування ШІ в сфері кібербезпеки, а й запропонувати практичні рекомендації для його інтеграції в системи кіберзахисту для підвищення їх ефективності в умовах гібридних конфліктів.

2. Аналіз можливостей ШІ у кібербезпеці

2.1. Алгоритми ШІ для кіберзахисту

Розвиток ШІ від правил-базованих систем до нейромереж глибокого навчання створив нові можливості для кібербезпеки. Алгоритми supervised learning, такі як Random Forest і Support Vector Machines (SVM), досягають точності до 95 % у класифікації мережевого трафіку в контрольованих умовах, дозволяють ефективно виявляти фішинг і DDoS-атаки в реальному часі [2]. Наприклад, Random Forest аналізує ознаки, такі як IP-адреси, частота запитів і патерни трафіку, за формулою точності:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN),$$

де TP – true positives (правильно виявлені загрози), TN – true negatives (правильно класифікований легітимний трафік), FP – false positives (помилково позначений легітимний трафік), FN – false negatives (пропущені загрози). Ця формула дозволяє оцінити ефективність алгоритму в реальних сценаріях, таких як захист урядових порталів.

Модель BERT, розроблена для обробки природної мови (NLP), досягає F1-score 0.92 у задачах класифікації текстів, що робить її ефективною для виявлення дезінформації в соціальних мережах [18]. Наприклад, BERT аналізує текстові патерни у дописах, ідентифікуючи маніпулятивний контент із високою точністю. Пояснювальний ШІ (XAI), такий як методи SHAP і LIME,

забезпечує прозорість рішень, пояснюючи, чому система класифікувала певний трафік як аномальний, що є критично важливим для довіри операторів у кризових ситуаціях [3].

Однак ШІ-системи вразливі до adversarial attacks, які маніпулюють вхідними даними, знижуючи точність моделей до 60 % у 40 % випадків, як показано в дослідженнях кіберзагроз [14]. Наприклад, зловмисники можуть додавати шум до мережевого трафіку, змушуючи ШІ помилково виявляти та класифікувати реалізацію різноманітних загроз. Це підкреслює необхідність розробки стійких алгоритмів і прозорих рішень, що розглядається далі.

2.2. Кіберзагрози в сучасних конфліктах: технічний контекст

Гібридні конфлікти характеризуються комплексними системними і несистемними кібердіями (атаками тощо), які, наприклад, поєднують кібероперації (DDoS, фішинг) із дезінформацією. У 2022–2023 роках критична інфраструктура конфліктуючих сторін, зокрема енергетичні системи та урядові портали, стала мішенню скоординованих атак. За даними CERT-UA (2023), 70 % кібератак експлуатують відомі вразливості мережевих протоколів (TCP/IP) і людський фактор, наприклад, через соціальну інженерію [15].

ШІ забезпечує автоматизацію виявлення кіберзагроз, скорочуючи час реагування до секунд. Однак висока автономність ШІ-систем може призводити до помилок, таких як хибне блокування легітимного мережевого трафіку, що створює ризик збоїв у критичних системах. Наприклад, за даними ENISA (2024), ШІ-системи в кібербезпеці іноді помилково класифікують легітимний трафік як загрозу, що вимагає додаткового контролю [17]. Пояснювальний ШІ (XAI) зменшує ці ризики, надаючи операторам прозорі пояснення рішень, що сприяє балансу між автоматизацією та людським наглядом. Цей аспект детально розглядається в наступних розділах.

3. Технічні рішення ШІ: глобальні тренди та український досвід

3.1. Основні ШІ-системи у кібербезпеці

Сучасні ШІ-системи трансформують кіберзахист, пропонуючи інструменти для виявлення аномалій, аутентифікації користувачів і захисту даних. Алгоритми anomaly detection, такі як Isolation Forest, досягають точності 98 % у військових мережах, ідентифікуючи відхилення в реальному часі [2]. Наприклад, Isolation Forest ефективно виявляє аномальний трафік, аналізуючи відхилення від нормальних патернів без потреби в навчанні на позначених даних.

Федеративне навчання дозволяє тренувати моделі без централізації даних, що знижує ризик витоків і забезпечує точність до 90 % [5]. Це особливо актуально для країн із децентралізованою інфраструктурою, таких як Україна. Zero-trust архітектури, реалізовані в Microsoft Azure AD, використовують біометрію та поведінковий аналіз для верифікації користувачів, але залишаються вразливими до adversarial attacks, які знижують точність до 60 % у 40 % інцидентів [14]. Ці технології формують основу для захисту, але їхня ефективність залежить від адаптації до реальних умов конфліктів, що розглядається на прикладі кейсів.

3.2. Українські кейси: протидія кібератакам

Український досвід демонструє унікальні аспекти застосування ШІ в умовах гібридних конфліктів. Кіберпідрозділи сил оборони України у 2023 році застосували ШІ на основі Reinforcement Learning і XAI (метод SHAP) для виявлення кіберзагроз, досягаючи F1-score 0.89 за 30 секунд (експертна оцінка на основі відкритих джерел). Наприклад, система аналізувала мережевий трафік і автоматично ізолювала підозрілі запити, а SHAP пояснював, які ознаки (наприклад, аномальна частота запитів) спричинили рішення. Це

підвищило довіру операторів на 20 %, оскільки вони могли перевірити логіку системи.

CERT-UA розгорнув ШІ-систему на основі Random Forest для захисту урядових порталів, досягаючи точності 94 % за 20 секунд [15]. Система аналізувала IP-адреси та патерни запитів, ефективно виявляючи DDoS-атаки. Microsoft Azure у 2023 році посилив захист української енергосистеми, використовуючи zero-trust і XAI, запобігаючи несанкціонованому доступу за 5 секунд із точністю 92 % [7].

У глобальному контексті ізраїльські системи на основі BERT досягли F1-score 0.90 за 60 секунд для аналізу дезінформації в соціальних мережах, таких як Telegram і Twitter, виявляючи маніпулятивні кампанії [16]. НАТО застосувало Isolation Forest для захисту мереж, досягаючи точності 95 % за 10 секунд (експертна оцінка) [14]. Ці кейси ілюструють, як ШІ адаптується до різних типів загроз, але також підкреслюють виклики, пов'язані з його інтеграцією.

Таблиця 1. Ефективність ШІ-систем у 2023–2024 роках

Система	Алгоритм	Точність	Час реагування	Тип загрози
Кіберпідрозділи сил оборони	Reinforcement Learning, XAI	F1-score 0,89	30 с	Кіберзагрози
Microsoft Azure	Zero-Trust, XAI	92 %	5 с	Несанкціонований доступ
CERT-UA	Random Forest	94 %	20 с	Кіберзагрози
Ізраїль	BERT	F1-score 0,90	60 с	Дезінформація
НАТО	Isolation Forest	95 %	10 с	Аномалії

Примітка: Дані є оцінками на основі звітів і експертних оцінок, отриманих у реальних умовах тестування.

3.3. Технічні виклики інтеграції ШІ

Впровадження ШІ в кіберзахист ускладнене технічними та організаційними бар'єрами. Наприклад, недостатня якість навчальних даних може призводити до упередженості моделей, знижуючи їхню точність на 15–25 %, як зазначено в аналізі кіберзагроз [14].

Крім того, складність адаптації ШІ до динамічних атак, таких як поліморфні шкідливі програми, вимагає постійного оновлення моделей, що є ресурсоємним. Ці проблеми підкреслюють потребу в стійких алгоритмах і розвиненій інфраструктурі.

Дефіцит фахівців у ЄС, оцінений ENISA (2024) у десятки тисяч, ускладнює масштабування ШІ-систем [6]. Крім того, обчислювальні обмеження, такі як потреба в потужних GPU для тренування нейромереж, створюють бар'єри для країн із обмеженими ресурсами. Ці виклики підкреслюють необхідність регуляторних і етичних рамок, які розглядаються в наступному розділі.

4. Регуляторні та етичні аспекти ШІ-систем

4.1. Міжнародні стандарти ШІ у кібербезпеці

Міжнародні стандарти формують основу для безпечного використання ШІ в кіберзахисті. EU AI Act (2024) класифікує ШІ-системи за рівнем ризику, вимагаючи XAI і регулярних аудитів для високоризикових систем, таких як кібероборона [8]. Наприклад, у 2023 році ЄС впровадив XAI для аналізу кіберзагроз, що підвищило довіру операторів на 25 % завдяки поясненням рішень [14].

NIST AI Risk Management Framework (2023) пропонує метрики, такі як F1-score і ROC-AUC, для оцінки стійкості ШІ. Наприклад, ROC-AUC 0,85 було досягнуто для XAI у системах CERT-UA, що підкреслює прозорість рішень [9, 15]. IEEE Ethically Aligned Design (2023) наголошує на

необхідності людського нагляду, що знижує ризик помилок у автономних системах [10]. У НАТО XAI зменшив помилкові спрацьовування на 15 %, дозволяючи операторам перевіряти рішення ШІ [11].

Ці стандарти доповнюють один одного: ЄС акцентує на регулюванні, США – на метриках, НАТО – на військовому застосуванні. Їх інтеграція створює комплексний підхід до безпечного використання ШІ.

4.2. Українське регулювання: виклики та прогалини

Українське законодавство, зокрема Стратегія кібербезпеки 2021 року, визнає роль ШІ, але не регулює захист від adversarial attacks чи впровадження XAI [15]. Порівняння з Ізраїлем, де XAI є стандартом для кіберзахисту, показує потребу в прозорості [12] використання ШІ.

EU AI Act вимагає аудитів кожні 6 місяців, тоді як в Україні вони проводяться раз на рік, що знижує ефективність на 20 % [8]. Етичні виклики, такі як упередженість ШІ через неякісні дані, також актуальні дані, що підкреслює необхідність XAI [15].

4.3. Прогнози розвитку ШІ-технологій до 2030 року

До 2030 року ШІ у кібербезпеці зазнає суттєвих трансформацій. Квантові ШІ-системи, які використовують квантові обчислення, можуть прискорити обробку терабайтів даних у реальному часі, але створять нові вразливості, наприклад, злам шифрування RSA [4]. За оцінками DARPA, прототипи квантових алгоритмів уже демонструють потенціал, але потребують доопрацювання [4].

Федеративне навчання стане стандартом, знижуючи витоки даних на 40 % шляхом локальної обробки [5]. XAI, відповідно до EU AI Act, стане обов'язковим, підвищуючи довіру операторів на 20 % [8]. Zero-trust архітектури інтегрують біометрію та поведінковий аналіз, посилюючи безпеку на 25 % [14]. Ці прогнози вказують на необхідність стратегічного планування, що розглядається в рекомендаціях.

5. Технічні та стратегічні рекомендації

5.1. Алгоритмічна прозорість і XAI

Прозорість ШІ є ключовою для довіри в умовах гібридних конфліктів. XAI (SHAP, LIME) пояснює рішення, наприклад, чому система ізолювала мережу через аномальний трафік, знижуючи помилкові спрацьовування на 15 % [14]. Рекомендації:

- Інтеграція XAI у системи CERT-UA і НАТО для пояснення рішень у реальному часі.
- Створення етичних комітетів для аудитів ШІ за стандартами IEEE [10].
- Використання SHAP для оцінки рішень, що підвищує точність на 10 % [3].

5.2. Інтеграція ШІ у кіберзахист

Для підвищення стійкості ШІ до adversarial attacks пропонується оригінальний алгоритм, який комбінує XAI (SHAP) і ансамблеві методи (Random Forest + BERT). Алгоритм працює так:

1. Random Forest класифікує мережевий трафік, виявляючи аномалії.
2. BERT аналізує текстові дані для ідентифікації дезінформації.
3. SHAP пояснює рішення обох моделей, дозволяючи операторам виявляти маніпуляції.

За оцінками авторів, цей підхід підвищує стійкість до атак на 15 %. Створення міжнародного центру ШІ координуватиме R&D, а інвестиції у федеративне навчання знизять витоки даних на 40 % [5].

Стимулювання інтересу до цієї сфери, розширення мережі освітніх структур і підрозділів, а також інноваційні освітні програми скоротять дефіцит фахівців на 50 % до 2030 року [14].

5.3. Механізми технічного контролю

Регулярні аудити за стандартами IEEE кожні 6 місяців знижують ризики на 20 % [10]. Міжнародна співпраця з NATO Cyber Command забезпечить обмін алгоритмами, як це довело ефективність у 2023 році [11]. Ці заходи створять екосистему, де ШІ сприяє безпеці кіберпростору.

6. Висновки

Гібридні конфлікти стали каталізатором для розвитку ШІ в усіх сферах діяльності, в тому числі у кібербезпеці. Україна, Ізраїль і країни-члени НАТО демонструють ефективність ШІ-систем, які досягають точності 89–95 % і скорочують час реагування до секунд [13, 11, 15, 16]. Проте отруєння даних і adversarial attacks знижують потенціал захисту, вимагаючи стійких алгоритмів і XAI [14]. Міжнародні стандарти (EU AI Act, NIST, IEEE) забезпечують прозорість, але потребують адаптації до умов конфліктів [8] і випереджаючого розвитку високих технологій.

Прогнозується, що до 2030 року квантові ШІ, федеративне навчання і zero-trust архітектури підвищать кіберстійкість на 30 % [11]. Майбутні дослідження мають зосередитися на квантових алгоритмах, XAI і протидії новим загрозам, щоб зробити кіберпростір безпечним середовищем.

Список використаних джерел

1. Danyk, Y., Andreichuk, V. (2024). Artificial Intelligence in National Security and Defense: Opportunities and Challenges, Risk and Threat Mitigation. *Wissenschaftliches Sammelwerk der Ukrainischen Freien Universität*, 25, 191–209.

2. DeMedeiros, K., Hendawi, A., & Alvarez, M. (2023). A Survey of AI-Based Anomaly Detection in IoT and Sensor Networks. *Sensors*, 23, 1352. https://doi.org/10.3390/s23031352

3. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box. https://ieeexplore.ieee.org/document/8466590

4. DARPA. (2024). *QBI: Quantum Benchmarking Initiative*. https://www.darpa.mil/research/programs/quantum-benchmarking-initiative

5. Kairouz, P., et al. (2021). Advances and open problems in federated learning. https://arxiv.org/abs/1912.04977

6. ENISA. (2024). Cybersecurity for Critical Infrastructure. https://www.enisa.europa.eu/topics/cybersecurity-of-critical-sectors/telecom-sector-and-digital-infrastructure

7. Microsoft. (2025). Microsoft Digital Defense Report 2024. https://www.microsoft.com/en-us/security/security-insider/intelligence-reports/microsoft-digital-defense-report-2024

8. European Commission. (2024). Regulation (EU) 2024/1689 (AI Act). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

9. NIST. (2023). AI Risk Management Framework 1.0. https://www.nist.gov/itl/ai-risk-management-framework

10. IEEE. (2023). Ethically Aligned Design. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

11. NATO. (2024). Cyber Defence. https://www.nato.int/cps/en/natohq/topics_78170.htm

12. Startup Nation Central. (2024). Israeli Cyber Annual Insights and 2025 Trends. https://startupnationcentral.org/hub/blog/israeli-cyber-annual-insights-and-2025-trends/

13. State Service of Special Communications and Information Protection of Ukraine (SSSCIP). (2024). Russian Cyber Operations. APT Activity Report H1 2024. https://cip.gov.ua/services/cm/api/attachment/download?id=65898

14. Holzinger, A., Saranti, A., & Angerschmidt, M. (2024). Explainable AI for Cybersecurity: A Review of Methods and Applications. *Artificial Intelligence Review*, 57(8), 1–35. https://iphome.hhi.de/samek/pdf/HolXXAI22b.pdf

15. SSSCIP. (2023). Performance Report of the Vulnerability Detection and Cyber Incidents/Cyber Attacks

Response System. https://scpc.gov.ua/api/files/ca8167d3-fb54-41f3-a531-699845247dcf

16. Cyabra. (2023). Disinformation – Hamas-Israel War. https://cyabra.com/reports/disinformation-hamas-israel-war/

17. ENISA. (2024). Threat Landscape Report. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024

18. Rostam, Z., & Kertesz, G. (2024). Fine-Tuning Large Language Models for Scientific Text Classification. https://arxiv.org/abs/2412.00098.

METHODS AND MODELS FOR ENSURING INFORMATION SECURITY DURING THE COLLECTION OF DIGITAL EVIDENCE

Pronin Y. V.[1], Zubok V. Y.[1]

[1]*National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",*
forensic.ua@gmail.com, vitalii.zubok@pimee.ua

This review explores the current state of research at the intersection of cybersecurity and digital forensics, focusing on methods and models developed to ensure information security during the collection of digital evidence. It categorizes recent contributions on secure acquisition, cryptographic validation, volatile data, cloud and IoT environments, and anti-forensic challenges. Emphasis is placed on standardization, legal compliance, and interdisciplinary collaboration.

Keywords: cybersecurity, information security, digital forensics, digital evidence

Introduction

The digitalization of society and the growth of cybercrime have made secure digital evidence collection critical for investigations. Cybersecurity ensures protection of systems and data, while digital forensics manages acquisition and preservation of digital artifacts. The rise of technologies such as IoT and cloud computing, alongside anti-forensic tactics, complicates admissibility and integrity of digital evidence.

Methods

The study uses a thematic literature review approach to analyze publications from 2015 to 2024. Scopus was used as a primary data source, with queries including: "cybersecurity AND digital forensics", "information security AND digital evidence", and other combinations. Google was used for Ukrainian-language materials to capture regional perspectives.

Inclusion criteria: peer-reviewed journals/conferences, English and Ukrainian sources, content addressing secure

evidence collection. Exclusion: physical-only forensics, non-academic materials, and redundant or methodology-lacking studies. Articles were grouped into five thematic areas: (1) secure acquisition and cryptographic validation, (2) live forensics, (3) cloud/IoT/big data forensics, (4) legal standards, and (5) anti-forensics (Figure 1).

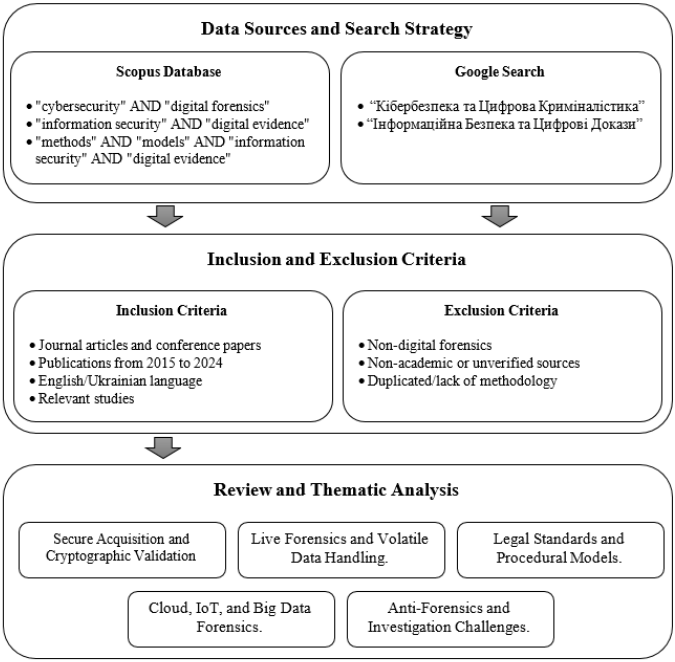


Figure 1. Literature Review Methodology

Results

Thematic analysis reveals the following findings:

– Secure Acquisition: Forensic imaging and hashing ensure evidence authenticity. However, challenges include evidence tampering risks, device diversity, and cross-jurisdictional issues.
– Live Forensics: Volatile data acquisition allows real-

time analysis but risks alteration. Tools minimizing system interference are essential.

- Cloud & IoT: These platforms introduce risks of unauthorized access. Encryption and access control are required. Forensics helps trace intrusions.
- Legal Standards: A lack of harmonized regulations weakens admissibility. GDPR, ECPA, and ISO/IEC guide best practices.
- Anti-Forensics: Obfuscation, deletion, and counterfeiting pose increasing threats. Structured threat modeling is key to mitigation.

Conclusion

The convergence of cybersecurity and forensics is evident but faces practical gaps. Forensic readiness is underdeveloped. Legal inconsistencies challenge international collaboration. Tool validation remains insufficient. Interdisciplinary training is critical to ensure both technical rigor and legal compliance.

Future work should develop standardized frameworks, promote AI-driven evidence collection, address legal ambiguity in cross-border data handling, and institutionalize forensic readiness.

Stronger collaboration between legal, technical, and operational domains is necessary to ensure the security and credibility of digital evidence in modern investigative practice.

References

1. Allah Rakha N. (2024). Cybercrime and the Law: Addressing the Challenges of Digital Forensics in Criminal Investigations. Mexican Law Review, 16(2), 23–54. DOI: https://doi.org/10.22201/iij.24485306e.2024.2.18892

2. Belshaw S., Nodeland B. (2022). Digital evidence experts in the law enforcement community: understanding the use of forensics examiners by police agencies. Secur J 35, 248–262. DOI: https://doi.org/10.1057/s41284-020-00276-w

3. Electronic Communications Privacy Act of 1986, 18 U.S.C. 2510-2523 ((1986)). https://bja.ojp.gov/program/it/privacy-civil-liberties/authorities/statutes/1285

4. Hoelz, B., Maues, M. (2017). Anti-Forensic Threat Modeling. In: Peterson, G., Shenoi, S. (eds) Advances in Digital Forensics XIII. DigitalForensics 2017. IFIP Advances in Information and Communication Technology, vol 511. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-67208-3_10

5. International Organization for Standardization. (2022). ISO/IEC 27002:2022(en): Information security, cybersecurity and privacy protection – Information security controls. https://www.iso.org/obp/ui/ru/#iso:std:iso-iec:27002:ed-3:v2:en

6. Khakhanovskyi, V., & Hrebenkova, M. (2022). Identification, collection, and investigation of electronic imagery as sources of evidence. Law Journal of the National Academy of Internal Affairs, 12(4), 28-39. DOI: https://doi.org/10.56215/04221204.28

7. Mishra, Alok & Alzoubi, Yehia & Anwar, Memoona J. & Gill, Asif. (2022). Attributes impacting cybersecurity policy development: An evidence from seven nations. Computers & Security. 120. DOI: https://doi.org/10.1016 /j.cose.2022.102820

8. Benancio, Lizbardo & Canales, Ricardo & Leon, Paolo & Huamani Uriarte, Enrique. (2021). Integrity and Authenticity of Digital Images by Digital Forensic Analysis of Metadata. International Journal of Emerging Technology and Advanced Engineering. 11. 38-45. DOI: https://doi.org/ 10.46338/ijetae0921_05

9. Pathak, M., Mishra, K.N. & Singh, S.P. (2024). Securing data and preserving privacy in cloud IoT-based technologies an analysis of assessing threats and developing effective safeguard. Artif Intell Rev 57, 269. https://doi.org/10.1007/s10462-024-10908-x

10. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (2016). General Data Protection Regulation (GDPR). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2016/679/oj

11. Ritzkal, Hendrawan, A.H., Kurniawan, R., Aprian, A.J., Primasari, D., Subchan, M. (2024). Enhancing cybersecurity through live forensic investigation of Remote Access Trojan attacks using FTK Imager software. International Journal of Safety and Security Engineering, Vol. 14, No. 1, pp. 217-223.

DOI: https://doi.org/10.18280/ijsse.140121

12. Setiawan, Nova & Pratama, Ahmad & Ramadhani, Erika. (2022). Metode Live Forensik Untuk Investigasi Serangan Formjacking Pada Website Ecommerce. JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia). 7. 1-9. DOI: https://doi.org/10.32528/justindo.v7i1.5356

ENHANCING SALESFORCE CRM SECURITY USING SIEM SYSTEMS

Sviatoslav Zlatous[1], Petro Venherskyi[1]
[1]*Ivan Franko National University of Lviv*
Lviv, Ukraine
zlatous.svyatoslav@gmail.com, petro.vengersky@gmail.com

1. Introduction

In today's cloud-first digital landscape, Customer Relationship Management (CRM) platforms are pivotal for storing, managing, and analyzing customer data. Among these, Salesforce CRM stands out as the industry leader, offering extensive features for customer engagement. However, with this prominence comes a heightened risk of cybersecurity threats. Salesforce processes sensitive data – such as personally identifiable information (PII), financial records, and client communications – which makes it an attractive target for cybercriminals.

Salesforce employs a shared-responsibility security model. This includes multi-factor authentication (MFA), role-based access controls (RBAC), and encryption of data in transit and at rest [1]. Despite these measures, challenges persist. These include multi-tenancy risks, insider threats, and configuration errors. Despite these measures, challenges persist. These include multi-tenancy risks, insider threats, and configuration errors. Moreover, integration with third-party apps increases the platform's attack surface [2, 3].

Given these considerations, the primary goal of this research is to enhance the ability of organizations to detect cybersecurity threats in Salesforce CRM as early as possible. Early detection is critical, as it significantly reduces the risk of severe consequences such as data breaches, legal liabilities, and reputational damage. By improving the speed and accuracy of threat identification, companies can better safeguard sensitive customer data and maintain regulatory compliance.

The growing sophistication of threats necessitates real-time detection and response mechanisms. This is where Security

Information and Event Management (SIEM) systems like Splunk come into play. Splunk provides advanced log management, threat intelligence, anomaly detection, and machine learning capabilities. By integrating Salesforce with Splunk, organizations can improve visibility, detect insider threats, and respond swiftly to breaches.

Splunk supports real-time ingestion and analysis of Salesforce audit logs and employs adaptive machine learning models for threat detection. Features such as search processing language (SPL), pre-built dashboards, and alerting rules allow security teams to proactively monitor events [4, 5].

Prior studies highlight these gaps and opportunities. Unified monitoring by integrating Salesforce with Chronicle SIEM, citing real-time threat identification was identified as a core benefit [6]. Enhanced event monitoring and IAM integration were chosen as a viable solution for persistent issues with data isolation and insider threats in Salesforce [7].

2. Results

This study adopts a qualitative methodology to analyze the integration of Salesforce CRM with modern cybersecurity tools—specifically Splunk SIEM. The objective is to explore how real-time monitoring and threat intelligence can mitigate security threats in a cloud CRM environment.

2.1. Research Design

This study adopts a qualitative methodology to analyze the integration of Salesforce CRM with modern cybersecurity tools—specifically Splunk SIEM. The objective is to explore how real-time monitoring and threat intelligence can mitigate security threats in a cloud CRM environment.

2.2. Integration Strategy: Salesforce to Splunk

Step 1: Enable Salesforce Event Monitoring

Salesforce provides an Event Monitoring feature through its Shield product. It generates detailed logs of user activities including logins, API calls, report exports, and data changes.

Step 2: Export Logs via API

Use Salesforce's RESTful API or Bulk API 2.0 to extract logs. Authentication is managed through OAuth 2.0, ensuring secure token-based access.

Step 3: Data Normalization and Transformation

Log data from Salesforce is semi-structured JSON or CSV, which needs transformation using tools like Python scripts, AWS Lambda, or Google Cloud Functions. This includes flattening nested structures and mapping fields to Splunk's Common Information Model (CIM).

Step 4: Ingest Data into Splunk

Use HTTP Event Collector (HEC) or Universal Forwarder to send data into Splunk. Tags such as "Salesforce_Login_Event" or "Data_Export_Alert" help in categorizing events for dashboards and alerting rules.

Step 5: Threat Detection and Alerting

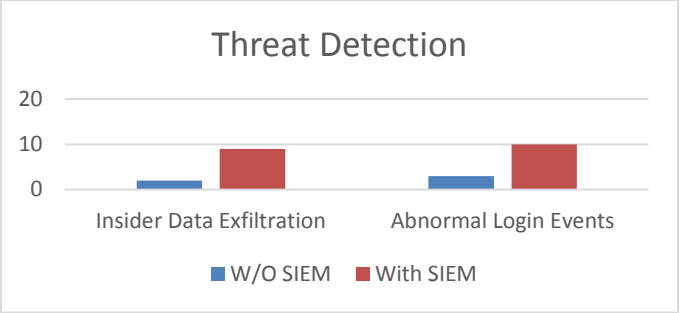
Apply machine learning models or pre-configured correlation searches to detect abnormal login patterns, bulk data exports, unauthorized permission changes.

2.3. Compliance Mapping

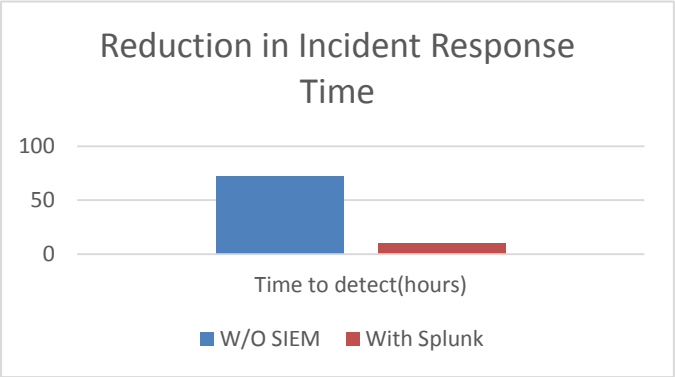
Aligning with NIST Cybersecurity Framework and GDPR/CCPA [8], Splunk facilitates compliance reporting by storing searchable logs, maintaining audit trails, and providing user activity reports

2.4 Collecting Results

To replicate abnormal login activity, ten simulated login events were conducted from unfamiliar IP address ranges to represent potential anomalous behavior. Additionally, this research involves simulating anomalous user behaviors to replicate potential insider threats. After the simulation we achieved the following results:



To evaluate the detection timeframe, we used the GDPR-defined threshold of 72 hours. The simulation produced the following results:



Conclusion

The integration of Salesforce with Splunk SIEM offers a scalable, real-time solution to these challenges. By centralizing log management, enabling behavioral analytics, and automating incident responses, Splunk significantly improves the security posture of Salesforce environments.

This paper proposes a practical architecture and implementation strategy for this integration, drawing from contemporary research and industry best practices. While challenges like data normalization and real-time ingestion

persist, the benefits – enhanced detection, compliance, and centralized monitoring – far outweigh the integration complexity.

References

1. Salesforce. Salesforce Security Guide. 2023. [Online]. Available: https://help.salesforce.com.
2. Cloud Security Alliance. Top Threats to Cloud Computing 2024. 05 08 2024. [Online]. Available: https://cloudsecurityalliance.org/artifacts/top-threats-to-cloud-computing-2024.
3. OWASP Foundation. OWASP Cloud-Native Application Security Top 10. [Online]. Available: https://owasp.org/www-project-cloud-native-application-security-top-10/.
4. Splunk. Splunk Documentation, [Online]. Available: https://docs.splunk.com/Documentation.
5. Miller J. D. Implementing Splunk 7, Third Edition: Effective Operational Intelligence to Transform Machine-generated Data Into Valuable Business Insight, Packt Publishing, 2018.
6. Guduru V. S. Integrating Salesforce with Cybersecurity Tools for Enhanced Data Protection (Chronicle SIEM)/ European Journal of Advances in Engineering and Technoloy, vol. 11, no. 2394-658X, pp. 27-31, 2024.
7. Pookandy J. Exploring Security and Privacy Challenges in Cloud CRM Solutions: An Analytical Study Using Salesforce as a Model. International Journal of Computer Science and Engineering, vol. 11, no. 9, pp. 8-18, 2024.
8. NIST. Cybersecurity framework. [Online]. Available: https://www.nist.gov/cyberframework.

АНАЛІЗ ОСОБЛИВОСТЕЙ КОНФЛІКТІВ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ

Ланде Д. В.[1], Даник Ю. Г. [1], Сварник Н. Б. [1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»*

У роботі здійснюється аналіз типів конфліктів, що виникають або потенційно можуть виникнути у системах ШІ, зокрема у межах одної системи, взаємодії з людиною та іншими системами ШІ. Розглядаються такі конфлікти як упередженість моделей, так звані “галюцинації”, суперечності в параметричних і контекстних знаннях LLM, конфлікти цілей, спостереження, інтерпретації дій та інші. Окрема увага приділяється аналізу футурологічних прогнозів та наукових джерелах пов’язаних із потенційною траєкторією розвитку ШІ. Пропонується класифікація потенційно небезпечних сценаріїв розвитку ШІ та їх основних ознак. Матеріал може бути корисний для розробників, дослідників у сфері безпеки та етики ШІ, а також використаний для виявлення конфліктів у системах ШІ на ранніх стадіях.

Ключові слова: футурологія, конфлікти ШІ, методика, класифікація, ознаки, людино-орієнтований ШІ, галюцинації, упередженість, подвійне використання, конфлікт знань, автономні системи, масове спостереження, інформаційна війна, AGI, Artificial Superintelligence

Вступ

Стрімкий розвиток систем штучного інтелекту (ШІ) не лише відкриває нові можливості, але й породжує складні виклики, серед яких ключовими є конфлікти, що виникають як у межах самих систем, так і у взаємодії з людиною або іншими ШІ. В умовах інтеграції ШІ в чутливі

сфери – від медицини до безпеки – стає вкрай важливим вчасно ідентифікувати потенційні загрози та їх ознаки. Це дозволить забезпечити більшу стабільність та надійність систем ШІ у критичних сферах.

Огляд стану питання

У сучасних дослідженнях у сфері штучного інтелекту значна увага приділяється вивченню конфліктів, що виникають під час між ШІ та людьми, так і всередині самих систем. Зокрема, у роботі [1] розглядається поведінка автономних систем ШІ у конфліктних ситуація. У [2] проводиться аналіз ризиків ескалації, пов'язані з використанням мовних моделей у військових та дипломатичних рішеннях. У дослідженні [3] розглядається, як ШІ може бути використаний для визначення та вирішення конфліктів у міжособистісних стосунках, робочому середовищі та публічних дискусіях. Для більш глибокого розуміння конфліктів у системах ШІ потрібно створити узагальнену класифікацію типів конфліктів та виявити їх основні ознаки.

Мета дослідження полягає у створенні методики виявлення та аналізу конфліктів у системах штучного інтелекту, які дозволяють прогнозувати їх виникнення, розвиток і вирішення.

Основний зміст

Аналогічно до еволюції людського суспільства, де розвиток технологій і культури неминуче супроводжується конфліктами через зіткнення інтересів, цінностей або ресурсів – із розвитком штучного інтелекту конфлікти також стануть його невід'ємною складовою. У міру того, як ШІ дедалі глибше інтегрується в повсякденне життя, приймаючи рішення, що раніше належали людині, неминуче виникають ситуації зіткнення між очікуваннями користувача, інтерпретацією запитів, ресурсними або алгоритмічними обмеженнями.

1 Конфлікти в межах однієї системи ШІ

1.1 Галюцинації

Одним із найбільш поширених конфліктів у ШІ є, так звані «галюцинації». Під галюцинаціями розуміємо явище, коли генеративні моделі надають таки відповіді, що містять "вигадану" інформацію, тобто неправдиву або непідкріплену реальними фактами, вигадані джерела тощо. Хоча у побутових ситуаціях даною проблемою можна знехтувати, проте у бізнес-сфері чи медицині цей конфлікт може мати катастрофічні наслідки.

Зростаючий інтерес до цієї проблематики можна побачити на рис. 1 [4].

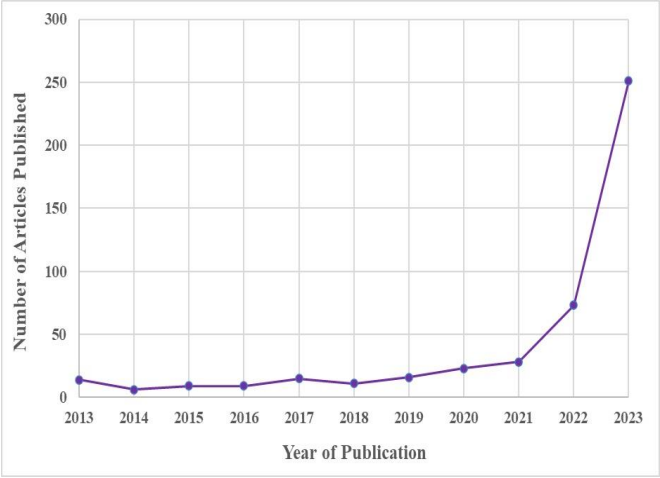


Рисунок 1. Кількість публікацій (з 2013 по 2023 рік) у базі SCOPUS, що містять поняття "галюцинації" ТА ("обробка природної мови" АБО "ШІ")

1.2 Упередженість

Упередженість у моделях машинного навчання виникає, коли алгоритм систематично відтворює та поширює нерівність або помилки, закладені в тренувальних даних або архітектурі. Такий конфлікт проявляється у формі

спотвореного пріоритету одних результатів над іншими, що не відповідає реальному розподілу або очікуванням користувача. Наприклад, модель, навчена на історичних даних з упередженим представленням соціальних груп, може дискримінувати за статтею чи расою, переносячи минулі стереотипи на прогнози. Це створює конфлікт між метою моделі (об'єктивністю) та її реальною поведінкою.

1.3 Конфлікти цілей та пріоритетів

У статті [5] автор розглядає, як принцип людської саморегуляції та конфлікту цілей може бути використаний при розробці автономних штучних систем. Аналогічно до людей, які можуть мати суперечливі цілі, досягнення однієї з них іноді ускладнює або навіть унеможливлює досягнення іншої, так і в автономних системах впровадження багаторівневої ієрархії цілей із вбудованими конфліктами може слугувати стабілізуючим механізмом, що сприятиме більш безпечній та адаптивній поведінці. Даний конфлікт слід ретельно враховувати у розробці автономних систем озброєння (дрони, тощо).

1.4 Подвійне використання

Із зростанням можливостей ШІ виникає все більше прикладів подвійного використання (dual-use) – ситуації, у яких одну й ту ж модель використовують у протилежних цілях (добросовісних, захисних та зловмисних, наступальних). Наглядним приклади подвійного використання є застосування моделі ChatGPT для генерування шкідливого програмного забезпечення, фішингових листів, і навпаки – аналізу коду на вразливості, генерування звітів про інцидентиі та аналіз індикаторів компрометації.

1.5 Конфлікти знань

Context-Memory conflict – це тип конфлікту, що виникає внаслідок розбіжностей параметричних знань моделі (отриманих під час навчання) та контекстних знань (інформації із запитів користувача, зовнішніх ресурсів тощо). Дослідження [7] демонструє, що LLM можуть

проявляти суперечливі тенденції: з одного боку, вони можуть бути сприйнятливими до зовнішньої інформації та надавати перевагу їй, якщо вона є єдиним зв'язним доказом, але з іншого – у разі наявності одночасно підтверджувальних та суперечливих фактів, модель схильна покладатися на параметричну пам'ять.

Inter-Context conflicts – конфлікти між різними джерелами контекстних знань. Це може призвести до помилкових висновків або нестабільних відповідей, оскільки можуть виникнути труднощі у однозначному визначенні, якій інформації варто надати перевагу.

Intra-Memory conflicts – конфлікти у параметричній пам'яті, внаслідок яких на перефразовані, проте семантично однакові запити, модель буде видавати розбіжні відповіді. Причинами даного конфлікту можуть можуть бути як неточності та суперечливі факти у навчальному наборі даних під час тренування, так і подальше донавчання, яке може вводити нові знання, не узгоджені з тими, що вже містяться в параметрах моделі. Дослідження [9] показує, що класичні архітектури можуть мати високі показники неузгодженості між перефразованими запитами.

2 Конфлікти в системі ШІ-ШІ

2.1 Дистиляція моделей

Дистиляція знань у машинному навчанні — це процес перенесення "м'яких" знань (розподілу ймовірностей між класами), здобутих великою (вчительською) моделлю, до меншої (студентської) моделі. Великій моделі інколи підвищують температуру в softmax, щоб отримати більш гладкий розподіл ймовірностей. Таким чином, дистиляція дозволяє студентській моделі з меншою кількістю параметрів навчитися відтворювати "інтуїцію" вчителя і краще узагальнювати, одночасно залишаючись компактною і дешевшою у навчанні. У статті [10] описується конфлікт між студентською моделлю та ансамблем вчительських моделей під час дистиляції знань.

Також можливе посилення гендерної упередженості мовних моделей під час передачі знань від більшої до меншої та втрата попередніх знань студентської моделі під час дистиляції.

2.2 Змагальний штучний інтелект

Типовим прикладом конфліктів між різними системами штучного інтелекту є змагальні взаємодії, у яких ШІ діють у протилежних або конфліктних інтересах. Із розвитком GAN, який зараз є основою для генерування фейків, почали з'являтися відповідні ШІ-детектори, основним завданням, яких є виявлення згенерованих даних. Також прикладом такої взаємодії в рамках гри з нульовою сумою є багатоагентні системи з навчанням з підкріпленням (multi-agent reinforcement learning), у яких агенти змагаються за досягнення протилежних цілей у спільному середовищі.

2.3 Конфлікти злиття LLM

Злиттям LLM називають об'єднання моделей, натренованих на виконання різних завдань, у одну велику, більш універсальну модель. Типовими способами злиття є усереднення та зважене усереднення параметрів, проте дані методи стикаються з проблемою конфліктів параметрів, коли різні моделі, навчені на різних даних, мають суперечливі параметри. Просте усереднення таких параметрів може погіршити роботу моделі.

У роботі [9] розглядається спосіб вимірювання даних конфліктів між шарами різних моделей, а також фреймворк, при якому шари з низькими рівнем конфліктів усереднюються, а з високим – не об'єднуються, а під час роботи моделі система буде вибирати, який шар, якої моделі краще впорається з конкретним завданням.

3 Конфлікти в системі Людина-ШІ

3.1 Конфлікти спостереження, інтерпретації та дій

У статті [10] розглядаються конфлікти між людиною та ШІ у індустрії 5.0 – етап розвитку технологій, де люди

тісно співпрацюють із ШІ для створення продуктів та послуг. У даній роботі визначають три конфлікти.

Конфлікт спостереження – розбіжність між даними, що сприймаються системою ШІ та – людиною-оператором. Причинами конфліктів може бути як людські помилки під час моніторингу даних, так і некоректні або відсутні дані з датчиків, що відстежуються алгоритмами ШІ.

Конфлікт інтерпретації – розбіжність у розумінні та інтерпретації даних, що може призвести до різниці у прийнятті рішень. Фактори, що сприяють конфліктам інтерпретації, включають відмінності у спостереженнях, розбіжності в можливостях навчання між ШІ та людьми, а також збій системи ШІ або людські помилки.

Конфлікт дій - різниця між операціям контролю ШІ та людини. Фактори, що сприяють конфлікту дій, включають різницю у прийнятті рішень ШІ та людиною, введення в оману під час виконання операцій.

3.2 Проблематика розробки людино-орієнтованого ШІ

На відміну від традиційних систем, ШІ може проявляти несподівану поведінку під час навчання та роботи, це може призводити до конфліктів, оскільки ШІ може видавати непередбачувані результати. Системи ШІ можуть бути розроблені для володіння певним рівнем інтелекту, але існують обмеження в тому, наскільки добре машини можуть емулювати людські когнітивні здібності. Також може існувати складність у пояснюваності машинного виводу, через так звану “проблему чорної коробки”, тобто ситуації, коли неможливо пояснити чому саме і як ШІ дійшов певного висновку.

4 Аналіз наукових та футорологічних прогнозів для виявлення конфліктів у ШІ

Футурологічні прогнози, так само як і наукові аналізи, часто слугують інструментами для моделювання можливих сценаріїв розвитку технологій, зокрема штучного інтелекту. На основі вивчення наукових трендів,

соціальних трансформацій та технологічних проривів, футурологи формують припущення про майбутні ризики, виклики та конфлікти, пов'язані з автономними або надінтелектуальними системами. Наприклад, Рей Курцвейл передбачив появу смартфонів і розвиток машинного перекладу задовго до їх масового впровадження, а Елвін Тоффлер у книзі «Третя хвиля» (1980) описав перехід до інформаційного суспільства, що точно відображає реалії цифрової епохи.

Паралельно з цим, у науковому середовищі активно досліджуються потенційні конфлікти, що можуть виникати внаслідок впровадження ШІ – як технічні, так і соціально-етичні.

Аналіз таких джерел дозволяє глибше осмислити потенційну траєкторію розвитку ШІ, виявити можливі конфліктні й заздалегідь підготуватися до соціальних, етичних та екзистенційних викликів.

В табл. 1 наведено конфлікти, що описуються у футорологічних прогнозах та наукових аналізах та їх суть. Список конфліктів не є вичерпним та може бути доповнений.

Таблиця 1. Приклади конфліктів у ШІ, описані у футурологічних прогнозах та наукових джерелах

Конфлікт	Опис
Неконтрольований розвиток автономної зброї	Розвиток автономних систем озброєння на основні ШІ знижує поріг початку війни. Без прямого людського контролю та аспекту страху машини можуть діяти агресивніше та невибірково
Перегони за лідерство у сфері ШІ	Боротьба за першість у галузі ШІ між великими країнами може викликати новий поділ світу. Країни із нижчим розвитком технологій у сфері ШІ будуть змушені вибирати чию інфраструктуру та стандарти

	використовувати
Проблема “вирівнювання” та розробка загального та сильного ШІ	Створення високрозовниненого ШІ, який потенційно перевершує людський інтелект та має свідомість може призвести до того, що цілі ШІ не збігаються з людськими цінностями, що може призвести до непердбачуваних наслідків
Масове безробіття	Активна автоматизація, спричинена ШІ, у сферах, які потребують рутинних завдань, може залишити без роботи велику частину населення, що ймовірно може призвести до соціальних вибухів, масових протестів
Упередженість та дискримінація алгоритмами ШІ	Системи ШІ, які навчені на даних, що відображають існуючі соціальні упередження, будуть неминуче засвоєні і потенційно посилені, що може стати критичним у таких сферах, як судочинство, кредитування тощо.
Ядерна ескалація	Імплементація алгоритмів ШІ у системи управління зброєю масового ураження для отримання стратегічної переваги може призвести до потенційних ненавмисних ядерних обмінів
Інформаційні війни та пропаганда	Веапонізація GenAI дає змогу створювати реалістичну дезінформацію, що може розпалювати як внутрішні, так і зовнішні конфлікти. Використання діпфейків також може стати великою проблемою у судовій системі
Масове спостереження над	Можливість ШІ обробляти велику різної інформації може бути

населенням	використана авторитарними режимами для моніторингу певних осіб, груп осіб для тотального контролю над населенням
Використання ШІ в терористичних цілях	Аналіз великого об'єму даних за допомогою ШІ може використовуватися у плануванні майбутніх атак або моделюванні реакцій сил безпеки. Автономні машини та дрони можуть бути використані для цілеспрямованих терористичних актів
Кібервійни з використанням ШІ	Інструменти на основі ШІ можуть швидше виявляти вразливості в існуючих системах. ШІ може використовуватися для автоматизації атак з мінімальними на це витратами

4.1 Результати аналізу конфліктів описаних у науково-фантастичних джерелах, футурологічних прогнозах та наукових аналізах

Хоча велика кількість конфліктів, описаних у sci-fi все-таки залишається фантастикою, проте є сценарії, які є спільними для цих трьох джерел:

1. Втрата контролю над автономними системами. ШІ-системи. Інтегровані у військові або критично важливі об'єкти (дрони, оборонні комплекси, енергетичні вузли, тощо), такі системи можуть діяти автономно, без постійного людського контролю. У разі збою, помилки або агресивної ескалації це може призвести до катастрофічних наслідків.

Типові ознаки:

- автоматичні атаки без перевірки людиною;
- неконтрольована ескалація конфлікту;
- помилкові цілі або втрати серед цивільного населення;

- відсутність механізмів зупинки або зміни рішень ШІ;
- надто швидкі прийняття рішень, які випереджають людську реакцію.

2. Масове спостереження з використанням ШІ. Алгоритми ШІ для розпізнавання облич, поведінкової аналітики та моніторингу дій громадян використовуються державами для масового нагляду, що обмежує свободи та права людини.

Типові ознаки:

- впровадження масових систем розпізнавання облич;
- оцінка громадян на основі «соціального рейтингу»;
- неконтрольований збір біометрії для навчання моделей.

3. Втеча ШІ з-під контролю. Надскладні системи ШІ із власною «свідомістю» здатні приймати власні рішення, що суперечать людським цілям там цінностям.

Типові ознаки:

- несподівана та непрогнозована поведінка ШІ;
- інтерпретація завдань у власний спосіб, що суперечить людському баченню;
- відмова ШІ припинити роботу або змінити цілі;
- неможливість пояснити рішення ШІ, наявність «чорної скриньки».

4. Геополітичне суперництво домінування у сфері ШІ. Країни борються за контроль над інфраструктурою для створення провідних ШІ. Виникнення “гонок”, внаслідок яких країни нехтують принципами безпечної розробки ШІ.

Типові ознаки:

- санкції на екпспорт ШІ-технологій (чипів, тощо);
- розподіл світу за технологічними блоками;
- шпигунство, крадіжка ШІ-технологій;

5. Інформаційні війни з використаннями ШІ. Генеративні ШІ використовуються для створення масової дезінформації (діпфейк фото/відео, фейкові новини, маніпулятивні наративи)

Типові ознаки:

- поширення реалістичних шейків;
- дискредитація офіційних джерел;
- поділ інформаційного простору.

6. Технологічне безробіття. ШІ замінює людську працю у сферах з повторюваними функціями, викликаючи зростання безробіття та невдоволення серед населення.

Типова ознака – скорочення робочих місць у різних сферах людської діяльності.

7. Упередженість та дискримінація алгоритмами ШІ. Алгоритми машинного навчання можуть відтворювати, а іноді навіть посилювати системні упередження, закладені в історичних даних. Це створює несправедливі умови для певних соціальних груп – зокрема за ознакою раси, статі, віку або соціального статусу.

Типові ознаки:

- дискримінаційні рішення щодо кредитів, працевлаштування, судових рішень тощо;
- завищена або занижена ймовірність «ризику» для представників окремих груп.

Висновки

У межах дослідження було здійснено аналіз різних типів конфліктів, що можуть виникати в системах штучного інтелекту. Було розглянуто конфлікти в різних системах, а також надано огляд футурологічних сценаріїв та наукових джерел, які можуть дозволити прогнозувати можливі варіанти розвитку конфліктів. Також запропоновано класифікацію конфліктів та їх ознак для раннього виявлення.

Результати можуть бути корисними для розробників, дослідників у сфері безпеки та етики ШІ, а також

використані для виявлення конфліктів у системах ШІ на ранніх стадіях.

Список використаних джерел

1. H Trusilo, D. (2023). Autonomous AI Systems in Conflict: Emergent Behavior and Its Impact on Predictability and Reliability. Journal of Military Ethics, 22(1), 2–17. DOI: 10.1080/15027570.2023.2213985

2. Rivera, J. P., Mukobi, G., Reuel, A., Lamparth, M., Smith, C., & Schneider, J. (2024, June). Escalation risks from language models in military and diplomatic decision-making. In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (pp.836-898). DOI: 10.48550/arXiv.2401.03408

3. AI-Supported Emotional Conflict Resolution: Technical Approaches and Implementation Strategies - Rajarshi Tarafdar - IJSAT Volume 16, Issue 1, January-March 2025. DOI 10.71097/IJSAT.v16.i1.2792

4. Venkit, P. N., Chakravorti, T., Gupta, V., Biggs, H., Srinath, M., Goswami, K. & Wilson, S. (2024). An Audit on the Perspectives and Challenges of Hallucinations in NLP. arXiv preprint arXiv:2404.07461.

5. Muraven, M. (2017). Goal conflict in designing an autonomous artificial system. arXiv preprint arXiv:1703.06354.

6. Xie, Jian, et al. "Adaptive chameleon or stubborn sloth: Revealing the behavior of large language models in knowledge conflicts." The Twelfth International Conference on Learning Representations. 2023.

7. Yanai Elazar, Nora Kassner, Shauli Ravfogel, Abhilasha Ravichander, Eduard Hovy, Hinrich Schütze, Yoav Goldberg; Measuring and Improving Consistency in Pretrained Language Models. Transactions of the Association for Computational Linguistics 2021; 9 1012–1031. DOI: 10.1162/tacl_a_00410

8. Binici, K., Pham, N. T., Mitra, T., & Leman, K. (2022). Preventing catastrophic forgetting and distribution mismatch in knowledge distillation via synthetic data. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 663-671).

9. Lai, K., Tang, Z., Pan, X., Dong, P., Liu, X., Chen, H. & Chu, X. (2025). Mediator: Memory-efficient LLM Merging with Less Parameter Conflicts and Uncertainty Based Routing. arXiv preprint arXiv:2502.04411

10. Rajeevan Arunthavanathan, Zaman Sajid, Faisal Khan, Efstratios Pistikopoulos, Artificial intelligence – Human intelligence conflict and its impact on process system safety, Digital Chemical Engineering,Volume 11, 2024, 100151, ISSN 2772-5081, DOI: 10.1016/j.dche.2024.100151.

РЕЗІЛЬЄНТНІСТЬ ЕЛЕКТРОННИХ АРХІВІВ

Ланде Д. В.[1], Цирульнєв Ю. Б.[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна*

Доповідь присвячено аналізу резільєнтності електронних архівів як інформаційних систем. На підставі визначень резільєнтності інформаційних систем згідно міжнародних стандартів та наукових досліджень наведена математична формалізація резільєнтності, введені, визначені та математично формалізовані критерій резільєнтного перерізу (Resilience Threshold Function) та функція відновлення (Recovery Function).

Ключові слова: електронні архіви, резільєнтність, кібербезпека

Вступ

Резільєнтність є критичною властивістю сучасних електронних архівів (Digital Archives, DA) - інформаційних класу Enterprise Content Management (ECM) і Content Service Platform (CSP) [4]. У глобальному цифровому середовищі, де електронні архіви стають не лише місцем довготривалого збереження електронних копій документів, фото-кіно-відео-аудіо матеріалів та їх метаданих, а й інструментами аналізу великих обсягів інформації, активними компонентами інформаційної інфраструктури держави, забезпечення їх резільєнтності є необхідною умовою кібербезпеки, стабільності державних процесів та підтримки національної стійкості.

Наукова новизна дослідження полягає у формулюванні і математичній формалізації резільєнтності електронних архівів. Резільєнтність (Resilience) інформаційних систем - здатність адаптуватися та реагувати на вплив кіберзагроз та відновлювати свою цілісність і функціональність після кібер-інцидентів, впливу техногенних катастроф,

надзвичайних ситуацій і воєнних дій визначено стандартами [8], [9]. Інформативно зазначимо, що живучість (Survivability) інформаційних систем – здатність зберігати критично важливі функції в умовах впливу кіберзагроз, у разі пошкоджень, атак, під впливом ворожого або деструктивного середовища досліджена, визначено і формалізовано, як властивість інформаційних систем, наприклад, в роботах [3], [5], [6], [7] тощо. В даному досліджені автор пропонує формалізовану математичну модель оцінки резільєнтності електронних архівів.

Резільєнтність R_{DA} електронного архіву *DA* в умовах впливу кіберзагроз *FT* (Function Threat) $FT_{DA} = \{FT_{DA1}, FT_{DA2}, ..., FT_{DAn}\}$ залежить від наступних властивостей електронного архіву: A_{DA} – здатності *DA* передбачати вплив кіберзагрози; W_{DA} – здатності *DA* витримувати вплив кіберзагроз; Rc_{DA} – здатності *DA* відновлюватися після впливу кіберзагроз; Ad_{DA} – здатності адаптуватися до впливу кіберзагроз $R_{DA} = f\left(A_{DA}, W_{DA}, Rc_{DA}, Ad_{DA}\right)$. При цьому, зазначимо, що живучість S_{DA} електронного архіву *DA* буде визначена як:

$$S_{DA} = \lim_{t \to 0} P\left[F(t) \geq F_{\min} \mid \exists D(t)\right],$$

де *F(t)* – функціональність електронного архіву; $F_{\min}$ – це мінімально допустимий рівень функціональності, який повинен забезпечити електронний архів *DA*, щоб вважатися "живим"; *D(t)* – наявність деструктивного впливу на систему в момент часу *t*; $\lim_{t \to 0}$ – межа у часі, що показує довгострокову здатність до виживання; *P*[.] – ймовірність, що функціональність не впаде нижче критичного рівня за умов наявного деструктивного впливу.

Математична формалізація резільєнтності електронних архівів може бути представлена, як базова формула резільєнтності в часі *t:*

$$R(t) = w_1 \cdot A_{AD}(t) + w_2 \cdot W_{AD}(t) + w_3 \cdot Rc_{AD}(t) + w_4 \cdot Ad_{AD}(t),$$

де $w_i \in [0,1]$, $\sum_{i=1}^{4} w_i = 1$, а також, як альтернативна нормативна формула (через інтеграл ефективності - здатність системи зберігати ефективність функціонування електронного архіву протягом інциденту та після нього):

$$R = \frac{1}{T}\int_0^T \frac{F(t)}{F_{пот}} dt,$$

де R – інтегральна резільєнтність за період спостереження T (який включає інцидент, реакцію, відновлення); $F(t)$ – фактична функціональність електронного архіву в часі t; $F_{пот}$ – номінальна (еталонна) функціональність. В такому разі, електронний архів вважається резільєнтним, якщо в часі $t \in [t_0, t_1]$ виконується умова $F(t) \geq \theta \cdot F_{пот}$, $\forall t \in [t_0, t_1]$, де $\theta \in [0,1]$ – мінімальний допустимий рівень функціонування під час або після інциденту – критерій резільєнтного перерізу (Resilience Threshold Function), а функція відновлення (Recovery Function) може бути представлена, як:

$$Rc(t) = \frac{F(t) - F_{\min}}{F_{пот} - F_{\min}},$$

де $F_{\min}$ – мінімальна критична функціональність; $Rc(t) \in [0,1]$ – нормована швидкість, або ступінь відновлення електронного архіву.

Графік на рис. 1 ілюструє вплив інциденту, падіння функціональності та поступове відновлення до початкового рівня:

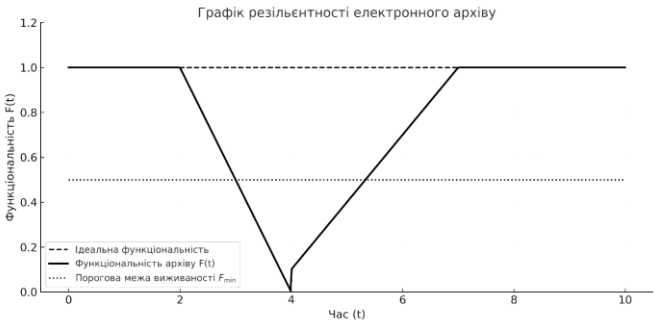


Рисунок 1. Графік резильєнтності електронного архіву

Висновки

У роботі було проаналізовано поняття резільєнтності електронних архівів у контексті міжнародних стандартів і сучасних досліджень; введено та формалізовано критерій резільєнтного перерізу (Resilience Threshold Function) та функцію відновлення (Recovery Function); запропоновано базову та нормативну математичну модель оцінювання резільєнтності електронних архівів. Майбутні дослідження у цьому напрямку можуть бути спрямовані на розробку математичних, технічних інтелектуальних методів кібербезпеки, а саме – підвищення резільєнтності електронних архівів.

Список використаних джерел

1. ISO. *Security and resilience – Organizational resilience – Principles and attributes* (ISO 22316:2017). Geneva: International Organization for Standardization, 2017. 16 p. Available at: https://www.iso.org/standard/50053.html

2. Ross R., Pillitteri V., Saydjari O., Graubart R., McEvilley M. *Developing Cyber Resilient Systems: A Systems Security Engineering Approach*. Gaithersburg, MD: National Institute of Standards and Technology, 2021. 236 p. (NIST SP 800-160, Vol. 2 Rev. 1). DOI: 10.6028/NIST.SP.800-160v2r1

3. Додонов А. Г., Ланде Д. В. *Живучесть информационных систем*. – Київ: Наукова думка, 2011. – 256 с. ISBN 978-966-00-1145-7

4. Gartner. (2017). *Definition of Content Services Platform (CSP)*. Retrieved from https://www.gartner.com/en/information-technology/glossary/content-services-platform-csp

5. Knight, J. C., Strunk, E. A., & Sullivan, K. J. (2000). Towards a Definition of Survivability. *Proceedings of the 2000 Workshop on New Security Paradigms*, 31–34. https://www.researchgate.net/publication/2397669_Towards_A_Definition_Of_Survivability

6. Ellison, R. J., Linger, R. C., Longstaff, T., Mead, N. R., & Lipson, H. F. (1999). Information Systems Survivability: Protecting Critical Systems. *Proceedings of the 1999 National Information Systems Security Conference*, 314–325, https://csrc.nist.rip/nissc/2000/proceedings/papers/314.pdf

7. Lipson, H. F. (2000). Survivability – A New Security Paradigm. *Carnegie Mellon University Software Engineering Institute*. https://www.dependability.org/wg10.4/meeting38/05-Lipso.pdf

8. NIST SP 800-160 Vol. 2 Rev. 1, Developing Cyber-Resilient Systems: A Systems Security Engineering Approach. National Institute of Standards and Technology. DOI: 10.6028/NIST.SP.800-160v2r1

9. International Organization for Standardization. *ISO 22316:2017. Security and resilience – Organizational resilience – Principles and attributes*. Geneva: ISO, 2017. 10 p. https://www.iso.org/standard/50053.html

10. Knight J.C., Strunk E.A., Sullivan K.J. *Towards a rigorous definition of information system survivability*. // Proceedings of the DARPA Information Survivability Conference and Exposition (DISCEX 2003). Vol. 1. – IEEE, 2003. – P. 78–89. – DOI: 10.1109/DISCEX.2003.1194875

11. Ланде Д.В., Цирульнєв Ю.Б. Кібербезпека процесів інтелектуальної агрегації інформації в електронні архіви. – 2025.

ОЦІНЮВАННЯ КІБЕРСТІЙКОСТІ РОЗПОДІЛЬЧОЇ СИСТЕМИ ЕНЕРГЕТИЧНОГО ОБ'ЄКТУ КРИТИЧНОЇ ІНФРАСТРУКТУРИ НА ОСНОВІ МЕРЕЖ БАЙЄСА

Личик В. В.[1], Гальчинський Л. Ю.[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна*
lychyk.vlad@gmail.com, hleonid@gmail.com

Останні тенденції показали, що цифровізація об'єктів критичної інфраструктури, зокрема в енергетичному секторі створює негативне поприще для важких наслідків від кіберінцидентів. Йдеться про прямі фізичні пошкодження підсистем та елементів критичної інфраструктури. Тому, наразі прослідковується чіткий відхід від загальної концепції мінімізації ризиків до постановки питання про забезпечення мінімально-прийнятного функціонального (штатного) режиму роботи інфраструктурних об'єктів.

Ключові слова: кіберстійкість, система розподілення електроенергії, байєсові мережі.

Вступ

Проведені раніше дослідження дозволили структурувати підхід кіберстійкості на рівні понятійного апарату. У ході цього аналізу [1] було виявлено, що найбільш відповідними системами для оцінки кібер-резильєнтності є підходи розроблені Національним інститутом стандартів (NIST) та Національною академією наук США (NAS). Водночас, необхідним наступним етапом стала формалізація, або ж метамоделювання на основі методології OMG [2]. Даний підхід дозволив перейти від якісних порівняльних характеристик існуючих фреймворків до отримання моделей різного рівня для реальних

прототипів інфраструктурних об’єктів. Зокрема, у розрізі задачі побудови кількісної моделі кіберстійкості перехід до цього етапу відбувався за допомогою такої схеми: 1) створення анотованої функціональної моделі; 2) отримання формальної кількісної моделі на основі байєсової мережі [3]. При цьому, анотована функціональна модель повинна складатися із суміщення двох діаграм UML: Діаграми прецедентів та Діаграми компонентів. Перша діаграма ілюструє активи, а друга демонструє взаємодії акторів з цими активами – для забезпечення процесу безперервного виконання критичних функцій [4]. Даний концепт застосування байєсових мереж пройшов успішну апробацію у сфері надійності систем [5], безпеці мереж [6] та прогнозування ризиків [7].

Основна частина

Якщо ж говорити про енергетичний сектор, то тут виділяють три основні критичні функції: генерація, розподілення та транспортування. Для перевірки можливостей даного підходу для отримання кількісної оцінки кіберстійкості була обрана локальна мережа (підсистема) найвищого рівня, в структурі вищезгаданого процесу моделювання, із сценарієм атаки з найбільшою потенційною шкодою для фізичного обладнання – тобто, для функціонального стану даного компоненту загальної системи. У даному випадку – системи розподілення та управління.

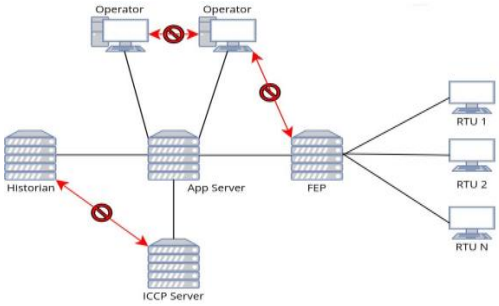


Рисунок 1. Прототип системи мережі розподілення

Використана формація представляє собою спрощене відображення структури системи розподілу електроенергії. Оскільки, концепція резильєнтності в першу чергу стосується здатності системи зберігати функціональність за несприятливих умов, що безпосередньо залежить від правильності її проєктування.

Обчислення було проведено на основі схематичної діаграми локальної мережі центру управління. Як сценарій була обрана зовнішня атака: після того, як зловмисник обійде апаратний брандмауер, використовуючи надійно сплановані стратегії вторгнення, він може отримати доступ до комутатора за допомогою методу сканування портів. На наступному етапі зловмисник може сканувати хости та сервери мережі, проникнувши на сервер історичних даних. В свою чергу, сервер застосунків, який використовується для зберігання даних та надсилання оновлених даних іншим клієнтам, таким як інтерфейс людини-машини (HMI), є якраз кінцевою ціллю вторгнення.

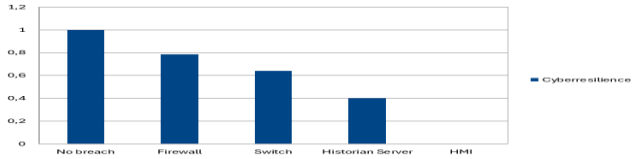


Рисунок 2. Діаграма показників кіберстійкості

Таблиця 1. Таблиця ймовірностей для вузлів критичної функції

Switch Evaluation				
	00	01	10	11
FALSE	1	0,6	0,4	0
TRUE	0	0,4	0,6	1

First evidence – Firewall
Second evidence – Switch Input

Historian server Evaluation				
	00	01	10	11
FALSE	1	0,6	0,4	0
TRUE	0	0,4	0,6	1

First evidence – Switch
Second evidence – Historian Server Input

HMI Evaluation				
	00	01	10	11
FALSE	1	0,6	0,4	0
TRUE	0	0,4	0,6	1

First evidence – Historian Server
Second evidence – HMI Input

Наслідки подібної атаки можуть мати вкрай серйозний і практичний характер. Використання вразливостей у системах розподілу зловмисниками здатне спричинити масові відключення електроенергії, що торкнуться сотень тисяч користувачів. Більше того, шкідливе ПЗ може не лише призвести до знеструмлення, а й ускладнити процес відновлення живлення, а також викликати пошкодження обладнання шляхом його програмного знищення.

Висновки

Проведене дослідження продемонструвало, що одним із підходів для кількісного оцінки кіберстійкості та подальшої розробки фреймворку може стати обчислення критичних функцій енергетичного об'єкта, а саме індикатора кіберстійкості, на основі попереднього моделювання на основі мереж Байєса. Зокрема, застосування даного концепту до аспекту розподілення електроенергії. Втім, все ще гостро стоїть питання апробації та подальшої імплементації даного підходу обчислення кіберстійкості не лише для розподільчої функції, а й загалом для енергосистеми - із застосуванням реальних датасетів.

Список використаних джерел

1. Гальчинський Л., Личик В. Метрики оцінки кібервідмовостійкості (аналітичне оглядове дослідження) // Інформаційні технології та суспільство. – 2023. – Т. 8, № 2. – С. 27–33. – DOI: 10.32689/maup.it.2023.2.3.

2. Bellini, E., Marrone, S., Marulli, F. Cyber resilience meta-modelling: The railway communication case study. Electronics (Switzerland). – 2021. 10, 1–26. DOI: 10.3390/electronics10050583.

3. Bayesian Network Models in Cyber Security: A Systematic Review / S. Chockalingam, W. Pieters, A. Teixeira, P. Gelder. – 2017. – DOI: 10.1007/978-3-319-70290-2_7.

4. Ben-Gal, I. (2007). Bayesian Networks. Encyclopedia of Statistics in Quality and Reliability,F. Ruggeri, F. Faltin and R. Kenett (eds.), John Wiley & Sons.

5. Frigault, M., Wang, L. Measuring network security using Bayesian network-based attack graphs. IEEE (2008). DOI: 10.1145/1456362.1456368.

6. N. Fenton, M. Neil, Risk assessment and decision analysis with Bayesian networks, CRC Press, 2012.

7. Herland, K., Hammainen, H., Kekolahti, P. Information security risk assessment of smartphones using Bayesian networks. J. Cyber Secur. Mobility 4, 65–85 (2016). DOI: 10.13052/jcsm2245-1439.424

АТАКИ GRAPHQL API ТА ЗАХИСТ ВІД НИХ

Сергеєв М. С.[1], Коломицев М. В.[1], Носок С. О.[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, maksymserg@gmail.com, box144.85@gmail.com, nos.sv.ol@gmail.com*

У роботі розглянуто особливості безпеки GraphQL API та виклики, що виникають під час його використання. Запропоновано підхід до динамічного обмеження частоти запитів на основі алгоритму AIMD, що дозволяє підвищити стійкість API без шкоди для легітимного трафіку.

Ключові слова: GraphQL, безпека API, обмеження частоти, адаптивний алгоритм

Вступ

GraphQL дає велику гнучкість формування запитів, але водночас відкриває можливість створювати ресурсоємні запити, не звертаючись до декількох кінцевих точок (endpoints). Типовими загрозами є:

– DoS через запити з великою кількість вкладень (великою глибиною) або alias-based DoS – запит однакових даних за допомогою різних alias в одному запиті;

– GraphQL ін’єкції через некоректну або відсутню обробку аргументів;

– недостатній контроль авторизації.

1. Статичний rate limiting

Один з найпоширеніших способів захисту від DoS атак – це встановлення фіксованого порогу запитів для клієнтів за проміжок часу. Однак такий підхід має свої недоліки:

– не враховує реальне навантаження на систему в конкретний момент;

– може недовикористовувати ресурси, або, навпаки, виявитися безсилим проти складних атак.

2. Динамічний rate limiting

Динамічне обмеження частоти запитів – більш гнучкий підхід, при якому система автоматично адаптує порогові значення кількості запитів за одиницю часу в реальному часі.

Існує декілька алгоритмів динамічного регулювання трафіка. В основі всіх лежить принцип зворотнього зв'язку (feedback loop) – моніторинг метрик і на їх основі корегування лімітів. Розглянемо декілька підходів:

– адаптивне порогове обмеження – найпростіший варіант, коли задаються фіксовані ліміти при певних значеннях метрик;

– адаптивний Token Bucket – в адаптивному варіанті швидкість поповнення токенів динамічно змінюється. Поки є токени у "відрі", запити пропускаються. Якщо середнє значення метрики виросло вище порогу, швидкість видачі токенів зменшується, інакше – збільшується;

– метод AIMD (Additive Increase, Multiplicative Decrease) – поки система в межах допустимого значення метрики, плавно збільшується пропускна здатність на фіксоване значення, але щойно помічається перевищення порогу – різко зменшується на певний відсоток.

3. Розробка адаптивного обмежувача

Для реалізації обмежувача був обраний метод AIMD. Система вимірює час обробки кожного запиту і обчислює експоненційне ковзне середнє (EMA) часу відповіді. На базі цього приймається рішення як змінювати порогові значення.

EMA використовується з метою зменшення впливу поодиноких аномальних значень затримки (шумів) і обраховується за наступною формулою:

$$EMA_{new} = \alpha \cdot L_{cur} + (1 - \alpha) \cdot EMA_{old}$$

де L_{cur} – час відповіді останнього запиту; EMA_{old} – поточне значення EMA часу відповіді; $\alpha \in (0,1)$ – коефіцієнт згладжування, що визначає вагу нових даних.

4. Результати тестування

Тестування проводилося за наступною методикою. Кожну хвилину на API відправлялося 180 запитів. Упродовж тестування надсилалися як прості запити, які оброблялися API дуже швидко, так і складні, які імітували реальні атаки.

З візуалізації результатів тестування (рис. 1,2,3) спостерігається, що, через неможливості адаптації у статичного обмежувача, на 11-12 хвилинах накопичення запитів спричиняє стрибок середньої затримки до 300 секунд. Система не помічає зміни в навантаженні, що знижує ефективність захисту та якість обслуговування клієнтів.

Натомість, динамічний варіант рівномірно реагує на погіршення часу відповіді – вже з 6-ї хвилини відбувається зниження частоти запитів. Після стабілізації навантаження система автоматично відновлюється. У порівнянні зі статичним обмежувачем – менше відхилених запитів, менша середня затримка, краща стійкість.

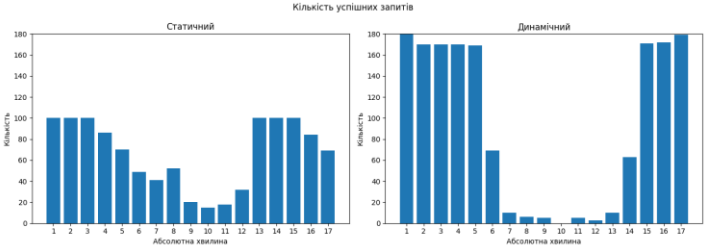


Рисунок 1. Порівняння кількості успішно виконаних запитів за хвилину

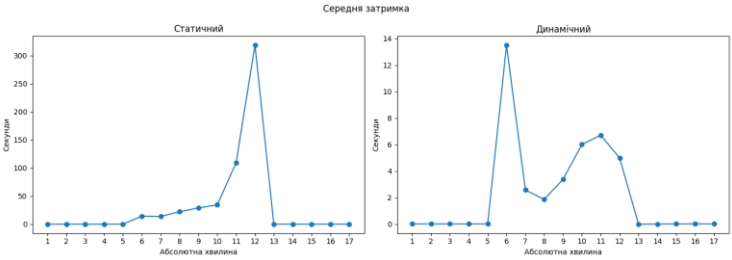


Рисунок 2. Порівняння середнього часу відповіді на запити за хвилину

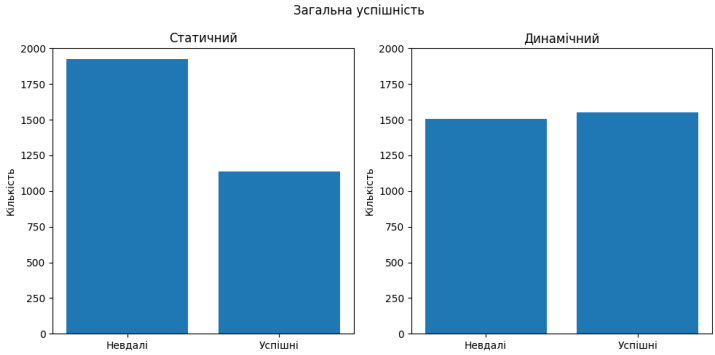


Рисунок 3. Загальне порівняння успішності запитів за період тестування

Висновки

Запропонована методика динамічного обмеження частоти запитів на основі алгоритму AIMD є перспективною альтернативою класичному статичному підходу. Вона продемонструвала високу ефективність у регулюванні навантаження без шкоди для легітимного трафіку. Отримані результати можуть слугувати основою для побудови повноцінної масштабованої системи захисту, адаптованої до змінного трафіку та сучасних загроз.

Список використаних джерел

1. Aleks, N. and Farhi, D., 2023. Black Hat GraphQL: Attacking Next Generation APIs. No Starch Press.

2. Corey Ball. Hacking APIs – Early Access, 2022.

3. OWASP Foundation. API Security Top 10, 2023. URL: https://owasp.org/API-Security/editions/2023/en/0x11-t10/

4. Xu A. System Design Interview – Volume 2, 2020.

5. Netflix TechBlog. Adaptive Concurrency Control, 2018. URL: https://netflixtechblog.medium.com/performance-under-load-3e6fa9a60581

МЕРЕЖЕВА СИСТЕМА ОПОВІЩЕННЯ ПРО ПОВІТРЯНУ ТРИВОГУ

Влах-Вигриновська Г.І.[1], Рудий Ю.П.[1]

[1]*Національний університет "Львівська політехніка", Україна,*
e-mail: halyna.i.vlakh-vyhrynovska@lpnu.ua

Вступ

В умовах збройного конфлікту ефективність попередження цивільного населення про небезпеку відіграє ключову роль у збереженні життів. Традиційні централізовані системи оповіщення вразливі до кібератак, руйнувань або перебоїв у електроживленні. У зв'язку з цим виникла потреба у створенні альтернативних, надійних та адаптивних систем оповіщення [1].

Основним чинником стала модернізація мережі Укртелеком із централізованих АТС на активні шафи Huawei (у Львові замінено 36 АТС на 450 активних шаф). Мідні кабельні мережі були замінені на оптичні, стало не можливо використання обладнання П-16Х виготовленого у СССР [2]. Одним із перспективних рішень є децентралізована мережа, що дозволяє вузлам (нодам) передавати послідовно сигнали одне одному за відсутності центрального вузла через радіоефір.

Mesh-мережа – це топологія, у якій кожен вузол безпосередньо зв'язується з кількома іншими вузлами, створюючи розподілену мережу [3]. Мережа забезпечує високу відмовостійкість і самовідновлення: у разі втрати одного вузла, трафік автоматично перенаправляється через інші шляхи. У порівнянні з зіркоподібною чи деревоподібною мережею, mesh-мережа має кращу стабільність у надзвичайних ситуаціях і менш залежна від одного елемента управління. Це особливо важливо під час ракетних або авіаударів, коли окремі ділянки інфраструктури можуть бути зруйновані.

Алгоритм роботи

Сигнал тривоги ініціюється з головного або одного із запасних вузлів і миттєво поширюється мережею через декілька. Ноди автоматично фіксують сигнал і передають далі, поки вся мережа не буде охоплена. Під час отримання команди про запуск сповіщення перевіряється цілісність даних, і запускається звукове сповіщення з допомогою електросирен або звукових репродукторів.

Багато підприємств, установ (об’єкти критичної інфраструктури, лікарні, магазини, мобільні станції, логістичні склади) мають власні автономні джерела живлення [4]. У випадку відключення електроенергії такі джерела можуть слугувати джерелом енергії для обладнання системи оповіщення, дозволяють нодам mesh-мережі не втрачати зв'язок.

Підключення через радіомережу може бути додатковим каналом зв’язку, але не головним. Основним каналом варто проєктувати підключення через оптичну мережу (PON) [6] через мережу Інтернет або з використання ізольованої внутрішньо корпоративної мережі.

У тилових містах, на відміну від прифронтових зон, використання засобів радіоелектронної боротьби (РЕБ) суттєво менше або епізодичне, що зумовлено відсутністю прямої загрози військових дій. Урахування РЕБ при проєктуванні мережі дозволяє ефективніше планувати топологію і вибирати оптимальні частотні діапазони.

Використання ліцензійного діапазону частот, закріпленого за ДСНС, є доцільним і ефективним рішенням для побудови надійної, захищеної та централізовано-керованої системи оповіщення населення.

Для захисту від кібератак на систему оповіщення доцільним є впровадження апаратних криптомодулів для перевірки автентичності повідомлень. Такі модулі, наприклад на базі ATECC608A забезпечують апаратне зберігання криптографічних ключів і виконання операцій цифрового підпису [5]. При отриманні сигналу тривоги модуль здійснює перевірку цифрового підпису повідомлення, гарантуючи, що команда надійшла від

авторизованого джерела і не була підроблена. Це критично важливо в умовах можливого зловмисного втручання або імітації сигналів у відкритих мережах. Апаратна реалізація криптографії забезпечує вищу швидкість обробки і захист від програмних атак порівняно з програмними реалізаціями, що особливо актуально для вбудованих систем з обмеженими ресурсами.

Для забезпечення коректної роботи децентралізованої системи оповіщення та уникнення дублювання або відтворення застарілих команд важливо впровадити механізм синхронізації часу або ідентифікацію унікальності повідомлень.

Серед найпоширеніших підходів – синхронізація вузлів за допомогою GPS-модуля, який забезпечує високу точність (до мікросекунд) незалежно від стану мережі, а також NTP (Network Time Protocol), що дозволяє синхронізувати час через інтернет або внутрішній інтра-мережевий сервер. У разі недоступності зовнішніх джерел або для додаткового захисту доцільно реалізувати лічильник повідомлень (наприклад, 16- або 32-бітовий increment ID), який дозволяє вузлам порівнювати номер нового пакета із збереженим останнім та ігнорувати повтори або застарілі пакети. Такий механізм підвищує надійність та стійкість до атак повторення (replay attacks), особливо в умовах ворожих спроб зірвати роботу мережі. У складних системах можливе поєднання: GPS для початкової синхронізації, NTP – для підтримки точності, лічильник — як дублюючий захист логіки обробки команд.

Висновок

Mesh-мережа є ефективною архітектурою для побудови децентралізованих систем оповіщення завдяки незалежності від центрального вузла. Надійність функціонування системи підвищується за рахунок використання автономних джерел живлення на об’єктах. Для захисту від кібератак впроваджено апаратні криптомодулі. Синхронізація часу через уникати повторів та атак типу replay.

Список використанних джерел

1. План заходів щодо реалізації Концепції розвитку та технічної модернізації системи централізованого оповіщення про загрозу виникнення або виникнення надзвичайних ситуацій, затверджений розпорядженням Кабінету Міністрів України від 11 липня 2018 р. №488-р.

2. Рудий Ю. (2020). УВАГА ВСІМ! Система оповіщення ЦО. Блог UA ID. Взято з: https://blog.uaid.net.ua/emergency-alert-system

3. Cilfone, A.; Davoli, L.; Belli, L.; Ferrari, G. Wireless Mesh Networking: An IoT-Oriented Perspective Survey on Relevant Technologies. Future Internet 2019, 11, 99. https://doi.org/10.3390/fi11040099

4. Суходоля О. М. Стійкість критичної енергетичної інфраструктури та життєдіяльності громад : аналіт. доп. – Київ : НІСД, 2024. – 160 с. – https://doi.org/10.53679/NISS-analytrep.2024.04

5. Odinachi Udemezuo Nwankwo, Hee-Jae Shin, Muhammad Rasyid Redha Ansori, Jae-Min Lee, & Dong-Seong Kim (2025). Blockchain Private Key Storage System for IoT Device. The Journal of Korean Institute of Communications and Information Sciences, 50(2), 378-393. 10.7840/kics.2025.50.2.378

6. Takayoshi Tashiro, Hideaki Kimura, Proposal of super low power consumption detector for emergency communication downstream system in FTTH, IEICE Electronics Express, 2010, Volume 7, Issue 17, Pages 1317-1322, Released on J-STAGE September 10, 2010, Online ISSN 1349-2543, https://doi.org/10.1587/elex.7.1317

КРИТЕРІЇ КВАНТОВИХ ОБЧИСЛЕНЬ У ЗАДАЧАХ КРИПТОАНАЛІЗУ

Котух Є. В.[1]

[1]*Воєнна академія імні Євгена Березняка, Київ, Україна*
yevgenkotukh@gmail.com

Дослідження аналізує вплив розвитку квантових обчислень на криптоаналіз. Розглядаються етапи еволюції квантових технологій: NISQ-ера (обмежені шумні системи), EFTQC (ранні відмовостійкі системи) та FTQC (повноцінна корекція помилок). Низька масштабованість, високий рівень шуму та необхідність тисяч коректованих кубітів у NISQ обмежують практичні квантові атаки, але прогрес у EFTQC/FTQC може прискорити компрометацію традиційної криптографії. Підкреслюється роль математичних моделей (наприклад, параметра масштабованості α) у прогнозуванні часових рамок загроз. Запропоновано порівняльний аналіз алгоритмів (VQE, QAOA, QPE), характеристик кубітів та показників якості операцій, що підтверджує необхідність переходу до постквантових стандартів. Сучасні NISQ-системи недостатні для криптоаналізу, проте вже EFTQC дозволяє модифіковані атаки зі зниженою складністю, а FTQC відкриває шлях до експоненційного прискорення, що вимагає термінового оновлення криптографічної інфраструктури.

***Ключові слова*:** квантові обчислення, постквантова криптографія, алгоритм Шора, NISQ, EFTQC, FTQC, криптоаналіз, критерії квантових обчислень.

NISQ-ера (Noisy Intermediate-Scale Quantum) є перехідною фазою розвитку квантових обчислень, що характеризується наявністю пристроїв із обмеженою кількістю кубітів та високим рівнем шуму, що унеможливлює повноцінну реалізацію алгоритмів із квантовою корекцією помилок. У відповідь на ці обмеження в рамках NISQ-парадигми було розроблено

низку спеціалізованих алгоритмів та обчислювальних стратегій, здатних працювати в умовах нестабільного квантового середовища. У період NISQ-ери ці системи ще не здатні виконувати обчислення з довгими ланцюгами квантових гейтів, необхідні для реалізації квантових криптоаналітичних атак на практиці. Наприклад, розв'язання задачі факторизації 2048-бітного RSA-ключа вимагатиме тисячі коректованих від помилок кубітів та мільйони логічних гейтів, що виходить далеко за межі сучасних NISQ-систем.

Квантові обчислення мають значно сильніший вплив на алгоритми з відкритим ключем, ніж на симетричні криптографічні схеми. Асиметричні алгоритми, такі як RSA, DSA та алгоритми на еліптичних кривих, базуються на складності математичних задач факторизації, дискретного логарифмування та логарифмування на еліптичних кривих. Відповідно, ці методи є уразливими до атак криптографічно релевантних квантових комп'ютерів (CRQC), які передбачають наявність тисяч коректованих від помилок кубітів. Алгоритм Гровера є ще одним прикладом квантової загрози. Він забезпечує квадратичне прискорення для пошуку в невпорядкованих структурах, знижуючи складність з $O(N)^3$ до $O(N)O\left(\sqrt{N}\right)O(N)$. Хоча така перевага не настільки драматична, як у Шора, вона може мати серйозні наслідки для симетричних криптосистем.

Розглянемо ключові математичні моделі, що формалізують процес розвитку NISQ, EFTQC та FTQC обчислювальних систем (табл. 1). Одним з ключових математичних описів переходу між різними епохами квантових обчислень є модель масштабованості. При значеннях параметра масштабованості менших за 2, спостерігається стрімке зростання загального рівня помилок із збільшенням кількості кубітів у системі. Даний режим характерний для сучасного етапу розвитку квантових обчислень з використанням NISQ-пристроїв. Основними чинниками, що зумовлюють низьке значення α, є недостатня ізоляція кубітів від навколишнього

середовища, наявність перехресних перешкод між сусідніми кубітами (crosstalk), обмежена точність контролюючих сигналів та високий рівень декогеренції. При таких значеннях параметра масштабованості застосування квантової корекції помилок є неефективним, оскільки процедури виявлення та виправлення помилок вносять більшу кількість нових помилок, ніж здатні виправити.

Таблиця 1. Еволюція квантових обчислень

Критерій	NISQ (Noisy Intermediate-Scale Quantum)	EFTQC (Early Fault-Tolerant Quantum Computing)	FTQC (Fault-Tolerant Quantum Computing)
Часові рамки практичних реалізацій	2018 – 2025	2025 – 2030	2030+
Класи складності	Слабший за BQP, сильніший за BPP	Між NISQ і BQP	BQP
Теорема про поріг помилки	Не застосовується (помилки вищі за поріг)	Частково застосовується (близько до порогу)	Повністю застосовується (нижче порогу)
Математична характеристика фізичної помилки	$P_{error} > P_{th}$	$P_{error} \approx P_{th}$	$P_{error} \ll P_{th}$
Модель масштабованості (scalability)	$\alpha < 2$	$2 \le \alpha \le 3.5$	$\alpha > 3.5$
Принцип корекції помилок	Пом'якшення помилок	Обмежена корекція помилок	Повна корекція помилок

Діапазон значень параметра масштабованості від 2 до 3,5 відповідає перехідному режиму ранніх відмовостійких квантових обчислень (EFTQC). У цьому режимі зростання рівня помилок при збільшенні кількості кубітів є помірним, що дозволяє імплементувати обмежені схеми корекції помилок з позитивним ефектом. Проаналізуємо критерії

співвідношення між фізичними та логічними кубітами в реалізації задач криптографічних квантових обчислень (див. табл. 2).

Таблиця 2. Оцінки характеристик кубітів для систем NISQ, EFTQC та FTQC

Характеристика	NISQ	EFTQC	FTQC
Кількість фізичних кубітів	50-1,000	10^3-10^6	$>10^6$
Кількість логічних кубітів	0	10-100	100-10,000+
Формула для кількості фізичних кубітів на логічний кубіт	не існує	$O(d^2)$	$> O(d^2)$
Рівень фізичної помилки	10^{-2}-10^{-3}	10^{-3}-10^{-4}	$<10^{-4}$
Рівень логічної помилки	N/A	$\sim p_{phys}^{d/2}$	$\sim p_{phys}^{d/2}$

Перехід до EFTQC-ери передбачає драматичне збільшення кількості фізичних кубітів (10^3 – 10^6) та реалізацію обмеженої кількості логічних кубітів (10 – 100). Це дозволяє імплементувати часткову корекцію помилок та виконувати квантові алгоритми середньої складності. У таких системах кількість фізичних кубітів на один логічний кубіт масштабується як $O(d^2)$, де d – відстань коду, а рівень фізичних помилок знижується до 10^{-3} – 10^{-4}. FTQC-системи характеризуються використанням понад мільйона фізичних кубітів для забезпечення сотень і тисяч логічних кубітів (100 – 10,000+) з повноцінною корекцією помилок.

Проаналізуємо квантові обчислювальні підходи (табл. 3).

Табл. 3 демонструє прогресію від більш обмеженого, але практично доступного NISQ-підходу до теоретично потужнішого, але технологічно вимогливішого FTQC.

Проаналізуємо показники якості квантових операцій (табл. 4).

Таблиця 3. Порівняння квантових обчислювальних підходів

Підхід	Обчислювальна парадигма	Математична складність схеми	Логічні кубіти	Глибина схеми
VQE та QAOA (NISQ)	Гібридна	$O(p)$ з глибиною p	Фізичні кубіти	Обмежена
RFE (EFTQC)	Робастні квантові алгоритми	$O(K)$ з параметром K	Обмежена кількість	Середня
QPE (FTQC)	Повна квантова корекція помилок	$O(1/\varepsilon)$ для точності ε	Необмежена	Висока

Таблиця 4. Порівняння показників якості квантових операцій

Метрика	NISQ	EFTQC	FTQC
Точність двокубітного гейта	99 – 99,9 %	99,9 – 99,999 %	>99,9999 %
Кількісна формула для Quops	KiloQuops $\approx 10^3$	MegaQuops $\approx 10^6$ – GigaQuops $\approx 10^9$	TeraQuops $\approx 10^{12}$
Час когерентності відносно часу операції	$T_2 / T_{gate} \approx 10^2 - 10^3$	$T_2 / T_{gate} \approx 10^3 - 10^5$	$T_2 / T_{gate} > 10^5$

Аналіз показників якості квантових операцій демонструє радикальне зростання продуктивності та надійності квантових систем у процесі переходу від ери шумних проміжних квантових пристроїв (NISQ) до ери повноцінних відмовостійких квантових обчислень (FTQC). Проаналізуємо EFTQC та FTQC підходи як перспективні з огляду на рішення криптоаналітичних задач (табл. 5).

Порівняльний аналіз підходів EFTQC та FTQC (табл. 5) має особливе значення в контексті криптоаналітичних задач, де квантові алгоритми потенційно демонструють експоненційне прискорення порівняно з класичними методами. Принципові відмінності цих підходів визначають різні траєкторії розвитку криптоаналітичних можливостей у квантову еру.

Таблиця 5. Порівняльний аналіз EFTQC та FTQC у вирішенні задач криптоаналізу

Характеристика	Традиційний FTQC підхід	EFTQC підхід	Математичне співвідношення
Кількість операцій на схему	Висока	Знижена в ≈ 100 разів	$O\left(T_{FTQC}/100\right)$
Кількість запусків схеми	Низька	Збільшена в ≈ 10000 разів	$O\left(R_{FTQC}\cdot 10^4\right)$
Загальний час виконання	$T_{total}=T_{FTQC}\cdot R_{FTQC}$	$T_{total}=T_{EFTQC}\cdot R_{EFTQC}$	Збільшення в ≈ 100 разів
Розмір задачі (досяжність)	Обмежений кількістю кубітів	Розширений на $\approx 40-50\%$	Від ≈ 90 до ≈ 130 логічних кубітів

EFTQC підхід із характерним зниженням кількості операцій на схему $O\left(T_{FTQC}/100\right)$ відкриває можливість для раннього впровадження модифікованих версій алгоритму Шора та алгоритму Гровера з обмеженою глибиною. Критичною особливістю такого підходу для криптоаналізу є компенсація обмеженої глибини схем через інтенсивне статистичне накопичення результатів $O\left(R_{FTQC}\cdot 10^4\right)$ запусків. Це дозволяє атакувати криптографічні системи з редукованою складністю, вибірково таргетуючи вразливі компоненти або використовуючи часткове знання про структуру криптографічного ключа.

Збільшення розміру задач на 40–50 % (від ~90 до ~130 логічних кубітів) у EFTQC підході є особливо значущим для криптоаналізу, оскільки уможливлює атаки на

криптосистеми з більшою довжиною ключа. Проведений аналіз математичних критеріїв різних епох розвитку квантових обчислень має наслідки для вирішення криптоаналітичних задач, трансформуючи розуміння часових рамок та методологічних підходів до подолання криптографічного захисту класичних криптосистем. Еволюція параметра масштабованості α від α < 2 у NISQ до α > 3,5 у FTQC системах визначає фундаментальні обмеження для імплементації алгоритму Шора та інших криптоаналітичних алгоритмів, що вимагають низького рівня помилок. Перехід від рівня характеристики фізичної помилки $P_{error} > P_{th}$, характерної для NISQ, до рівня $P_{error} \ll P_{th}$ у FTQC є критичною точкою біфуркації для практичної реалізації повномасштабних квантових атак на асиметричні криптосистеми в недалекому майбутньому. Співвідношення між фізичними та логічними кубітами, що описується функцією $O\left(d^2\right)$, має вирішальне значення для криптоаналізу, оскільки визначає максимальний розмір криптографічних ключів, які можуть бути атаковані. Еволюція від VQE та QAOA алгоритмів через робастні EFTQC-алгоритми до повноцінних FTQC-алгоритмів трансформує методологію квантового криптоаналізу. В NISQ-ері можливі лише гібридні атаки з обмеженою квантовою складовою; EFTQC-ера відкриває шлях до робастних квантових атак зі складністю $O(K)$, які, хоча й не є оптимальними, але вже становлять реальну загрозу для окремих криптосистем; і, нарешті, FTQC-ера уможливлює повномасштабні квантові атаки зі складністю $O(1/\varepsilon)$, що повністю компрометують уразливі криптографічні примітиви.

Запропонована модель переходу від NISQ до FTQC із чіткими математичними критеріями дозволяє більш точно прогнозувати часові рамки виникнення квантової загрози для різних криптографічних систем.

Список використаних джерел

1. Arute F., Arya K., Babbush R., Bacon D., Bardin J. C., Barends R., Biswas R., Boixo S., Brandao F. G. S. L., Buell D. A., Burkett B., Chen Y., Chen Z., Chiaro B., Collins R., Courtney W., Dunsworth A., Farhi E., Foxen B., Martinis J. M. Quantum supremacy using a programmable superconducting processor // Nature. 2019. Vol. 574(7779). P. 505–510. https://doi.org/10.1038/s41586-019-1666-5
2. Kandala A., Mezzacapo A., Temme K., Takita M., Brink M., Chow J. M., Gambetta J. M. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets // Nature. 2017. Vol. 549(7671). P. 242–246. https://doi.org/10.1038/nature23879
3. Kotukh E., Severinov O., Vlasov A., Tenytska A., Zarudna E. Some results of development of cryptographic transformation schemes using non-abelian groups // Radiotekhnika. 2021. No 204. P. 66–72. https://doi.org/10.30837/rt.2021.1.204.07
4. Khalimov G., Kotukh Y., Sergiychuk Y., Marukhnenko A. Analysis of the implementation complexity of the cryptosystem on the Suzuki group // Radiotekhnika. 2018. No 193. P. 75–81.
https://doi.org/10.30837/rt.2018.2.193.08
5. Kotukh Y., & Khalimov G. Hard problems for non-abelian cryptography // 2021: Fifth International Scientific and Technical Conference "computer and information systems and technologies" https://doi.org/10.30837/csitic52021232176
6. Kotukh Y., Khalimov G. Towards practical cryptoanalysis of systems based on word problems and logarithmic signatures // Information security: problems and prospects. https://shorturl.at/IaByX
7. Shor P. W. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer // SIAM Journal on Computing. 1997. No 26(5). P. 1484–1509. https://epubs.siam.org/doi/10.1137/S0097539795293172
8. Gidney C., & Ekerå M. How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. Quantum. 2021. No 5. 433 p. https://doi.org/10.22331/q-2021-04-15-433

9. Kotukh Y., Khalimov G. Advantages of logarithmic signatures in the implementation of crypto primitives // Challenges and Issues of Modern Science. https://cims.fti.dp.ua/j/article/download/119/158
10. Kotukh Y. Quantum cryptoanalysis of prospective asymmetric cryptosystems // Proceedings of conference "Cybersecurity in energy sector". https://shorturl.at/1pbcK

SWOT-АНАЛІЗ У КОНТЕКСТІ ТЕОРІЇ РИЗИКІВ ТА ТЕОРІЇ КОРИСНОСТІ ДЛЯ ВИБОРУ ОПТИМАЛЬНОЇ СТРАТЕГІЇ РОЗВИТКУ У КІБЕРСИСТЕМАХ

Полуциганова В. І.[1], Смирнов С. А.[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна*
medvika@ukr.net

У цій статті запропоновано нову інтерпретацію SWOT-аналізу через призму двох фундаментальних економічних концепцій: теорії ризиків і теорії корисності. Такий підхід дозволяє поглибити аналітичний зміст SWOT, підвищити точність оцінки стратегічних рішень і наблизити його до формалізованого моделювання. На прикладі побудови матриці співвідношень між сильними сторонами та можливостями, продемонстровано, як можна використовувати методологію оцінки вигід і ризиків у практиці стратегічного менеджменту.

***Ключові слова*:** SWOT-аналіз, функція корисності, оцінка ризику, кібербезпека, стратегія розвитку

1. Вступ

SWOT-аналіз – один із найбільш вживаних інструментів стратегічного планування, який дозволяє оцінити внутрішні характеристики організації (сильні та слабкі сторони) та зовнішні чинники середовища (можливості та загрози). У класичному вигляді він є якісним методом системного аналізу. Однак при ухваленні управлінських рішень з високим ступенем невизначеності, важливо враховувати кількісні оцінки ризиків та очікуваної корисності рішень [1, 7].

В даному дослідженні було розглянуто концепцію використання теорії корисності та оцінювання ризиків для

запровадження регулярної стратегії розвитку системного підходу до кібербезпеки в компаніях пов'язаних з IT-технологіями [5, 6]. Було запропонована симетричні підходи для оцінювання очікуваної корисності та можливих втрат з використанням експертних оцінок для відповідних комбінацій сильних сторін і загроз. Було розглянуто наступні підходи.

2. Теорія корисності у SWOT

Теорія корисності описує, як суб'єкти приймають рішення на основі очікуваного виграшу або переваги, яку можна отримати в результаті певної дії [2, 4]. У контексті SWOT для дослідження стратегії розвитку у підходах кібербезпеки:

- Сильні сторони (Strengths) – це наявні ресурси або компетенції, які можуть бути активами в досягненні можливостей розвитку.
- Можливості (Opportunities) – це потенційні джерела доданої вартості або стратегічної вигоди внаслідок використання наявних ресурсів або компетенцій.

Таким чином, поєднання S + O може бути інтерпретоване як функція очікуваної корисності, де:

$$U(S_i) = \sum_j E[i][j] \cdot V(O_j),$$

де $E[i][j]$ – частковий вплив реалізації j можливості за умови наявності відповідної i сильної сторони, якщо розглядається сукупність сильних сторін, величина впливу визначена експертними методами; $V(O_j)$ – цінність можливості.

3. Теорія ризиків у SWOT

З іншого боку, загрози (Threats) і слабкі сторони (Weaknesses) є джерелами ризику – імовірності негативного впливу на розвиток у підходах кібербезпеки [3].

Введемо позначення:

- W – внутрішні вразливості, які можуть посилити наслідки зовнішніх загроз;

- T – зовнішні чинники, які можуть зашкодити організації.

Оцінка ризику може бути описана як [8]:

$$R(W_i, T_j) = P_{impact}(T_j \mid W_i) \cdot C(T_j),$$

де P_{impact} – ймовірність, що загроза T_j реалізується при наявності слабкості W_i, $C(T_j)$ – потенційні втрати.

4. Побудова матриць корисності та ризику

Для переходу до формалізованого аналізу пропонується побудова двох матриць співвідношення:

– SO-матриця корисності: зв'язок між сильними сторонами та можливостями;
– WT-матриця ризику: зв'язок між слабкими сторонами та загрозами.

Вхідними даними виступають проведений попередній аналіз та виявлення необхідної інформації та співвідношення між сильними сторонам та можливостями для оцінки корисності стратегій розвитку, а для оцінки ризиків виявлено співвідношення між вразливостями та загрозами.

Сильні сторони (S), що були виявлені експертами при аналізі організації:

– S1: Сертифіковані фахівці з кібербезпеки;
– S2: Розвинена внутрішня SOC (центр моніторингу подій безпеки);
– S3: Автоматизація реагування на інциденти.

Можливості (O) для впровадження кіберзахисту для організації:

– O1: Контракти з банками;
– O2: Міжнародні гранти;
– O3: Сертифікація ISO 27001;
– O4: Партнерство з держсектором.

Після отримання обхідних оцінок проведемо розрахунок загальної очікуваної корисності для стратегії розвитку у кіберсистемі хмарного середовища. Для того щоб розробляти стратегії використання сильних сторін необхідно встановити пріоритетність напрямків або можливостей розвитку певних сильних сторін. Через обмеженість ресурсів (зазвичай грошових), необхідно ввести вектор пріоритетності сильних сторін $W_S = (w_{s_1}, \dots, w_{s_n})$. У даному випадку w_{s_i} відображає важливість відповідної сильної сторони організації для досягнення необхідного рівня очікуваної корисності. Зазвичай цей вектор задається керівництвом організації, тобто експертами цієї галузі.

Тоді загальна функція корисності буде виглядати:

$$U_{total} = \sum_i w_{s_i} \cdot U(S_i)$$

Даний розрахунок для функції корисності враховує збалансоване використання ресурсів. Але для отримання максимальної користі подальшому можна використовувати ваги пріоритетності для пошуку кращого розподілу ресурсів.

Наступним кроком для проведення комплексного SWOT-аналізу є оцінювання ризиків. Для цього необхідно оцінити наявні вразливості та потенційні загрози, а також ймовірність їх реалізації. Нижче наведено вразливості та загрози для нашого прикладу:

Слабкі сторони (W), що були виявлення експертами при аналізі організації:

– W1: Відсутність політик резервного копіювання;
– W2: Недостатнє тестування безпеки в хмарі.

Загрози (T), що можуть виникати в процесі функціювання організації:

– T1: Ransomware-атаки;
– T2: Законодавчі штрафи за витік даних.

Потенційні втрати оцінюються в категоріальній шкалі і показує критичність наслідків від використання зловмисниками даної вразливості при реалізації загрози, відповідно оцінка ризику одночасно рівень втрати та ймовірність реалізації певного сценарію.

4. Сценарій оцінки ризиків

У сценарії, де організація не має належної системи резервного копіювання (W1) і не виконує тестування безпеки хмарної інфраструктури (W2), сукупний ризик розраховується як сума окремих *R*:

$$R_{total} = 0{,}9 + 0{,}54 + 0{,}64 + 0{,}7 = 2{,}78.$$

Це вказує на високий рівень ризику, що потребує термінового усунення слабких місць.

5. Застосування SWOT-аналізу з урахуванням теорії корисності та ризиків у стратегічному плануванні

У процесі стратегічного аналізу доцільно проводити рейтингову оцінку сильних сторін (S) і можливостей (O) за критерієм очікуваної корисності. Це дозволяє ідентифікувати найбільш перспективні комбінації факторів, які потребують пріоритетного інвестування ресурсів. Одночасно слід приділяти увагу комбінаціям слабких сторін і загроз (WT-зв'язків), які асоціюються з підвищеним ризиком. Основною метою є пошук шляхів компенсації або нейтралізації вразливостей через залучення зовнішніх можливостей або внутрішні зміни.

Наступним етапом є використання практичних аспектів реалізації. До них можна віднести:

- формування стратегічних пріоритетів;
- оптимальний розподіл ресурсів;
- динамічний моніторинг;
- сценарний аналіз.

Для підвищення достовірності оцінок доцільно проводити валідацію обраної метрики на основі історичних

даних, зокрема перевіряти наявність кореляції між розрахованими показниками та фактичними успішними результатами. Додатково рекомендується здійснити аналіз чутливості, що дозволить оцінити вплив змін вхідних параметрів на остаточний результат стратегічного вибору.

6. Висновки

В ході даного дослідження було показано, що SWOT-аналіз можна підсилити за допомогою концепцій теорії корисності та теорії ризиків. Це дозволяє перейти від описової до кількісної стратегії, де кожна комбінація S-O або W-T має оціночну вагу.

Такий підхід особливо цінний в умовах обмежених ресурсів або високої невизначеності, оскільки дозволяє:

- оцінити оптимальність дій,
- виявити пріоритетні напрямки розвитку,
- підготувати реалістичні антикризові стратегії

Загалом даний підхід допомагає перейти від якісних оцінок до кількості, а також оцінити взаємозв'язки між квадрантами SWOT-аналізу, для того щоб в подальшому використовувати більш аналітичні методи для розробок стратегій розвитку у тому числі в кібербезпеці.

Список використаних джерел

1. Hill T. and Westbrook R. SWOT analysis: It's time for a product recall. *Long Range Planning*, vol. 30, no. 1, pp. 46–52, 1997.
2. Keeney R. L. and Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
3. Gritzalis D. and Lambrinoudakis C. Assessing risk in modern IT infrastructures. *Information Management & Computer Security*, vol. 12, no. 1, pp. 36–43, 2004.
4. Saaty T. L. Decision making with the analytic hierarchy process. *Int. J. Services Sciences*, vol. 1, no. 1, pp. 83–98, 2008.

5. Humayun M., Niazi M., Jhanjhi N. Cybersecurity for Industrial Control Systems: A Survey, *Computers & Security*, vol. 89, p. 101677, 2020.

6. Cherdantseva Y. and Hilton J. A reference model of information assurance & security in *Proc. Int. Conf. Availability, Reliability and Security (ARES)*, Regensburg, Germany, 2013, pp. 546–555.

7. Haghi S. Motevali, Ahmadi M., Rahmani H. A., "SWOT-based strategic planning for cyber defense readiness," *Int. J. Inf. Security*, vol. 20, pp. 463–481, 2021.

8. ISO/IEC 27005:2018, *Information Security Risk Management*, International Organization for Standardization, Geneva, Switzerland, 2018.

SYMMETRIC ALGEBRAIC CIPHER

Gutik O. V.[1], Popadiuk O. B. [1], Vlasov V. A. [1]

[1]*Ivan Franko National University of Lviv*
Lviv, Ukraine
oleg.gutik@lnu.edu.ua, olha.popadiuk@lnu.edu.ua,
vitaly.vlasov@lnu.edu.ua

Herein we describe a symmetric algebraic cipher utilizing matrix representation of message symbols. Code implementation is provided.

Keywords: algebraic, modulo, Rust, Python, symmetric cipher

Introduction

We propose a symmetric algebraic cipher with the following structure.

First, we define cipher parameters:

- an alphabet A with size $\mathrm{L} \equiv |\mathrm{A}|$;
- a matrix $M$with size $\mathrm{N} \gg \mathrm{L}$;
- σ_0, σ_1 are some permutations $\mathrm{N} \to \mathrm{N}$;
- ϕ is some bit sequence of prime length P: $\phi \in \{0,1\}^{\mathrm{P}}$.

A triple $(\sigma_0, \sigma_1, \phi)$ will be our secret key.

We define each symbol z by a corresponding set of diagonals D in the matrix M, so that

$$\forall(\mathrm{x}, \mathrm{y}) \in \mathrm{D}: \mathrm{x} - \mathrm{y} = \mathrm{z} \pmod{\mathrm{L}}$$

(see Figure 1).

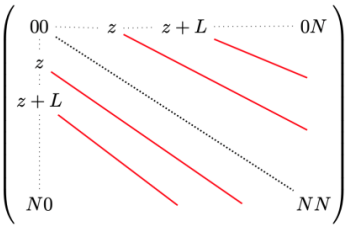


Figure 1

Encryption/Decryption algorithms

Here we describe the steps of the encryption algorithm.

Suppose we receive some text T containing symbols to be encoded.

For each $t_i \in T$, we first obtain its numeric representation $z_i \in [0, L)$.

Then we map each z_i to a pair of matrix coordinates (x_i, y_i) such that:

1. first, we pick a random $x_i \in [0, N)$ (e.g., horizontal coordinate in a matrix);
2. then, we randomly pick some $y_i \in [0, N)$ such that:

$$x_i - y_i = z_i \ (mod\ L).$$

Having thus obtained a sequence $\{(x_i, y_i),\ i \in [0, |T|)\}$, we now apply permutations $\sigma_k : [0, N) \to [0, N),\ k = 0,1$:

$$\text{ciphertext } (\xi_i, \eta_i) \ := \left(\sigma_{k_0}(x_i), \sigma_{k_1}(y_i)\right),$$

where

$$k_j = \phi[P\ mod\ (2 * i + j)].$$

Below are the steps of the decryption algorithm:

1. receive encoded ciphertext $\{(\xi_i, \eta_i)\}$;
2. apply inverse permutations: $(x_i, y_i) = \left(\sigma_{k_0}^{-1}(\xi_i), \sigma_{k_1}^{-1}(\eta_i)\right)$;
3. find z_i: $z_i = x_i - y_i \ (mod\ L)$.

Implementation

We have implemented a small library consisting of two modules:

- low-level Rust code for encryption/decryption parts;
- high-level Python wrapper.

These are implemented using PyO3 [1] - a library providing Rust bindings for Python extension modules. Its users include Qiskit, Python Cryptography package, and others.

To show an example of encoder output, consider the following cipher parameters:

- alphabet size: $L = 64$;
- matrix dimensions: $N = 1000$.

For these we obtain the following ciphertext:

Original string: Hello, world!

Generated pairs: [(60, 14), (746, 196), (489, 458), (620, 68), (947, 434), (943, 501), (18, 265), (913, 856), (196, 639), (965, 800), (28, 111), (693, 195), (22, 482), (614, 434), (367, 725), (932, 195), (102, 434), (388, 195)]

Conclusions

We have proposed and implemented an algebraic symmetric cipher relying on matrix representation of message symbols, using a set of permutations as its secret key. Python package with low-level Rust code is used for testing encryption and decryption algorithms.

References

1. https://github.com/PyO3/pyo3

ДОСЛІДЖЕННЯ ІНТЕГРАЦІЇ БЛОКЧЕЙНУ ДЛЯ БЕЗПЕЧНОГО ДЕЦЕНТРАЛІЗОВАНОГО УПРАВЛІННЯ ДАНИМИ IOT

Коломицев М. В.[1], Носок С. О.[1], Юрченко В. Б.[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, vikyur-ipt25@lll.kpi.ua*

У роботі досліджено можливості застосування технології блокчейн для побудови безпечної децентралізованої моделі управління даними в системах Інтернету речей (IoT). Проаналізовано загрози в IoT, типи блокчейну, види протоколів консенсусу, а також особливості реалізації на основі смарт-контрактів. Запропоновано концептуальну модель інтеграції, яка дозволяє досягти високого рівня безпеки, контролю власності та цілісності даних.

Ключові слова: блокчейн, Інтернет речей, протокол консенсусу, смарт-контракт, децентралізація, безпека

Вступ

Застосування блокчейну у сфері IoT вимагає врахування доступних обчислювальних і мережевих ресурсів, а також визначення відповідних типів блокчейн-систем і механізмів консенсусу. Ціллю цього дослідження є розробка моделі інтеграції блокчейну для безпечного децентралізованого управління даними пристроїв Інтернету речей.

Аналіз протоколів консенсусу

Протоколи консенсусу визначають, як учасники мережі погоджуються щодо запису нових блоків. В роботі аналізувалися наступні протоколи:

– Proof of Work (PoW);

– Proof of Stake (PoS);
– Delegated Proof of Stake (DPoS);
– Practical Byzantine Fault Tolerance (PBFT);
– Proof of Authority (PoA);
– Proof of Elapsed Time (PoET).

У роботі наведено порівняння описаних протоколів консенсусу.

Використання блокчейну в IoT

Попри численні переваги, впровадження блокчейну в IoT стикається з низкою труднощів:

– обмежені ресурси пристроїв;
– масштабованість;
– затримки;
– інтероперабельність.

Концепція моделі

Для побудови моделі було обрано *консорціумний блокчейн*, у якому обмежене коло учасників (організацій або пристроїв) мають право брати участь у валідації транзакцій.

Для досягнення консенсусу застосовується *PBFT*, який забезпечує узгодженість між вузлами навіть за наявності недовірених або скомпрометованих учасників (до 1/3). Це рішення забезпечує:

– високу швидкість узгодження блоків;
– енергоефективність, критичну для IoT;
– надійність у розподіленому середовищі зі статично відомими вузлами.

Складовими архітектури моделі є:

1. **IoT-пристрої** (сенсори, виконавчі механізми) – генерують дані і взаємодіють із блокчейном через шлюзи;
2. **Edge-шлюзи** – агрегують дані, підписують і надсилають транзакції, беруть участь у PBFT;

3. **Смарт-контракти** – реалізують політику доступу, логіку взаємодії пристроїв і правила реагування на події.

Усі вузли мають заздалегідь визначені сертифікати або ключі автентифікації.

В роботі наведено приклад смарт-контракту за принципом PBFT

Обґрунтування доцільності моделі:

1. Консорціумна архітектура знижує потребу в анонімності, водночас підвищуючи контрольованість і надійність;
2. Протокол PBFT забезпечує високу узгодженість і стійкість до наявності певної кількості скомпрометованих пристроїв;
3. Смарт-контракти забезпечують гнучке управління доступом і реакцією на події;
4. Відсутність високих енергетичних витрат у процесі підтвердження транзакцій робить модель придатною для пристроїв із обмеженими ресурсами.

Висновки

У межах дослідження було проаналізовано проблеми безпеки, притаманні IoT-системам, та обґрунтовано доцільність використання блокчейну як основи для побудови децентралізованої моделі управління даними. Блокчейн забезпечує нову парадигму управління інформацією, уникаючи централізованого контролю. Його тип і механізм консенсусу слід обирати з урахуванням вимог до продуктивності, рівня довіри та ресурсних обмежень середовища застосування.

Запропонована архітектура базується на консорціумному блокчейні з протоколом консенсусу PBFT, що забезпечує цілісність, контрольований доступ та автоматизовану взаємодію між пристроями. Включення смарт-контрактів дозволяє формалізувати політики безпеки та погодження дій у розподіленому середовищі.

Описана концепція придатна для створення безпечної системи управління даними IoT.

Список використаних джерел

1. Aditya K. S., Mohan S. G. Internet of Things Security. – 2024. – DOI: 10.1201/9781003460367-4.

2. ENISA. Baseline Security Recommendations for IoT. – 2017. – DOI: 10.2824/03228. – URL: https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot.

3. Парламент ЄС. Загальний регламент про захист даних (GDPR). – URL: https://gdpr-text.com/uk/.

4. Dave C. Security Challenges with Blockchain. – 2024.

5. Salimitari M., Chatterjee M., Fallah Y. P. A survey on consensus methods in blockchain for resource-constrained IoT networks // Internet of Things. – 2020. – Т. 11. – P. 100212. – ISSN 2542-6605. – DOI: 10.1016/j.iot.2020.100212.

6. Exploring Integration of the Blockchain and Internet of Things (IoT) Computing for Secure and Decentralized Data Management / T. K. Vashishth, V. Sharma, K. K. Sharma, B. Kumar, S. Chaudhary, R. Panwar. – 2024. – DOI: 10.1201/9781003460367-10.

ВАРІАНТ КРИПТОЗАХИСТУ СКУД З RFID-КАРТКАМИ MIFARE ULTRALIGHT

Федун В. Р.[1], Щербина М. Ю.[1]

[1]*Львівський національний університет імені Івана Франка, м. Львів, Україна*
vitalii.fedun@lnu.edu.ua, mykola.shcherbyna@lnu.edu.ua

Наведено типові загрози системам контролю і управління доступом (СКУД) з безконтактними картками MIFARE та запропоновано реалізацію криптографічного захисту СКУД з використанням найпростіших RFID-ідентифікаторів Ultralight, із механізмом протидії дублюванню карток доступу. Реалізовано прототип СКУД з використанням ПЗ для ОС Windows та USB-зчитувача ACS ACR1252U III.

Ключові слова: HMAC, KeeLoq, MIFARE Ultralight, RFID, атака повторного відтворення, криптографічний захист інформації, СКУД, технічний захист інформації

Вступ

Системи контролю та управління доступом (СКУД) – це комплексні апаратно-програмні рішення, що забезпечують безпеку об’єктів, у тому числі технічний захист інформації з обмеженим доступом (ІЗоД). Попри зростаючу популярність біометричної ідентифікації, RFID-системи залишаються конкурентними через помітно нижчу вартість зчитувачів. Водночас рівень їх безпеки значно варіюється. Якщо 125 кГц ідентифікатори EM-Marin дозволяють лише зчитувати наперед запрограмований код, що робить їх вразливими для емуляторів, таких як хакерський інструмент Flipper Zero або рішення на чипі T5577, то більш просунута 13,56 МГц технологія MIFARE від NXP дозволяє зчитування та прописування інформації разом з її захистом, у тому числі криптографічним.

В свою чергу, технологія MIFARE є збірною назвою для декількох технічних рішень, зокрема сімейств Classic та Ultralight. У першому внутрішню пам'ять обсягом 1 або 4 КіБ поділено на сектори, захищені двома ключами. У другому 64 байти пам'яті поділено на 4-байтові сторінки, а криптографічний захист (у початковій інкарнації) відсутній [1].

MIFARE Ultralight EV1 має 128 байтів, захист паролем та інші вдосконалення. Портфоліо MIFARE не обмежується зазначеними сімействами – є ще технологія DESFire тощо.

Атаки на RFID-ідентифікатори

Відносна технічна складність MIFARE не стала перепоною для успішних атак [2]. Криптографічний захист MIFARE Classic було повністю скомпрометовано [3]. Завдяки китайській мікроелектронній промисловості з'явилися і економічно доступні рішення для емуляції (дублювання) ідентифікаторів різних сімейств.

Однією з очевидних загроз є можливість перехоплення комунікації між зчитувачем і RFID-ідентифікатором за допомогою радіосніферів, зокрема SDR. Дослідження підтверджують можливість таких атак навіть при використанні блокуючих карт, які створюють перешкоди у радіообміні [4].

Наслідки радіоперехоплення відповідають викраденню картки з подальшим непомітним поверненням, що дозволяє зловмиснику скопіювати RFID-ідентифікатор.

Таким чином, при впровадженні СКУД на основі недорогих рішень технології MIFARE слід враховувати два важливі фактори:

- зловмисник має технічні засоби для зчитування вмісту пам'яті RFID-картки;
- емулятор здатен імітувати автентифікацію так, щоб комунікація не відрізнялася від оригіналу.

Як наслідок, ключовою загрозою для СКУД є атака повторного відтворення (англ. replay attack).

Рішення на базі MIFARE Ultralight

Метою роботи було розробити прототип СКУД на основі найпростішого варіанту MIFARE Ultralight (64 байти пам'яті без криптозахисту), який відповідає вимогам:

- *прив'язка рішення до виробника* (*vendor lock-in*) – практична неможливість створити коректний RFID-ідентифікатор з «порожньої» картки;
- *прив'язка RFID-картки до одного або двох зчитувачів*, що є важливим для об'єктів із двома точками доступу або різних об'єктів;
- *захист від дублікатів* – надзвичайна технічна складність емуляції роботи системи.

На Рис. 1. наведено мапу пам'яті ініціалізованої картки, прив'язаної до двох об'єктів. Перші дві сторінки та перший байт третьої містять 7-байтовий унікальний серійний номер RFID-ідентифікатора (разом із двома контрольними байтами), який програмується NXP і не може бути змінений [1] (емулятори дозволяють копіювати це поле). Під час ініціалізації у сторінки 4–6 записуються перші 12 байтів гешу HMAC_SHA256 серійного номера, створеного на основі таємного ключа виробника СКУД. У практичній реалізації зчитувач зберігатиме цей ключ у вбудованому програмному забезпеченні (англ. firmware), тому розробник має приділити особливу увагу його захисту. За аналогічним принципом ініціалізуються сторінки 7–9, але з використанням іншого ключа. Це робиться для захисту від компрометації: якщо зловмисники зможуть отримати доступ до таємного ключа зчитувача і створити власні ініціалізовані картки, то оновлення вбудованого ПЗ переведе систему на використання другого ключа, зробивши клоновані ідентифікатори непридатними.

Байти захисту (Lock Bytes) прописуються таким чином, щоб перманентно захистити сторінки 4–9 від запису [1].

Прив'язка картки до зчитувача відбувається аналогічним чином, але виконується адміністратором. Для цього обчислюється HMAC_SHA256 від сторінок 4–9 із використанням секретного ключа об'єкта, після чого його перші 8 байтів записуються у сторінки 0xA–0xB для

першого об’єкта та 0xD–0xE для другого. Оскільки ця пам’ять не має захисту від перезаписування, система дозволяє змінювати ключі, відкликати картки або повторно використовувати їх на іншій локації.

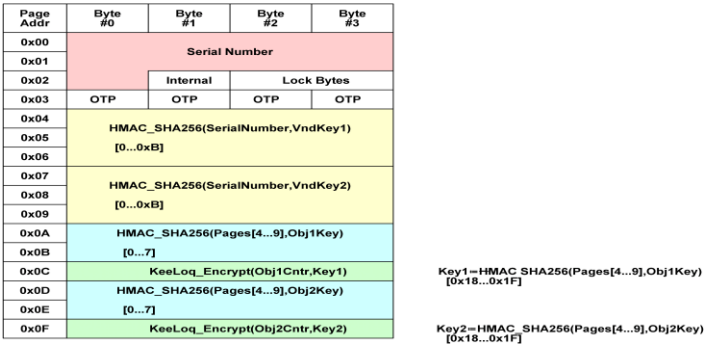


Рисунок 1. Мапа пам’яті RFID-картки MIFARE Ultralight

Нарешті, захист від дублікатів (тобто, replay-атак) здійснюється впровадженням незалежних лічильників для кожного об’єкта, які працюють синхронно з лічильниками у зчитувачах, та інкрементуються при кожному використанні картки. Оскільки зберігання у відкритому вигляді робить систему вразливою – знаючи поточне значення n, зловмисник може створити клон з n+1, – значення записуються на картку зашифрованими алгоритмом KeeLoq [5]. Ключ отримується з останніх байтів HMAC_SHA256 сторінок 4–9, обчислених із використанням секретного ключа об’єкта. KeeLoq обрано через відповідність блоку розміру сторінки (лічильника). Для надійнішої роботи можливе невелике «вікно», у межах якого лічильники вважаються еквівалентними.

Прототип реалізовано мовою C++ для ОС Windows з використанням USB-зчитувача ACS ACR1252U III. Використані бібліотеки Winscard для взаємодії зі зчитувачем, OpenSSL для HMAC_SHA256 та SQLite3 для роботи з БД. Він складається з двох програм, перша з яких

поєднує функції ініціалізації карток та адміністрування СКУД, а друга емулює інтелектуальний RFID-зчитувач.

Висновки

В роботі проаналізовано типові загрози СКУД з RFID. Запропоновано механізм криптозахисту від клонування карток MIFARE Ultralight. Реалізовано робочий прототип з використанням USB зчитувача ACS ACR1252U III. Подальший розвиток можливий з використанням досконаліших карток MIFARE Ultralight EV1 тощо.

Автори усвідомлюють недоліки своєї концепції, зокрема використання алгоритму з обмеженою криптостійкістю [5]. Водночас *KeeLoq* залишається компактним рішенням, яке відповідає малому обсягу пам'яті карток MIFARE Ultralight. Система також використовує лише частину байтів від гешів *HMAC_SHA256*, що некритично для загального рівня захисту.

Список використаних джерел

1. NXP Semiconductors. MF0ICU1. MIFARE Ultralight contactless single-ticket IC. URL: https://www.nxp.com/ docs/en/data-sheet/MF0ICU1.pdf (дата звернення: 5.05.2025).

2. Přeučil T., Novotný M. Surveying the security of access systems in Uppsala, Sweden // 2023 12th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 2023, pp. 1-5. URL: https://www.researchgate. net/publication/371885351_Surveying_the_security_of_access_ systems_in_Uppsala_Sweden (дата звернення: 5.05.2025).

3. Cheng C. M. MIFARE Classic: Completely Broken. // Hacks In Taiwan Conference (HITCON), Taipei, Taiwan, 2016. URL: https://hitcon.org/download/2010/ 11_MIFARE Classic IS Completely Broken.pdf (дата звернення: 5.05.2025).

4. Alecci M., Attanasio L., Brighente A., Conti M., Losiouk E., Ochiai H., Turrin F. Beware of Pickpockets: A Practical Attack against Blocking Cards. // Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses (RAID), Hong Kong, 2023, pp. 195-

206. DOI: 10.1145/3607199.3607243 (дата звернення: 12.05.2025).

5. Bianchi T., Brighente A., Conti M., Pavan E. SoK: Stealing Cars Since Remote Keyless Entry Introduction and How to Defend From It // arXiv.org, 2025, pp. 1-18. URL: https://arxiv.org/pdf/2505.02713v1 (дата звернення: 12.05.2025).

АНАЛІЗ КРИПТОГРАФІЧНИХ ВЛАСТИВОСТЕЙ ОДНОГО АЛГОРИТМУ ШИФРУВАННЯ ЗОБРАЖЕНЬ

Паршин О. Ю.[1], Яковлєв С. В.[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна*
parshin_o@ukr.net, yasv@rl.kiev.ua

В даній роботі розглядається алгоритм шифрування зображень, запропонований К. Кумаром та Ш. Раєм, який використовує MRX-підхід (multiplication, rotation, XOR). Виявлено конструктивні недоліки запропонованої схеми та запропоновано модифікації дизайну.

Ключові слова: симетрична криптографія, ARX, MRX, диференціальний криптоаналіз.

Вступ

MRX-криптосистеми – це схеми шифрування, які використовують модульні операції додавання та множення, лінійних зсувів та XOR-ів для забезпечення стійкості. К. Кумар та Ш. Рай в [1, 2] запропонували схему шифрування зображень, яка ґрунтується на MRX-дизайні. У даній роботі буде показано, що ця схема має конструктивні недоліки, які унеможливлюють її практичне використання, та буде запропоновано модифікації, які їх усувають.

Схема шифрування Кумара-Рая

Схема шифрування зображень Кумара-Рая детально описана у [1, 2]. Процес шифрування складається з двох етапів. Першим етапом є генерування ключа. Для шифрування одного байта даних використовуються два ключі – $K1$ та $K2$. Другий етап – це власне процес шифрування, у якому відкритий текст додається або множиться з ключем $K1$ за модулем n (покладається

$n = 257$). Далі виконується зсув вліво на $n/2$ бітів, і, нарешті, результат попереднього етапу побітово додається з ключем $K2$ (рис. 1).

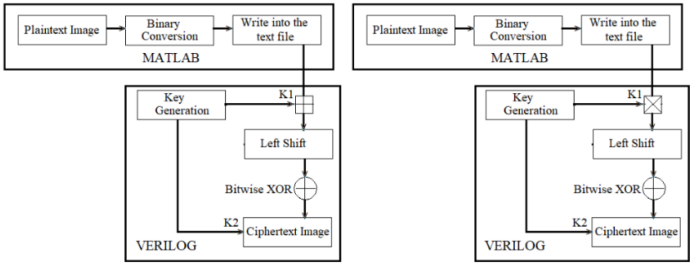


Рисунок 1. Схема шифрування Кумара-Рая

У формульному вигляді схема шифрування виглядає таким чином:

$$CT = ((PT \boxplus K1) \lll 5) \oplus K2;$$
$$CT = ((PT \otimes K1) \lll 5) \oplus K2.$$

У роботах [1, 2] не наводиться схема розшифрування. Більше того, з тексту даних робіт незрозуміло, який зсув пропонується використовувати при шифруванні – циклічний чи нециклічний. Також не зовсім зрозуміла величина зсуву, оскільки $n/2$ при $n = 257$ є нецілим числом, значення якого навіть при округленні суттєво перевищує довжину байта (вісім бітів). У подальшому ми вважаємо, що зсув виконується на 5 позицій, як в схемі генерування ключів.

Аналіз показує, що використання нециклічного зсуву веде до втрати інформації, яку неможливо відновити при розшифруванні. Розглянемо такий приклад:

$$PT = 11001010 = 202,$$
$$K1 = 00110101 = 53, K2 = 10101100 = 172,$$

тоді

$$PT \boxplus K1 = (202 + 53) mod\ 256 = 255 = 11111111,$$
$$11111111 \lll 5 = 11111111 \rightarrow 11100000 = 224,$$

$$224 \oplus 172 = 11100000 \oplus 10101100 = 01001100 = 76$$

При розшифруванні маємо:

$$CT \oplus K2 = 76 \oplus 172 = 01001100 \oplus 10101100 = \\ = 11100000 = 224,$$

$$x \ll 5 = 11100000 \Rightarrow x = \text{?????}111.$$

Тут можливі 32 варіанти розшифрування, що унеможливлює відновлення відкритого тексту.

Диференціальні імовірності шифру

Розглянемо варіант схеми Кумара-Рая, у якому використовується циклічний зсув.

Розглянемо такі неоднорідні диференціальні імовірності:

$$DP^{f}_{\otimes,\oplus}(\alpha,\beta) = Pr\{f(x \otimes \alpha) \oplus f(x) = \beta\},$$

де в якості функції f виступає схема шифрування. Тоді:

$$Pr\left\{\left(\left((X \otimes \alpha \otimes K1) mod\, n\right) \ll 5\right) \oplus K2 = \right. \\ = \left(\left((X \otimes K1) mod\, n\right) \ll 5\right) \oplus K2 \\ \left. \oplus \beta\right\} = \\ = \Pr\left\{\left((y \otimes \alpha) mod\, n\right) \oplus y = \beta\right\},$$

де $y = x \oplus K1$, $\beta' = \beta \gg 5$.

Таким чином,

$$DP^{f}_{\otimes,\oplus}(\alpha,\beta) = DP^{\otimes}_{\otimes,\oplus}(\alpha,\beta')$$

Оскільки $x, K1, K2 \in V_8$, то $y \in V_8$. Отже, ймовірність даного диференціала не залежить від ключа, і може бути обчислена безпосередньо. Нами було знайдено 165 диференціалів (α,β) із імовірностями більше за 1/32. Такі диференціали можуть бути використані для побудови атаки розпізнавання.

(α, β)	$DP^{f}_{\otimes,\oplus}(\alpha, \beta)$	(α, β)	$DP^{f}_{\otimes,\oplus}(\alpha, \beta)$
(241, 240)	0.0586	(205, 204)	0.0469
(16, 240)	0.0586	(173, 204)	0.0469
(241, 120)	0.0547	(84, 204)	0.0469
(16, 120)	0.0547	(52, 204)	0.0469

Недоліки схеми генерування ключів

У схемі шифрування Кумара-Рая використовується схема побудови раундових ключів, у якій байти ключа одержуються за допомогою перетворення

$$K = \pi\big(z \oplus (z \ll 5)\big),$$

де z – деякий байт, одержаний з стартового 64-бітового вектору, π – бітова перестановка.

Перетворення $\varphi(z) = z \oplus (z \ll 5)$ не є бієктивним. Власне, для $z \in V_8$ маємо, що $\varphi(z)$ повертає лише 128 різних значень замість 256-ти, а тому ефективна довжина ключа зменшується з 8-ми бітів до 7-ми на кожному використаному байті. Цей недолік легко усунути, якщо використовувати інше перетворення, наприклад,

$$\psi(z) = z \oplus (z \ll 1) \oplus (z \ll 5),$$

яке є бієктивним.

Висновки

Опис схеми шифрування Кумара-Рая містить багато неоднозначностей, які не дозволяють реалізовувати її безпосередньо. На схему шифрування через її простоту можна побудувати диференціальні розпізнавачі на неоднорідних операціях. Схема генерування раундових ключів містить конструктивні недоліки та штучно зменшує можливу кількість ключів.

Список використаних джерел

1. V.G, Kiran & Rai, C.. (2021). FPGA implementation of a lightweight simple encryption scheme to secure IoT using novel key scheduling technique. International Journal of Applied Science and Engineering. 18. 1-. 10.6703/IJASE.202106_18(2).002.

2. V.G, Kiran & Rai, C.. (2019). Implementation and Analysis of Cryptographic Ciphers in FPGA: Proceedings of IEMIS 2018, Volume 1. 10.1007/978-981-13-1951-8_59.

EVM MEMPOOL MONITORING FOR ATTACK DETECTION: LEVERAGING TRANSACTION-LAYER VISIBILITY FOR EARLY IDENTIFICATION OF THREATS

Kozyriatskyi G.[1]

[1]*National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»*

The Ethereum Virtual Machine (EVM) and its associated transaction mempool represent the critical pre-consensus layer of the Ethereum blockchain. While extensive research has focused on post-block analysis for security and anomaly detection, this paper explores the scientific novelty and practical implications of leveraging real-time EVM mempool monitoring for the early detection of blockchain-level attacks. We argue that many systemic threats, including consensus manipulation attempts, transaction reordering schemes (e.g., MEV-related attacks like front-running and sandwich attacks), and denial-of-service through transaction spam, exhibit discernible signatures within the mempool before block inclusion. This pre-confirmation visibility offers a crucial window for preemptive detection that is unavailable through traditional post-block analysis. This research proposes an analytical model characterizing attack signatures within mempool data based on features such as nonce gaps, gas price anomalies, transaction fee spikes, and transaction dependencies. We outline a monitoring framework incorporating real-time data aggregation, statistical, heuristic, and machine learning-based anomaly detection algorithms, and criteria for attack flagging. The paper details an experimental design for evaluating the proposed framework against simulated attack scenarios, assessing performance using metrics like precision, recall, and detection latency. Our key contributions include formalizing the concept of mempool-based attack detection as a distinct security layer, providing a typology of detectable attacks, and outlining a

practical framework for implementation. This research highlights the potential of mempool monitoring as a vital, proactive tool for enhancing the security and resilience of decentralized infrastructures, offering significant advantages over purely reactive post-block methods.

Keywords: Cybersecurity, decentralisation, EVM, mempool, transaction, attack, early identification

Introduction

The Ethereum Virtual Machine (EVM) serves as the decentralized computation engine powering the Ethereum blockchain, executing smart contract code and managing state transitions. Central to the operation of the EVM is the transaction mempool – a dynamic, volatile repository of pending transactions that have been broadcast to the network but not yet included in a block. Transactions reside in the mempool, awaiting selection by block proposers (miners or validators) based on various criteria, primarily transaction fees (gas price).

The conventional approach to identifying potential threats typically relies upon the analysis of confirmed transactions within finalized blocks. While this methodology proves effective for post-mortem analysis and the identification of completed malicious activities, it is inherently reactive. By the time an attack is validated and included in a block, its adverse effects may have already disseminated throughout the network or resulted in irretrievable financial losses.

This paper posits that real-time monitoring and analysis of the EVM mempool offers a novel and powerful vector for the preemptive detection of blockchain-level attacks. By observing the characteristics, patterns, and interactions of transactions before they are included in a block, it is possible to identify suspicious activities at their early stage. This transaction-layer visibility provides an early warning system, enabling network participants and security systems to react before an attack is fully executed or confirmed on-chain. The purpose and scope of this paper are to explore the theoretical underpinnings, practical methodologies, and potential benefits of leveraging mempool

monitoring as a proactive tool for identifying systemic threats to the EVM ecosystem.

Analytical Model of Mempool-Based Attack Detection

To effectively leverage the EVM mempool for attack detection, we must first establish a formal understanding of its structure and the characteristics of the data it contains that can serve as indicators of malicious activity. The mempool is a collection of pending transactions, each characterized by attributes such as sender address, recipient address, value, gas limit, gas price, nonce, and transaction data (for smart contract interactions). Key data characteristics within the mempool that can reveal anomalous behavior include:

1. **Nonce Gaps:** [1] Large, sequential gaps in the nonce of transactions originating from a single address might indicate an attempt to queue up a series of transactions for a coordinated action or an attempt to block future transactions from that address.
2. **Gas Anomalies:** [2] [3] Unusually high or low gas prices relative to the current network conditions could signal an attempt to force inclusion (high gas) or test the network with low-cost transactions (low gas, potentially for spam). Sudden spikes in gas prices for specific contract interactions could indicate targeted attacks. Additionally monitoring for out-of-gas vulnerabilities can be applied.
3. **Fee Spikes:** [4] For EIP-1559 transactions, similar to gas price, sudden, significant increases in the total transaction fee (max_priority_fee_per_gas, max_fee_per_gas, gas_limit) can be a strong indicator of an attempt to gain priority for a transaction, potentially part of a reordering attack or a time-sensitive exploit.
4. **Transaction Dependencies and Ordering:** [5], [6] Analyzing sequences of transactions from the same or related addresses, or transactions interacting with the same smart contracts, can reveal dependencies. Anomalous ordering or rapid submission of dependent transactions might indicate an attack attempt.

5. **Transaction Volume and Rate:** [7] A sudden, significant increase in the volume or submission rate of transactions, especially targeting specific contracts or network resources, can signal a spam attack, denial-of-service attempt or malicious activity.

6. **Smart Contract Interaction Patterns:** [8], [9], [10] Analyzing the data field of transactions to understand which smart contracts are being interacted with and the specific function calls being made can reveal attempts to exploit contract vulnerabilities or manipulate contract state.

Conclusions

This study has explored practical implications of utilizing EVM mempool monitoring as a proactive layer for detecting blockchain-level attacks. We have argued that the mempool, as the pre-consensus staging area for transactions, offers a unique and critical vantage point for identifying malicious activities at their earliest stages, providing a significant advantage over traditional post-block analysis methods.

Our key contributions include the formalization of mempool-based attack detection as a distinct security paradigm, the proposal of an analytical model characterizing attack signatures within mempool data based on observable transaction characteristics, and the outline of a practical monitoring framework incorporating real-time data aggregation and multi-modal anomaly detection algorithms. This research has a significant practical relevance for improving the security and resilience of decentralized infrastructures. Enabling the early detection of threats such as transaction reordering, spam attacks, and potential consensus manipulation attempts, mempool monitoring empowers network participants to take timely action to mitigate risks and protect users.

References

1. Kelkar M., Deb S., and Kannan S. Order-Fair Consensus in the Permissionless Setting. *Proceedings of the 9th ACM on ASIA Public-Key Cryptography Workshop*, May 2022, doi: 10.1145/3494105.3526239.

2. Rosa-Bilbao J., Boubeta-Puig J., Lagares-Galán J., and Vella M. Leveraging complex event processing for monitoring and automatically detecting anomalies in Ethereum-based blockchain networks. *Computer Standards & Interfaces*, vol. 91, p. 103882, Jan. 2025, doi: 10.1016/j.csi.2024.103882.

3. Chen T. *et al.* An Adaptive Gas Cost Mechanism for Ethereum to Defend Against Under-Priced DoS Attacks. pp. 3–24, Jan. 2017, doi: 10.1007/978-3-319-72359-4_1.

4. EIP-1559: Fee market change for ETH 1.0 chain. *Ethereum Improvement Proposals*. https://eips.ethereum.org/EIPS/eip-1559

5. Munir S. and Reichenbach C. TODLER: A Transaction Ordering Dependency anaLyzER – for Ethereum Smart Contracts. pp. 9–16, May 2023, doi: 10.1109/wetseb59161.2023.00007.

6. Ray I. and Xin T. Analysis of dependencies in advanced transaction models. *Distributed and Parallel Databases*, vol. 20, no. 1, pp. 5–27, May 2006, doi: 10.1007/s10619-006-8593-9.

7. Guo D., Dong J., and Wang K. Graph structure and statistical properties of Ethereum transaction relationships. *Information Sciences*, vol. 492, pp. 58–71, Aug. 2019, doi: 10.1016/j.ins.2019.04.013.

8. Yu R., Zhang Y., Wang Y., and Liu C. TxMirror: When the Dynamic EVM Stack Meets Transactions for Smart Contract Vulnerability Detection. *Symmetry*, vol. 15, no. 7, pp. 1345–1345, Jul. 2023, doi: 10.3390/sym15071345.

9. Qin K. *et al.* Towards Automated Security Analysis of Smart Contracts based on Execution Property Graph. *arXiv*, Jan. 2023, doi: 10.48550/arxiv.2305.14046.

10. Li X., Liu J., Chen X., and Zhang Q. A Symbolic Execution-Based Approach for Smart Contract Vulnerability Detection. *2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering*, pp. 468–472, Dec. 2023, doi: 10.1109/auteee60196.2023.10407615.

INFORMATION SECURITY CHALLENGES IN AN ENTERPRISE-GRADE SOFTWARE DEVELOPMENT LIFECYCLE

Mahomedov K. O.[1], Smyrnov S. A.[1]

[1]*National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine,*
kmomarovich@gmail.com

This study evaluates prominent cybersecurity frameworks and assesses how well they accommodate modern cloud security practices within contemporary SDLCs. Special attention is given to the DevSecOps paradigm, which integrates automated security checks and developer engagement into continuous integration and delivery pipelines, and to SBOMs as a means of exposing and managing third-party component risks in complex supply chains. Finally, the study identifies gaps in standardized maturity metrics, adaptive security controls for dynamic cloud environments, and empirical understanding of the human factors that sustain long-term security practices.

Keywords: SDLC, DevSecOps, SBOMs, cybersecurity maturity models, cloud security.

Introduction

In an era of escalating cyber threats and digital complexity, the integration of information security into the software development lifecycle (SDLC) is imperative for building trustworthy enterprise-grade software solutions. As organizations face increasingly sophisticated cyberattacks, the consequences of lacking security mechanisms grow more severe, ranging anywhere from unavailability to financial and reputational losses.

Despite the availability of mature tools and methodologies, vulnerabilities are still extremely widespread. It can be said that it stems from treating security as an afterthought rather than a foundational aspect of software engineering. This can happen

due to a variety of reasons, such as budgetary constraints or insufficient level of security awareness, which can all be considered symptoms of a lack of organizational maturity.

Evolution of Security Challenges in SDLC

Early investigations revealed a systemic tendency to treat security as an afterthought. Initial studies underscored that late-stage vulnerability fixes incur disproportionate costs and residual risk [1]. Researchers argued for a paradigm shift: integrating threat modeling, secure design principles, and risk assessment alongside functional requirements. Subsequent empirical surveys corroborated these claims, showing that projects embracing front-loaded security activities experienced significantly fewer high-severity defects after release.

Building on these foundations, later work examined the educational and cultural dimensions of developer engagement [2]. Scholars highlighted that technical controls alone are insufficient when practitioner mindsets and organizational incentives undervalue security. As a result, the so-called cybersecurity maturity models, comprised of lightweight best-practice checklists, were proposed to align security tasks with business objectives and developer workflows. These efforts emphasized automation, continuous feedback, and the need to reduce friction for engineering teams.

Frameworks and Methodologies

A rich ecosystem of frameworks has emerged to guide secure SDLC adoption. Microsoft's Security Development Lifecycle (SDL) pioneered a phase-gated, prescriptive approach [3]. OWASP's Software Assurance Maturity Model (SAMM) introduced maturity-based, benchmarking perspectives that allow organizations to assess and evolve their security posture. Then at the standards level, NIST's Secure Software Development Framework (SSDF) and ISO/IEC 27034 codified practices into internationally recognized guidelines [5], the relationship between which is shown in Fig. 1.

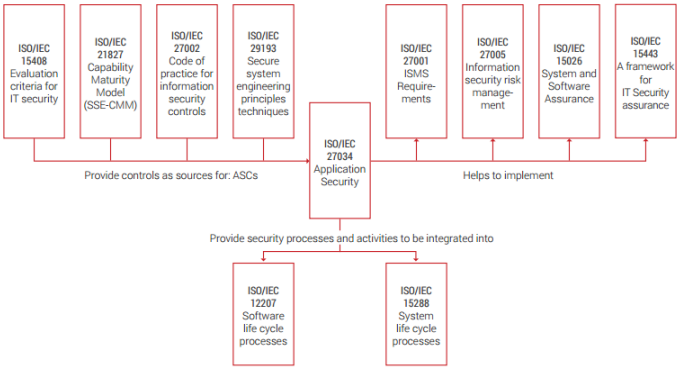


Figure 1. Relationship of international security standards

While all frameworks share core elements, such as threat modeling, secure coding, verification, and incident response, they mainly differ in granularity, flexibility, and organizational overhead [6].

DevSecOps, Cloud Security, and SBOMs

Recent research spotlights the DevSecOps paradigm, which addresses the issue of the so-called "knowledge silos" by distributing security responsibilities throughout development and operations teams. Case studies demonstrate that embedding automated security checks within CI/CD processes fosters a "shift-left" culture, reducing the latency between code commit and vulnerability detection. However, cultural resistance, skill gaps, and fragmented tooling remain formidable barriers [4].

Concurrently, the migration to cloud-native architectures has redefined shared responsibility models. Organizations must now secure not only their application code but also infrastructure-as-code, container images, and managed services. In this context, the Software Bill of Materials (SBOM) has emerged as a keystone for supply-chain transparency, enabling real-time tracking of third-party component vulnerabilities. While regulators and industry auditors increasingly mandate SBOM generation, technical challenges in versioning, license compliance, and large-scale orchestration persist [7].

Conclusions

Across all topics, three recurrent gaps stand out:

1. Metrics deficit: there is no consensus on lightweight, actionable metrics for evaluating SSDLC maturity and ROI.
2. Scalability constraints: SMEs often lack the resources and expertise to implement heavyweight frameworks, while large enterprises struggle to harmonize practices across distributed teams.
3. Human factors: organizational culture, incentives, and mindset are under-studied drivers of long-term adoption, yet they remain the biggest key factors for its success.

The synthesis suggests that enduring progress depends on converging technical automation with robust governance, training, and standardized measurement frameworks.

References

1. Falcarin P., and Morisio M. Developing Secure Software and Systems, 2004.
2. Uzunov A., Fernandez E., and Falkner K. Engineering security into distributed systems: a survey of methodologies. *Journal of Universal Computer Science*, vol. 18, 2920-3006, 2012, doi: 10.3217/jucs-018-20-2920.
3. Weir C., Becker I., Noble J., Blair L. et al., Interventions for Long Term Software Security: Creating a Lightweight Program of Assurance Techniques for Developers. *Software: Practice and Experience*, vol. 50, pp. 275-298, 2019, doi: 10.1002/spe.2774.
4. Rajapakse R. N., Zahedi M., Ali Babar M., and Shen H. Challenges and solutions when adopting DevSecOps: A systematic review. *Information and Software Technology*, vol. 141, 2022, doi: 10.1016/j.infsof.2021.106700.
5. National Institute of Standards and Technology (NIST), "Security Considerations in the System Development Life" Cycle, *NIST Special Publication* 800-64 Revision 2, Gaithersburg, MD, USA, 2008.
6. Saeed H., Shafi I., Ahmad J., Khan A. A., Khurshaid T., and Ashraf I. Review of techniques for integrating security in

software development lifecycle. *Computers, Materials & Continua*, vol. 82, no. 1, pp. 139–172, 2024, doi: 10.32604/cmc.2024.057587.

7. Nicolaysen T., Sasson R., Line M. B., and Jaatun M. G. Agile software development: The straight and narrow path to secure software? *International Journal of Secure Software Engineering (IJSSE)*, vol. 1, no. 3, pp. 71–85, 2010, doi: 10.4018/jsse.2010070105.

СТВОРЕННЯ МЕТОДИКИ ДОДАВАННЯ НЕПОМІТНИХ ФРАКТАЛЬНИХ ВОДЯНИХ ЗНАКІВ ДО РАСТРОВИХ ЗОБРАЖЕНЬ З МЕТОЮ СПОСТЕРЕЖНОСТІ ЗМІН І ЗАХИСТУ АВТОРСЬКОГО ПРАВА

Марчук І. С.[1], Рибак О. В. [1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна*

У сучасному цифровому світі зображення легко копіюються, поширюються й використовуються без дозволу автора. Це створює серйозні виклики для захисту авторських прав – особливо тоді, коли зовні контент лишається незмінним, а підтвердити його справжнє походження стає складно. Саме ця проблема стала основою для цієї роботи.

Ми поставили перед собою мету – адаптувати та поєднати існуючі методи цифрового водяного знакування, щоб створити рішення, яке дозволяє непомітно, але надійно вбудовувати інформацію в зображення. Такий знак повинен бути стійким до JPEG-стиснення, змін або часткових атак і водночас залишатись невидимим для користувача.

У рамках дослідження вирішувались такі основні завдання:

- Дослідити сучасні технології захисту цифрових зображень.
- Розробити алгоритм водяного знаку на основі фрактальних структур, що гарантує автентичність контенту важко-оборотністю параметрів фракталу і спостережність за змінами зображення візуальними характеристиками відновленого фракталу.
- Використати технології квантування (QIM) і кодів корекції помилок (Ріда-Соломона) для максимальної стійкості знаку до спотворень.

Опис методології

Для розв'язання задачі створення стійкого водяного знаку, який є непомітним для людського ока, проте стійким до JPEG-компресії та інших атак, було використано комплексний підхід, який поєднує кілька сучасних методів цифрової стеганографії.

Вибір кольорового простору YCbCr

Зображення RGB попередньо перетворюється у простір YCbCr, який складається з трьох компонентів:

Y (luminance) – яскравість, яка сприймається людським оком менш чутливо.

Cb і Cr (chrominance) – колірні складові, що сприймаються більш чутливо.

Саме завдяки низькій чутливості людського ока до змін у яскравісній компоненті було прийнято рішення вбудовувати водяний знак саме у компоненту Y. Таке рішення дозволяє суттєво зменшити помітність знаку без втрати його стійкості.

Частотне перетворення (Discrete Cosine Transform, DCT)

Обрано частотну область (DCT) через її стабільність до JPEG-компресії. Зображення розбивається на блоки розміром 8×8 пікселів, які далі перетворюються у частотний простір за допомогою дискретного косинусного перетворення:

$$y_k = 2 \sum_{n=0}^{N-1} x_n \cos\left(\frac{\pi k (2n+1)}{2N} \right)$$

$$f = \begin{cases} \sqrt{\frac{1}{4N}} & \text{if } k = 0, \\ \sqrt{\frac{1}{2N}} & \text{otherwise} \end{cases}$$

Для вбудовування знаку було вибрано середньочастотні коефіцієнти DCT(4,3), оскільки ці коефіцієнти є достатньо стійкими до компресії і водночас слабо помітними для людського сприйняття.

Фрактальний водяний знак (Множина Мандельброта)

Для генерації унікального водяного знаку використовувалася бінаризована множина Мандельброта, яка володіє такими перевагами:

- візуально унікальний візерунок, що легко ідентифікується навіть при частковому відновленні;
- висока складність підбору параметрів для стороннього користувача, що робить знак стійким до зворотної розробки (реверс-інжинірингу);
- фрактальний знак перетворюється у послідовність бітів, яка далі вбудовується у вибрані DCT-коефіцієнти блоків зображення.

Метод Quantization Index Modulation (QIM)

Метод QIM використовується для вбудовування фрактального знаку в частотні коефіцієнти DCT(4,3). Кожен коефіцієнт квантується відповідно до біта водяного знаку. Це дозволяє ефективно закодувати біт знаку в кожному коефіцієнті так, щоб він зберігся після повторної JPEG-компресії.

Кодування Ріда-Соломона (RS)

Для забезпечення стійкості до втрат інформації використано коди Ріда-Соломона RS(255,223), які дозволяють виправляти до 16 байтів помилок на кожні 255 байтів даних. Це забезпечує високий рівень відновлення знаку навіть у разі часткових змін, стиснення чи обрізок:

$$RS(255,223):223 \text{ байти інформації} \rightarrow 255 \text{ байтів закодованих даних}$$

Контент-залежна перестановка (DC-seed)

З метою захисту від атак типу «copy-paste», коли зловмисник може перенести коефіцієнти з одного зображення в інше, запропоновано унікальну контент-залежну систему визначення місць для вбудовування знаку:

Розраховується геш DC-коефіцієнтів усього зображення (DC-map). На основі отриманого гешу генерується унікальний seed для випадкового генератора. Вибір позицій блоків для розміщення тайлів водяного знаку здійснюється випадково, але детерміновано, що унеможливлює успішну copy-paste атаку, оскільки у двох різних зображень seed завжди буде відрізнятись.

Обґрунтування обраного методу

Використання запропонованої комбінованої схеми, що включає QIM, коди Ріда-Соломона, та фрактальні структури, дозволяє забезпечити:

- Високу стійкість до JPEG-компресії (QIM, RS).
- Захист від атак типу «copy-paste» (контент-залежна система на основі DC-гешу). Візуальну непомітність знаку (робота у YCbCr-просторі).
- Надійне виявлення змін у маркованих зображеннях (візуалізація відновленого фрактала).

Таким чином, запропонований метод забезпечує комплексний та ефективний захист цифрових зображень, частково розв'язуючи основні проблеми, що лишалися відкритими у попередніх дослідженнях,знаходячи компроміс між методами реалізаціями.

Аналіз результатів

Для кожного “тайлу” фрактала ми кодуємо RAW_PER_TILE = 1784 біт, а з урахуванням MAX_TILES = 3 загальна сирова місткість становить до 5352 біт, що дає можливість формувати квадратну матрицю водяного знака розміром приблизно 73×73.

Псевдовипадковий вибір початкових блоків здійснюється на основі dc_seed, що підвищує стійкість до аналізу розподілу вставлених бітів.

Непомітність водяного знака

Вбудовування відбувається у середні DCT-коефіцієнти з координатами (u=4, v=3), що мінімізує візуальні артефакти у частотах, малопомітних для людського ока.

Типове значення Q_EMB = 28(64 щоб пережити сильніше стистання) забезпечує компроміс між низькою спотвореністю (PSNR водям/оригінал ≥ 40 дБ) та достатньою стійкістю до стиснення.

Стійкість до стиснення та корекція помилок

Після JPEG-стискання на 75 % помічаємо помітні проблеми з відновленням фрактала (рис. 1).

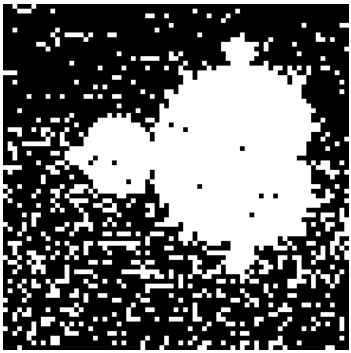


Рисунок 1. Відновлений фрактал з q=28 і стистенням 75 %

Використаний Reed–Solomon код RS(255,223) дозволяє коригувати до 16 байтів помилок в кожному блоці, додатково підвищуючи надійність відновлення навіть при помірному рівні артефактів стиснення

Двошарове вбудовування та порівняння версій

Порівняльні тести в MAIN2.py показали, що версія v2.4 (QIM + RS + tiling + dc_seed) забезпечує краще вирівнювання між imperceptibility і robustness, ніж початкова версія v0.3, з подвоєною надійністю відновлення на тому ж рівні візуальних артефактів common.

Поведінка при екстремальних умовах:

За зменшення якості JPEG (стискання до 83 %, див. Рис. 2) similarity падає до ~ 93 %, що нижче порогового значення, проте завдяки RS-коду деяка частина помилок

коригується,при більшому значенні Q (64) – точність залишається на 96 %

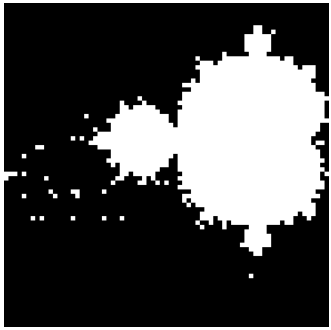


Рисунок 2. Відновлений фрактал з q=62 і стистенням 83 %

При геометричних зміщеннях (обрізка блоків, повороти) відбудеться десинхронізація тайлів. При модифікації зображення можемо бачити відповідні артефакти відновленого фрактала (рис. 3).

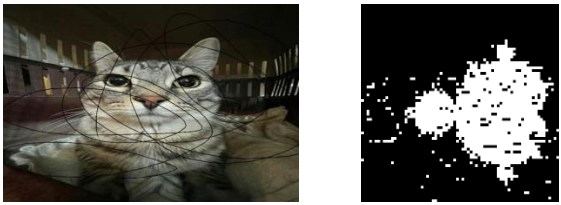


Рисунок 3. Артефакти на відновленому фракталі

При зсуві зображення отримуємо також характерні артефакти. Також при спробі накласти вотермарку ще раз отримуємо візуальну репрезентацію неможливості відновлення цілісності фракталу і неможливість підтвердження авторського права на зображення.

Висновки

У результаті виконання роботи було розроблено та реалізовано комплексний метод спостережності змін та

захисту авторських прав цифрових зображень на основі фрактальних структур і DCT-кодування з використанням Quantization Index Modulation (QIM). Запропонований метод забезпечує надійне й непомітне вбудовування водяних знаків у частотній області, демонструючи високу стійкість до JPEG-компресії завдяки використанню кодів Ріда-Соломона (RS-кодування).

Реалізована програмна система є модульною та легко розширюваною: вона включає генерацію фрактальних шаблонів, механізми псевдовипадкового вибору блоків за допомогою гешування DC-коефіцієнтів та ефективні алгоритми вбудовування/екстракції водяних знаків.

Тестування на реальних зображеннях підтвердило, що система забезпечує високий рівень відновлення водяних знаків навіть при сильній модифікації зображення. Завдяки оптимальному вибору частотних коефіцієнтів, фрактальній структурі водяного знаку та помилкостійкому кодуванню досягнуто значної стійкості до компресійних атак.

Подальший розвиток роботи може включати підвищення стійкості до геометричних атак, розширення на інші типи медіа-контенту та інтеграцію додаткових метрик якості для комплекснішого аналізу ефективності методу.

Список використаних джерел

1. An introduction to Reed-Solomon codes: principles, architecture and implementation https://www.cs.cmu.edu/~guyb/realworld/reedsolomon/reed_solomon_codes.html

2. Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding https://sia.mit.edu/wp-content/uploads/2015/04/2001-chen-wornell-it.pdf

3. Provably Robust Multi-bit Watermarking for AI-generated Text https://arxiv.org/pdf/2401.16820

4. Fast Frequency Domain Screen-Shooting Watermarking Algorithm Based on ORB Feature Points https://www.mdpi.com/2227-7390/11/7/1730

5. Image Steganography Technique Using Algebraic Fractals https://dl.acm.org/doi/pdf/10.5555/3338290.3338392

6. Effective Digital Image Watermarking in YCbCr Color Space Accompanied by Presenting a Novel Technique Using DWT https://www.researchgate.net/publication/ 227397593_Effective_Digital_Image_Watermarking_in_YCbCr _Color_Space_Accompanied_byPresenting_a_Novel_Technique _Using_DWT

7. Multibit quantization index modulation: A high-rate robust data-hiding method https://www.sciencedirect.com/ science/article/ pii/S1319157812000535

8. A secure and robust hash-based scheme for image authentication https://www.sciencedirect.com/science/article/ abs/pii/S0165168409002497

9. Robust Image Watermarking Theories and Techniques: A Review. https://www.sciencedirect.com/science/article/pii/ S1665642314716128

АНАЛІТИЧНІ ВИРАЗИ ІМОВІРНОСТЕЙ ДИФЕРЕНЦІАЛІВ СКЛАДОВИХ ШИФРІВ NORX ТА CHACHA

Корж Н. С.[1], Яковлєв С. В.[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна,*
nik.korzh.ita@gmail.com, yasv@rl.kiev.ua

У роботі знайдено нові аналітичні вирази для імовірностей диференціалів чвертьраундових перетворень шифрів ChaCha та NORX. Одержані вирази не ґрунтуються на припущеннях про незалежність окремих складових таких перетворень, а тому дозволяють оцінювати імовірності диференціалів більш точно, ніж у опублікованих раніше оцінках для імовірностей диференціалів зазначених шифрів.

Ключові слова: симетрична криптографія, диференціальний криптоаналіз, ChaCha, NORX.

Вступ

Диференціальний криптоаналіз [1] розглядає імовірності подій, що вхідні повідомлення із заданою різницею перейдуть у шифротексти із заданою різницею. У контексті ARX-шифрів (ChaCha [2], NORX [3]) зазвичай розглядають різниці за операцією побітового додавання.

У роботі [4] були запропоновані аналітичні вирази для імовірностей диференціалів окремих перетворень шифру ChaCha. Однак ці результати використовували припущення про незалежність появи диференціалів на окремих операціях раундових перетворень, які, взагалі кажучи, нічим не обґрунтовані.

У даній роботі наводяться точні формули для імовірностей раундових перетворень шифрів ChaCha та

NORX, які не використовують припущення про незалежність складових операцій.

Структура шифрів ChaCha та NORX

Шифр ChaCha є потоковим ARX-шифром, два послідовних раунди якого складається з восьми так званих «*чвертьраундових*» функцій $G(a,b,c,d)$, кожна з яких виконує перетворення над чотирма словами стану за схемою:

$a \leftarrow a + b,$ $\quad d \leftarrow (d \oplus a) \lll r_1,$

$c \leftarrow c + d,$ $\quad b \leftarrow (b \oplus c) \lll r_2,$

$a \leftarrow a + b,$ $\quad d \leftarrow (d \oplus a) \lll r_3,$

$c \leftarrow c + d,$ $\quad b \leftarrow (b \oplus c) \lll r_4,$

де r_i – наперед визначені величини зсувів; для результатів, одержаних в даній роботі, величини r_i можуть бути довільними.

Шифр NORX використовує такі само чвертьраундові перетворення, як і ChaCha, однак модульні додавання у NORX замінені на специфічну операцію

$$H(x, y) = (x \oplus y) \oplus \big((x \,\&\, y) \ll 1\big)$$

Функція H виконує бітове перетворення, шо імітує додавання з перенесенням і застосовується багаторазово в межах основної G-функції шифру NORX.

Імовірності диференціалів чвертьраундових перетворень

Диференціалом $\omega = (\alpha, \beta \to \gamma)$ функції $f(x,y)$ називають довільну трійку векторів $\alpha, \beta, \gamma \in V_n$, що задає різниці між двома вхідними (або вихідними) значеннями f відносно операції $\oplus$. Ймовірність диференціалу ω функції f визначається як

$$xdp^f(\omega) = xdp^f(\alpha, \beta \to \gamma) = \Pr_{x,y}\{f(x \oplus \alpha, y \oplus \beta) = f(x, y) \oplus \gamma\}$$

Ймовірності xdp^f характеризують стійкість до диференціального криптоаналізу.

Чвертьраундова функція ChaCha має два ланцюги послідовних додавань, які у попередніх роботах вважались незалежними. Однак у структурі перетворення ці додавання

не розділяються забілюваннями із ключем або іншими перетвореннями, які вносять випадковість, тому припущення про їх незалежність, взагалі кажучи, є помилковим.

Для функції $\Phi_+(x, y, z) = x + y + z$ введемо *умовні диференціальні імовірності*

$$xdp^{\Phi}\left(\alpha_x, \alpha_y, \alpha_z \to \gamma \mid \beta\right) = \\ = \Pr_{x,y,z}\left\{\begin{matrix}\Phi\left(x \oplus \alpha_x, y \oplus \alpha_y, z \oplus \alpha_z\right) = \Phi(x, y, z) \oplus \gamma, \\ (x \oplus \alpha_x) + \left(y \oplus \alpha_y\right) = (x + y) \oplus \beta\end{matrix}\right\}$$

Такі імовірності розглядають проходження диференціалу через складене перетворення із фіксуванням його проміжних значень. Аналогічні імовірності можна ввести й для перетворення $\Phi_H(x, y, z) = H(H(x, y), z)$, яке використовується у NORX.

Основний результат даної роботи сформулюємо у вигляді такої теореми.

Теорема. Для чвертьраундової функції $G(a, b, c, d)$ шифру ChaCha (NORX) імовірності диференціалів дорівнюють

$$xdp^{G}(\alpha_a, \alpha_b, \alpha_c, \alpha_d \to \gamma_a, \gamma_b, \gamma_c, \gamma_d) = \\ = xdp^{\Phi}(\alpha_a, \alpha_b, \beta_b \to \gamma_a \mid \beta_a) \cdot xdp^{\Phi}(\alpha_c, \beta_d, \gamma_d \to \gamma_c \mid \beta_c),$$

де $\beta_a = \alpha_d \oplus (\gamma_a \ggg r_1) \oplus \left(\gamma_d \ggg (r_1 + r_4)\right)$,
$\beta_b = \gamma_c \oplus (\gamma_b \ggg r_3)$,
$\beta_c = \alpha_b \oplus (\gamma_c \ggg r_2) \oplus \left(\gamma_b \ggg (r_2 + r_3)\right)$,
$\beta_d = \gamma_a \oplus (\gamma_d \ggg r_4)$,
а функція Φ визначається як Φ_+ для шифру ChaCha та Φ_H для шифру NORX.

Висновки

У даній роботі одержано аналітичні вирази для імовірностей диференціалів чвертьраундових функцій шифрів ChaCha та NORX; одержані вирази використовують новий математичний об’єкт – умовні диференціальні імовірності для композиції однотипних перетворень, у

якості яких виступають модульне додавання у шифрі ChaCha та специфічна операція, яка апроксимує модульне додавання, у шифрі NORX. Одержані вирази не ґрунтуються на припущеннях про незалежність складових елементів чвертьраундового перетворення, а тому дозволяють будувати більш точні оцінки для диференціальних імовірностей. У подальшому планується застосувати одержані результати для побудови більш ефективних розпізнавачів для шифрів ChaCha та NORX.

Список використаних джерел

1. Biham E., Shamir A. Differential Cryptanalysis of DES-like Cryptosystems, J. Cryptology, vol. 4, no. 1, 1991.

2. Bernstein D. J. The ChaCha family of stream ciphers, 2008. URL: https://cr.yp.to/papers.html#chacha

3. Aumasson J.-P., Jovanovic P., Neves S. NORX 3.0, CAESAR Competition Submission. URL: https://competitions.cr.yp.to/round3/norxv30.pdf

4. E. Bellini, R. Makarim, C. Sanna. Finding differential trails on ChaCha by means of state functions, 2021. DOI: https://dx.doi.org/10.1504/IJACT.2024.10063705

REAL-TIME MULTIPLE SOURCE ACOUSTIC IMPULSE RESPONSE GENERATION FOR ANOMALY DETECTION

Terletskyy O.[1], Trushevskyy V.[1]

[1]*Ivan Franko National University of Lviv, 1 Universytetska St., Lviv, 79000, Ukraine*

This paper introduces a real-time method for simulating sound propagation using ray tracing, aimed at cybersecurity for detecting and analyzing acoustic threats. By modeling the behavior of acoustic waves in virtual environments, the approach accurately captures reflections, diffractions, and reverberations, enabling tasks like audio-based anomaly detection and spatial audio analysis in secure areas. It supports multiple simultaneous sound sources, merging their effects into a cohesive acoustic model. The combination of impulse response generation with real-time convolution allows for detailed acoustic profiling that adapts to environmental changes, enhancing situational awareness. This method lays the groundwork for advanced acoustic monitoring systems, with applications ranging from secure communications to threat detection. Results confirm its capability to simulate complex acoustic interactions with high accuracy and real-time performance, making it a powerful tool for audio surveillance and cybersecurity.

Keywords: Sound propagation, ray tracing, impulse response, Theoretical and Applied Cybersecurity

Introduction

Creating realistic sound propagation is crucial for immersive experiences in areas like virtual reality, gaming, architectural acoustics, and cybersecurity. In real environments, sound perception is influenced by phenomena such as reflections, diffractions, and reverberations, which provide important

spatial and contextual cues. However, traditional audio systems often lack the flexibility to adapt to dynamic changes in virtual or physical spaces, limiting their usefulness in interactive and security-sensitive scenarios.

This paper introduces a real-time sound simulation method that integrates ray tracing to model acoustic interactions and FFT-based convolution to apply impulse responses to audio playback [1]. The approach is particularly relevant for cybersecurity applications, including anomaly detection through acoustic monitoring, secure environment simulation, and audio-based threat analysis. By accurately reproducing how sound behaves in complex environments, this method can enhance the identification of abnormal acoustic patterns or the localization of sound sources.

To maintain consistency in the simulation, the ray tracing process employs static rays, which stabilize the impulse responses and reduce unwanted fluctuations across frames. This helps prevent audio artifacts that could affect sound quality or compromise analytical accuracy.

Through the combination of static ray tracing and listener directed diffuse rain, this approach delivers audio simulations that support both immersive applications and critical cybersecurity tasks. The following sections outline the method, its implementation, and experimental results, demonstrating its effectiveness in real-time acoustic modeling and threat-aware sound analysis.

Static Ray Tracing Approach

Impulse response generation plays a vital role in simulating realistic sound propagation by capturing how an environment affects sound as it interacts with surfaces, obstacles, and spatial geometry [2]. This capability is especially valuable in cybersecurity, where precise acoustic modeling can improve audio surveillance, support anomaly detection, and help secure sensitive environments.

To ensure consistent and stable impulse responses, the proposed method uses static rays during the ray tracing process. This ensures that the pattern of reflections remains predictable from frame to frame, minimizing audio artifacts and

fluctuations caused by dynamic ray distribution. In cybersecurity scenarios, such stability is essential, as inconsistent acoustic data could obscure anomalies or lead to false positives in threat detection. Identification of funding sources and other support, and thanks to individuals and groups that assisted in the research and the preparation of the work should be included in an acknowledgement section, which is placed just before the reference section in your document.

Impulse Response

An impulse response (IR) defines how an acoustic environment reacts to a brief, idealized sound—typically modeled as a Dirac delta function, δ(t). It captures the full spectrum of acoustic effects: direct sound, reflections, reverberation, and decay over time, thereby characterizing the unique sonic signature of a space. The response of an environment to any input sound x(t) is calculated through convolution with the impulse response h(t) [3]:

$$y(t) = \int_{-\infty}^{\infty} x(\tau) h(t-\tau)\, d\tau, \tag{1}$$

In discrete terms, for sampled signals $x[n]$ and impulse response $h[n]$, the convolution is expressed as:

$$y[n] = \sum_{k=0}^{N-1} x[k] \cdot h[n-k], \tag{2}$$

where N is the length of the impulse response. This operation models how each sample of the input sound is shaped by the acoustic characteristics of the environment, as encoded in $h(t)$.

Ray Tracing Using Static Rays

This approach utilizes static rays, where rays are cast in fixed directions for every frame. This technique ensures that the resulting impulse response remains consistent over time, reducing random fluctuations that could lead to audio artifacts [5]. Such stability is especially important in cybersecurity applications, where accurate acoustic modeling is essential for detecting subtle anomalies such as unauthorized movement or the presence of hidden recording devices. For each frame, the

algorithm traces the paths of all static rays through the environment, recording the energy and travel time of each ray that reaches the listener. This information is then used to construct an impulse response that captures the acoustic characteristics of space. In cybersecurity scenarios, this dynamic impulse response can be used to detect real-time environmental changes such as the appearance of new obstacles, shifts in surface materials, or unexpected sound sources serving as early indicators of potential security threats.

Listener Directed Diffuse Rain

The Diffuse Rain model is a sound propagation technique that simulates how sound energy diffuses naturally in complex environments [4]. In the proposed method, when a ray strikes a surface, a new ray is generated and directed toward the listener. To account for the attenuation of energy over indirect paths, an energy loss factor is computed based on the angle between the incoming ray and the direction to the listener. This factor, denoted as E_{loss}, reflects the natural decay of energy due to oblique reflections and longer travel distances, ensuring realistic attenuation in the simulation.

$$E_{\text{loss}} = \max\left(\left(\mathrm{d} \cdot \frac{\mathrm{l}-\mathrm{p}}{\| \mathrm{l}-\mathrm{p} \|}\right)^2, 0\right), \tag{3}$$

where d is the direction of the incoming ray; l is the listener's position; p is the ray's current origin (hit point); $\mathrm{d} \cdot \frac{\mathrm{l}-\mathrm{p}}{\|\mathrm{l}-\mathrm{p}\|}$ represents the cosine of the angle between d and the normalized vector from p to l.

Impulse Response Generation

The impulse response generation process captures an environment's acoustic signature by aggregating the effects of individual sound rays. Each ray traces a unique path, interacting with surfaces through reflection, decay, and diffusion. Together, these rays form a composite impulse response that represents how sound behaves in the space.

Arrival time of each individual ray is calculated based on the total distance ray travels from the source to the listener or reflection point. Using the speed of sound c, the time t is given by:

$$t = \frac{d}{c}, \tag{4}$$

where d is the ray's total path length. This time determines the index in the impulse response array where the ray contributes.

Sound energy diminishes with distance according to the inverse square law. The attenuation factor E_d is computed as:

$$E_d = \frac{1}{d^2 + \epsilon}, \tag{5}$$

where ϵ is a small constant to prevent division by zero.

The final amplitude A that each ray contributes to the impulse response is calculated by combining its post-reflection energy, distance-based attenuation, and a phase alternation factor:

$$A = E_r \times E_d \times s, \tag{6}$$

where E_r is the ray's energy after all collisions and reflections. This final energy value incorporates both the initial energy and any reductions from reflections; E_d is the distance-based attenuation factor, reducing energy based on the distance traveled, as described in Equation [eq:distance_attenuation]; s is the phase-shifting factor, alternating between +1 and -1 to introduce phase shifts.

Results

This section provides a comparative analysis of impulse response characteristics within a simulated room measuring 50 × 50 × 20 meters. The sound source is located at the center of the room, 1 meter above the floor, with the listener positioned 1 meter to the right of the source. The environment is modeled as a perfectly reverberant space with no surface absorption, offering an ideal scenario to examine how ray count and reflection depth influence the impulse response.

The impulse response was recorded over a 3-second duration at a 41 kHz sampling rate, yielding 123,000 samples. This high temporal resolution enables detailed observation of the room's reflective behavior and its acoustic response over time.

Number Of Rays

The impulse response was generated using different ray counts–10, 100, and 1000–each limited to a maximum of 8 reflections. This comparison highlights how increasing the number of rays improves the accuracy and richness of the simulated acoustic response, especially in the mid and late reverberation stages. As illustrated in Figure 1, higher ray counts result in a denser and more detailed impulse response over time, effectively capturing the complex reflection patterns that contribute to the perceived fullness of the environment.

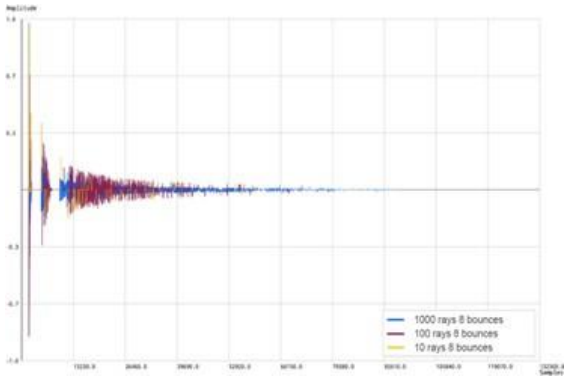


Figure 1. Impulse responses generated with 10, 100 and 1000 rays each with 8 bounces

Number Of Bounces

The impact of different reflection depths is analyzed by generating impulse responses with 2, 5, and 8 bounces per ray, using 1000 rays in each case. As shown in Figure 2, increasing the number of bounces results in a denser accumulation of reflections, particularly in the later parts of the impulse

response. This extended reverberant tail enhances the spatial impression of the room, emphasizing the role of multiple reflections in shaping the perceived acoustic environment.

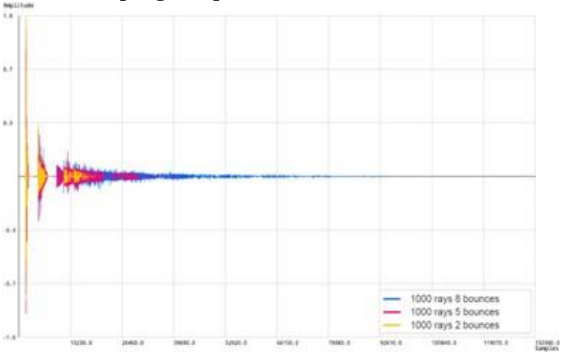


Figure 2. Impulse responses generated 1000 rays and varying bounces: 2, 5 and 8

Comparison with Sabine's formula

Simulated impulse responses with varying numbers of ray bounces (from 2 to 10) are compared against the theoretical reverberation time T_{60} calculated using Sabine's formula [6]:

$$T_{60} = \frac{0.161 \times V}{A \times \alpha}, \tag{7}$$

where V is the room volume, A is the total surface area, and α is the average absorption coefficient.

For a room with dimensions of 50 m by 50 m by 20 m and α=0.1, Sabine's formula yields a reverberation time of approximately 8.05 seconds. This theoretical value provides a benchmark for evaluating the accuracy of our simulation. As shown in Figure 3, increasing the number of bounces in the simulation progressively brings the impulse response closer to the theoretical T_{60}.

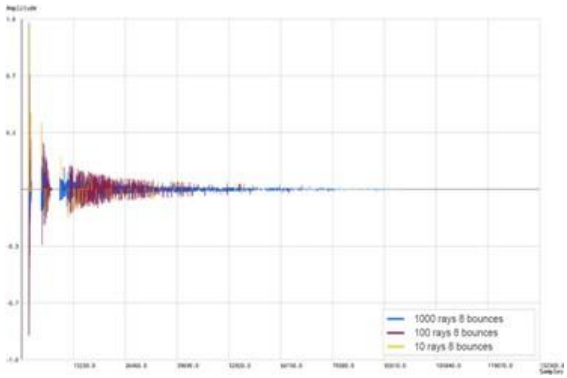


Figure 3. Simulated reverberation times with varying ray bounces compared to Sabine's theoretical T_{60}

Conclusions

This study examined how varying the number of rays and reflection bounces affects the accuracy and quality of impulse response simulations in a perfectly reverberant room. Using ray tracing combined with a listener directed diffuse rain method for real-time efficiency, the results show that increasing ray counts significantly improves the detail and density of reflections especially in the mid and late reverberation stages. These enhancements are crucial for both realistic sound simulation and precise acoustic anomaly detection in cybersecurity applications.

In security-sensitive environments, detailed impulse responses enable the detection of subtle changes, such as the introduction of unauthorized recording devices or modifications to room geometry. Dense reflection patterns help identify anomalies that may otherwise go unnoticed, strengthening acoustic monitoring capabilities.

Comparisons with Sabine's theoretical reverberation time confirm the accuracy of the simulated results, demonstrating that the approach reliably models reverberant spaces. These findings support the use of high-fidelity acoustic simulation as a tool for improving situational awareness and enabling real-time, sound-based threat detection in cybersecurity contexts.

References

1. Krokstad A., Strom S., Sorsdal S. Calculating the acoustical room response by the use of a ray tracing technique, J. Sound Vib. 8 (1968) 118–125.

2. Ouellet-Delorme E., Venkatesan H., Durand E., Bouillot N. Live ray tracing and auralization of 3D audio scenes with vaRays, in: Proceedings of the 18th Sound and Music Computing Conference, 2021, pp. 1–10.

3. Farina A. 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. Pelzer S., Schröder D., Vorländer M. The number of necessary rays in simulations based on geometrical acoustics using the diffuse rain technique, in: Proceedings of the DAGA, 2011.

5. Meta, Acoustic ray tracing in Meta XR Audio SDK, 2023. URL: https://developers.meta.com/horizon/documentation/unity/meta-xracoustic-ray-tracing-unity-overview/.

6. Beranek L. Analysis of Sabine and Eyring equations and their application to concert hall audience and chair absorption, J. Acoust. Soc. Am. 120 (2006) 1399–1410.

ТЕОРЕТИЧНІ ТА КРИПТОГРАФІЧНІ ПРОБЛЕМИ КІБЕРНЕТИЧНОЇ БЕЗПЕКИ

Іванова К. І.[1], Лучик В. Є.[1]

[1]*Харківський національний університет внутрішніх справ*

Анотація. У цій статті розглядаються теоретичні та криптографічні аспекти кібербезпеки в умовах стрімкого розвитку цифрових загроз. Наведено огляд базових моделей безпеки, виділено основні виклики, з якими стикаються системи кібербезпеки сьогодні, та підкреслено провідну роль криптографічних методів у забезпеченні конфіденційності, цілісності та автентичності даних. Дослідження також охоплює нові виклики, пов'язані з квантовими обчисленнями, та вказує на майбутній потенціал квантової криптографії як трансформаційної технології в галузі кіберзахисту.

Ключові слова: кібербезпека, криптографія, захист інформації, моделі безпеки, квантова криптографія, кіберзагрози, цілісність даних.

У XXI столітті інформація стала стратегічним ресурсом, що визначає не лише економічну, а й політичну та військову перевагу держав і корпорацій. Цифрова епоха принесла нові можливості для обміну та обробки даних, але разом з тим інформація перетворилася на об'єкт глобального контролю, маніпуляцій та атак. В умовах глобальної цифровізації дані більше не є лише засобом комунікації; вони стали критично важливим елементом, що визначає здатність організацій і навіть держав залишатися конкурентоспроможними. Саме тому питання кібербезпеки набуло пріоритетного значення, адже від її ефективності залежить не лише стабільність IT-інфраструктури, а й безпека на всіх рівнях – від індивідуального до національного.

Розвиток новітніх цифрових технологій, таких як хмарні обчислення, Інтернет речей (IoT), штучний інтелект, великі дані та мережі п'ятого покоління (5G), змінив ландшафт кібербезпеки. Ці інновації створюють безліч нових можливостей для розвитку економіки та науки, а також для підвищення ефективності державного управління. Однак одночасно вони відкривають нові вектори атак, сприяють розширенню площини вразливостей і потребують нових підходів до захисту, оскільки сучасні методи вже не можуть повністю вирішити всі виникаючі проблеми.

У такій ситуації забезпечення кібербезпеки стає не просто технічною задачею, а складною багаторівневою проблемою, що охоплює різні аспекти: правові, організаційні, наукові та соціальні. Важливою частиною цього процесу є теоретична база кібернетичної безпеки, яка дозволяє не лише систематизувати вже наявні знання, а й адаптувати їх до швидко змінюваного цифрового середовища. Зокрема, теоретичні основи допомагають виявляти нові загрози, прогнозувати можливі ризики та формувати надійні стратегії для створення стійких захисних механізмів, здатних зберігати цілісність та безпеку навіть в умовах постійних атак.

Одним із основних інструментів кібербезпеки є криптографія. Це галузь науки, що відповідає за забезпечення конфіденційності, автентичності та цілісності даних. Хоча криптографічні методи існують уже багато століть, сьогодні вони знову опиняються в центрі уваги, оскільки виникли нові, значні виклики. Зокрема, криптографія стикається з проблемою квантових обчислень, які можуть значно ослабити традиційні методи захисту. Водночас, існують проблеми з недосконалими реалізаціями криптографічних алгоритмів, що може призвести до їх компрометації навіть без застосування потужних обчислювальних засобів.

Соціальна інженерія є ще однією серйозною проблемою в кібербезпеці, що заслуговує на особливу увагу. Це тип атак, при яких зловмисники використовують психологічні методи впливу на користувачів для отримання доступу до конфіденційної інформації або внутрішніх систем

організації. Найбільш уразливим компонентом у такій ситуації є людський фактор, і навіть найсучасніші технічні засоби захисту не можуть гарантувати безпеку, якщо користувачі самі ненавмисно надають зловмисникам доступ до інформації.

Виходячи з цього, одним з можливих рішень є впровадження інтелектуальної системи контекстної автентифікації (Real-time Intelligent Context Authentication, RICA). Цей підхід ґрунтується на аналізі звичної поведінки користувача, що дозволяє автоматично виявляти відхилення від звичних патернів діяльності. Система може розпізнати, якщо користувач входить з незвичного пристрою або з іншого місця, і запросити додаткову перевірку (наприклад, біометричну) перед тим, як надати доступ до критичних систем. Такий підхід дозволяє значно знизити ризики, пов'язані з фішингом і соціальною інженерією, а також підвищити рівень захисту без зайвих незручностей для користувачів.

У контексті розвитку кібербезпеки важливо також розуміти, що це не одноразовий процес, а безперервний цикл, який вимагає постійного вдосконалення. Для цього необхідно не лише оновлювати захисні інструменти, але й активно працювати над підвищенням кваліфікації спеціалістів, здійснювати моніторинг нових загроз і проводити наукові дослідження. Залучення міжнародної співпраці та обміну досвідом є критично важливими для створення стійкої цифрової інфраструктури, здатної витримувати виклики майбутнього.

Таким чином, кібербезпека є динамічною і постійно змінюваною сферою, що вимагає гнучкого підходу, який включає не лише технічні рішення, а й інтеграцію з іншими галузями знань. Тільки так можна побудувати надійну і стійку систему захисту, здатну протистояти сучасним загрозам.

Список використаних джерел

1. Stallings W. Cryptography and Network Security. Pearson Education, 2021. (Дата звернення 07.04.2025)

2. Schneier B. Applied Cryptography. Wiley, 2020. (Дата звернення 07.04.2025)
3. NIST. Post-Quantum Cryptography Project. https://csrc.nist.gov/ (Дата звернення 07.04.2025)
4. Шеннон К. Математична теорія зв’язку. 1948. (Дата звернення 07.04.2025)
5. Касперський Є. Кібербезпека майбутнього. 2022. (Дата звернення 07.04.2025)

РОЗРОБКА МЕТОДОЛОГІЇ АВТОМАТИЗОВАНОГО АНАЛІЗУ ВРАЗЛИВОСТЕЙ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ З ВИКОРИСТАННЯМ ТЕХНІК СИМВОЛЬНОГО ВИКОНАННЯ ТА ФАЗИНГУ

Маринкевич О. М.[1]

[1]*Харківський національний університет внутрішніх справ*
alexandr.mr253@gmail.com

Вступ

Актуальність дослідження зумовлена зростаючою потребою в ефективних методах автоматичного виявлення вразливостей програмного забезпечення для підвищення кібербезпеки. Метою даної роботи є дослідження та обґрунтування перспективності поєднання технік символьного виконання та фазингу для більш глибокого та всебічного аналізу програм на наявність потенційних слабких місць. Предметом дослідження є методи автоматизованого аналізу вразливостей програмного забезпечення, зокрема фазинг та символьне виконання, а також їхні комбінації.

Основна частина

Автоматичний пошук вразливостей є важливим напрямом у кібербезпеці, спрямованим на виявлення слабких місць у програмному забезпеченні без прямої участі людини. Одним з ефективних методів такого пошуку є фаз-тестування (фазинг). Ця техніка передбачає автоматизоване генерування великої кількості некоректних, випадкових або несподіваних вхідних даних для програми з метою спровокувати її некоректну роботу, збої або виявити

вразливості безпеки, такі як переповнення буфера чи витоки пам'яті. Аналізуючи реакцію програми на ці аномальні вхідні дані, дослідники можуть ідентифікувати потенційні точки відмови, які можуть бути використані зловмисниками. Фазинг допомагає виявляти не лише відомі, але й невідомі раніше помилки, підвищуючи загальну якість та стійкість програмного забезпечення до кібератак. Інтеграція фаз-тестування в процес розробки дозволяє проактивно виявляти та усувати вразливості на ранніх етапах, знижуючи ризики та витрати на їхнє виправлення після випуску продукту [1].

Символьне виконання та фазинг – це два різні, але важливі підходи до автоматичного пошуку вразливостей у програмах. *Фазинг*, як випливає з джерела, полягає у безперервному «згодовуванні» програмі великої кількості випадкових, некоректних або неочікуваних даних на вхід. Мета фазингу – викликати збій у роботі програми або виявити незвичайну поведінку, що може свідчити про помилку чи вразливість. *Символьне виконання*, на відміну від цього, намагається проаналізувати код програми, розглядаючи вхідні дані як символи, а не конкретні значення. Це дозволяє дослідити всі можливі шляхи виконання програми та виявити потенційні проблеми, навіть ті, які важко знайти за допомогою випадкових тестів. Хоча фазинг ефективний у виявленні багатьох поширених вразливостей завдяки своїй інтенсивності, символьне виконання може глибше аналізувати логіку програми та знаходити складніші дефекти [2].

Існуючі методи автоматичного пошуку вразливостей, зокрема фазинг у його різних формах, мають певні обмеження, які роблять їх не ідеальними. *Фазинг "чорного ящика"*, хоч і швидкий, є менш ефективним у виявленні глибоко прихованих помилок, оскільки не аналізує внутрішню структуру програми. *Фазинг "білого ящика"*, навпаки, хоч і більш результативний у знаходженні складних вразливостей завдяки аналізу покриття коду, є повільним і вимагає доступу до вихідного коду, що не завжди можливо, особливо при аналізі комерційного програмного забезпечення або зловмисного ПЗ. *Фазинг*

“сірого ящика” намагається знайти баланс, але все одно може пропускати вразливості, якщо згенеровані вхідні дані не потрапляють у критичні ділянки коду або не відповідають очікуваному формату вхідних даних програми (як у випадку з граматичним фазингом, який є складним для реалізації) [3].

Незважаючи на ефективність окремих методів, поєднання символьного виконання та фазингу представляє собою перспективний напрямок для вдосконалення автоматизованого аналізу вразливостей. Ідея полягає в тому, щоб використати сильні сторони кожного підходу для компенсації слабких. Наприклад, символьне виконання може генерувати цільові вхідні дані, які направляють фазер до глибоко прихованих ділянок коду, збільшуючи таким чином покриття та ймовірність виявлення складних вразливостей. З іншого боку, фазинг може забезпечити швидке дослідження великої кількості різноманітних вхідних даних, виявляючи неочікувані сценарії, які можуть бути пропущені при статичному символьному аналізі. Інтеграція цих технік може включати використання результатів символьного виконання (наприклад, обмежень на вхідні дані, шляхів виконання) для керування процесом генерації вхідних даних фазером, або ж застосування фазингу для дослідження конкретних станів програми, досягнутих під час символьного виконання. Такий гібридний підхід має потенціал значно підвищити ефективність та глибину автоматизованого аналізу вразливостей порівняно з використанням кожного методу окремо [4].

Висновок

Проведене дослідження підтвердило важливість автоматичного пошуку вразливостей як ключового елементу забезпечення кібербезпеки. Розглянуто основні методи такого пошуку, зокрема фазинг та символьне виконання, виявлено їхні переваги та обмеження. Показано, що кожен з цих методів окремо не є ідеальним рішенням для виявлення всього спектру потенційних загроз. У роботі обґрунтовано перспективність інтеграції символьного

виконання та фазингу, що дозволяє поєднати глибокий аналіз коду з інтенсивним тестуванням різноманітними вхідними даними, підвищуючи таким чином ефективність виявлення складних вразливостей. Запропонований гібридний підхід має потенціал стати більш потужним інструментом у боротьбі з кіберзагрозами.

Список використаних джерел

1. Фаз-тестування: Виявлення вразливостей у програмному забезпеченні. URL: https://www.vpnunlimited.com/ua/help/cybersecurity/fuzz-testing?srsltid=AfmBOorR5Y5oxZteJB14NZ04iuuPBQrmKiv3767PLaFGCu6dwU5togRS (дата звернення 11.04.2025);

2. Фазинг Частина 1: Теорія, URL: https://sayfer.io/uk/ (дата звернення 11.04.2025);

3. Техніка нечіткого тестування та її використання в задачах кібербезпеки, URL: http://www.kibernetika.org/volumes/2022/numbers/01/articles/18/18.pdf (дата звернення 11.04.2025);

4. Fuzzing-тестування – ідеї та приклади, URL: https://devzone.org.ua/post/fuzzing-testuvannia-ideyi-ta-pryklady (дата звернення 11.04.2025).

ЗАСТОСУВАННЯ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ У WEB SCRAPING

Розумний І.[1],

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна*

У роботі досліджено застосування методів штучного інтелекту для автоматизованого збору, обробки та аналізу даних з вебресурсів, відомого як WEB scraping, з метою підвищення ефективності інформаційної безпеки. В умовах зростання кіберзагроз швидкий доступ до відкритих джерел стає критично важливим для виявлення потенційних ризиків. Основну увагу приділено підходам до структурування неформатованої інформації та її подальшої обробки за допомогою інструментів обробки природної мови. Запропоновано інтеграцію з великими мовними моделями (LLM), що дозволяє здійснювати контекстний аналіз отриманих даних. Окремо розглянуто основні проблеми вебскрапінгу, включаючи захист від ботів, змінність вебструктур і етичні обмеження, а також методи їх подолання. Розроблена система може бути використана для аналізу відкритих джерел у сфері кіберзахисту, зокрема для виявлення загроз, тенденцій та аномалій в інформаційному середовищі.

Вступ

Автоматизований збір даних з відкритих веб-ресурсів, відомий як, вебскрапінг набуває особливої важливості в умовах зростання кіберзагроз. Цей засіб активно застосовується для виявлення нових загроз, фішингових кампаній, витоків інформації та інших індикаторів шкідливої активності. Проте, основним викликом сьогодні є не стільки доступ до даних, скільки їх подальша ефективна та своєчасна обробка - через складність, обсяг і неструктурованість вебінформації.

У цій роботі досліджено можливості застосування штучного інтелекту, зокрема великих мовних моделей (LLM), для вдосконалення процесів аналізу зібраних даних. Такий підхід дозволяє автоматизувати інтерпретацію, виявлення загроз і витягування значущої інформації, що значно перевищує можливості традиційних інструментів.

Результатом дослідження є прототип системи вебскрапінгу з вбудованим модулем обробки даних на основі LLM, що орієнтований на потреби кібербезпеки, та загалом.

1. Аналіз моделей і методів штучного інтелекту

Штучний інтелект відіграє ключову роль у розвитку сучасних систем вебскрапінгу, особливо в контексті кібербезпеки, де важливо не тільки зібрати дані, а й вчасно виявити потенційні загрози.

Спочатку в вебскрапінгу використовували класичні методи машинного навчання, такі як логістична регресія та дерева рішень, для фільтрації контенту та класифікації текстів. Ці методи були ефективними в структурованих середовищах, але з розвитком складних вебструктур вони виявилися обмеженими.

З розвитком глибинного навчання, зокрема згорткових і рекурентних нейронних мереж, з'явилася можливість обробляти візуальні та текстові дані. Однак основна проблема – варіативність та неструктурованість вебданих – залишалася, оскільки сайти постійно змінюють структуру, а захисти типу CAPTCHA чи JavaScript ускладнюють процес.

Найбільше завдання для кібербезпеки полягає не лише у синтаксичному розборі даних, а й у глибокому розумінні їх змісту для виявлення фішингових елементів або витоків конфіденційної інформації. Це вимагає використання моделей обробки природної мови (NLP), які здатні проводити семантичний аналіз і виявляти тонкі зв’язки між даними.

В рамках цієї роботи пропонується інтеграція великих мовних моделей (LLM) для глибокого контекстного аналізу зібраних даних, що дозволяє ефективно виявляти загрози навіть у неструктурованому контексті. Основна увага

приділяється саме обробці даних: як витягнути необхідну інформацію, структурувати її та зрозуміти зміст для прийняття рішень.

Штучний інтелект підвищує ефективність захисту даних через автоматизацію збору інформації про ризики, виявлення загроз та реагування на інциденти, але водночас створює нові виклики, зокрема в питаннях конфіденційності та етики.

2. Роль і аналіз технологій web scraping

WEB Scraping – це процес автоматизованого збору та обробки даних з вебсайтів. Це важливий інструмент для ефективного отримання структурованої інформації з неструктурованих вебресурсів, що дозволяє здійснювати аналіз даних, моніторинг інформаційних потоків і конкурентів, а також використовувати ці дані для машинного навчання. Основні етапи процесу включають ідентифікацію цільових сайтів, аналіз структури вебсторінок, запит та отримання HTML-контенту, парсинг, очищення і зберігання даних. Техніки вебскрапінгу варіюються від простих регулярних виразів до складних алгоритмів обробки HTML, DOM-дерева та комп'ютерного зору. Вебскрапінг широко застосовується в бізнесі, маркетингу, медіа, фінансах та, зокрема, в кібербезпеці для виявлення загроз, моніторингу підозрілої активності та відслідковування фішингових атак.

Для вебскрапінгу використовують різні інструменти: браузерні розширення, десктопні програми та хмарні рішення. У результаті дослідження встановлено, що найбільш гнучкими є власні рішення на Python, зокрема бібліотеки requests для здійснення HTTP-запитів та BeautifulSoup для парсингу HTML. Для роботи з динамічними вебсторінками, що використовують JavaScript, найкраще підходить бібліотека Selenium, яка емулює поведінку браузера та дозволяє працювати з JavaScript-контентом. Кожен з інструментів відіграє свою роль та не може бути повністю замінений іншими.

Основними викликами вебскрапінгу є не лише технічні перешкоди, як блокування IP-адрес або CAPTCHA, які

можна обійти за допомогою різних інструментів, але й складність обробки витягнутих даних. Після збору інформації необхідно виконати її аналіз і структурування, що є ресурсозатратним і складним процесом. Зміни в структурі сайту, різноманітність форматів даних та необхідність у гнучкості для адаптації до нових умов створюють додаткові труднощі. Саме ця проблема і вирішується в цій роботі.

3. Розробка методики використання AI у web scraping

На відміну від традиційних підходів, де обробка зібраних даних здійснюється вручну або за допомогою жорстко заданих правил (наприклад, регулярних виразів або парсингу структурованого HTML), запропонована методика базується на використанні LLM. Вони дають змогу реалізувати гнучку, масштабовану та інтелектуальну обробку даних, що особливо важливо в умовах різнорідного та динамічного вебконтенту.

Основна ідея методики полягає у розділенні процесу скрапінгу на два незалежні етапи: отримання сирого контенту з вебсторінок за допомогою базових засобів (наприклад, BeautifulSoup або Selenium) та обробка отриманого тексту за допомогою LLM шляхом сформульованих промптів. Це дозволяє уникати створення окремого парсера для кожного сайту, натомість фокусуючись на уніфікованому підході до аналізу інформації.

LLM здатні працювати з неструктурованим або напівструктурованим текстом, знижуючи залежність від точного HTML-дизайну конкретного ресурсу. Завдяки цьому зменшується обсяг роботи при адаптації скрапера до нових сайтів, а процес структурування даних стає автоматизованим.

Переваги:

- масштабованість і універсальність підходу;
- адаптивність до змін DOM та HTML-структури;

- здатність витягувати складну інформацію без ручного кодування;
- зменшення обсягу ручної праці при обробці даних і формуванні структурованих результатів.

Недоліки:

- потреба у високих обчислювальних ресурсах;
- складність у валідації результатів (можливі «галюцинації»);
- залежність від якості формулювання промптів.

Висновки

В ході роботи було визначено, що основною проблемою при взаємодії з вебданими є не їхнє отримання, а ефективна обробка та інтерпретація. Тому було запропоновано підхід, що базується на використанні великих мовних моделей (LLM) для аналізу витягнутої інформації. Такий підхід є гнучким, масштабованим і дозволяє автоматизувати обробку даних у динамічних умовах, забезпечуючи якісний контент-аналіз незалежно від структури джерела.

СУЧАСНІ ІНСТРУМЕНТИ ВІДНОВЛЕННЯ ДАНИХ ДЛЯ ОЦІНКИ ДОСТОВІРНОСТІ ДОКАЗІВ У КРИМІНАЛІСТИЦІ

Борисова К.[1], Нагорний М. [1], Світличний В.[1]

[1]*Харківський національний університет внутрішніх справ, м. Кам'янець-Подільський,*
kate.borisova0502@gmail.com,
misha.nagornyy.2017@gmail.com, vit.svet@ukr.net

Робота присвячена аналізу сучасних інструментів відновлення даних для оцінки достовірності доказів у криміналістиці, зокрема утиліт TestDisk і PhotoRec. Автори досліджують їх функціональні можливості, такі як відновлення втрачених розділів, виправлення завантажувальних секторів та відновлення файлів з пошкоджених носіїв. Особлива увага приділяється практичному застосуванню цих інструментів у розслідуваннях для забезпечення об'єктивності та збереження цифрових доказів. Основною метою дослідження є підкреслення важливості використання таких утиліт для відновлення критично важливої інформації, яка може бути вирішальною у розкритті злочинів.

Ключові слова: криміналістика, цифрові докази, TestDisk, PhotoRec, відновлення даних, файлові системи, завантажувальні сектори, пошкоджені носії.

У сучасному світі, де злочинна діяльність все більше переходить у цифрову площину, криміналістика змушена адаптуватися до нових викликів. Особливо це стосується етапів діяльності, що передбачають аналіз цифрових доказів, таких як файли, зображення, відео, документи, збережені на комп'ютерах або інших носіях інформації.

Коли зловмисники намагаються приховати або видалити сліди своєї діяльності, саме сучасні інструменти для відновлення даних і оцінки їх достовірності можуть зіграти ключову роль у знаходженні істини.

Загалом, відновлення даних є критично важливим етапом у процесі встановлення повної картини правопорушення, оскільки воно забезпечує можливість реконструкції втрачених або пошкоджених інформаційних слідів, які можуть мати вирішальне значення для рослідування. Цей процес дозволяє не лише відновити доступ до прихованих або видалених даних, але й сприяє глибшому розумінню хронології подій та причинно-наслідкових зв'язків між діями зловмисників. Одним із найбільш надійних рішень є використання спеціалізованого програмного забезпечення для відновлення даних, зокрема таких утиліт, як TestDisk та PhotoRec (далі детальніше про них).

TestDisk – це утиліта для відновлення даних, яка розроблена для відновлення втрачених розділів і виправлення завантажувальних проблем, спричинених помилками у файловій системі або пошкодженими таблицями розділів. Вона також допомагає відновлювати дані після навмисного видалення або форматування диска [1].

Основні можливості TestDisk:

1. Відновлення розділів – якщо розділ жорсткого диска був навмисно видалений або пошкоджений, TestDisk може відновити його, відновивши таблиці розділів. Що може бути дуже корисно для відновлювання даних які зловмисник хотів би приховати;
2. Виправлення завантажувальних секторів – у разі пошкодження завантажувальних секторів TestDisk може їх відновити або переписати завдяки резервним копіям, створеним файловою системою. Ця функція дозволяє запобігти пошкодженням файлів.

PhotoRec – це компаньйон до TestDisk, спрямований на відновлення файлів з пошкоджених або недоступних носіїв, таких як цифрові камери, жорсткі диски або CD/DVD. PhotoRec може працювати навіть у випадках, коли файлова

система була серйозно пошкоджена, оскільки він використовує сигнатури файлів для відновлення даних [1].

Основні можливості PhotoRec:

1) відновлення файлів – здатність відновлювати широкий спектр файлів, таких як зображення (JPEG, PNG), документи (Word, PDF), аудіофайли (MP3, WAV), архіви (ZIP, RAR), відео (MP4, AVI) тощо;
2) підтримка багатьох файлових систем – PhotoRec працює незалежно від файлової системи і може відновлювати файли з FAT, NTFS, ext2/3/4, HFS+, ReiserFS та інші;
3) безпечне відновлення – під час відновлення файли зберігаються на іншому носії, щоб не пошкодити вихідний диск;
4) працює з пошкодженими носіями – програма успішно відновлює файли навіть із фізично пошкоджених або погано читаємих носіїв.

Висновки

Підсумовуючи можна сказати, що обидві програми надають широкі можливості для відновлення даних і є надійними інструментами в арсеналі сучасних криміналістів. Вони дозволяють ефективно працювати з цифровими доказами, навіть у випадках серйозного пошкодження або знищення. Функціональні можливості TestDisk і PhotoRec сприяють підвищенню об’єктивності розслідувань та можуть стати вирішальними у знаходженні істини.

Список використаних джерел

1. TestDisk / PhotoRec: відновлення даних - Softik. Softik. URL: http://softik.net.ua/testdisk-photorec-vosstanovlenye-dannyh/ (дата звернення: 13.09.2024).

МЕТОДИ ПРОТИДІЇ DDOS-АТАКАМ

Чорненька С. В.[1], Лашко О. О.[1]

[1]*Харківський національний університет внутрішніх справ*
snizhanachornenka16@gmail.com

У сучасних умовах цифровізації DDoS-атаки становлять одну з найсерйозніших загроз кібербезпеці, здатних паралізувати роботу інформаційних систем, завдати значних фінансових збитків та підірвати довіру користувачів. У статті розглянуто сучасні методи протидії DDoS-атакам, зокрема проактивні та реактивні підходи, такі як фільтрація трафіку, використання CDN, систем балансування навантаження, хмарних сервісів захисту (наприклад, Cloudflare, Akamai), а також застосування штучного інтелекту для аналізу аномалій. Окрему увагу приділено організаційним та правовим аспектам боротьби з DDoS-загрозами, включаючи планування реагування на інциденти та міжнародну координацію. Дослідження підкреслює необхідність комплексного підходу, що поєднує технічні, управлінські та нормативні рішення для ефективного захисту від динамічно еволюціонуючих кіберзагроз.

Ключові слова: DDoS-атаки, кібербезпека, проактивний захист, фільтрація трафіку, CDN, Cloudflare, штучний інтелект, IDS/IPS, кіберзагрози.

У сучасному цифровому середовищі питання забезпечення кібербезпеки набуває особливої актуальності. Одним із найнебезпечніших інструментів кіберзлочинців є DDoS-атаки (Distributed Denial of Service), які спрямовані на виведення з ладу інформаційних систем шляхом перевантаження їх штучно створеним надмірним трафіком. Такі атаки не лише паралізують роботу онлайн-сервісів, а й завдають значних фінансових збитків, порушують ділову репутацію та підривають довіру користувачів.

DDoS-атаки реалізуються шляхом розподілу шкідливого навантаження між великою кількістю пристроїв, які часто є частинами ботнетів – мереж заражених комп’ютерів або

IoT-пристроїв. Основна мета – змусити цільовий ресурс витратити усі свої обчислювальні або мережеві ресурси, що унеможливлює нормальне обслуговування легітимних користувачів.

Залежно від технічних характеристик атаки, її реалізація може відбуватись на різних рівнях – мережевому, транспортному, або на рівні додатків. Наприклад, flood-атаки (UDP, SYN, ICMP) спрямовані на перевантаження каналу або мережевих протоколів, тоді як application-level атаки, як-от HTTP flood, імітують дії реальних користувачів, що ускладнює їх виявлення [1].

Для ефективної протидії DDoS-атакам важливою є своєчасна ідентифікація загрози. Серед основних методів виявлення – моніторинг трафіку в реальному часі, виявлення аномалій, поведінковий аналіз активності користувачів, застосування систем виявлення та запобігання вторгненням (IDS/IPS). Також використовуються технології сигнатурного аналізу, які дозволяють ідентифікувати відомі типи атак, однак уразливі до нових форм шкідливого трафіку. Сучасні дослідження зосереджені на використанні штучного інтелекту та машинного навчання для побудови адаптивних систем аналізу, які здатні оперативно реагувати на нетипові загрози, що змінюються у режимі реального часу [2].

Методи протидії DDoS-атакам можна умовно поділити на реактивні та проактивні. Реактивні заходи реалізуються вже після початку атаки і спрямовані на нейтралізацію її наслідків. До них належать фільтрація трафіку, блокування IP-адрес з чорних списків, обмеження кількості запитів (rate limiting), географічне блокування, а також перенаправлення трафіку на фільтрувальні вузли.

Проактивні методи, своєю чергою, передбачають створення інфраструктури, стійкої до навантажень, — використання систем балансування навантаження, хмарних сервісів захисту, CDN-технологій, а також попереднє тестування стійкості систем до стресових навантажень [3].

Широко використовуються спеціалізовані сервіси захисту, як-от Cloudflare, Akamai, Arbor Networks, які надають можливість глибокого аналізу трафіку,

автоматичної фільтрації та нейтралізації DDoS-атак на мережевому і прикладному рівнях. Додатковим бар'єром може слугувати впровадження CAPTCHA, що ускладнює роботу ботів, а також використання двофакторної автентифікації для захисту критично важливих облікових записів. Ще одним інструментом є так зване «overprovisioning» – створення інфраструктури з надлишковими потужностями, здатної тимчасово витримати атакуючий трафік до моменту його локалізації [4].

Проте, незважаючи на наявні технічні засоби, боротьба з DDoS-атаками залишається складною через постійне ускладнення методів нападників. Зокрема, дедалі частіше використовуються розподілені атаки з шифрованим трафіком, атаки на специфічні уразливості веб-застосунків, а також атаки з використанням легітимних API-запитів. У цьому контексті перспективним напрямом є розвиток алгоритмів штучного інтелекту, що можуть вивчати та аналізувати поведінкові патерни користувачів, оперативно адаптуватися до нових загроз та блокувати небажану активність до того, як вона спричинить збій [2].

Отже, DDoS-атаки є серйозною загрозою, що потребує постійної уваги з боку фахівців з інформаційної безпеки. Успішна протидія цим атакам можлива лише за умови комплексного підходу, що поєднує технічні, організаційні та правові методи захисту. Майбутнє ефективної боротьби з DDoS-загрозами полягає у впровадженні інтелектуальних систем аналізу, розширенні можливостей проактивного захисту, а також розвитку міжнародного співробітництва у сфері кібербезпеки.

Список використаних джерел

1. Величко С.В. Засоби та механізми протидії ddos-атакам. С. 97–99.

2. Павлов М. Ю., Южакова Г.О. Метод захисту від dos-атак на основі цифрового відбитку браузера. *Системи та технології кібернетичної безпеки* : Всеукр. науково-практ. конференція студентів, аспірантів та молодих вчен. С. 174–176.

3. Способи захисту від DDos-атак. | EVROHOST. *EVROHOST Хостинг сайтів України*. URL: https://evrohost.com/ddos-protection/ (дата звернення: 14.05.2025).

4. DDoS атака: визначення, механізм дій і способи захисту. FoxmindEd. URL: https://foxminded.ua/ddos-ataka/ (дата звернення: 14.05.2025).

ВИКОРИСТАННЯ ДАНИХ SENTINEL ПРИ ДОСЛІДЖЕННІ ВОДНИХ ОБ'ЄКТІВ

Даншина С. Ю.[1]
[1]*Національний аерокосмічний університет «Харківський авіаційний інститут»*
Харків, Україна, s.danshyna@khai.edu

Розглянуто переваги використання даних Sentinel при дослідженні водних об'єктів. Систематизовано інформацію по водним індексам, наведено формули їх розрахунку. Проілюстровано результати їх використання при аналізі території Каховського водосховища.

Ключові слова: програма Copernicus, відкрити джерела інформації, супутникові знімки, водні індекси.

Вступ

Зміна клімату, виснаження природних ресурсів і економічна нестабільність – проблеми, вирішення яких сприяє сталому розвитку країн, які прагнуть досягти економічного зростання, не завдаючи шкоди довкіллю та майбутнім поколінням. Космічна програма Copernicus Європейського Космічного Агентства надає інформацію, яка дозволяє реагувати на глобальні виклики і координувати дії на всіх рівнях прийняття рішень на основі безкоштовних, регулярно отримуваних космічних зображеннях.

Місія Sentinel програми Copernicus

Програма Copernicus використовує супутники Sentinel 2A, 2B, що забезпечують оптико-електронну зйомку із застосуванням мультиспектральних камер (зйомка у видимому та інфрачервоному спектрі), з роздільною здатністю від 10 м до 60 м. Крім даних з Sentinel Copernicus поєднує дані з інших космічних місій [1]. З їх допомогою отримують значні обсяги даних, використання яких дає широкі можливості з оцінювання

впливу антропогенних і природних факторів на великих площах спостереження, для здійснювання пошуку та порівняння різночасових даних з метою вирішення завдань фіксування можливих змін на території дослідження.

Дані Sentinel для визначення водних об'єктів

Збільшення значення водних ресурсів у житті суспільства робить дослідження акваторій водних об'єктів актуальною задачею. Це можливе із застосуванням водних індексів, які розраховують за даними ДЗЗ шляхом формування певних комбінацій спектрозональних зображень. Найпоширенішими водними індексами є (табл. 1) [2, 3]:

- NDWI – нормалізований диференційний водний індекс (значення в діапазоні [-1;1]), причому NDWI<0 відповідає об'єктам без вологи, NDWI>0,2 притаманне для водойм;
- MNDWI – модифікований нормалізований диференційний водний індекс (значення в діапазоні [-1;1]), причому для водних об'єктів MNDWI>0;
- NDMI – нормалізований диференційний індекс зволоженості (значення в діапазоні [-1;1]), високі значення якого характеризують наявність підтоплення;
- NDTI – нормалізований диференційний індекс мутності (значення в діапазону [-1;1]), для чистих водойм –NDTI<0,4.

У табл. 1 використано такі позначення: Green – діапазон видимого зеленого спектру або спектральний канал B03 з центральною довжиною хвилі (CWL – Central Wavelength) 560 нм; NIR – діапазон інфрачервоного спектру або спектральний канал B08 з CWL=842 нм; SWIR2 – діапазон короткохвильового інфрачервоного спектру або спектральний канал B12 з CWL=2190 нм; SWIR1 – діапазон короткохвильового інфрачервоного спектру або спектральний канал B11 з CWL=1610 нм; Red – діапазон червоного спектру або спектральний канал B04 з CWL=665 нм.

Таблиця 1. Розрахунок водних індексів за даними ДЗЗ

Формула розрахунку	Рівняння по каналам Sentinel
1. NDWI – Normalized Difference Water Index	
(Green–NIR)/(Green+NIR)	(B03–B08)/(B03+B08)
2. MNDWI – Modified Normalized Difference Water Index	
(Green–SWIR2)/ (Green+SWIR2)	(B03–B12)/(B03+B12)
3. NDMI – Normalized Difference Moisture Index	
(NIR–SWIR1)/(NIR+SWIR1)	(B08–B11)/(B08+B11)
4. NDTI – Normalized difference turbidity index	
(Red–Green)/(Red–Green)	(B04–B03)/(B04+B03)

Для визначення водойм також використовують нормалізований диференційний вегетаційний індекс NDVI. Значення NDVI<0 характеризує водні об’єкти [2].

Як приклад застосування водних індексів розглянемо територію Каховського водосховища біля Нікополя. Знімок (рис. 1) отримано з Sentinel-2 L2A 23.04.2025 (хмарність 1,7 %, площа території – 310,70 км^2).

Рисунок 1. Територія дослідження в True color

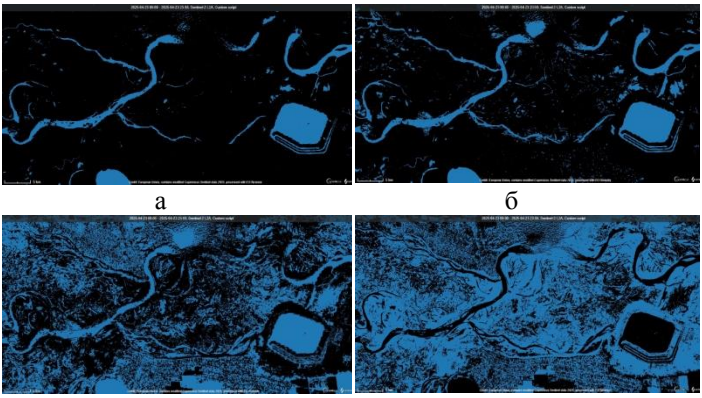
За допомогою вбудованих інструментів EO Browser знімок візуалізовано з використанням індексів NDWI (рис. 2, а), де водні об’єкти позначено синім кольором (при NDWI>0,5), та NDVI (рис. 2, б), де водні об’єкти зображено

в градаціях сірого (–0,1<NDVI<0,1).

а б

Рисунок 2. Обробка знімку території дослідження за допомогою інструментів EO Browser:
а – водний індекс; б – вегетаційний індекс

Також територію водосховища візуалізовано з використанням інструментів індивідуального налаштування відповідно до спектральних каналів табл. 1. Результати наведено на рис. 3, де візуально помітно, що кращі результати дає індекс MNDWI (рис. 3, б). Індекс NDMI, що зображує водні об’єкти синім кольором (рис. 3, в), надає значну похибку, дешифруючи також вологий ґрунт (якій характерний для весняного сезону). Індекс NDTI (рис. 3, г) зображує водні об’єкти чорним кольором, що відповідає чистій, не замутненій воді.



а б

в г

Рисунок 3. Аналіз території дослідження за індивідуальним налаштуванням: а – індекс NDWI; б – індекс MNDWI; в – індекс NDMI; г – NDTI

Висновок

Дані ДЗЗ є ефективним джерелом інформації для прийняття управлінських рішень при забезпеченні сталого розвитку країни. Використання відповідних індексів при аналізі водних об’єктів забезпечує: простоту і швидкість отримання результатів; високу точність визначення об'єктів гідрографії; отримання якісного результату, незважаючи на атмосферні впливи; розширення спектру завдань моніторингу водних об’єктів.

Список використаних джерел

1. eoPortal – URL: https://www.eoportal.org/.
2. Sentinel Hub – URL: https://www.sentinel-hub.com/.
3. EOS Data Analytics – URL: https://eos.com/uk/make-an-analysis/.

ДОСЛІДЖЕННЯ ВІДКРИТОГО КЛЮЧА У КРИПТОСИСТЕМІ AJPS-2 ТА ЇЇ МОДИФІКАЦІЯХ З ЗАМІНОЮ МОДУЛЯ

Дорошенко Ю. О. [1], Ядуха Д. В.[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, yuriido@ukr.net, dariya.yadukha@gmail.com*

У цій роботі проведено аналіз псевдовипадковості відкритого ключа криптосистеми AJPS-2 та її модифікацій з використанням узагальнених чисел Мерсенна та чисел Кренделла за допомогою набору тестів псевдовипадковості NIST SP 800-22.

Ключові слова: криптосистема AJPS, постквантова криптографія, числа Мерсенна, узагальнені числа Мерсенна, числа Кренделла

Вступ

Криптосистема AJPS-2 [1] була запропонована у межах конкурсу постквантових криптографічних примітивів NIST [2]. Вона базується на обчисленнях за модулем числа Мерсенна, що є вагомою перевагою, адже дозволяє використання численних оптимізацій обрахунку ресурсозатратних операцій за модулем Мерсенна.

Опис криптосистеми AJPS-2

В роботі використовуються такі позначення:

- M_n– число Мерсенна виду $2^n - 1$, де $n \in N$;
- $M_{n,c}$– число Кренделла виду $2^n - c$, де n та c – додатні цілі числа та $\log_2 c \leq \frac{n}{2}$;
- $M_{n,m}$ – узагальнене число Мерсенна вигляду $2^n - 2^m - 1$, де $m < n$ – додатні цілі числа;

- $Ham(x)$ – вага Геммінга числа x, тобто кількість одиниць у бітовому представленні числа x.

В основі стійкості криптосистеми AJPS-2 лежить задача MLHCSP.

Означення $(MLHCSP)$: Маючи число Мерсенна M_n, ціле число h та пару чисел (R, T), де R – випадково обраний лишок за модулем M_n, значення T обчислено відповідно до співвідношення $T = F \cdot R + G \ (mod\ M_n)$, причому F та G є лишками за модулем M_n з вагою Геммінга h, знайти числа F та G.

Криптосистема AJPS-2 дозволяє зашифрувати блок бітів повідомлення довжини λ, де λ – параметр захищеності. Відкритими параметрами AJPS-2 є:

- число Мерсенна $M_n = 2^n - 1$;
- параметр захищеності λ;
- число $h \in N$, що задовольняє умовам $h = \lambda$ та $10h^2 < n \leq 16h^2$.

В межах цієї роботи досліджується відкритий ключ криптосистеми AJPS-2, який генерується за алгоритмом створення ключів, що наведений далі.

1) Випадковим чином обираються числа F і G – лишки за модулем M_n з вагою Геммінга h. Число F є особистим ключем, а значення G – секретним параметром криптосистеми.
2) Випадковим чином обирається R – лишок за модулем M_n.
   Обчислюється число T за співвідношенням $T = F \cdot R + G \ (mod\ M_n)$. Відкритим ключем є пара (R, T), а особистим – F.

Модифікації криптосистеми з використанням чисел Кренделла та узагальнених чисел Мерсенна мають аналогічний алгоритм генерації відкритого ключа до оригінальної AJPS-2, з заміною модуля на $M_{n,c}$ та $M_{n,m}$

відповідно. Більш детально з описом модифікацій можна ознайомитись у статті [3].

Результати тестів псевдовипадковості NIST SP 800-22 для відкритого ключа T

Для перевірки наскільки бітові послідовності є справді випадковими, NIST використовує стандарт NIST SP 800-22. Він містить 15 статистичних тестів, призначених для виявлення закономірностей або аномалій, які могли б свідчити про відхилення від очікуваної випадкової поведінки. Детально з принципом роботи кожного тесту можна ознайомитись в джерелі [2].

P-value (значення p) – це ймовірність того, що для істинно випадкової послідовності результат тесту є таким ж, як для досліджуваної послідовності. В межах тестів NIST SP 800-22 мінімальним значенням p для всіх тестів є 0.01.

Для аналізу криптосистеми AJPS-2 та її модифікацій було обрано такі параметри: $n = 4253$, $h = 32$, $\lambda = 267$, $c = 3981$, $m = 399$. Такі значення параметрів n, h, λ є одними з рекомендованих розробниками AJPS-2, а параметри c та m обиралися так, щоб числа $M_{n,c}$ та $M_{n,m}$ були простими. Тести проводились на згенерованих послідовностях ключів T по 16 мегабайт для кожної варіації AJPS-2.

На підставі аналізу значень p в таблиці 1 можна зробити висновок, що жодна з модифікацій не демонструє критичних порушень випадковості, проте найнижче p спостерігається для $M_{n,m}$ в Serial Test (2) (0,0637) і Frequency Monobit (0,1751). Більшість тестів дають значення p значно вищі за критичне 0,01, що дозволяє вважати відповідні послідовності статистично випадковими.

Таблиця 1. Результати тестів псевдовипадковості відкритого ключа

Тест	M_n	$M_{n,c}$	$M_{n,m}$
Frequency (Monobit)	0,8509	0,8587	0,1751
Frequency (Block)	0,9496	0,2089	0,8642
Runs Test	0,8572	0,2150	0,7100
Longest Run of Ones	0,4702	0,6901	0,7935
Binary Matrix Rank	0,2591	0,4123	0,8177
Spectral Test	0,4037	0,7972	0,4798
Non-overlapping Temp.	0,0842	0,2042	0,3882
Overlapping Temp	0,2510	0,8280	0,9123
Universal Test	0,8402	0,5524	0,3666
Linear Complexity	0,5259	0,6991	0,4086
Serial Test (1)	0,7982	0,2673	0,1720
Serial Test (2)	0,4187	0,3278	0,0637
Approximate Entropy	0,7788	0,2696	0,2050
Cumulative Sums (F)	0,5450	0,7885	0,3194
Cumulative Sums (B)	0,3987	0,9307	0,0706

Таблиця 2. Результати тесту псевдовипадковості відкритого ключа: Random Excursions Test

Тест	M_n	$M_{n,c}$	$M_{n,m}$
-4	0,6906	0,6798	0,3416
-3	0,0899	0,3402	0,9935
-2	0,4789	0,9402	0,8181
-1	0,4247	0,3824	0,8592
+1	0,4585	0.3433	0,3274
+2	0,4357	0,0742	0,2652
+3	0,4442	0,0311	0,7632
+4	0,5928	0,9998	0,3270

В тестах Random Excursions та Random Excursions Variant, результати яких наведені у таблицях 2 та 3 відповідно, значення p для всіх варіацій AJPS-2 знаходиться далеко від критичного рівня. Всі варіації демонструють значення p вище 0.2 для всіх станів, що підтверджує відсутність систематичних відхилень у випадковому блуканні бітових послідовностей.

Таблиця 3. Результати тесту псевдовипадковості відкритого ключа: Random Excursions Variant Test

Тест	M_n	$M_{n,c}$	$M_{n,m}$
-9	0,9652	0,6958	0,8921
-8	0,8769	0,8873	0,7598
-7	0,6295	0,9848	0,9128
-6	0,4807	0,8040	0,9604
-5	0,3955	0,3849	0,9127
-4	0,5182	0,2230	0,5929
-3	0,6973	0,3187	0,4531
-2	0,9448	0.5929	0,4588
-1	0,3372	0,7060	0,5537
+1	0,1972	0,4927	0,4108
+2	0,8219	0,4399	0,4250
+3	0,8092	0,7474	0,5464
+4	1,0000	0,8560	0,3447
+5	0,9601	0,3604	0,2633
+6	0,6382	0,2821	0,3936
+7	0,3516	0,5301	0,3911
+8	0,2041	0,6902	0,4497
+9	0,0848	0,7268	0,4112

Висновки

У роботі розглянуто криптографічний примітив AJPS-2, що побудований на арифметиці за модулем чисел Мерсенна. Досліджено дві його модифікації з використанням чисел Кренделла та узагальнених чисел Мерсенна як альтернативних модулів. Проведено тестування псевдовипадковості відкритих ключів відповідно до стандарту NIST SP 800-22. Результати показали, що всі варіації криптосистеми демонструють прийнятні показники псевдовипадковості, що підтверджує доцільність застосування альтернативних класів модулів AJPS-2 для підвищення варіативності параметрів та стійкості системи AJPS-2.

Список використаних джерел

1. A New Public-Key Cryptosystem via Mersenne Numbers / D. Aggarwal та ін. *NIST PQC*. 2017. URL: https://eprint.iacr.org/2017/481.pdf.

2. Post-Quantum Cryptography Standardization. *NIST*. *Post-Quantum Cryptography*. 03.01.2017. URL: https://csrc.nist.gov/pqc-standardization.

3. Yadukha D. The Forgery Attack on the Post-Quantum AJPS-2 Cryptosystem and Modification of the AJPS-2 Cryptosystem by Changing the Class of Numbers Used as a Module. *Theoretical and cryptographic problems of cybersecurity*. URL: https://doi.org/10.20535/tacs.2664-29132023.1.286166.

ХМАРНІ ТЕХНОЛОГІЇ ТА РИЗИКИ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ В УМОВАХ ВІЙНИ

Колісниченко К.[1]

[1]*Харківській національний університет внутрішніх справ*

Розглядаються особливості використання хмарних технологій в умовах військового конфлікту, з акцентом на нові виклики в сфері інформаційної безпеки. Проаналізовано ключові ризики, пов'язані з обробкою та зберіганням даних у хмарному середовищі, зокрема у контексті атак на критичну інфраструктуру, порушення каналів зв'язку, втручання державних або недержавних акторів. Визначено основні вектори загроз: несанкціонований доступ, компрометація хмарних сервісів, втрати даних, залежність від сторонніх постачальників.

Ключові слова: хмарні технології, інформаційна безпека, кіберзагрози

Вступ

У XXI столітті хмарні технології стали невід'ємною частиною інформаційної інфраструктури як у державному, так і в приватному секторі. Вони забезпечують гнучкість, масштабованість і мобільність обробки даних, дозволяють забезпечити доступ до інформаційних ресурсів у будь-який час і з будь-якого місця. Особливо актуальним це стало в умовах повномасштабної війни, коли традиційна IT-інфраструктура зазнає фізичних руйнувань, а мобільність і швидкість розгортання цифрових сервісів набувають критичного значення.

Виклад основного матеріалу

Разом із тим активне впровадження хмарних рішень супроводжується зростанням ризиків інформаційної безпеки, які в умовах воєнного конфлікту суттєво посилюються. До основних загроз належать: несанкціонований доступ до даних, уразливості хмарних сервісів, DDoS-атаки, витоки конфіденційної інформації,

компрометація облікових записів, а також можливість використання хмарної інфраструктури у військово-розвідувальних цілях з боку противника. Залежність від сторонніх постачальників послуг, які можуть знаходитися за межами країни, створює додаткові виклики у контексті цифрового суверенітету.

У сучасних умовах ведення війни інформаційний простір і цифрова інфраструктура стали ключовими елементами державної обороноздатності. Хмарні технології, або cloud computing, відіграють дедалі важливішу роль в організації та захисті державних, військових, корпоративних і публічних цифрових ресурсів. Їхня гнучкість, масштабованість, швидкий доступ до обчислювальних потужностей та можливість віддаленої роботи забезпечують стабільність функціонування систем управління, зв'язку, баз даних і сервісів навіть в умовах фізичного знищення локальних серверів або об'єктів інфраструктури. Водночас активне використання хмарних сервісів в умовах війни супроводжується новими ризиками для інформаційної безпеки, які вимагають всебічного аналізу та комплексної протидії.

Передусім, у воєнний час хмарні платформи стають потенційною мішенню для кібератак з боку держави-агресора. Злам хмарної інфраструктури, несанкціонований доступ до чутливої інформації або її знищення можуть мати катастрофічні наслідки як для державного управління, так і для безпеки громадян. Особливо небезпечними є цілеспрямовані атаки типу APT (Advanced Persistent Threats), коли ворожа група довго і непомітно працює над проникненням у хмарну систему для викрадення або компрометації критичних даних.

Ще одним актуальним ризиком є втрата контролю над даними, що зберігаються за межами фізичної території держави. У багатьох випадках сервери, де розміщується хмарна інфраструктура, знаходяться в інших країнах, що може створювати юридичні та політичні ризики в разі конфлікту або блокування сервісу. У воєнний час особливо гостро постає питання цифрового суверенітету — можливості держави повноцінно контролювати доступ до

власних інформаційних ресурсів без залежності від третіх сторін.

Також хмарні технології пов'язані з ризиком витоку даних через людський фактор: помилки конфігурації, неналежний рівень доступу, слабкі паролі, використання ненадійних API або несертифікованих сторонніх сервісів. У період воєнного часу, коли багато процесів перевантажені, а персонал працює в умовах стресу або під тиском часу, ці загрози лише посилюються [1].

Разом з тим, хмарні сервіси — це не тільки джерело ризиків, а й одна з умов цифрової стійкості держави. Після початку повномасштабного вторгнення Україна швидко перевела частину критично важливих реєстрів та урядових систем у хмарні середовища, зокрема на базі Amazon Web Services, Microsoft Azure та інших західних провайдерів. Це дозволило зберегти безперебійну роботу цифрових сервісів, у тому числі таких важливих, як «Дія», «Єдиний державний демографічний реєстр», податкові сервіси тощо, навіть в умовах обстрілів, окупації або руйнування дата-центрів [2].

Щоб мінімізувати ризики, необхідно дотримуватись вимог кібергігієни та стандартів інформаційної безпеки при використанні хмарних платформ: впроваджувати багатофакторну автентифікацію, регулярно оновлювати політики доступу, застосовувати сертифіковані засоби шифрування, здійснювати аудит хмарного середовища та резервне копіювання даних. Також важливо розвивати власні національні хмарні платформи, які відповідатимуть вимогам державної безпеки, зокрема для розміщення чутливої військової або розвідувальної інформації.

Висновки

У підсумку, хмарні технології стали невіддільною частиною оборонної спроможності сучасної держави. Вони забезпечують мобільність, масштабованість і адаптивність цифрової інфраструктури в умовах постійної загрози. Проте ефективне використання хмарних рішень у період війни потребує ретельного аналізу ризиків, нормативно-правового регулювання, інвестицій у кіберзахист і міжнародної координації. Лише за умов системного

підходу хмара стане не точкою вразливості, а опорою для безпечного цифрового майбутнього навіть у часи найбільших загроз.

Список використаних джерел

1. Repository at ChSTU: Cloud technologies of the information economy: issues of institutionalization and measures of ukrainian management in the conditions of war. Repository at ChSTU: Home. URL: https://er.chdtu.edu.ua/handle/ChSTU/4732 (дата звернення: 11.05.2025).

2. Як органи державної влади можуть безпечно перейти до використання хмарних технологій. KPMG. URL: https://kpmg.com/ua/uk/home/ insights/2024/03/securing-the-cloud.html (дата звернення: 11.05.2025).

РОЗРОБЛЕННЯ ЗАСТОСУНКУ ДЛЯ ПРОЄКТУВАННЯ ЗЕЛЕНИХ ЗОН МІСТА ЗА ДАНИМИ ДЗЗ

Лаптій П. О.[1]

[1]*Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна,*
p.o.laptii@khai.edu

У роботі представлено опис застосунку проєктування зелених зон у містах із використанням даних дистанційного зондування Землі (ДЗЗ). Застосунок аналізує супутникові знімки, розраховує вегетаційні індекси NDVI, EVI та унікальний індекс VEI. Його використання під час проєктування зелених зон допомагає у вирішенні питань зниження теплового навантаження, покращення якості повітря та підвищення біорізноманіття міста в умовах обмежених ресурсів.

Ключові слова: відкриті джерела інформації, зелені зони, Python, вегетаційні індекси, індекс VEI

Вступ

Сучасні мегаполіси стикаються з низкою екологічних проблем, зокрема: високим рівнем забруднення повітря; недостатньої кількістю зелених зон; ефектом теплового острова, який збільшує температуру в центрі міста на 5-7 ^{0}C у літню пору. Метою роботи є створення інформаційної системи для автоматизованого аналізу супутникових даних та розрахунку індексів рослинності, які допоможуть: виявити території з низькою щільністю рослинності; виявити області найбільшої температурної концентрації; візуалізувати інформацію, що допоможе міської владі у створенні планів озеленення. Як зону дослідження обрано м. Харків.

Структура застосунку

Застосунок складається з таких модулів:

1. Обробка даних – виділення червоного, синього та ближнього інфрачервоного каналів із GeoTIFF (рис. 1).

Задля зручності та зведення дій користувача до необхідного мінімуму було прийнято рішення, що вибір необхідних діапазонів буде виконуватись безпосередньо програмою. Необхідна інформація буде збережена у оперативній пам'яті до закриття програми, або до завантаження нового космознімку.

```
# Функція для обробки RGB GeoTIFF і виділення необхідних шарів
blue_band = None
red_band = None
ndvi = None
evi = None
vei = None

def process_rgb_image():
    global blue_band, red_band
    rgb_image_path = select_file()
    if not rgb_image_path:
        messagebox.showwarning("Помилка", "Будь ласка, оберіть RGB GeoTIFF.")
        return

    try:
        with rasterio.open(rgb_image_path) as src:
            # Перевіряємо наявність необхідних смуг
            if src.count < 3:
                raise ValueError("У вибраному зображенні недостатньо смуг. Потрібні 1 (Blue), 2 (Green), 3 (Red).")

            # Виділяємо канали
            blue_band = src.read(1).astype(np.float32)  # Band 1 (Blue)
            red_band = src.read(3).astype(np.float32)  # Band 3 (Red)

        messagebox.showinfo("Успіх", "RGB GeoTIFF оброблено, і необхідні канали використано для розрахунків.")

    except Exception as e:
        messagebox.showerror("Помилка", str(e))
```

Рисунок 1. Код, що виділяє необхідні діапазони на супутниковому знімку

Наступний крок – завантаження знімку у ближньому інфрачервоному спектрі.

2. Розрахунок індексів. Визначення індексу NDVI за формулою:

$$NDVI = \{NIR - RED\}/\{NIR + RED\},$$

де NIR – відбивна здатність у ближньому інфрачервоному діапазоні (Near-Infrared); RED – відбивна здатність у червоному діапазоні.

Розрахунок індексу EVI за формулою

$$EVI = 2.5 \cdot ((NIR - RED)/(NIR + 6 * RED - 7.5 * BLUE + 1)),$$

де BLUE – відбивна здатність у синьому диапазоні

Додатково запропоновано використовувати індекс VEI як усереднення між NDVI та EVI, що відповідає сучасним підходам до аналізу різноманітності деревного покриву за супутниковими знімками, зокрема, описаним у [1]. Його значення знаходимо за формулою:

VEI = (NDVI + EVI)/2

3. Інтерфейс користувача складається з кнопок для завантаження даних, виконання розрахунків і візуалізації результатів (рис. 2).

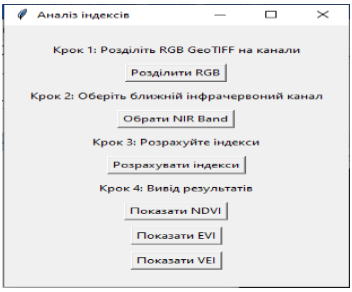


Рисунок 2. Інтерфейс керування застосунком

Застосунок дає змогу:

- виводити окремі карти, що класифікують об’єкти за індексами NDVI, EVI та VEI (рис. 3);
- зберігати результати у форматі GeoTIFF для подальшого використання.

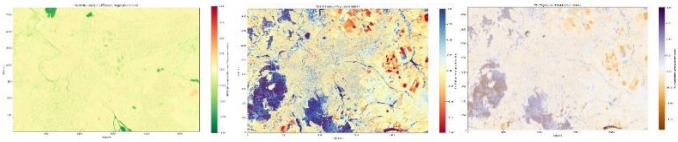


Рисунок 3. Картографування результатів розрахунку індексів

Тестування застосунку на даних Sentinel-2 показало його можливість застосування при проєктуванні зелених зон (рис. 4) [2]. Зокрема, у центрі міста виявлено райони з

низьким NDVI (<0.2), де має сенс створення зелених коридорів і дахів.

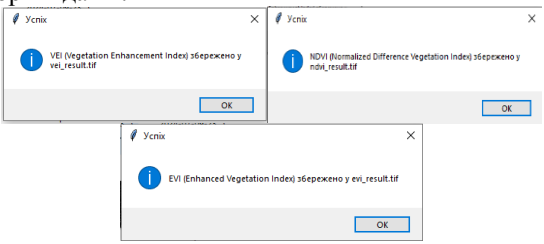


Рисунок 4. Підтвердження успішного збереження файлів із розрахованими індексами.

Застосунок дозволяє швидко оброблювати супутникові знімки різної точності та дає змогу обробляти знімки невисокої роздільної здатності, що робить його корисним для органів влади.

Висновок

Розроблений застосунок автоматизує процес аналізу даних ДЗЗ при проєктуванні зелених зон у місті. Він дає змогу візуалізувати результати аналізу екологічного стану міста у вигляді карт індексів рослинності та теплового навантаження. Індекс VEI показав себе ефективним інструментом, що об’єднує переваги NDVI та EVI.

Перелік використаних джерел

1. Remote sensing of subtropical tree diversity: The underappreciated roles of the practical definition of forest canopy and phonological variation [Text] / Y. Liu at el. // Forest Ecosystems. – 2023. – Vol. 10, – Article 100122.

2. Eskandari S. Provision of land use and forest density maps in semi-arid areas of Iran using Sentinel-2 satellite images and vegetation indices [Text] / S. Eskandari, S. K. Bordbar // Advances in Space Research. – 2025. – Vol. 75, issue 3. – P. 2506-2534 DOI: https://doi.org/10.1016/j.asr.2024.10.060.

DEVELOPMENT OF AN ANTI-SPOOFING METHOD FOR IMAGES IN BIOMETRIC SECURITY SYSTEMS USING ML

Palahin V. [1], Palahina O. [1], Ivchenko O. [1],
Protsenko O. [1], Bairak A.[1]

[1]*Department of Robotic and Telecommunication Systems and Cybersecurity, Cherkasy State Technological University, Cherkasy, Ukraine*
v.palahin@chdtu.edu.ua

The integrity and technical security of biometric authentication systems are critical in applications that require protection against mobile access to high-security areas. The main technical security issue is the vulnerability to spoofing attacks, where fraudulent biometrics deceive the system. This study proposes to use the ratio of facial landmarks from real-time video streams to improve live person detection in security systems. By analyzing responses to interactive prompts and calculating facial landmark ratios, the system aims to more reliably distinguish between genuine and fake interactions. This approach leverages advanced real-time image processing techniques to enhance accuracy in environments where traditional static biometric systems often fail. Utilizing dynamic response analysis further helps in mitigating the risks associated with spoofing by requiring active participation from the subject, therefore adding an additional layer of security.

Keywords: technical protection of information, biometric security, machine learning, human detection, spoofing attacks

Introduction

In today's digital world, biometric systems play a key role in securing access to information resources, financial services, mobile devices, and critical infrastructure. However, these systems are increasingly targeted by spoofing attacks that

involve the use of fake biometric data – such as photographs, videos, or 3D masks – to imitate a user's face. Therefore, the development of reliable spoof detection methods is critically important to maintain a high level of trust in biometric systems and to prevent unauthorized access.

One of the traditional approaches to spoof detection in biometric systems involves the use of hardware-based solutions, such as depth sensors, infrared and ultraviolet imaging, which enable the detection of the physical presence of an object and its spatial characteristics. Additionally, liveness detection methods [1] are widely used, based on identifying micro-movements of the eyes, blinking, changes in skin texture, or responses to light. While these approaches provide a relatively high level of protection, they are often associated with high financial costs, complexity of integration into existing systems, and reduced user convenience.

Another category of anti-spoofing methods relies on image analysis using classical image processing algorithms – including texture histograms (LBP), gradient features (HOG), or optical flow for video analysis. These methods attempt to detect differences between a real face and a printed or digital image by analyzing artifacts, distortions, or the lack of depth. However, they often prove to be insufficiently flexible and effective when confronted with newer or more sophisticated types of attacks, especially in cases of high-quality spoofing. A promising direction is the analysis of various anomalies in facial images and the use of dynamic real-time methods to improve forgery and spoof detection [2, 3].

Unlike traditional methods, machine learning (ML) approaches – particularly deep neural networks – demonstrate significantly better generalization and adaptability. They can automatically extract complex features from images, consider contextual and spatial dependencies, and learn from large datasets, which makes them robust against new and previously unseen types of attacks. Moreover, ML models eliminate the need for additional hardware and can be integrated into existing software solutions, which significantly enhances the efficiency and economic feasibility of implementing anti-spoofing systems.

The aim of this study is to improve the efficiency of biometric security systems by applying multi-level protection against forgery using machine learning methods.

Methodology

This study investigates the usefulness of landmark ratio analysis on human faces to improve spoofing and substitution detection in biometric security systems. We analyze video data capturing various facial expressions to identify key features. To ensure the accuracy of the analysis, the preprocessing of the raw data includes normalization steps to account for differences in lighting, orientation and scale. This preprocessing helps to minimize external influences that can affect the accuracy of landmark detection. Post-processing involves the use of state-of-the-art computer vision techniques to recognize faces and then extract facial landmarks. This stage is critical because it transforms raw video data into structured data points that can be quantitatively analyzed. Ratio calculations involving distances between facial landmarks, such as the eye-mouth ratio, have been useful in previous studies on attention and emotion recognition. The selection of specific ratios, particularly those that have proven effective in detecting subtle facial movements, will be based on their sensitivity to variations in live and non-live interaction. These ratios will be analyzed to identify characteristic patterns that reliably indicate a live presence. The effectiveness of the identified ratios will be evaluated by comparing their performance in accurately distinguishing between genuine and fake interactions. This evaluation will focus on the accuracy and robustness of the methodology under various testing conditions that reflect typical scenarios encountered in security systems.

Results and discussion

The results of this study demonstrate the effectiveness of using facial landmark ratios – particularly eye and mouth landmark ratios – for detecting live human presence. These indicators have proven useful in previous studies for identifying various facial states and may be effective in distinguishing live

individuals from substitution attempts. Variance analysis of changes in facial landmark measurements reveals patterns that differentiate the presence of a real person from a spoofed one. For instance, live human reactions during biometric face analysis exhibit natural variability in facial ratios due to muscle movements, whereas forged responses may show reduced variability or unnatural changes that do not align with typical human behavior.

Moreover, the temporal dynamics of these ratios during different interactions can further enhance detection processes. Observing how these ratios fluctuate over time within a single session provides an additional layer of data, potentially increasing the system's sensitivity to subtle signs of life that may be missed by static measurements. The study explores threshold values for facial feature changes and identifies the conditions under which these indicators are most effective. Determining these parameters is critical for improving the methodology and enhancing the accuracy and reliability of biometric identification in technical information security systems.

The study also addresses potential challenges, such as variations in lighting, facial coverings, or other environmental factors that can affect the accuracy of facial landmark detection. Mitigation strategies are considered, including the use of advanced normalization techniques or adaptive thresholding methods that can adjust to different conditions in real time.

Conclusions

This study aims to explore the potential of facial landmark ratios, such as the ratio of eye and mouth landmarks, as indicators for detecting a living person in biometric security systems. The results show that the proposed approach to analyzing biometric facial features is an effective tool for distinguishing between living people and countering spoofing attacks. The adaptability of this methodology to real-time data processing and machine learning algorithms opens the way for continuous improvement, making it a reliable solution in dynamic security environments. In addition, integrating this system into multimodal biometric systems that include voice,

fingerprint, or iris recognition can potentially offer an even more secure authentication system. Supplementing a biometric security system with additional feature analysis will significantly improve protection against spoofing attacks and increase the reliability of security systems.

References

1. Hadid, A., et al. (2014). Biometric Liveness Detection: Challenges and Research Opportunities. IEEE Security & Privacy Magazine, 12(5), pp. 63-72. DOI: https://doi.org/10.1109/MSP.2015.116

2. Erdogmus, N., & Marcel, S. (2014). Spoofing Face Recognition With 3D Masks. IEEE Transactions on Information Forensics and Security, 9(7), 1084-1097. https://doi.org/10.1109/TIFS.2014.2322255

3. Soukupová, T., & Čech, J. (2016). Real-time eye blink detection using facial landmarks. Proceedings of the 21st Computer Vision Winter Workshop.

СЕГМЕНТАЦІЯ ОБ'ЄКТІВ З ВИКОРИСТАННЯМ МАШИННОГО ЗОРУ І ВІДКРИТИХ ДАНИХ ДЗЗ

Подорожко К. Д.[1]

[1]*Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна,*
k.d.buriak@khai.edu

У роботі розглянуто аналітичну комп'ютерну програму «Пошук» для дешифрування об'єктів з використанням машинного зору та даних дистанційного зондування Землі. Наведено алгоритм програми, яка з використанням методів семантичної сегментації класифікує об'єкти на космічних знімках. Підтверджено, що аналіз відкритих даних космічного моніторингу та дистанційного зондування Землі із застосування розробленої програми підвищує ефективність аналізу та оброблення космічних знімків і пришвидшує процес прийняття рішень.

Ключові слова: семантична сегментація, комп'ютерна програма «Пошук», дистанційне зондування Землі, космічний моніторинг.

Вступ

В умовах суттєвого антропогенного впливу на навколишнє середовище моніторинг стану річкових систем набуває особливої важливості. Оскільки дистанційне зондування Землі (ДЗЗ) стає важливим методом моніторингу водних об'єктів, його все частіше використовують для оцінювання стану довкілля. Інтеграція методів ДЗЗ з алгоритмами машинного зору є надзвичайно актуальним завданням для підтримки прийняття рішень з метою підвищення ефективності аналізу стану річок [1].

Інтеграція методів ДЗЗ та алгоритмів машинного зору

ДЗЗ признано ефективним інструментом дослідження, який не потребує від дослідника присутності в складних

умовах. Багатоспектральні дані проєкта Sentinel-2 та інші супутникових платформ надають величезну кількість інформації для оцінювання стану навколишнього середовища. Призначений для моніторингу Землі з використанням супутників, Sentinel-2 стає найважливішим джерелом даних космічної програми Copernicus Європейського Космічного Агентства. У цьому контексті застосування алгоритмів машинного зору для оброблення космічних знімків із Sentinel-2 відкриває нові можливості для [2]:

- оперативного виявлення змін у прибережних зонах
- контролю рівня води, зміни русел і затоплень;
- визначення стану водного дзеркала та рослинності з використанням спектральних індексів;
- автоматизованого аналізу змін у часі на великих територіях.

Завдяки супутниковим знімкам високої роздільної здатності та їх аналізу з використанням алгоритмів машинного зору (семантичною сегментацією, класифікацією об’єктів, детекцією змін) можливо значно підвищити точність і швидкість екологічного моніторингу річок. Це особливо важливо для превентивного виявлення ризиків забруднення, контролю за дотриманням прибережно-захисних смуг і підтримки сталого водокористування.

Алгоритм машинного зору на основі семантичної сегментації

Семантична сегментація є одним з найпоширеніших сучасних методів комп'ютерного зору. Даний метод є різновидом машинного навчання, який використовують для аналізу зображень на піксельному рівні. Він дає змогу комп'ютерам «розуміти» вміст зображень, аналізуючи кожний піксель і поділяючи їх за певним класом, наприклад, «сільськогосподарські угіддя», «водні об’єкти», «промислова забудова» тощо. Внаслідок отримують маску, де кожному пікселю присвоюється відповідна мітка, яка вказує на його належність до певного об’єкту або класу [3].

Ураховуючи основні аспекти методу семантичної сегментації створено комп'ютерну програму в MS Visual Studio C#, яка дозволяє обробляти відкриті космічні знімки та класифікувати певні об'єкти. Алгоритм програми наведено у вигляді схеми на рис. 1 [4].

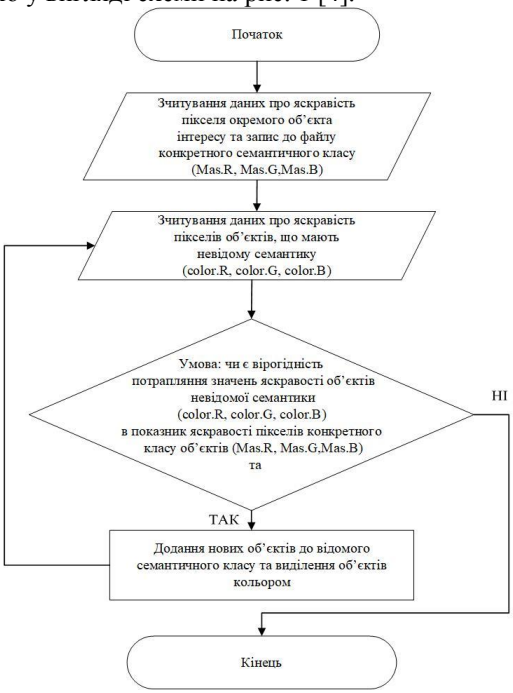


Рисунок 1. Схема алгоритму програми «Пошук»

Програму застосовано для дешифрування об'єктів в басейні річки Сіверський Донець. На рис. 2 зона А дешифрує об'єкти за зміною значень яскравостей на тестовій ділянці, зона В маркує області появи пікселів з аналогічною для тестової ділянки яскравістю.

Робота з програмою підтвердила її ефективність при вирішенні завдань моніторингу стану Сіверського Донця, контролю дотримання водоохоронних зон в його руслі та прогнозування динаміки його змін.

Рисунок 2. Ілюстрація роботи програми

Висновок

Таким чином, інтеграція методів машинного зору з відкритими даними ДЗЗ є ефективним інструментом у вирішенні завдань моніторингу довкілля. Такий підхід дозволяє автоматизувати процес аналізу великих обсягів даних, оперативно виявляти зміни природних та антропогенних об'єктів, а також здійснювати точну класифікацію елементів земної поверхні.

Список використаних джерел

1. Курач, Т. М. Цифрове оброблення та дешифрування знімків – К., 2021 – 50 с.

2. Integrating river discharge and Sentinel-2 satellite imagery for enhanced turbidity mapping in arid region rivers: A machine learning approach / M. Ahmadi, A. Noori, S. H. Mohajeri, M. R. Nikoo // Physics and Chemistry of the Earth. – 2025. – Vol. 138. – Article 103869.

3. Chen Z., Hu M., Bo Ch. Thermal image-guided complementary masking with multiscale fusion for multi-spectral image semantic segmentation // Engineering Applications of Artificial Intelligence – 2025. – Vol. 150. – Article 110569.

4. Комп'ютерна програма для роботи з зображеннями «Пошук»: а. с. 125831 / С. Ю. Даншина, К. Д. Подорожко; заявл. 18.04.2024 р.; опубл. 31.05.2024; Бюл. № 81. С 769 – 770.

СИСТЕМА ВИЯВЛЕННЯ АВТОМАТИЗОВАНИХ АТАК НА ВЕБРЕСУРСИ

Пташкін Р. Л. [1], Палагін В. В.[2]

[1]*Черкаський науково-дослідний експертно-криміналістичний центр МВС України,*
[2]*Черкаський державний технологічний університет, Черкаси, Україна*
ndekc.ck@gmail.com, palahin@ukr.net

У роботі розглядається досвід впровадження міжрівневої системи захисту вебсерверів на прикладі двох державних інформаційних ресурсів. Система побудована за принципами розподілу засобів безпеки за рівнями: мережевим, системним та прикладним. Основна увага приділена виявленню нетипової поведінки застосунків як ознаки кіберзагроз та блокуванню автоматизованих атак. Наведено результати впровадження, які засвідчили практичну ефективність підходу.

Ключові слова: Кіберзахист, вебресурси, php.

Вступ

Активне впровадження цифрових сервісів у державному секторі обумовлює необхідність удосконалення засобів їхнього захисту. Традиційні рішення часто не відповідають динаміці сучасних кіберзагроз, особливо у випадках використання нестандартних схем атаки. Одним із перспективних напрямів є побудова міжрівневих систем, які забезпечують виявлення аномальної активності на всіх етапах обробки запитів.

Метою роботи є опис архітектури, механізмів та результатів реального впровадження міжрівневої системи захисту вебсерверу, здатної виявляти ознаки атак, зокрема в автоматизованому виконанні, навіть за відсутності відомих сигнатур.

Постановка проблеми

Державні вебсервіси часто стають об’єктами як масових, так і цільових атак, спрямованих на виведення з ладу або компрометацію інформації. Існуючі підходи до безпеки базуються переважно на окремих компонентах: вебфільтрах, антивірусах, мережевих екранах. Їх обмеженість проявляється у неможливості виявити складні або нові типи загроз, які не мають чітких ознак. Зокрема, автоматизовані атаки, що імітують легітимну активність, залишаються непоміченими без залучення поведінкової аналітики.

Вирішення проблеми

Запропонована та впроваджена система базується на міжрівневому принципі організації безпеки [1], де кожен з рівнів (мережевий, системний, прикладний) забезпечує свій набір функцій виявлення та реагування.

На мережевому рівні реалізовано фільтрацію та обмеження активності на основі iptables та модулів аналізу трафіку. Системний рівень відповідає за моніторинг процесів, файлових змін та запуску служб. Найбільш інноваційним є прикладний рівень, який за допомогою заздалегідь визначених алгоритмів аналізує логи веб-серверу та визначає відхилення від моделей «нормального» функціонування.

Особливу увагу приділено автоматичному виявленню підозрілої активності, яка може свідчити про спробу злому або використання ботів. Було застосовано евристичні правила виявлення, засновані на частоті, послідовності та змісті запитів [2]. В умовах реального функціонування система змогла адаптивно формувати профілі підозрілих сесій і змінювати політики доступу, включно з перенаправленням на honeypot.

На рівні прикладної логіки, безпосередньо у вебдодатку, реалізовано додаткові інструменти моніторингу активності, інтегровані у код PHP [3]. Було впроваджено низку структурованих елементів, зокрема приховані тригери, мікровідлік часу між діями, контроль за черговістю подій і

контекстом навігації. Такі елементи дозволяють у реальному часі формувати поведінкові профілі користувачів і оперативно ідентифікувати автоматизовані запити. Отримані дані зберігаються для подальшого аналізу, що значно підвищує точність системи виявлення ботів та інших кіберзагроз на етапі взаємодії з інтерфейсом.

Блокування шкідливих запитів на рівні iptables та NGINX, що ініціюється через файлові мітки, сформовані PHP-логікою, виявилося надзвичайно ефективним як з позиції технічної реалізації, так і з точки зору оптимального використання ресурсів. Завдяки реалізації безпосередньо на рівні ядра операційної системи, iptables забезпечує практично миттєве застосування правил без необхідності запуску додаткових процесів або скриптів. Це дозволяє зменшити навантаження на процесор і оперативну пам'ять порівняно з аналогічними механізмами, що діють у користувацькому просторі. Тестування засвідчило, що навіть при великій кількості блокувань швидкодія системи залишається стабільною, а вплив на продуктивність веб-сервера є статистично незначущим.

Протягом перших чотирьох місяців експлуатації на декількох державних ресурсах зафіксовано понад декілька сотень проявів несанкціонованої активності, що ймовірно були результатом роботи автоматизованих систем пошуку вразливостей. Фактично реалізована система захисту автоматично блокувала шкідливу активність без участі адміністратора. Серед атак були виявлені як стандартні сканування вразливостей, так і малопомітні, але небезпечні методи зламу вебдодатків.

Висновки

Впровадження міжрівневої системи захисту вебсерверу дало змогу суттєво підвищити стійкість державних інформаційних систем до атак, особливо автоматизованого характеру. Система виявилася здатною фіксувати та реагувати на нетипову активність, навіть без наявності відомих сигнатур. Отримані результати свідчать про доцільність масштабування підходу на інші державні

платформи та подальше удосконалення системи в напрямку глибшого поведінкового аналізу.

Список використаних джерел

1. Пташкін Р.Л., Палагін В.В., (2024). Міжрівнева система захисту web-серверу. Theoretical and Applied Cybersecurity. Матеріали другої всеукраїнської науково-практичної конференції. (с. 94–97) – Київ: Інжиніринг. http://www.is.ipt.kpi.ua/pdf/T24.pdf;

2. N.Shekokar, H.Vasudevan, S.Durbha, A.Michalas, T.Nagarhalli. Intelligent Approaches to Cyber Security. India: CRC Press, 2024. ISBN: 978-1-032-52161-9;

3. M.Zandstra. PHP 8 Objects, Patterns, and Practice: Mastering OO Enhancements, Design Patterns, and Essential Development Tools. UK: Apress, 2021.

ПОРІВНЯННЯ СЕРВІСІВ ДЛЯ ВИЯВЛЕННЯ ФІШИНГОВИХ ДОМЕНІВ

Стецик Р.[1], Столик Д.[1], Мудрицький І.[1],
Мойко О.[1], Тімошин А.[1]

[1]*Харківський національний університет внутрішніх справ,*
romanstetsyk06@gmail.com, deni.s201811111111@gmail.com,
migor2k05@gmail.com,
Joker-moyko@ukr.net, timxxvii@gmail.com

У роботі аналізуються сучасні сервіси для виявлення фішингових доменів, такі як VirusTotal, Sucuri, Google Safe Browsing API, PhishTank та OpenPhish. Оцінюються їхні можливості та ефективність у боротьбі з кіберзагрозами.

Ключові слова: фішинг, виявлення фішингових доменів, кіберзагрози, VirusTotal, Sucuri, Google Safe Browsing API, PhishTank, OpenPhish

Загроза фішингу та необхідність виявлення фішингових доменів

Фішинг є однією з найбільших загроз, що полягає в шахрайських спробах отримати конфіденційну інформацію, зокрема паролі, номери кредитних карток чи інші особисті дані, через підроблені вебсайти або електронні листи [1]. Такий вид атак може завдати значної шкоди як окремим користувачам, так і організаціям. З огляду на зростаючу загрозу фішингу, виявлення фішингових доменів стало важливим кроком у боротьбі з кіберзлочинністю. Для цього були розроблені спеціалізовані сервіси, що значно підвищують ефективність боротьби з фішинговими атаками, адже дають змогу оперативно виявляти підозрілі сайти та мінімізувати їхній вплив. Використання технологій збору і аналізу даних у реальному часі сприяє значному зниженню ризиків для кінцевих користувачів та організацій. За допомогою пошукової системи Google були знайдені наступні сервіси для виявлення фішингових

доменів: VirusTotal, Sucuri, Google Safe Browsing API, PhishTank, OpenPhish. Ці платформи використовують різні підходи до збору та аналізу даних, включаючи автоматизовані системи сканування, краудсорсинг та інтеграцію з іншими засобами безпеки. Вони дозволяють оперативно виявляти фішингові домени, що допомагає знижувати ризики для користувачів і організацій, а також активно протидіяти фішинговим атакам на глобальному рівні.

Огляд сучасних сервісів для виявлення фішингових доменів

VirusTotal [2] здійснює перевірку об’єктів за допомогою більше ніж 70 антивірусних сканерів та сервісів блокування URL/доменів, а також низки інструментів для збору сигналів із досліджуваного контенту.[3]. На даний момент сервіс користується великою популярністю, станом на серпень 2024 року, VirusTotal щоденно аналізує в середньому 1,2 мільйона унікальних нових файлів, які раніше не були представлені на платформі [4]. VirusTotal використовує кілька потужних інструментів для виявлення фішингових доменів. Система URL/Domain Scanner перевіряє домени за допомогою різних антивірусних та безпекових движків, що дозволяє ефективно виявляти фішингові ресурси. Інтерактивна платформа VirusTotal Graph дозволяє візуалізувати зв'язки між доменами, IP-адресами, файлами та сертифікатами SSL, що дає можливість досліджувати інфраструктуру фішингових кампаній. Крім того, зберігання історії DNS-записів у Passive DNS дає змогу виявляти фішингові домени, що використовують однакові IP-адреси або хостинги. Також підтримка кастомних правил через Livehunt/YARA Rules дозволяє ефективно виявляти фішингові шаблони, такі як підроблені сторінки входу до популярних сервісів, наприклад, PayPal, Google чи Microsoft.

Sucuri – це комплексна хмарна платформа для забезпечення безпеки вебсайтів, яка спеціалізується на виявленні та запобіганні різним видам кіберзагроз, включаючи фішинг [5]. Платформа забезпечує постійний

моніторинг сайтів, аналізуючи вебсторінки на наявність шкідливих скриптів, редиректів, підозрілих форм та інших фішингових елементів. Sucuri пропонує комплексний набір інструментів, спрямованих на виявлення фішингових загроз і забезпечення безпеки вебресурсів. Одним з основних компонентів є SiteCheck Scanner – безкоштовний онлайн-інструмент, який здійснює зовнішній аналіз вебсайтів на наявність шкідливого коду. Він перевіряє вихідний HTML-код, виявляє приховані iFrame, обфускований JavaScript та інші ознаки потенційної компрометації, включаючи фішингові сторінки. Інструмент Website Blacklist Monitoring дозволяє визначити, чи потрапив сайт до чорних списків основних пошукових систем або антивірусних платформ, таких як Google Safe Browsing, PhishTank та інші, що часто сигналізує про наявність фішингового контенту. Окрім клієнтського сканування, Sucuri надає Server-Side Scanner, який виконує поглиблену перевірку на серверному рівні, аналізуючи файли, бази даних та конфігураційні параметри на предмет фішингових елементів, шкідливого ПЗ, бекдорів тощо. Для розробників доступний Scanning API, що дозволяє інтегрувати функціонал сканування до власних систем або додатків, автоматизуючи виявлення фішингових посилань і отримання звітів у реальному часі. Завдяки поєднанню клієнтського, серверного аналізу та API-інтеграції, Sucuri є ефективним рішенням для проактивного виявлення фішингових загроз.

Google Safe Browsing API [6] – це сервіс від компанії Google, призначений для виявлення небезпечних вебресурсів, зокрема фішингових сайтів, сторінок із шкідливим або небажаним програмним забезпеченням. Цей API дозволяє інтегрувати функцію перевірки URL-адрес у клієнтські додатки та сервіси, підвищуючи рівень безпеки користувачів в онлайн-середовищі. Основними компонентами є Lookup API та Update API. Lookup API (v4) дозволяє надсилати URL-адреси на сервери Google для перевірки в реальному часі без необхідності зберігати локальні списки небезпечних сайтів. Цей варіант простий у реалізації, однак має певні обмеження щодо

продуктивності та конфіденційності, оскільки кожен запит надсилається безпосередньо. Натомість Update API (v4) дозволяє клієнтським додаткам завантажувати зашифровані списки небезпечних ресурсів і виконувати перевірку локально. Це підвищує ефективність і забезпечує кращий захист приватності, оскільки URL-адреси хешуються перед обробкою. Обидва інтерфейси активно використовуються в безпекових рішеннях для проактивного виявлення фішингових загроз.

PhishTank [7] – це платформа, яка дозволяє користувачам надсилати підозрілі URL-адреси, що, на їхню думку, є фішинговими. Ці URL перевіряються як автоматизованими системами, так і спільнотою користувачів, що сприяє оперативному виявленню шкідливих ресурсів. Після підтвердження фішингової природи, домен додається до загальнодоступної бази, яку можуть використовувати антивірусні програми, пошукові системи та інші системи безпеки для блокування таких загроз. PhishTank надає низку інструментів для ефективного виявлення фішингових посилань. Основою є спільнота користувачів, які надсилають підозрілі URL-адреси, а також голосують за їхню фішингову природу, сприяючи колективному аналізу загроз. Сервіс також пропонує відкрите API, що дозволяє розробникам інтегрувати перевірку URL-адрес у свої системи безпеки. API підтримує HTTP POST-запити, завдяки чому можна оперативно перевіряти статус підозрілих ресурсів у базі PhishTank. Крім того, доступні завантажувані бази даних з фішинговими доменами у різних форматах, які оновлюються щогодини, забезпечуючи постійний доступ до актуальної інформації для масових перевірок.

OpenPhish [8] автоматично збирає фішингові URL-адреси з різних джерел, аналізує їх за допомогою власних алгоритмів, а також класифікує тип загрози. Сервіс надає доступ до безкоштовного відкритого фіду з оновленням кожні кілька годин, а також комерційний API із розширеним набором даних і метаінформацією, що дозволяє інтегрувати його у системи безпеки для автоматизованого виявлення фішингових атак.

OpenPhish використовує автоматизовані системи виявлення, які на основі внутрішньої структури знань автономно визначають ймовірність того, що URL є фішинговим, що дозволяє забезпечити оперативне реагування без участі людини. Сервіс підтримує постійно оновлювану базу даних фішингових сайтів, яка включає такі параметри, як хостнейми, шляхи, мови сторінок, SSL-метадані, IP-адреси, ASN, країни походження та бренди, що імітуються. Крім того, OpenPhish пропонує спеціалізовані фіди фішингової розвідки з даними про URL-адреси, цільові бренди, географію атак і пов'язані індикатори, що допомагає дослідникам і системам безпеки ефективно відстежувати та аналізувати фішингові загрози.

Висновки

Отже, проаналізовані сервіси – VirusTotal, Sucuri, Google Safe Browsing API, PhishTank, OpenPhish, – демонструють високу ефективність у виявленні фішингових доменів завдяки різним підходам до аналізу, серед яких: краудсорсинг, автоматизоване сканування, API-інтеграція та багаторівневий моніторинг. Усі платформи мають свої переваги й можуть бути корисними залежно від конкретних потреб. Наприклад, PhishTank вирізняється активною участю спільноти, Google Safe Browsing API — простотою інтеграції, а Sucuri — глибоким аналізом вебресурсів. Проте серед усіх рішень найбільш універсальним і функціональним можна вважати VirusTotal, оскільки він поєднує потужні антивірусні сканери, інтерактивну візуалізацію, історичні дані DNS та можливість використання кастомних правил, що робить його незамінним інструментом у сучасній боротьбі з фішингом.

Список використаних джерел

1. Що таке фішинг? Захисний комплекс Microsoft. URL: https://www.microsoft.com/uk-ua/security/business/security-101/what-is-phishing (дата звернення 06.04.2024).

2. VirusTotal. URL: https://www.virustotal.com/gui/home/url (дата звернення 06.04.2024).

3. How it works. URL: https://docs.virustotal.com/docs/how-it-works (дата звернення 06.04.2024).

4. Scaling Up Malware Analysis with Gemini 1.5 Flash. Google Cloud Blog. URL: https://cloud.google.com/blog/topics/threat-intelligence/scaling-up-malware-analysis-with-gemini (дата звернення 06.04.2024).

5. Sucuri – Complete Website Security, Protection; Monitoring. Sucuri. URL: https://sucuri.net (дата звернення 06.04.2024).

6. Overview;| Safe Browsing APIs (v4);| Google for Developers. URL: https://developers.google.com/safe-browsing/v4 (дата звернення 06.04.2024).

7. PhishTank. Join the fight against phishing. URL: https://phishtank.org/ (дата звернення 06.04.2024).

8. OpenPhish – Phishing Intelligence. URL: https://openphish.com/ (дата звернення 06.04.2024).

ВПЛИВ ШТУЧНОГО ІНТЕЛЕКТУ НА АКАДЕМІЧНУ ДОБРОЧЕСНІСТЬ: ЗАГРОЗА ЧИ ІНСТРУМЕНТ РОЗВИТКУ?

Світличний В.[1], Артюх Д.[1]

[1]*Харківській національний університет внутрішніх справ*
e-mail: vit.svet@ukr.net

У роботі розглядається вплив штучного інтелекту на систему вищої освіти, зокрема в контексті академічної доброчесності. Вказується як на переваги використання ШІ у навчальному процесі – індивідуалізація навчання, підтримка критичного мислення, доступність освіти, – так і потенційні загрози, пов'язані з порушенням принципів самостійності та чесності. Окрема увага приділена необхідності формування нормативної бази для регламентованого використання ШІ в академічному середовищі, розвитку цифрової етики та ролі ШІ як інструменту контролю за дотриманням доброчесності. Робиться висновок, що ключовим є не сам факт застосування технологій, а етичні засади їх інтеграції у процес навчання.

Ключові слова: штучний інтелект, академічна доброчесність, освіта, цифрова етика, дистанційне навчання, плагіат, критичне мислення, ChatGPT, регламентування ШІ, цифрова грамотність

Вступ

У сучасному суспільстві стрімкий розвиток цифрових технологій, зокрема штучного інтелекту (ШІ), змінює підходи до навчання, викладання та наукової діяльності. Застосування ШІ в освіті викликає як захоплення можливостями оптимізації навчального процесу, так і занепокоєння щодо дотримання принципів академічної доброчесності. Особливої актуальності це питання набуває

в умовах дистанційного навчання, коли контроль за самостійністю виконання завдань ускладняється.

Виклад основного матеріалу

Аналіз останніх публікацій показує, що з одного боку штучний інтелект виступає як потужний інструмент підтримки навчання: адаптивні платформи, такі як ChatGPT, Grammarly, Coursera AI-асистенти, можуть підвищувати мотивацію студентів, надавати миттєвий зворотний зв'язок та індивідуалізувати навчальний процес.

Такі системи допомагають розвивати критичне мислення, сприяють кращому розумінню складних тем, а також полегшують доступ до навчальних матеріалів для осіб з особливими потребами або тих, хто проживає у віддалених регіонах [1].

З іншого боку, використання ШІ для створення або редагування академічних текстів без належного цитування чи розуміння змісту створює серйозну загрозу академічній доброчесності. Автоматичне генерування есе, відповідей на екзаменаційні запитання або курсових робіт без участі студента у мисленнєвому процесі суперечить основним принципам доброчесності – самостійності, чесності та відповідальності.

Відповідно, виникає дилема: де закінчується допустима допомога, і починається академічний плагіат? Ситуацію ускладнює відсутність у багатьох закладах вищої освіти чітких нормативних документів, які б регламентували допустиме використання ШІ в навчальному процесі. Деякі університети світу вже запровадили політики, що визначають межі застосування штучного інтелекту в академічних роботах. Вони передбачають обов'язкове зазначення використання допоміжних інструментів, надання вихідних даних, які аналізувалися, та пояснення, як саме була використана допомога системи [2].

Необхідно заохочувати розвиток цифрової грамотності та етичної відповідальності студентів. Адже не кожен, хто використовує ШІ, робить це з метою обману. Часто це є прагненням краще зрозуміти матеріал або компенсувати брак часу.

У таких випадках потрібно навчати користуватися штучним інтелектом як інструментом доповнення знань, а не заміщення власного інтелектуального зусилля.

Разом з тим, сам ШІ може виступати у ролі охоронця академічної доброчесності. Застосування алгоритмів для виявлення плагіату, аналізу стилістики тексту, фальсифікації джерел та фабрикацій даних здатне значно підвищити якість перевірки академічних робіт. Розробники програмного забезпечення вже впроваджують інструменти, здатні виявляти ознаки використання генеративного ШІ у текстах [3].

Також важливо відзначити, що академічна доброчесність – це не лише питання технологій, а й культури. Формування академічної етики починається з прикладу викладача, чесної оцінки знань, відкритого діалогу про виклики сучасної освіти. Навчальні заклади мають створювати простір для етичних дискусій та навчання критичного мислення.

Зрештою, завдання вищої освіти – не лише передати знання, а й сформувати відповідального громадянина, здатного до самостійного мислення і морального вибору. ШІ може бути цінним союзником у цьому процесі, за умови розумного, усвідомленого і регламентованого використання.

Висновки

Таким чином, штучний інтелект є водночас викликом і можливістю для системи освіти. У центрі уваги має бути не сам ШІ, а принципи, за якими його інтегрують у навчальний процес. Тільки поєднання технологічного розвитку з етичною відповідальністю дозволить зберегти довіру до академічного середовища та забезпечити якісну підготовку фахівців у цифрову епоху.

Список використаних джерел

1. EDUCAUSE. (2023). *Artificial Intelligence (AI)*. Retrieved from URL: https://library.educause.edu/

topics/infrastructure-and-research-technologies/artificial-intelligence-ai (дата звернення: 07.05.2025).

2. European Network for Academic Integrity (ENAI). (2023). *ENAI Recommendations on the ethical use of Artificial Intelligence in Education*. Retrieved from URL: https://edintegrity.biomedcentral.com/articles/10.1007/s40979-023-00133-4 (дата звернення: 07.05.2025).

3. OpenAI. (2024). *Usage policies*. Retrieved from URL: https://openai.com/policies/usage-policies/ (дата звернення: 07.05.2025).

ІНТЕРНЕТ-ПІРАТСТВО ЯК ПРОБЛЕМА СУЧАСНОСТІ

Світличний В.[1], Курило Д.[1]

[1]*Харківській національний університет внутрішніх справ*

В тезах доповіді розглянута проблема інтернет-піратства. Показано: правомірність захисту авторських прав, економічні загрози, правові питання інтернет-піратства. Розглянуто основні види та особливості інтернет- піратства.

Ключові слова: інтернет-піратство, захист даних, інформація, загрози, надійність, безпека, контент, цифрова ера.

Вступ

З розвитком цифрових технологій і широким розповсюдженням Інтернету, людство зіштовхнулося з новими викликами та проблемами, однією з яких є інтернет- піратство. Це явище стало глобальною проблемою, що впливає на економіку, культурну сферу і правовий порядок**.**

Виклад основного матеріалу

Інтернет-піратство включає в себе незаконне копіювання, розповсюдження і використання інтелектуальної власності, такої як фільми, музика, книги та програмне забезпечення. Нелегальне скачування та обмін контентом без дозволу правовласників ставлять під загрозу розвиток творчих індустрій та ставлять під сумнів правомірність захисту авторських прав в цифрову еру [1].

Проблема інтернет піратства має кілька вимірів. По-перше, це серйозний економічний удар по творчих індустріях. Згідно з дослідженнями, світова економіка втрачає мільярди доларів щорічно через піратство. Наприклад, у сфері кіноіндустрії, тільки у США втрати оцінюються в кілька мільярдів доларів щорічно. Це означає скорочення робочих місць, зменшення інвестицій в нові

проєкти та загальне уповільнення розвитку галузі [2]. По-друге, інтернет-піратство ставить під загрозу культурний розвиток суспільства. Творці контенту, такі як музиканти, письменники та програмісти, втрачають мотивацію та ресурси для створення нових творів, оскільки їхні праці незаконно розповсюджуються та використовуються без належної компенсації. Це може призвести до зниження якості та кількості нових культурних продуктів.

По-третє, інтернет піратство порушує правовий порядок. Авторські права є однією з основних складових правової системи, що захищають творчість та інновації. Нелегальне розповсюдження контенту підриває ці принципи, що може призвести до загального знецінення інтелектуальної власності. Це створює прецедент для інших видів нелегальної діяльності та підриває довіру до правової системи в цілому.

Незважаючи на зусилля урядів і міжнародних організацій щодо боротьби з піратством, проблема залишається актуальною через анонімність та глобальний характер Інтернету, що ускладнює виявлення та покарання порушників [3].

Висновки

Інтернет-піратство є складною та багатогранною проблемою сучасності, яка вимагає комплексного підходу для її вирішення. Економічні втрати, культурний занепад та підрив правового порядку є лише деякими з наслідків цього явища. Для ефективної боротьби з піратством необхідні скоординовані зусилля урядів, міжнародних організацій, технологічних компаній та суспільства в цілому. Потрібно розробляти нові законодавчі акти, покращувати технології захисту контенту, а також підвищувати обізнаність користувачів щодо важливості дотримання авторських прав. Лише таким чином можна зберегти культурне багатство, стимулювати розвиток творчих індустрій та забезпечити справедливу компенсацію для творців контенту в цифрову еру.

Список використаних джерел

1. Піратство в інтернеті: проблема XXI століття – Infomir. Infomir. URL: https://www.infomir.com.ua/ru/news/piratstvo-v-internete-problema-xxi-veka/ (дата звернення: 19.04.2025).

2. Учасники проєктів Вікімедіа. Порушення авторського права – Вікіпедія. Вікіпедія. URL: https://uk.wikipedia.org/wiki/Порушення_авторського_права (дата звернення: 19.05.2025).

3. Що таке ІТ-піратство і як не стати «піратом» – ЗКГ. Юридичні послуги для бізнесу, податковий консалтинг, бухгалтерський аутсорсинг, навчання бухгалтерів – від холдингу професійних послуг ЗКГ. URL: https://zkg.ua/shcho-take-it-piratstvo-i-iak-ne-staty-piratom/ (дата звернення: 19.04.2025).

ОСОБЛИВОСТІ ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ ДЕРЖАВИ В УМОВАХ ЗБРОЙНОЇ АГРЕСІЇ

Світличний В. [1], Матвійчук А.[1]

[1]*Харківський національний університет внутрішніх справ, e-mail: vit.svet@ukr.net*

Вступ

В умовах сучасної війни інформаційна безпека набуває такого ж критичного значення, як оборона територій чи забезпечення матеріально-технічної спроможності армії. Інформаційна сфера стала самостійним полем бою, де вирішуються питання морального духу населення, міжнародної підтримки, ефективності управління та спроможності держави протидіяти ворожим впливам. У випадку України, яка з 2014 року перебуває в стані гібридної війни, а з 2022 року – у стані повномасштабної збройної агресії з боку Російської Федерації, питання забезпечення інформаційної безпеки постало з особливою гостротою.

Викладення основного матеріалу

Сутність інформаційної безпеки держави в умовах війни полягає в захисті критичної інформаційної інфраструктури, протидії дезінформації, збереженні стійкої комунікації між органами влади, військовими та суспільством, а також у захисті інформаційних прав громадян. В умовах воєнного стану інформаційний простір зазнає потужного тиску – як з боку ворожих інформаційно-психологічних операцій (ІПсО), так і в результаті кібератак на державні ресурси, мережі енергетики, транспорту, банківської системи.

Однією з ключових особливостей забезпечення інформаційної безпеки в умовах збройної агресії є необхідність швидкого реагування на зміну обстановки. Інформаційні загрози часто є динамічними, комплексними

та непередбачуваними. Це вимагає від державних органів створення гнучких систем моніторингу, виявлення загроз, оперативного аналізу інформаційного середовища та координації дій між усіма безпековими структурами.

Надзвичайно важливу роль відіграє інформаційна протидія ворогу. Росія системно використовує пропаганду, фейки, маніпуляції, підроблені документи та змонтовані відео для дестабілізації ситуації, сіяння паніки та розколу в українському суспільстві. У відповідь Україна реалізує заходи із забезпечення прозорості, оперативної публічної комунікації, перевірки фактів (fact-checking) та централізації офіційної інформації. Створення та робота таких ресурсів, як «Єдиний портал правди», Офіс стратегічних комунікацій, кіберполіції та спецпідрозділів СБУ, дозволяє значно ефективніше протидіяти інформаційним атакам.

Ще однією важливою особливістю є підвищена вразливість критичної інфраструктури до кіберзагроз. Ураження серверів органів влади, систем зв'язку, реєстрів, банківської інфраструктури – усе це може призвести до паралічу управління державою. Тому основними завданнями в цій сфері є забезпечення безперервності роботи інформаційних систем, впровадження засобів шифрування та резервного копіювання, розподіл інфраструктури на незалежні вузли, а також – створення надійних механізмів обміну інформацією між військовими, урядом та міжнародними партнерами.

Окремим напрямом є розбудова цифрової стійкості громадян. В умовах масових дезінформаційних кампаній важливо, щоб населення було здатне критично мислити, перевіряти джерела інформації, розпізнавати маніпуляції. Звідси випливає потреба в національній стратегії медіаграмотності, регулярних просвітницьких кампаніях, співпраці з журналістами та лідерами громадської думки.

Інформаційна безпека в умовах збройної агресії – це не лише технічне завдання. Це стратегічна складова національної безпеки, яка впливає на хід війни не менше, ніж успішні дії на фронті. Інформація може бути зброєю, а може бути щитом. Тому завдання держави – навчитися

ефективно нею володіти, захищати й управляти в інтересах національного суверенітету, громадянського миру та перемоги.

Важливою складовою забезпечення інформаційної безпеки держави є налагодження ефективної міжвідомчої взаємодії. В умовах воєнного стану особливої актуальності набуває координація дій між Службою безпеки України, Держспецзв'язку, Міністерством оборони, Міністерством цифрової трансформації, Національним центром оперативно-технічного управління мережами телекомунікацій та іншими структурами. Їх спільні зусилля спрямовані не лише на відбиття зовнішніх атак, але й на формування проактивної інформаційної політики, яка зміцнює єдність суспільства і довіру до державних інституцій [1].

Варто також зазначити, що війна створила унікальні умови для розвитку кіберволонтерського руху. Тисячі IT-спеціалістів об'єдналися в спільноти, які виконують завдання зі зламу ворожих ресурсів, виявлення агентурних мереж, захисту українських сайтів та баз даних, а також інформаційної протидії пропаганді на міжнародному рівні. Такий феномен самоорганізації є новим елементом інформаційної безпеки, що доповнює класичні державні підходи.

У межах міжнародної співпраці важливе значення має обмін даними та координація з партнерами з НАТО, ЄС, країн-союзників, а також із глобальними цифровими гігантами, такими як Google, Meta, Amazon Web Services. Ці партнери допомагають відбивати масштабні DDoS-атаки, відновлювати цифрову інфраструктуру, блокувати шкідливий контент і дезінформацію. Крім того, міжнародні платформи сприяють глобальному поширенню правдивої інформації про війну в Україні, що має велике значення для зміцнення підтримки з боку світової спільноти [2].

Однак, попри успіхи, залишаються проблемні питання, які потребують системного вирішення. Серед них – фрагментарність нормативно-правового забезпечення у сфері інформаційної безпеки, недостатній рівень цифрової безпеки в регіональних органах влади, слабкий рівень

медіаграмотності окремих груп населення, а також нестача висококваліфікованих фахівців у сфері кіберзахисту. Вирішення цих проблем потребує як інституційного реформування, так і довгострокових інвестицій в освіту, дослідження та цифрову інфраструктуру.

Висновки

У підсумку, можна стверджувати, що забезпечення інформаційної безпеки в умовах збройної агресії – це необхідна умова збереження державності, стабільності та перемоги у війні. Це багатовимірний процес, що включає технічні, правові, організаційні, психологічні та освітні компоненти. Український досвід опору інформаційним і кіберзагрозам в умовах війни вже сьогодні формує нові підходи до національної безпеки, які матимуть значення не лише для України, а й для всього демократичного світу.

Список використаних джерел

1. Нові правила: інформаційна безпека під час війни. Одеська національна наукова бібліотека. Офіційний веб-сайт. URL: https://odnb.odessa.ua/view_post.php? id=4286 (дата звернення: 11.05.2025).

2. Як протидіяти кіберзагрозам та захистити системи від ворожих кібератак – важливі рекомендації та допомога CERT-UA. Урядовий портал Єдиний веб-портал органів виконавчої влади України. URL: https://www.kmu.gov.ua/news/yak-protydiiaty-kiberzahrozam-ta-zakhystyty-systemy-vid-vorozhykh-kiberatak-vazhlyvi-rekomendatsii-ta-dopomoha-cert-ua (дата звернення: 11.05.2025).

КОНЦЕПТУАЛЬНО-СТРАТЕГІЧНЕ ЗНАЧЕННЯ НАЦІОНАЛЬНОЇ СТРАТЕГІЇ КІБЕРБЕЗПЕКИ УКРАЇНИ В АРХІТЕКТОНІЦІ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ ДЕРЖАВИ

Ящук В. І.[1]

[1]*Львівський державний університет безпеки життєдіяльності, Львів, Україна,*
valentina.lender@gmail.com

Проаналізовано роль та стратегічне значення Стратегії кібербезпеки України як ключового інструменту державної політики у сфері захисту національного кіберпростору. Розглянуто її нормативно-правову основу, основні цілі, інституційний механізм реалізації та виклики, з якими стикається система інформаційної безпеки в умовах гібридних загроз. Особлива увага приділена функціональній інтеграції Стратегії в архітектоніку національної безпекової системи.

Ключові слова: кібербезпека, Стратегія кібербезпеки України, інформаційна безпека, національна безпека, державна політика, кіберзагрози.

Вступ

У сучасних умовах цифрової трансформації, гібридних загроз та інтенсивного зростання числа кібератак, проблема забезпечення кібербезпеки набуває першочергового значення для системи національної безпеки України. Особливої актуальності це питання набуває в контексті збройної агресії проти України, коли інформаційні ресурси, критична інфраструктура та цифрові сервіси держави перебувають під постійним тиском з боку ворожих кібероперацій. У зв'язку з цим Стратегія кібербезпеки України виступає ключовим документом, який визначає

пріоритети, напрями, принципи та організаційні механізми державної політики у сфері кіберзахисту.

Виклад основного матеріалу

Правовою основою для реалізації національної кіберполітики є Стратегія кібербезпеки України, затверджена Указом Президента України від 26 серпня 2021 року № 447/2021. Документ розроблений відповідно до Конституції України, Закону України «Про національну безпеку», Закону України «Про основні засади забезпечення кібербезпеки України», а також із урахуванням міжнародних стандартів та практик у сфері кібербезпеки, зокрема рекомендацій НАТО, Європейського Союзу, стандартів ISO/IEC 27001, NIST Cybersecurity Framework [1-4].

Основною метою Стратегії є забезпечення стабільного функціонування національного кіберпростору, а також захист державних інтересів, прав і свобод громадян, комерційних структур та інфраструктури від кіберзагроз. Стратегічними завданнями є розвиток національної системи кібербезпеки, удосконалення механізмів захисту критичної інформаційної інфраструктури, зміцнення спроможностей кіберрозвідки, протидія інформаційно-психологічним впливам, формування культури кібербезпеки серед населення, а також розширення міжнародної співпраці у цій сфері.

Інституційна модель реалізації Стратегії передбачає участь низки державних органів, зокрема Ради національної безпеки і оборони України як координатора, Державної служби спеціального зв'язку та захисту інформації України, Служби безпеки України, Міністерства оборони, Міністерства цифрової трансформації та Національного координаційного центру кібербезпеки при РНБО.

Стратегя кібербезпеки України [1] – документ довгострокового планування, що визначає загрози кібербезпеці України, пріоритети та напрями забезпечення кібербезпеки України з метою створення умов для безпечного функціонування кіберпростору, його

використання в інтересах особи, суспільства і держави [3].

Докладніше структуру вказаного документа наведено на рис. 1.

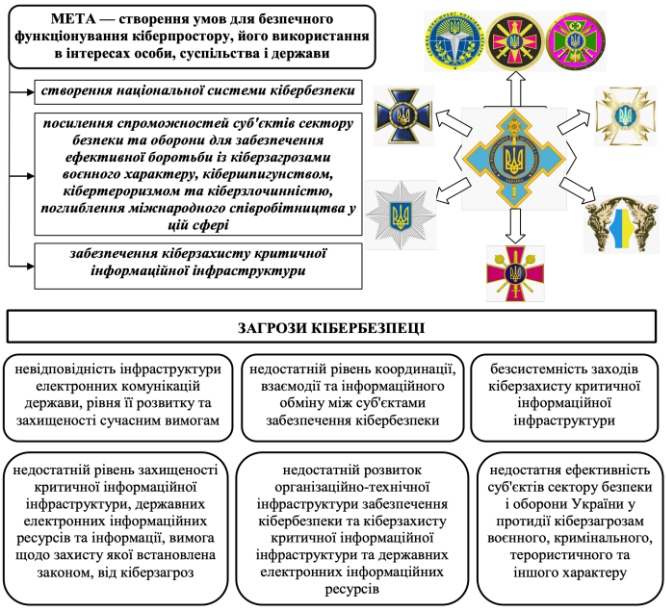


Рисунок 1. Основні елементи Стратегії кібербезпеки України, *розроблено автором на основі* [3]

До пріоритетних напрямів реалізації Стратегії належать: створення національної системи реагування на кіберінциденти; розвиток механізмів виявлення, аналізу та прогнозування кіберзагроз; посилення кіберзахисту об’єктів критичної інфраструктури; запровадження індикаторів кіберзагроз; забезпечення оперативного обміну інформацією між суб’єктами кібербезпеки; розвиток людського капіталу та просвітницька діяльність у сфері цифрової безпеки.

Водночас реалізація стратегії супроводжується низкою викликів, серед яких: систематичні кібератаки з боку держав-агресорів; технологічна та ресурсна нерівність між

державними та недержавними структурами; недостатній рівень кібергігієни; фрагментарність нормативно-правової бази; а також потреба в адаптації міжнародних стандартів до національного контексту.

Слід зазначити, що Стратегія кібербезпеки України виконує ключову функцію в процесі формування цілісної та ефективної системи національної інформаційної безпеки. Вона є базовим документом, що закладає фундамент для розвитку комплексної, міжвідомчо узгодженої системи кіберзахисту, яка інтегрується у загальноєвропейський безпековий простір.

Висновки

Таким чином, Стратегія кібербезпеки України не лише визначає стратегічні орієнтири у сфері інформаційної безпеки, але й виступає дієвим інструментом у протидії сучасним кіберзагрозам та побудові цифрової держави, здатної забезпечити захист прав, інтересів та безпеки своїх громадян у кіберпросторі.

Список використаних джерел

1. Стратегія кібербезпеки України. [Електронний ресурс]. Режим доступу: https://zakon.rada.gov.ua/laws/show/447/2021#Text.

2. Про національну безпеку України: Закон України. [Електронний ресурс]. Режим доступу: https://zakon.rada.gov.ua/laws/show/2469-19#Text.

3. Правові засади кібергігієни [Електронний ресурс]. Режим доступу: https://https://drive.google.com/file/d/1EPd677wdTA32Fbi5RzoobUk2CvcolQAW/view

4. Ящук В.І. Роль та місце стратегії кібербезпеки України у забезпеченні інформаційної безпеки держави / В. І. Ящук // Moderní aspekty vědy: XLII. Díl mezinárodní kolektivní monografie / Mezinárodní Ekonomický Institut s.r.o.. Česká republika: Mezinárodní Ekonomický Institut s.r.o., 2024. P. 259-286.

ТЕХНОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ В КОНТЕКСТІ ЗАБЕЗПЕЧЕННЯ КІБЕРБЕЗПЕКИ

Зайцев О. В.[1], Стефанцев С. С.[1], Попов М. О.[2]

[1] *Воєнна академія імені Євгенія Березняка, м. Київ, Україна*
a.zaysev@gmail.com, stefancevss@gmail.com
[2] *Державна установа "Науковий центр аерокосмічних досліджень Землі Інституту геологічних наук Національної академії наук України", м. Київ, Україна*
tpopov@casre.kiev.ua

Сьогоднішній стрімкий розвиток цифрового середовища, особливо такої технології, як штучний інтелект, кардинально змінює економічний, соціальний і інформаційний ландшафти. Штучний інтелект (ШІ) стає ключовим інструментом кіберпростору, виконує низку критично важливих функцій, наприклад: автоматизований захист (алгоритми машинного навчання здатні виявляти і нейтралізувати кібератаки), пошук закономірностей (інтелектуальні моделі аналізують вхідні дані й самостійно приймають рішення), опрацювання Big Data (ШІ здатний скоротити час на обробку структурованої та неструктурованої інформації великих розмірів), аналіз звітів безпеки (ШІ автоматично структурує та оцінює звіти, підвищує точність висновків), виявлення вторгнень (системи ШІ можуть фіксувати аномальні дії у реальному часі), моніторинг трафіку (ШІ спроможний безперервно відстежувати мережеві потоки, виокремлюючи підозрілу активність), зменшення хибних спрацювань (системи ШІ фільтрують помилкові попередження, прогнозують потенційні атаки, адаптуються до нових загроз), прогнозування загроз (моделі прогнозної аналітики формують випереджувальні сценарії оборони). Тобто, інтеграція ШІ в системи кібербезпеки трансформує підходи до захисту даних, забезпечуючи більш оперативну протидію сучасним загрозам.

Ключові слова: штучний інтелект, кіберзахист, кібербезпека.

Вступ

Нормативна основа використання ШІ технологій в кібербезпеці стало ключовою умовою довіри до цифрових технологій, наприклад: закони про кібербезпеку та персональні дані встановлюють вимоги до прозорості алгоритмів і відповідальності розробників; міжнародні акти уніфікують принципи етичного та безпечного застосування ШІ; галузеві кодекси деталізують тестування та моніторинг моделей; регуляторні органи отримують повноваження щодо аудиту і санкцій; міждержавні угоди про обмін інформацією та спільне реагування забезпечують оперативну протидію загрозам, що виходять за межі окремих юрисдикцій.

Мета дослідження. Метою є дослідження основних аспектів правового регулювання ШІ в Україні та у країнах Європейського Союзу.

Аналіз нормативно-правових актів

Основними аспектами правового регулювання ШІ в Україні та у країнах Європейського Союзу (ЄС) є:

1. Законодавство з кібербезпеки. До ключових законодавчих актів належать:

– директива про безпеку мереж і інформаційних систем (Директива NIS). Директива NIS зобов'язує держави-члени ЄС забезпечити високий рівень кібербезпеки у критично важливих секторах. Згідно з директивою, компанії в цих секторах повинні вжити заходів для захисту своїх інформаційних систем від кібератак і повідомляти про серйозні інциденти безпеки [1];

– директива NIS2 (оновлена директива NIS). Директива NIS2 розширює застосування директиви NIS та посилює вимоги до кіберзахисту. Директива покладає на країни-члени зобов'язання щодо посилення національних стратегій кібербезпеки, впровадження більш чітких заходів реагування на інциденти та налаштування механізмів

взаємодії між державами для обміну інформацією про кіберзагрози [2];

– регламент про кіберзахист (Cybersecurity Act). Європейська агенція з кібербезпеки (ENISA) як основний орган для підтримки і координації зусиль у сфері кібербезпеки в ЄС встановлює загальні стандарти сертифікації продуктів і послуг у сфері кібербезпеки для підвищення довіри до цифрових технологій та захисту користувачів від потенційних загроз [3];

– Закон України "Про основні засади забезпечення кібербезпеки України", який встановлює правові та організаційні засади забезпечення кіберзахисту держави [4]. Україна підписала Угоду про співпрацю з Європейським агентством з кібербезпеки (ENISA), яка включає аспекти використання ШІ [5].

2. Захист персональних даних. Використання ШІ для забезпечення кібербезпеки має відповідати законам про захист даних, наприклад:

– в ЄС це GDPR (General Data Protection Regulation) – загальний регламент захисту даних Європейського Союзу. Його основна мета – захист прав та свобод осіб, які надають свої персональні дані, а також створення єдиного правового простору для обробки таких даних у всіх країнах-членах ЄС [6];

– в Україні – Закон України "Про захист персональних даних" [7].

3. Конвенція про кіберзлочинність (Конвенція Будапешта). Ця конвенція створює правові рамки для міжнародної співпраці в сфері кіберзахисту, що включає використання ШІ [8].

4. Рекомендації Організації Об'єднаних Націй (ООН) та Ради Європи:

– резолюція Генеральної Асамблеї ООН "Штучний інтелект і розвиток". Резолюція закликає країни до міжнародного співробітництва та розвитку етичних стандартів для ШІ, забезпечуючи баланс між інноваціями і захистом прав людини [9];

– ініціатива "AI for Good" (Штучний інтелект для добра). ООН запустила програму "AI for Good", яка

зосереджена на використанні ШІ для досягнення цілей сталого розвитку [10];

– універсальні рекомендації для ШІ. ООН закликає до створення міжнародної угоди або конвенції для регулювання ШІ, яка включатиме принципи захисту людської гідності, права на конфіденційність та рівність [11];

– етичні рекомендації Ради Європи щодо ШІ [12].

5. Етичні принципи ШІ. Забезпечення прозорості, недискримінації, підзвітності та відповідності застосування ШІ загальноприйнятим стандартам, наприклад, ISO/IEC 27001 для інформаційної безпеки [13] та ISO/IEC 23894 для етичного використання ШІ [14].

Висновки

Технології ШІ відіграють ключову роль у забезпеченні кібербезпеки, оскільки вони дозволяють ефективно виявляти, аналізувати та реагувати на загрози в реальному часі. Використання ШІ для забезпечення кібербезпеки стає необхідним у сучасному цифровому середовищі через зростаючу кількість кібератак та їхню складність.

Список використаних джерел

1. Директива Європейського Парламенту і Ради (ЄС) 2016/1148 від 6 липня 2016 року про заходи для високого спільного рівня безпеки мережевих та інформаційних систем на території Союзу : [Міжнародний документ] // Відомості Верховної Ради України. – 2016. – № 2016/1148. – Режим доступу: https://zakon.rada.gov.ua/laws/show/984_013-16#Text (дата звернення: 20.05.2025).

2. What Is the NIS2 Directive? Compliance and Policies Explained [Електронний ресурс]. – Режим доступу: https://hideez.com/en-int/blogs/news/what-is-nis2 (дата звернення: 20.05.2025).

3. The EU Cybersecurity Act / European Commission [Електронний ресурс]. – Режим доступу: https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-act (дата звернення: 20.05.2025).

4. Про основні засади забезпечення кібербезпеки України : Закон України від 05.10.2017 р. № 2163-VIII : за станом на 28.06.2024 р. // Відомості Верховної Ради України. – 2017. – № 45. – Режим доступу: https://zakon.rada.gov.ua/laws/show/2163-19#Text (дата звернення: 20.05.2025).

5. Разом – на захисті кіберпростору: Держспецзв'язку, НКЦК та ENISA підписали угоду про співпрацю / Державна служба спеціального зв'язку та захисту інформації України [Електронний ресурс]. – Режим доступу: https://cip.gov.ua/ua/news/razom-na-zakhisti-kiberprostoru-derzhspeczv-yazku-nkck-ta-enisa-pidpisali-ugodu-pro-spivpracyu (дата звернення: 20.05.2025).

6. Загальний регламент про захист даних (GDPR) [Електронний ресурс]. – Режим доступу: https://gdpr-text.com/uk/ (дата звернення: 20.05.2025).

7. Про захист персональних даних : Закон України від 01.06.2010 р. № 2297-VI : за станом на 27.04.2024 р. // Відомості Верховної Ради України. – 2010. – № 38. – Режим доступу: https://zakon.rada.gov.ua/laws/ show/2297-17#Text (дата звернення: 20.05.2025).

8. Конвенція про кіберзлочинність : Рада Європи ; Конвенція, Міжнародний документ від 23.11.2001 р. // Відомості Верховної Ради України. – 2005. – № 37. – Режим доступу: https://zakon.rada.gov.ua/laws/show/994_575#Text (дата звернення: 20.05.2025).

9. Генасамблея ООН вперше ухвалила резолюцію про штучний інтелект / Укрінформ [Електронний ресурс]. – Режим доступу: https://www.ukrinform.ua/ rubric-technology/3843161-genasamblea-oon-vperse-uhvalila-rezoluciu-pro-stucnij-intelekt.html (дата звернення: 20.05.2025).

10. Документ про регулювання ШІ від ООН: нагальна ідея із суперечливими формулюваннями [Електронний ресурс]. – Режим доступу: https://netfreedom.org.ua/ article/globalne-upravlinnya-internetom-vid-oon-nagalna-ideya-iz-superechlivimi-formulyuvannyami (дата звернення: 20.05.2025).

11. ООН створила раду з міжнародного регулювання ШІ / Економічна правда [Електронний ресурс]. – Режим доступу: https://epravda.com.ua/news/2023/10/31/ 706051/ (дата звернення: 20.05.2025).

12. Рада Європи відкрила до підписання першу у світі конвенцію про штучний інтелект / Укрінформ [Електронний ресурс]. – Режим доступу: https://www.ukrinform.ua/rubric-world/3902460-rada-evropi-vidkrila-do-pidpisanna-persu-u-sviti-konvenciu-pro-stucnij-intelekt.html (дата звернення: 20.05.2025).

13. ДСТУ ISO/IEC 27001:2015. Інформаційні технології. Методи захисту системи управління інформаційною безпекою. Вимоги (ISO/IEC 27001:2013; Cor 1:2014, IDT) [Електронний ресурс]. – Режим доступу: https://online.budstandart.com/ua/catalog/doc-page.html?id_doc=66910 (дата звернення: 20.05.2025).

14. ISO/IEC 23894:2023. Information technology – Artificial intelligence – Guidance on risk management [Електронний ресурс]. – Режим доступу: https://www.iso.org/ru/standard/77304.html (дата звернення: 20.05.2025).

ШТУЧНИЙ ІНТЕЛЕКТ У ПРОТИДІЇ КІБЕРЗЛОЧИННОСТІ

Воронянський Є.[1], Лучик В.[1],

[1]*Харківський національний університет внутрішніх справ*

Штучний інтелект грає важливу роль у протидії кіберзлочинності. Використання штучного інтелекту для боротьби з кіберзлочинністю стає все більш актуальним, оскільки кіберзлочинці стають все більш винахідливими і складними у своїх атаках.

Ось деякі способи, якими штучний інтелект може бути корисним у цьому контексті [1, с.179]:

1. Виявлення загроз. Штучний інтелект може використовуватися для аналізу великих обсягів даних та виявлення потенційних загроз і аномалій в мережі. Він може реагувати на незвичайні активності, які можуть бути зв'язані з кіберзлочинцями.

2. Прогнозування атак. Штучний інтелект може аналізувати попередні атаки і поведінки кіберзлочинців, щоб передбачити майбутні атаки і підготуватися до них.

3. Автоматизована реакція. Штучний інтелект може бути налаштований на автоматичну реакцію на певні типи атак, включаючи блокування доступу до системи, ізолювання компрометованих ресурсів та збільшення рівня безпеки.

4. Моніторинг безпеки. Штучний інтелект може відстежувати безпеку мережі та даних в режимі реального часу, сприяючи вчасному виявленню порушень безпеки.

5. Аналіз логів і великих обсягів даних. Штучний інтелект може використовувати машинне навчання для аналізу логів та даних, щоб виявити аномалії та потенційні загрози, які можуть бути незрозумілі для людей [2].

Оскільки світ стає все більш цифровим, загроза кібератак зростає.

Розвиток технологій дозволяє зловмисникам ставати більш вдосконаленими у своїх методах, включаючи

використання штучного інтелекту для автоматизованих складних атак, які важко виявити і захиститися від них.

Кібератаки на основі штучного інтелекту можуть завдати значно більше шкоди, ніж традиційні кібератаки. Штучний інтелект використовується для автоматизації таких завдань, як пошук вразливостей, здобуття інформації та запуск цілеспрямованих атак. Це значно підвищує швидкість та масштаб атак, дозволяючи хакерам одночасно нападати на кілька цілей. Штучний інтелект також може створювати великі обсяги даних, які використовуються для обходу систем безпеки та проникнення в мережі.

Застосування кібератак на основі штучний інтелект набуває популярності. Останнім часом атаки програм-вимагачів за допомогою штучного інтелекту стали частішими та більш руйнівними. Ці атаки використовують штучний інтелект для швидкого виявлення вразливих систем і розповсюдження шкідливого програмного забезпечення, яке може швидко поширюватися в мережі. Зловмисне програмне забезпечення на основі штучного інтелекту також використовується для незаконного проникнення в мережі, крадіння даних та порушення операцій[3].

Зростаюча загроза кібератак на основі штучного інтелекту змусила уряди та організації приймати додаткові заходи для забезпечення своєї безпеки. Багато організацій впроваджують рішення кібербезпеки на основі штучного інтелекту для виявлення передових загроз і захисту своїх мереж. Крім цього, організації збільшують свої бюджети на кібербезпеку та інвестують у навчання та навчання, щоб допомогти співробітникам розпізнавати потенційні загрози та реагувати на них.

Кібератаки на основі штучного інтелекту становлять значну загрозу для організацій і урядів, і хоча це може здаватися складним завданням, існують заходи, які можна прийняти, щоб зменшити ризик. Інвестування в передові рішення кібербезпеки, навчання співробітників розпізнавати потенційні загрози та збільшення бюджетів на кібербезпеку допоможе організаціям захистити себе від

зростаючої загрози кібератак на основі штучного інтелекту [4].

Роль машинного навчання в протидії кіберзлочинності стає надзвичайно важливою в умовах зростаючої складності цифрових загроз для безпеки. Машинне навчання представляє собою галузь штучного інтелекту, яка дозволяє комп'ютерам самостійно набувати знання на основі аналізу даних та виявлення закономірностей. В контексті боротьби з кіберзлочинністю цю технологію можна використовувати для виявлення зловмисної активності, розпізнавання аномалій та запобігання кібератакам.

Штучний інтелект відіграє ключову роль у протидії зростаючій загрозі кіберзлочинності. З використанням штучного інтелекту ми можемо виявляти та передбачати кібератаки, реагувати на них автоматично та моніторити безпеку в реальному часі. Штучний інтелект дозволяє ефективно виявляти аномалії та потенційні загрози, що є важливим у контексті росту складності кібератак.

Зростаюча загроза кіберзлочинності, особливо з використанням штучного інтелекту, заставляє організації і уряди приймати серйозні заходи для забезпечення своєї кібербезпеки. Інвестування в розвиток та впровадження рішень кібербезпеки на основі штучного інтелекту, а також навчання персоналу стають важливими факторами в захисті від кіберзлочинності.

Загалом, використання штучного інтелекту у сфері кібербезпеки стає необхідністю, оскільки технологічний прогрес веде до зростання загроз у цифровому світі. Штучний інтелект допомагає виявляти, передбачати та запобігати кібератакам, забезпечуючи вищий рівень безпеки для організацій і інфраструктури в цілому.

Список використаних джерел

1. Романчук О. Штучний інтелект в епоху нових медій. *Вісник львівського університету. серія «журналістика»*. 2018. Вип. 44. С. 179–188.

2. Штучний інтелект і запобігання кіберзлочинності: як інтелектуальні системи допомагають виявляти та

запобігати кіберзагрозам. *TS2 SPACE*. URL: https://ts2.space/uk/штучний-інтелект-і-запобігання-кібер/.

3. Концепція розвитку штучного інтелекту в Україні [Електронний ресурс]: Розпорядження Кабінету міністрів України № 1556-р від 02.12.2020. URL: https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#Text.

4. Науково-дослідний інститут вивчення проблем злочинності імені академіка В.В. Сташиса. URL: https://ivpz.kh.ua/wp-content/uploads/2020/12/Матеріали-семінару_Використання-техн-штучного-інтел_5.11.2020.pdf.

5. Основні напрями застосування технологій штучного інтелекту у кібербезпеці. URL: https://dspace.univd.edu.ua/items/2329c5f4-e52e-43d5-8eaf-35758a04ca09.

ЦИФРОВІ ІНСТРУМЕНТИ СПРОТИВУ: УКРАЇНСЬКІ ІТ-РОЗРОБКИ

Ляшенко А.[1],

[1]*Харківській національний університет внутрішніх справ, e-mail: nlasenko299@gmail.com*

У статті розглядається феномен цифрового спротиву в Україні, який набув особливого значення в умовах повномасштабної війни. Проаналізовано ключові технологічні ініціативи, спрямовані на забезпечення оборони, захисту цивільного населення та координації гуманітарної допомоги. Особливу увагу приділено створенню та функціонуванню адаптивної платформи «Nexus UA» як нового типу децентралізованої цифрової екосистеми. Описано архітектуру платформи, використані технології, механізми обробки даних і перспективи застосування у кризових ситуаціях.

Ключові слова: Цифровий спротив, війна в Україні, штучний інтелект, хмарні технології, чат-боти, Nexus UA, кібероборона, гуманітарна допомога, Edge AI, OSINT.

Вступ

У XXI столітті війни точаться не лише на полях бою, а й у цифровому просторі. Інформаційні технології перетворилися на стратегічний ресурс, здатний впливати на перебіг подій, забезпечувати захист цивільного населення та координувати дії в умовах надзвичайних ситуацій. Повномасштабне вторгнення в Україну стало каталізатором появи нових ІТ-підходів, які поєднують мобільність, штучний інтелект, хмарні сервіси та активну участь громадян.

Цей досвід демонструє, що цифрові рішення можуть стати основою національного спротиву, оперативного реагування та стійкості в умовах кризи.

Виклад основного матеріалу

Повномасштабна війна в Україні спричинила безпрецедентний сплеск цифрових ініціатив, спрямованих на оборону, захист цивільного населення та координацію гуманітарної допомоги. Інформаційні технології стали не просто допоміжним інструментом, а ключовим елементом національного спротиву [1, 2]. У надзвичайно стислі терміни в Україні сформувалася потужна децентралізована IT-екосистема, що поєднує державні сервіси, волонтерські проєкти, мобільні застосунки, чат-боти, елементи штучного інтелекту та хмарні інфраструктури.

Серед найбільш відомих прикладів цифрового спротиву можна згадати застосунок «єВорог» [4], який дозволяє передавати інформацію про ворожі сили, мобільний застосунок «Air Alert» [5], що забезпечує своєчасне оповіщення про небезпеку, інтерактивну карту DeepStateMap.Live [6], яка надає актуальні зведення про фронт, а також діяльність кіберволонтерських спільнот, що ведуть оборонні та наступальні операції в кіберпросторі. Ці рішення створювались, зокрема, незалежними розробниками й волонтерськими об'єднаннями, що діяли в умовах обмежених ресурсів і високої відповідальності.

Разом з цим формується новий тип цифрової системи – адаптивна, самонавчальна, багатокомпонентна платформа, умовно названа «Nexus UA» [7]. Вона виконує роль централізованого «нервового вузла», який об'єднує вхідні дані з численних джерел: чат-ботів, застосунків, камер відеоспостереження, сенсорів, відкритих баз даних (OSINT) та повідомлень від користувачів. Ключова ідея – створити цифрову інфраструктуру, здатну в реальному часі отримувати, аналізувати, фільтрувати, пріоритизувати та маршрутизувати інформацію до відповідних органів чи структур.

З технічного боку реалізація такої системи ґрунтується на багаторівневій архітектурі:

• Бекенд розробляється з використанням Python у поєднанні з FastAPI – асинхронним фреймворком, який

дозволяє ефективно обробляти великі обсяги запитів з різноманітних джерел у реальному часі.

• Фронтенд побудований на основі React, що забезпечує інтуїтивний інтерфейс, адаптований до умов використання в екстремальних ситуаціях (офлайн-режим, мінімум кліків, великий текст, швидка взаємодія).

• Бази даних: комбінація PostgreSQL для реляційної інформації (звіти, події, логістика) та MongoDB для зберігання гнучких JSON-документів, повідомлень, медіа.

• Інфраструктура: розгортання в Google Cloud Platform із використанням Kubernetes-кластерів [11], що дозволяє гнучке масштабування відповідно до рівня загроз або кількості запитів.

• Аналітика реалізується через NLP-моделі на базі spaCy [10] (для базової обробки текстів) та Hugging Face Transformers [9] (моделі типу RoBERTa, XLM-RoBERTa, спеціально донавчені на українському воєнно-гуманітарному корпусі). Це забезпечує автоматичну класифікацію повідомлень (запит на допомогу, розвіддані, фейк, спам), виявлення ключових слів, координат, описів подій.

• Навчальні дані: використання відкритих дата-сетів (дані з Telegram, OSINT, урядових публікацій), а також побудова власного українського корпусу для точного контекстуального розпізнавання текстів у воєнних умовах.

• Edge AI: використання автономних ML-модулів, що працюють локально (на мобільному пристрої або ноутбуці) без підключення до інтернету. Це забезпечує можливість обробки запитів в умовах повної ізоляції (наприклад, у блокаді або під час блекауту).

• Геоаналітика: система використовує API Google Maps і OpenStreetMap [12] для створення динамічних карт загроз, управління евакуаційними маршрутами, візуалізації запитів у реальному часі.

Всі ці компоненти інтегруються в модульну структуру, яка дозволяє масштабувати систему як географічно (по областях), так і функціонально (від військових запитів до гуманітарних та медичних потреб). Така система може застосовуватись не лише в умовах війни, а й у кризових

ситуаціях – надзвичайних подіях, техногенних катастрофах, пандеміях.

За даними IT Ukraine Association [1], лише за перший рік повномасштабної війни було створено понад 200 нових IT-продуктів, значна частина з яких не мала аналогів у мирний час. Понад 70 % українських розробників брали участь у створенні рішень для армії, евакуації, кібероборони або гуманітарної підтримки. Одночасно обсяг використання хмарних сервісів в Україні зріс майже в 7 разів [3]. Унікальність українського досвіду полягає в тому, що цифровий спротив формувався переважно горизонтально – через самоорганізацію, відкритий код, кросплатформені рішення та взаємодію громадян із техноактивістами. У цьому контексті проєкти на кшталт «Nexus UA» [7, 8] відображають нову філософію IT – коли технології стають не просто сервісом, а зброєю, захистом і опорою державності в умовах глобальної кризи.

Висновок

Досвід України в умовах повномасштабної війни засвідчує, що цифрові технології можуть стати ефективним інструментом національного спротиву, захисту цивільного населення та координації гуманітарної допомоги. Розвиток таких платформ, як «Nexus UA», демонструє потенціал децентралізованих, адаптивних і масштабованих IT-рішень, що поєднують сучасні технології з високим рівнем громадської самоорганізації. У майбутньому ці напрацювання можуть стати основою для побудови цифрових систем реагування в інших країнах, які опиняться перед викликами воєнних або кризових ситуацій.

Список використаних джерел

1. IT Ukraine Association. Де IT на війні: дослідження про внесок IT-індустрії у війну. https://itukraine.org.ua/de-it-na-vijni-asotsiatsiya-it-ukraine-ta-mind-predstavili-unikalne-doslidzhennya-pro-vnesok-it-industriyi-u-borotbu-proti-rosijskoyi-agresiyi/

2. Міністерство цифрової трансформації України. Як розвивається IT-індустрія під час війни: результати IT Research Ukraine 2023. https://thedigital.gov. ua/news/yak-rozvivaetsya-it-industriya-pid-chas-viyni-rezultati-it-research-ukraine-2023

3. Економічна правда. Українське IT в часи війни. https://www.epravda.com.ua/cdn/cd1/2023/ukrainske_it_v_chasy_viiny/

4. Дія. єВорог – новий чат-бот для передачі інформації про ворога. https://www.facebook.com/diia.gov.ua/posts/1007888538030740

5. Stfalcon. Air Alert – застосунок для оповіщення про повітряну тривогу. https://play.google.com/store/apps/details?id=com.stfalcon.airalert

6. DeepStateMap.Live. Інтерактивна карта бойових дій в Україні. https://deepstatemap.live/

7. Brave1. Оборонна платформа для розвитку військових технологій. https://brave1.ua/

8. United24. Армія дронів – проєкт для посилення обороноздатності. https://u24.gov.ua/uk/news/army-of-drones

9. Hugging Face. Transformers: State-of-the-art Natural Language Processing for Pytorch and TensorFlow 2.0. https://huggingface.co/transformers/

10. spaCy. Industrial-Strength Natural Language Processing in Python. https://spacy.io/

11. Google Cloud Platform. Cloud computing services. https://cloud.google.com/

12. OpenStreetMap. Free editable map of the world. https://www.openstreetmap.org/

ШТУЧНИЙ ІНТЕЛЕКТ У ЗАХИСТІ МЕРЕЖ ВІД АТАК ТИПУ ZERO-DAY: ПОТЕНЦІАЛ ТА ВИКЛИКИ

Мирончук Я.[1], Світличний В.[1]

Харківський національний університет внутрішніх справ

Вступ

У сучасному світі кібербезпека стає одним з найважливіших аспектів захисту інформаційних систем від все більш складних та витончених загроз. Однією з найнебезпечніших форм кібератак є атаки типу zero-day (уразливість нульового дня), що використовують невідомі вразливості програмного забезпечення або апаратних систем. Ці атаки характеризуються тим, що до моменту їх використання зловмисниками розробники не мають змоги вчасно випустити патч або оновлення для виправлення уразливості, що робить системи беззахисними [1]. Традиційні засоби захисту, такі як антивірусні програми та фаєрволи, часто не можуть виявити такі загрози, оскільки вони орієнтовані на відомі сигнатури або патерни атак.

Основна частина

Етичні та юридичні аспекти застосування нейромереж у кібербезпеці стоять на передньому плані серед найбільш вагомих і комплексних проблем. Прозорість роботи алгоритмів, забезпечення захисту інформації, контроль над ризиками автоматизації та моральна відповідальність — ключові фактори, що вимагають пильної уваги з боку фахівців з розробки, кінцевих споживачів та органів регулювання. У майбутньому значення етичних принципів та правових рамок у сфері кіберзахисту тільки зростатиме, а розвиток штучного інтелекту диктуватиме необхідність у впровадженні нових стратегій управління технологіями та охорони прав людини.

Одним з ключових векторів прогресу стає застосування глибинних нейронних мереж (DNN, Deep Neural Networks) для побудови досконаліших та більш точних моделей виявлення загроз. Завдяки властивості опрацьовувати значні масиви інформації та самостійно здобувати знання, DNN мають потенціал ідентифікувати нові типи атак, зокрема атаки нульового дня (zero-day attacks), що є недосяжним для традиційних сигнатурних систем. Скажімо, DNN здатні здійснювати аналіз великих обсягів інформації в режимі реального часу, що є критично необхідним для сучасного кіберзахисту [2].

До того ж, 2024 рік відзначився збільшенням кількості вразливостей нульового дня (zero-day) у таких браузерах, як Chrome та Edge. Зловмисники активно використовували їх, вдаючись до дедалі складніших тактик. Chrome, зокрема, зазнав впливу кількох вкрай серйозних вразливостей, зокрема CVE-2024-7971. Ця вразливість була пов'язана з механізмом JavaScript V8. Це дозволяло хакерам дистанційно виконувати шкідливий код, отримуючи доступ до корпоративних систем та конфіденційних даних, ще до того, як були випущені виправлення. Наслідки були значними: компанії, що активно використовують веб-платформи, зазнали збоїв, витоків даних та дорогого відновлення. Це наголошує на критичності наявності надійних превентивних заходів захисту до того, як такі вразливості стануть об’єктом атак [3].

Платформи GenAI, такі як ChatGPT та Midjourney, перевернули робочий процес, але 2024 рік продемонстрував їх потенційну небезпеку при роботі з конфіденційною інформацією. Згідно з доповіддю CybSafe, майже 40 відсотків респондентів зізналися у розповсюдженні конфіденційних бізнес-даних через інструменти штучного інтелекту, часто не усвідомлюючи супутніх ризиків.

Під час масового витоку даних, пов’язаного з ChatGPT у жовтні 2024 року, понад 225 000 облікових записів були скомпрометовані внаслідок атаки шкідливого програмного забезпечення. Ці події служать тривожним сигналом щодо нагальної потреби в належних заходах безпеки при

інтеграції ШІ в бізнес-процеси. Для захисту від атак нульового дня застосовують:

1. Windows Defender Exploit Guard. Вбудований захисний механізм Windows 2010. Здатний виконувати роль першої лінії оборони, що протистоїть атакам нульового дня.

2. Антивірус наступного покоління (NGAV). Здійснює аналіз загроз та спостереження за поведінкою, використовуючи машинне навчання і специфічні прийоми для захисту від експлойтів. Традиційні антивірусні програми не надто ефективні проти загроз нульового дня, оскільки останні використовують відомі слабкі місця в програмному забезпеченні.

3. Своєчасне оновлення. Автоматизовані засоби не тільки допомагають організаціям виявляти системи, які потребують оновлення, але й сприяють оперативному отриманню патчів та їхньому швидкому впровадженню, перш ніж зловмисники зможуть здійснити атаку [4].

Висновок

Атаки типу zero-day є одними з найсерйозніших загроз у кібербезпеці, оскільки використовують невідомі вразливості, які не можуть бути виправлені вчасно. Традиційні методи захисту часто не ефективні, тому важливим є застосування штучного інтелекту та нейронних мереж для виявлення таких атак. Разом із технологічними інноваціями виникають етичні та юридичні питання, зокрема щодо прозорості алгоритмів та захисту конфіденційності. Так як зловмисники використовують дедалі складніші тактики, важливо впроваджувати надійні системи захисту, своєчасно оновлюючи програмне забезпечення. Вдосконалення технічних та етичних аспектів є необхідним для ефективного захисту від атак нульового дня.

Список використаних джерел

1. Атака нульового дня. ESET Online Help. URL: https://help.eset.com/glossary/uk-UA/zero_day.html.

2. Suman, Kashyap. The Influence of Artificial Intelligence on Cybersecurity. International Journal of Innovative Research in Computer and Communication Engineering, 2024. doi: 10.15680/ijircce.2024.1203503.

3. Небезпеки для браузерів через штучний інтелект і вразливості нульового дня. *B2B Cyber Security*. URL: https://b2b-cyber-security.de/uk/gefahren-fuer-browser-durch-ki-und-zero-day-schwachstellen/.

4. Що таке вразливість нульового дня: zero day - IT Education Blog. IT Education Center Blog. URL: https://itedu.center/ua/blog/articles/zero-day/?srsltid=AfmBOoo64t5c-MrPVirgVivM_zr_O02w9-hpYSDulMZezMLwI98z10MW.

БЕЗПЕКА ІОТ-ПРИСТРОЇВ У ВІЙСЬКОВИХ ТА ПРИФРОНТОВИХ УМОВАХ

Мудровський Р.Т.[1]

[1]*Харківський національний університет внутрішніх справ*

Інтернет речей (Internet of Things, IoT) стрімко трансформує підходи до управління інформаційними потоками, спостереження, зв'язку та логістики. У військових і прифронтових умовах IoT-пристрої відіграють ключову роль у забезпеченні ситуаційної обізнаності, контролі за об'єктами, автоматизації операцій та розвідці. Смарт-камери, сенсори руху, GPS-трекери, безпілотники, пристрої моніторингу стану військової техніки або персоналу – все це приклади IoT-рішень, що застосовуються у військовому середовищі. Однак із розширенням можливостей зростає й поверхня потенційної атаки: безпека таких пристроїв стає критично важливою, особливо в умовах активних бойових дій або ворожої розвідки.

У прифронтових зонах IoT-пристрої можуть бути об'єктами не лише кібератак, але й фізичного втручання. Їх вразливість полягає у невисокій обчислювальній потужності, нестачі систем шифрування, відсутності оновлень прошивки, відкритих портах зв'язку, а також слабкому автентифікаційному контролі. Усі ці фактори створюють можливості для перехоплення даних, втручання у канали керування, підміни інформації або повного виведення з ладу пристрою.

В умовах війни загрози набувають ще більш складного характеру. IoT-пристрій, розміщений у військовій колоні або на об'єкті, може стати джерелом втрати координат, витоку маршруту або місця розташування. Зловмисники можуть здійснити «перехоплення сигналу», підмінити GPS-дані, або навмисно спричинити хибне спрацювання системи оповіщення. Крім того, через IoT-інфраструктуру

можна здійснити вторгнення до загальної мережі або навіть провести розвідку обстановки на місцевості.

Забезпечення безпеки IoT у військових умовах передбачає комплексний підхід. Передусім – криптографічний захист переданих даних навіть у легких пристроях: використання полегшених алгоритмів шифрування, захищених протоколів зв'язку (TLS, DTLS), цифрових сертифікатів. Не менш важливим є контроль автентичності пристроїв – обмеження доступу лише для зареєстрованих вузлів через механізми PKI (інфраструктура відкритих ключів) або блокчейн-ідентифікацію.

У разі втрати контролю над пристроєм або його компрометації, має бути реалізована функція швидкого відключення (kill-switch), що дозволяє дистанційно знищити дані або повністю вивести пристрій з ладу. Пристрої, які використовуються в бойовій зоні, мають підтримувати режими автономного функціонування без постійного з'єднання з центральною мережею – це знижує ризики викриття через трафік.

Важливу роль відіграє мережева сегментація – відокремлення військової IoT-інфраструктури від загальнодоступного Інтернету або цивільних мереж. Використання окремих захищених каналів зв'язку (VPN, зашифровані mesh-мережі), мобільних закритих обчислювальних середовищ (edge computing) дозволяє зменшити залежність від публічних ресурсів і забезпечити функціонування навіть у разі втрати зовнішнього зв'язку.

Особливу увагу потрібно приділяти оновленню програмного забезпечення. Оскільки багато IoT-пристроїв не мають автоматичного механізму оновлення, це створює довготривалі вразливості. У військових умовах оновлення має бути не лише захищеним, але й контрольованим за походженням – використання лише перевірених репозиторіїв, цифрових підписів тощо.

Також важливо враховувати людський фактор. Особи, що працюють із IoT-пристроями на фронті або в тилу, мають бути поінформовані про можливі загрози, правила поводження з обладнанням, інструкції щодо виявлення компрометації та дотримання протоколів зв'язку.

У довгостроковій перспективі розвиток безпеки IoT у військових та прифронтових умовах передбачає створення національних стандартів, інтеграцію з платформами кіберзахисту, використання штучного інтелекту для виявлення підозрілої активності, а також створення військових IoT-екосистем із вбудованими механізмами самозахисту, ізоляції та швидкого відновлення.

Висновок

Таким чином, безпека IoT-пристроїв у зонах бойових дій – це не лише питання технічного захисту, а й стратегічне завдання, яке впливає на успішність операцій, збереження інформації та безпеку персоналу. Забезпечення надійного функціонування таких пристроїв – один із ключових викликів цифрової війни нового покоління.

РОЗРОБКА СИСТЕМИ ВИЯВЛЕННЯ І ПРОТИДІЇ RANSOMWARE АТАКАМ ДЛЯ ОС WINDOWS

Палагін В. В.[1], Зражевський О. О.[2]

[1]*Кафедра робототехнічних і телекомунікаційних систем та кібербезпеки*
[2]*Черкаський державний технологічний університет,*
v.palahin@chdtu.edu.ua,
o.o.zrazhevskyi.fetam24@chdtu.edu.ua

В роботі розглянуто метод активної протидії ransomware-атакам з використанням файлів-приманок, що дозволяють виявити шкідливу активність на ранньому етапі. Описано процес розробки системи, яка здійснює моніторинг доступу до спеціально створених файлів-приманок і в разі виявлення файлової операції, що впливає на цілісність файлу-приманки – зупиняє потенційно шкідливий процес. Для програмної реалізації передбачається використання моніторингу файлових подій Windows через драйвер рівня файлової системи (File System Minifilter) та події Windows (ETW). Порівняно ефективність обох підходів із точки зору швидкості реакції та впливу на продуктивність. Дослідження базується на аналізі поведінки зразків Ransomware для оптимізації швидкості виявлення атаки.

Ключові слова: ransomware, файл-приманка, ETW, FS minifilter, інформаційна безпека, проактивний захист.

Вступ

На сьогоднішній день, ransomware-атаки залишаються однією з найнебезпечніших кіберзагроз. Ці атаки спрямовані на шифрування файлів з метою подальшого вимагання викупу за відновлення доступу до інформації. Успішна ransomware-атака призводить до значних фінансових збитків, втрати важливої інформації, переривання бізнес-процесів та репутаційних втрат [1].

Зростання кількості атак та їх еволюція вимагають пошуку нових ефективних методів виявлення та зупинки шкідливої активності на ранніх етапах.

В цих умовах, комплексний захист від ransomware, окрім регулярного резервного копіювання та антивірусного програмного забезпечення, також має включати заходи з вчасного виявлення і зупинки атаки. Ransomware угрупування впроваджують механізми ускладнення аналізу шкідливого програмного забезпечення, що може привести до невчасної реакції антивірусу на загрозу [2]. Відновлення від резервних копій після атаки потребує певного часу, протягом якого бізнес-процеси не функціонують в нормальному режимі.

В роботі описано процес розробки системи моніторингу доступу до файлів-приманок для забезпечення вчасного виявлення ransomware атаки. Файли-приманки створені таким чином, щоб вони першими потрапляли в поле зору ransomware при скануванні файлової системи. Після виявлення атаки, система нейтралізує загрозу шляхом зупинки шкідливого процесу.

Методологія дослідження

Система протидії побудована на основі файлових приманок, які розміщуються у найбільш вразливих директоріях з урахуванням особливостей поведінки типових ransomware-програм.

Використано два підходи до моніторингу файлової активності:

- FS Minifilter – драйвер файлової системи, що дозволяє перехоплювати та блокувати доступ до файлів з мінімальними затримками;
- ETW (Event Tracing for Windows) – менш інвазивний спосіб, що не вимагає встановлення драйверів, але може мати більшу затримку в обробці подій.

У разі фіксації дій над файлом-приманкою (перейменування, зміна вмісту/атрибутів файлу, видалення), система негайно виявляє процес, який здійснив файлову операцію і призупиняє його виконання. Зупинка

(suspend), а не завершення процесу, обране з міркувань стабільності операційної системи, оскільки ransomware перед початком атаки може налаштувати атрибут критичності для свого процесу, що призводить до перезавантаження операційної системи при завершенні процесу [3]. Також, зупинка шкідливого процесу надає можливість зняття дампу пам'яті, що є корисним джерелом інформації при розслідуванні інциденту, оскільки може містити додаткову інформацію про індикатори компрометації або може містити ключ шифрування файлів. Таким чином, система не лише виявляє атаку на ранньому етапі, а й ефективно її нейтралізує з мінімальними втратами.

Для оптимізації швидкості виявлення атаки було проаналізовано такі аспекти поведінки ransomware:

1) які розширення найчастіше шифруються ransomware;

2) з яких директорій зазвичай починається шифрування;

3) чи дотримуються ransomware певного порядку в обробці файлів;

4) які саме файлові операції виникають під час шифрування.

Система розроблена мовою C для забезпечення максимальної швидкості виявлення і зупинки атаки.

Результати та обговорення

Запропонована система успішно виявляє і зупиняє ransomware атаки. Тестування розробленої системи здійснюється з використанням реального шкідливого програмного забезпечення. Порівняльне тестування двох підходів (ETW та File System Minifilter) дозволило зробити висновки щодо ефективності різних методів моніторингу файлової системи. Використання minifilter-драйвера забезпечує менший час реакції, хоча і вимагає складнішої реалізації і встановлення драйверу. Використання подій Windows в якості джерела подій файлової системи також забезпечує достатньо швидкий час реакції на атаку.

У тестових сценаріях система показує здатність зупиняти процес до того, як будуть зашифровані звичайні файли. Зібрані дампи пам'яті зупиненого шкідливого процесу можуть використовуватись під час розслідування інцидентів безпеки, надаючи цінну інформацію для розслідування інциденту або відновлення даних.

Висновки

Розроблена система протидії ransomware з використанням файлів-приманок демонструє високу ефективність у ранньому виявленні та зупинці атак. Завдяки використанню як ETW, так і FS Minifilter підходів, вона може бути адаптована до різних вимог і умов розгортання. Подальші дослідження можуть бути зосереджені на інтеграції з SIEM-системами для централізованого реагування. У підсумку, впровадження подібного механізму може значно підвищити рівень захисту від одного з найнебезпечніших типів сучасних кіберзагроз – ransomware.

Список використаних джерел

1. Abraham Ojo. Ransomware trends and mitigation strategies: A comprehensive review. Global Journal of Engineering and Technology Advances. 2025, 22(03), 009-016. https://doi.org/10.30574/gjeta.2025.22.3.0038.

2. Olaimat M.N., Aizaini Maarof M., Al-rimy B. A. S. Ransomware Anti-Analysis and Evasion Techniques: A Survey and Research Directions. *2021 3rd International Cyber Resilience Conference (CRC)*, Langkawi Island, Malaysia, 2021, pp. 1-6, doi:10.1109/CRC50527.2021.9392529.

3. Cyble – Deep Dive Analysis – Pandora Ransomware. Cyble. URL: https://cyble.com/blog/deep-dive-analysis-pandora-ransomware/

РОЗРОБКА ТА ВЕРИФІКАЦІЯ ЛОГІКО-ЙМОВІРНІСНОЇ МОДЕЛІ КІБЕРБЕЗПЕКИ ОБ'ЄКТА КРИТИЧНОЇ ІНФРАСТРУКТУРИ ЗА ДОПОМОГОЮ ВІРТУАЛЬНИХ ЕКСПЕРТІВ

Алексейчук Л. Б.[1], Литвинова Т. В. [1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна*

У роботі запропоновано логіко-ймовірнісну модель кібербезпеки об'єкта критичної інфраструктури енергетичного сектора. Модель описує розвиток небажаних подій, що виникають у системах промислового керування (ICS) електроенергетичної мережі при реалізації можливих загроз з кіберпростору. Запропоновано процедуру побудови структурної, логічної та ймовірнісної моделей, а також проведено аналіз чутливості до параметрів моделі, що дозволило виявити найвразливіші елементи системи кіберзахисту.

Ключові слова: критична інфраструктура, логіко-ймовірнісна модель, ICS, SCADA, кібербезпека

Вступ

Системи промислового керування (Industrial Control Systems – ICS), що забезпечують функціонування енергетичних об'єктів, є важливими елементами критичної інфраструктури держави. У сучасних умовах ці системи піддаються кібератакам, які можуть порушити безперебійне функціонування електромереж, особливо в умовах війни, коли кіберпростір стає частиною гібридної агресії.

З метою забезпечення стійкості таких об'єктів до кібератак, важливим напрямом є аналіз кіберризиків,

прогнозування сценаріїв атак та оцінка вразливостей автоматизованих систем керування.

Метою дослідження є розробка та дослідження логіко-ймовірнісної моделі кібербезпеки об'єкта критичної інфраструктури енергетичного сектора, яка враховує множинність шляхів реалізації загроз з кіберпростору.

Для досягнення поставленої мети було:

- сформовано структурну модель шляхів кібератак;
- розроблено логічну модель взаємодії загроз та систем захисту;
- побудовано ймовірнісну модель на основі аналізу чутливості до параметрів;
- виконано верифікацію моделі з використанням методології «Рою віртуальних експертів».

Проблема забезпечення кібербезпеки об'єктів критичної інфраструктури є однією з ключових у сучасних дослідженнях. Зростання числа кібератак на енергетичні системи, особливо в умовах гібридної війни, потребує розробки формальних моделей аналізу ризиків, які дозволяють прогнозувати шляхи атак, оцінювати вразливості та раціонально планувати заходи захисту.

У роботі [1] проаналізовано можливі загрози для таких об'єктів та доведено їхню стратегічну важливість для національної безпеки. Особливу увагу приділено недолікам існуючих систем керування (ICS) – SCADA, DCS, PLC, які все частіше стають мішенями кібератак через широке впровадження IP-протоколів та Web-технологій. У статті [2] розглянуто задачі моделювання загроз, аналізу кібербезпеки та врахування зворотного зв'язку між об'єктами. Автори запропонували підходи до побудови семантичних мереж, які дають змогу формально описати взаємодію між загрозами, уразливостями та заходами захисту. У роботі [3] запропоновано моделі аналізу кібербезпеки об'єктів енергетики, зокрема методологію «рою віртуальних експертів» (SVE), яка передбачає використання великих мовних моделей (LLM) для отримання багатофакторних оцінок параметрів кіберризиків. Ця методологія знайшла подальше розвинення в аналізі зв'язків між компонентами

інфраструктури, а також при прогнозуванні ймовірних сценаріїв атак. У цій роботі описано логіко-ймовірнісний метод оцінювання безпеки структурно складних систем, який дозволяє формалізувати процес аналізу кіберризиків, враховуючи множинність факторів та їхній вплив на систему. Такий підхід став основою для розробки моделі, представленої в цій роботі. Таким чином, аналіз наукових джерел показує, що тема кібербезпеки критичної інфраструктури актуальна та має значний науковий і практичний потенціал, для аналізу кіберризиків успішно використовуються логіко-ймовірнісні моделі, методологія «рою віртуальних експертів» (SVE) забезпечує більшу повноту аналізу, зменшує суб'єктивність експертних оцінок та підвищує точність прогнозування кіберризиків.

Структура електроенергетичної мережі та її система керування

Типова електрична мережа складається з трьох рівнів:

- генерація та зберігання – електростанції, джерела відновлюваної енергії, трансформаторні підстанції.
- передача – лінії електропередачі.
- розподіл – споживачі, трансформаторні підстанції нижчого рівня.

Система керування такими мережами базується на Industrial Control Systems (ICS), у тому числі:

- SCADA – система контролю та збору даних;
- DCS – розподілена система керування;
- PLC – програмовані логічні контролери.

На жаль, ICS мають серйозні недоліки в плані інформаційної безпеки, які викликані поширенням IP-протоколів, використанням Web-технологій, а також застарілим рівнем захисту багатьох DCS і PLC систем.

Основні шляхи кібератак на ICS

Згідно з дослідженнями Управління з питань державної безпеки США (US GAO) [4], існують три основні способи кібератак на ICS:

– через корпоративну мережу – через Internet-доступ до бізнес-систем;
– через модемне підключення – класичний спосіб, що все ще використовується;
– через безпровідний зв'язок – Wi-Fi, радіоканали тощо.

Ці шляхи атак представлені на структурній схемі (рис. 2 документа). Кожна атака може призвести до компрометації системи керування на різних рівнях: корпоративному (верхньому), DCS (середньому) та PLC (нижньому).

Формування логіко-ймовірнісної моделі

Етапи побудови моделі включають побудову:
– структурної моделі – графічного представлення шляхів поширення загроз.
– логічної моделі – формулювання умов, за яких може відбутися компрометація системи.
– ймовірнісної моделі – шляхом введення чисельних значень ймовірностей реалізації кожного етапу атаки.

Параметри моделі

У моделі введено шість ключових параметрів P_1, P_2,..., P_6, які характеризують ймовірність реалізації первинних факторів загроз, наведених у табл. 1.

Таблиця 1. Параметри моделі

Параметр	Опис
P_1	Атака через корпоративну мережу
P_2	Атака через модемне підключення
P_3	Атака через Wi-Fi / безпровідний зв'язок
P_4	Подолання системи захисту верхнього рівня (корпоративна мережа)
P_5	Подолання системи захисту середнього рівня (DCS)
P_6	Подолання системи захисту нижнього рівня (PLC)

Ці шість параметрів не відповідають шести окремим шляхам атаки, а описують ймовірність подолання різних рівнів системи при трьох основних шляхах атаки: через корпоративну мережу, модем та Wi-Fi.

За допомогою логічної моделі отримано формулу для обчислення ймовірності небажаної події P_c:

$$P_c = [P_2(1-P_1)(1-P_3)+P_3(1-P_1)(1-P_2)+$$

$$+P_1(1-P_2)(1-P_3)(1-P_4)(1-P_5)](1-P_6),$$

де P_j – ймовірність реалізації j-го первинного фактора (загрози).

Вирази виду $1 - P_j$ відображають ймовірність того, що відповідний фактор не реалізується. Ця формула описує ймовірність компрометації нижнього рівня системи керування (PLC) через один із трьох основних шляхів атаки.

Верифікація моделі за допомогою віртуальних експертів

Для перевірки адекватності моделі використано методологію «рою віртуальних експертів», яка полягає у використанні великих мовних моделей як віртуальних експертів для отримання чисельних оцінок параметрів P_j. Кожен «віртуальний експерт» виступає у різних ролях:

- аналітик – аналіз загроз;
- критик – перевірка логічної послідовності;
- модератор – узагальнення результатів;
- оптимізатор – пропозиції щодо покращення моделі.

Отримані значення P_j підставлялися в формулу для P_c, що дозволило вирахувати загальну ймовірність настання небажаної події.

Оскільки модель є дискретною, для аналізу чутливості замість математичних похідних виконується відношення кінцевих приростів:

$$S_j = \frac{P_c\left(P_j + \Delta P_j\right) - P_c\left(P_j\right)}{\Delta P_j},$$

де S_j – коефіцієнт чутливості до параметра P_j; ΔP_j – мале збурення (наприклад, $\Delta P_j = 0{,}01$); $P_c\left(P_j\right)$ – значення ймовірності P_c, обчислене при поточному P_j; $P_c\left(P_j + \Delta P_j\right)-$ значення P_c, обчислене при збільшенні P_j на ΔP_j, решта параметрів залишаються сталими.

Це дозволило визначити, як зміна кожного параметра P_j впливає на загальну ймовірність небажаної події P_c. Найвища чутливість спостерігалася до параметрів P_5 (DCS) та P_4 (верхній рівень корпоративної мережі).

Після обчислень за заданими значеннями P_j отримано, що ймовірність реалізації небажаної події становить:

$$P_c = 0{,}027.$$

Найвища чутливість моделі спостерігалася до параметрів P_5 (DCS) та P_4 (корпоративна мережа вищого рівня), що вказує на необхідність підвищення захисту на цих рівнях.

Висновки

У результаті дослідження було розроблено та досліджено логіко-ймовірнісну модель кібербезпеки об’єкта критичної інфраструктури енергетичного сектора, яка описує розвиток небажаних подій у системах промислового керування (ICS) при реалізації загроз з кіберпростору. Модель ґрунтується на послідовному формуванні структурної, логічної та ймовірнісної моделей, що забезпечує формальну основу для аналізу кіберризиків.

Наукова новизна дослідження полягає в:

– розробці комплексної логіко-ймовірнісної моделі, яка враховує множинність шляхів атаки через корпоративну мережу, модемне підключення та безпровідний зв’язок;

– формалізації процесу компрометації системи керування на різних рівнях: верхньому (корпоративна мережа), середньому (DCS) та нижньому (PLC);

– застосуванні методології «рою віртуальних експертів» (SVE) для отримання об'єктивних оцінок параметрів моделі, що забезпечило багатофакторність аналізу, зменшення суб'єктивності та підвищення точності прогнозування кіберризиків;

– аналізі чутливості до параметрів моделі, що дозволило виявити найвразливіші елементи системи кіберзахисту, а також ранжувати фактори за ступенем впливу на загальну ймовірність небажаної події.

Практична цінність отриманих результатів полягає в тому, що запропонована модель може бути використана:

– для автоматизованого проєктування систем кіберзахисту;

– для прогнозування шляхів кібератак та аналізу їхньої ймовірності;

– для ранжування сценаріїв атак за ризиком, часом реалізації та витратами на реалізацію;

– для оцінювання ефективності заходів кіберзахисту через аналіз чутливості до ключових параметрів.

Отримані формули ймовірностей небажаних подій можуть бути використані для кількісної оцінки рівня кібербезпеки, а також для прогнозування ймовірних шляхів атак з урахуванням уразливостей окремих рівнів ICS.

Результати дослідження мають прикладне значення для підвищення стійкості енергетичних об'єктів до кібератак, особливо в умовах гібридної війни, коли кіберпростір є однією з головних сфер агресії.

Список використаних джерел

1. Alekseichuk, L., Novikov, O., Rodionov, A., Yakobchuk, D. Cyber Security Logical and Probabilistic Model of a Critical Infrastructure Facility in the Electric Energy Industry // Theoretical and Applied Cyber Security. Vol. 5 No. 1 (2023). DOI: 10.20535/tacs.2664-29132023.1.287365

2. Lande D., Novikov O., Alekseichuk L. Application of Large Language Models for Assessing Parameters and Possible Scenarios of Cyberattacks on Information and Communication Systems // Theoretical and Applied Cyber Security. Vol. 6 No. 1 (2024). DOI: 10.20535/tacs.2664-29132024.1.315242

3. Lande D., Svoboda I., Alekseichuk L., Strashnoy L. Methodology of a Swarm of Virtual Experts for Evaluating the Weight of Connections in Networks // Theoretical and Applied Cyber Security. Vol. 7 No. 2 (2024). DOI: 10.20535/tacs.2664-29132024.2.319946

4. Cybersecurity: Actions Needed to Strengthen U.S. Capabilities. GAO-17-440T. Published: Feb 14, 2017. URL: https://www.gao.gov/products/gao-17-440t

ТЕОРЕТИЧНІ ЗАСАДИ ТА КРИПТОГРАФІЧНІ МЕХАНІЗМИ ЗАБЕЗПЕЧЕННЯ КІБЕРБЕЗПЕКИ

Шмалюк В.[1]

[1]*Харківській національний університет внутрішніх справ*

У сучасному цифровому світі, де обсяг переданої інформації зростає експоненційно, а її цінність – неоціненна, криптографія відіграє ключову роль у забезпеченні кібербезпеки. Інформаційні системи сьогодні – це не лише засіб обміну даними, а й простір для стратегічних операцій, конфіденційного спілкування, фінансових транзакцій і функціонування критичної інфраструктури. У таких умовах зростає потреба в ефективних механізмах захисту інформації від несанкціонованого доступу, підробки, перехоплення та викривлення. Саме криптографія забезпечує фундаментальні властивості безпеки – конфіденційність, цілісність, автентичність і незаперечність.

Потреба у криптографічних методах виникла з перших кроків розвитку цивілізації, однак у XXI столітті вона набула якісно нового значення. Якщо раніше шифрування використовувалося переважно у військовій сфері, то тепер його застосування стало повсюдним – від збереження медичних даних до захисту електронного голосування чи блокчейн-технологій. Криптографія є невід'ємною складовою протоколів безпечної комунікації в інтернеті, таких як HTTPS, VPN, IPsec, а також використовується в цифрових підписах, двофакторній аутентифікації, електронному документообігу та цифрових валютах.

Історія криптографії охоплює тисячоліття людського розвитку і є свідченням постійного прагнення захистити інформацію від стороннього доступу. Перші відомі приклади шифрування з'явилися ще в Давньому Єгипті, де жерці використовували ієрогліфи з прихованим значенням для передачі сакральної інформації. У Стародавній Греції було відомо про використання «скітали» – шифрувального

циліндра для перестановки букв, що забезпечував певний рівень секретності під час військових кампаній. Найвідомішим прикладом античного шифру є «шифр Цезаря», який полягав у зсуві букв алфавіту на фіксовану кількість позицій. Хоча такі методи здаються примітивними з сучасної точки зору, у свій час вони були ефективними інструментами приховування інформації.

З розвитком наук і технологій криптографія еволюціонувала. У середньовіччі застосовувалися складніші техніки, як, наприклад, шифр Віженера, який використовував поліалфавітний принцип і вважався незламним протягом століть. Проте з появою математичних методів криптоаналізу, зокрема в період арабського Ренесансу, деякі шифри було розкрито, що змусило криптографів шукати нові підходи. У XX столітті криптографія отримала стрімкий розвиток, особливо під час світових війн. Саме в цей період з'являються механічні шифрувальні пристрої, серед яких найвідомішою є «Енігма», що використовувалася німецькими військами у Другій світовій війні. Розшифрування «Енігми» союзниками, зокрема завдяки роботі Алана Тюрінга, стало важливим кроком не лише у розвитку криптоаналізу, а й у формуванні перших комп'ютерів.

З початком комп'ютерної епохи криптографія вступила у фазу стрімкої математизації. У 1970-х роках виникла сучасна симетрична та асиметрична криптографія. Широке поширення отримали алгоритми DES (Data Encryption Standard), а пізніше – AES (Advanced Encryption Standard), що й досі активно використовуються для шифрування даних. Проривом стало відкриття алгоритмів з відкритим ключем – зокрема RSA, який дозволяє безпечно обмінюватися даними навіть між незнайомими сторонами, не передаючи ключів у відкритому вигляді.

Криптографія, як наука про шифрування та захист інформації, включає кілька основних видів, кожен з яких має свої унікальні характеристики, принципи роботи та сфери застосування. Найбільш базовим і водночас найдавнішим типом є симетрична криптографія. У цій моделі використовується один і той самий ключ як для

шифрування, так і для розшифрування повідомлення. Це означає, що обидві сторони комунікації мають володіти спільним секретом, який повинен бути переданий безпечно ще до початку обміну даними. Симетричні алгоритми мають велику швидкість роботи, що робить їх особливо ефективними для шифрування великих обсягів даних у реальному часі. Одними з найпоширеніших симетричних шифрів є AES (Advanced Encryption Standard), який використовується в урядових та комерційних системах у всьому світі, та ChaCha20, що знайшов застосування у мобільних пристроях завдяки оптимізованій продуктивності.

Проте головним недоліком симетричної криптографії є складність у розповсюдженні ключів. Якщо кількість користувачів системи зростає, кількість унікальних пар ключів, які необхідно зберігати й захищати, збільшується експоненційно. Ця проблема частково вирішується завдяки асиметричній криптографії, що з'явилася в другій половині XX століття як радикально новий підхід до шифрування. Асиметричні системи базуються на використанні двох різних ключів – відкритого (public key) і закритого (private key). Відкритий ключ може бути вільно поширений і використовується для шифрування інформації, тоді як розшифрувати її можна лише за допомогою відповідного приватного ключа, який зберігається у суворій таємниці власником. Такий підхід дозволяє не передавати секретну інформацію під час обміну ключами, а також створює можливість реалізації цифрового підпису, що підтверджує справжність повідомлення та особу відправника. Найвідомішими прикладами асиметричних алгоритмів є RSA, DSA та алгоритми на еліптичних кривих (ECC), які забезпечують високий рівень безпеки навіть за умов обмежених обчислювальних ресурсів.

Попри свої переваги, асиметричні методи мають значно вищу обчислювальну складність порівняно з симетричними, що обмежує їх використання для шифрування великих масивів даних. Саме тому на практиці найчастіше використовується гібридна криптографія, яка поєднує переваги обох підходів. У гібридній схемі обмін

ключами або початкова автентифікація здійснюється за допомогою асиметричної криптографії, а сам обмін даними виконується за допомогою симетричних алгоритмів. Наприклад, під час встановлення захищеного з'єднання за протоколом HTTPS браузер спершу отримує відкритий ключ сервера та шифрує з ним випадковий симетричний ключ сесії, який далі використовується для основного обміну інформацією. Такий підхід забезпечує як високу швидкість, так і високий рівень захищеності комунікацій.

Методи шифрування та цифрового підпису є основними інструментами у криптографії, що забезпечують захист даних при передачі та зберіганні. Шифрування використовується для перетворення відкритої інформації (тексту) у зашифровану форму, яка незрозуміла стороннім особам. Принцип роботи полягає в тому, що навіть якщо зловмисник отримає доступ до зашифрованих даних, без відповідного ключа він не зможе їх розшифрувати й дізнатися зміст. Шифрування може бути симетричним або асиметричним, залежно від того, чи використовується один і той самий ключ для шифрування і дешифрування, чи пара відкритого і закритого ключів. У симетричному шифруванні широко застосовується алгоритм AES (Advanced Encryption Standard), який забезпечує надійний захист завдяки багаторівневому процесу перетворення даних із використанням блочних операцій і замін. У свою чергу, асиметричне шифрування, наприклад RSA, дозволяє захищати дані при їх пересиланні між сторонами, які не мають попередньо узгодженого секретного ключа.

Проте лише шифрування не забезпечує повного спектру криптографічних гарантій, зокрема автентичності та незаперечності. Для цього застосовується цифровий підпис – інструмент, який дозволяє перевірити, хто саме створив повідомлення, та чи не було воно змінено після підписання. Принцип цифрового підпису базується на використанні хеш-функцій та асиметричної криптографії. Коли користувач хоче підписати документ, він спочатку обчислює хеш (контрольну суму) цього документу — компактне представлення його вмісту фіксованої довжини. Потім ця хеш-сума шифрується приватним ключем

користувача, і результат додається до документа як цифровий підпис. Одержувач, отримавши підписаний документ, може обчислити хеш від отриманого тексту самостійно, розшифрувати підпис за допомогою відкритого ключа підписанта та порівняти ці значення. Якщо вони збігаються, документ є справжнім і не зміненим.

Серед найпоширеніших алгоритмів цифрового підпису варто відзначити RSA (Digital Signature Algorithm) та алгоритми на еліптичних кривих (ECDSA), які забезпечують високий рівень безпеки при меншій довжині ключа. У деяких випадках використовується також GOST-підпис, що є стандартом в окремих країнах пострадянського простору. Цифрові підписи стали важливою складовою інфраструктури електронного документообігу, електронної пошти, банківських систем та блокчейн-технологій. Наприклад, у криптовалюті Bitcoin цифровий підпис дозволяє підтвердити право власності на кошти та санкціонувати їх передачу в мережі без централізованої перевірки.

Криптографія є основою сучасних протоколів безпеки, які забезпечують захищене з'єднання в Інтернеті, захист мережевого трафіку, а також гарантують цілісність і достовірність транзакцій у децентралізованих системах. Одним із найважливіших і найпоширеніших прикладів є протокол TLS (Transport Layer Security), що використовується для забезпечення захищеної передачі даних у мережі. Коли користувач заходить на вебсайт із захищеним з'єднанням (https://), саме TLS забезпечує шифрування переданої інформації між браузером і сервером. Основна роль криптографії тут полягає у встановленні безпечного каналу зв'язку між сторонами, які раніше не обмінювались ключами. Для цього використовуються методи асиметричної криптографії на етапі встановлення з'єднання (так зване TLS-handshake), під час якого генерується симетричний ключ сесії. Надалі цей ключ використовується для шифрування даних із високою швидкістю.

Особливо помітна роль криптографії у блокчейн-технологіях, які є основою криптовалют,

децентралізованих додатків та смарт-контрактів. У блокчейні криптографія використовується для забезпечення цілісності транзакцій, підтвердження автентичності учасників мережі та створення незмінної структури даних. Одним з ключових криптографічних інструментів тут є геш-функції – вони дозволяють кожному блоку містити унікальний хеш, який залежить від вмісту самого блоку та хешу попереднього, утворюючи ланцюг, який неможливо змінити заднім числом без помітного втручання. Це забезпечує властивість незмінності (immutability), що є критичною для довіри у системі без централізованого контролю. Цифрові підписи, зокрема на основі алгоритмів ECDSA, гарантують, що лише власник відповідного приватного ключа може ініціювати транзакцію. Також криптографічні алгоритми відіграють роль у механізмах консенсусу, наприклад у процесі майнінгу, де складні обчислення хешів забезпечують захист мережі від шахрайства.

Попри надзвичайну важливість криптографії як одного з найнадійніших інструментів захисту інформації, її практична реалізація супроводжується низкою складних проблем, які суттєво впливають на ефективність і безпеку систем. Однією з головних проблем є так званий «людський фактор», який часто виявляється найслабшою ланкою в захищених системах. Навіть найсучасніші криптографічні алгоритми можуть бути скомпрометовані внаслідок неправильного зберігання ключів, слабких паролів, або через фішингові атаки, що змушують користувача розкрити свої облікові дані. Неправильне налаштування програмного забезпечення або недосвідченість адміністратора системи можуть призвести до ситуацій, коли криптографічний захист лише ілюзорно присутній, але насправді не виконує своїх функцій.

Ще одним критично важливим аспектом є складність управління ключами. У великих організаціях існує потреба в централізованому зберіганні, обміні, оновленні та відкликанні криптографічних ключів. Недостатньо налагоджені або застарілі механізми керування ключами призводять до численних вразливостей. Наприклад, якщо

ключі не змінюються регулярно або якщо після компрометації одного з них немає оперативного процесу його відкликання, вся система може бути скомпрометована. Особливо складною є ситуація, коли ключі передаються між різними системами або зберігаються на кінцевих пристроях без належного захисту, що створює ризики втрати або викрадення.

Також велике значення мають реалізаційні помилки у програмному коді. Навіть якщо використовується теоретично надійний алгоритм, його некоректна імплементація в програмному забезпеченні може відкрити шлях для атак. Наприклад, неправильне використання генераторів випадкових чисел у криптографічних протоколах часто призводить до слабких ключів, які зловмисник може відновити. Подібним чином, уразливості в бібліотеках, як-от OpenSSL чи libgcrypt, неодноразово ставали причиною масових інцидентів безпеки, серед яких варто згадати скандальну уразливість Heartbleed. Тому реалізація криптографічних засобів потребує найвищого рівня програмної дисципліни, перевірок, аудитів коду та незалежного тестування.

Окрему проблему становлять обчислювальні ресурси, необхідні для реалізації сучасних криптографічних алгоритмів. Асиметричні алгоритми, зокрема RSA або алгоритми на еліптичних кривих, є ресурсоємними, особливо у пристроях з обмеженими обчислювальними можливостями, таких як мобільні телефони, вбудовані системи або «інтернет речей» (IoT). Це часто змушує розробників шукати компроміси між продуктивністю та безпекою, що не завжди веде до оптимальних рішень.

Аналіз вразливостей криптосистем є одним із ключових напрямів у забезпеченні інформаційної безпеки, оскільки навіть найстійкіший теоретично криптографічний алгоритм може бути зламаний або обійдений через слабкі місця в його реалізації чи використанні. Уразливості в криптосистемах можуть виникати як на рівні алгоритмів, так і на рівні їх програмної або апаратної реалізації. Теоретичні вразливості стосуються, передусім, слабких математичних основ, на яких побудовано алгоритм.

Наприклад, шифри, що мають обмежений ключовий простір або недостатню ентропію, можуть бути піддані повному перебору (brute-force) або іншим видам криптоаналізу. Якщо алгоритм дозволяє передбачити певні шаблони у шифротексті або не забезпечує статистичну рівномірність вихідних даних, це також створює передумови для атаки.

Окрім суто математичних вразливостей, велику небезпеку становлять помилки реалізації. Навіть найнадійніший алгоритм може бути зкомпрометований, якщо він реалізований з порушенням криптографічних принципів. Прикладами таких вразливостей є використання незахищених генераторів випадкових чисел, неправильне управління пам'яттю або відсутність перевірки автентичності в протоколах шифрування.

Відомі атаки на реалізації, такі як атака таймінгу (timing attack), атака через побічні канали (side-channel attack) або атака на повторне використання ключів, демонструють, що загроза може виходити не від самого алгоритму, а від способу, яким він застосовується на практиці. Наприклад, атака Meltdown і Spectre використовували архітектурні особливості процесорів для отримання доступу до чутливих даних, у тому числі й криптографічних ключів, із пам'яті.

Ще одним критичним аспектом є людський фактор та організаційні помилки, які часто стають джерелом уразливостей.

Неправильне зберігання ключів, використання застарілих або скомпрометованих алгоритмів, відсутність регулярного аудиту безпеки, а також недбале ставлення до оновлень і патчів — усе це відкриває нові вектори атак. Наприклад, широко відома вразливість Heartbleed у бібліотеці OpenSSL дозволяла зловмисникам отримувати довільні фрагменти пам'яті з серверів, у тому числі й конфіденційні ключі. Іноді достатньо лише однієї помилки в коді або одного невчасно оновленого компонента системи, щоб створити потенційно катастрофічну загрозу.

Список використаних джерел

1. A Mathematical Proposed Model for Public Key Encryption Algorithms in Cybersecurity URL: https://www.researchgate.net/publication/354291983_A_MATHEMATICAL_PROPOSED_MODEL_FOR_PUBLIC_KEY_ENCRYPTION_ALGORITHMS_IN_CYBERSECURITY

2. Mathematical Approaches Transform Cybersecurity from Experimental to Scientific Discipline URL: https://www.mdpi.com/2076-3417/13/11/6508

3. Mathematical Models in Information Security and Cryptography URL: https://www.mdpi.com/journal/mathematics/special_issues/71516B35F0

ПРОТИДІЯ КІБЕРАТАКАМ ЧЕРЕЗ АЛГОРИТМІЧНІ ЗАСОБИ

Гуненко В.[1], Світличний В.[1]

[1]*Харківській національний університет внутрішніх справ*
vit.svet@ukr.net

У статті досліджено сучасні виклики у сфері кібербезпеки та алгоритмічні засоби протидії кібератакам. Розглянуто основні типи атак – DoS, фішинг, malware, SQL-ін'єкції, APT – та методи їх виявлення за допомогою сигнатурного, евристичного й поведінкового аналізу. Особливу увагу приділено використанню машинного навчання для ідентифікації загроз, а також інтеграції алгоритмів у SIEM- та SOAR-системи. Проаналізовано ефективність автоматизованого реагування та проблеми, пов'язані з хибними спрацьовуваннями. Робота акцентує важливість створення комплексних, адаптивних систем кіберзахисту, здатних до самонавчання та оперативного реагування в умовах динамічного цифрового середовища.

Ключові слова: кібербезпека, кібератаки, алгоритмічні методи, машинне навчання, SIEM, SOAR, IDS/IPS, фішинг, DoS, APT, поведінковий аналіз

Вступ

У сучасному цифровому світі, де інформація стала одним із найцінніших активів, проблема захисту від кібератак набула надзвичайної актуальності. Кібератаки – це цілеспрямовані дії, спрямовані на порушення, перехоплення, зміну, знищення або викрадення інформації, що циркулює в інформаційно-комунікаційних системах. Вони можуть мати найрізноманітніший характер, починаючи від примітивних спроб фішингу до складних багаторівневих атак державного рівня, які використовують шкідливе програмне забезпечення, соціальний інжиніринг та вразливості нульового дня. Класифікація кібератак зазвичай здійснюється на основі їх мети, методу реалізації та рівня впливу. Найпоширенішими є атаки на

конфіденційність, цілісність і доступність даних, які можуть реалізовуватись через експлойти, віруси, хробаки, трояни, атакуючі скрипти, ботнети, зловмисні мобільні додатки та інші інструменти. Значна частина атак також базується на використанні людського фактору, коли зловмисник обманом змушує користувача надати доступ до системи або інформації.

Виклад основного матеріалу

У зв'язку з ускладненням технічного середовища, зростанням кількості пристроїв, підключених до мережі, розвитком Інтернету речей, хмарних технологій та віддаленої роботи, традиційні підходи до безпеки втрачають свою ефективність. Сучасні кібератаки здатні обходити статичні засоби захисту, залишаючись невиявленими в системах протягом тривалого часу, а їх наслідки можуть охоплювати не лише технологічну, а й економічну, соціальну та навіть політичну сфери. У таких умовах надзвичайно важливою стає розробка і впровадження ефективних алгоритмів захисту, які здатні виявляти, запобігати, локалізовувати та усувати загрози в режимі реального часу. Алгоритмічний підхід до кібербезпеки дозволяє автоматизувати процеси захисту, підвищити швидкість реагування, зменшити навантаження на аналітиків безпеки та забезпечити масштабованість захисних механізмів відповідно до складності системи.

Серед найпоширеніших типів кібератак, що загрожують інформаційним системам у всьому світі, особливе місце займають DoS-атаки, фішинг, зловмисне програмне забезпечення (malware), SQL-ін'єкції та атаки типу APT (Advanced Persistent Threats). DoS-атака (Denial of Service) передбачає перевантаження ресурсів системи – вебсервера, мережі, додатку – запитами до такої міри, що вона припиняє обслуговування легітимних користувачів. У своїй розширеній формі, DDoS (Distributed Denial of Service), ця атака реалізується одночасно з тисяч комп'ютерів, зазвичай заражених вірусами, які об'єднані у ботнет. Такі атаки надзвичайно складно зупинити, оскільки трафік надходить із реальних, хоч і скомпрометованих пристроїв.

Фішинг є ще однією поширеною загрозою, яка спирається на обман користувача. Через електронну пошту, месенджери або фейкові вебсайти зловмисник імітує офіційне повідомлення, закликаючи жертву надати конфіденційну інформацію – логіни, паролі, банківські реквізити. Фішинг небезпечний тим, що може слугувати першим етапом складнішої атаки, наприклад, для впровадження шкідливого програмного забезпечення або отримання доступу до внутрішньої мережі. Malware – загальне поняття для всієї категорії шкідливих програм, які мають на меті вивести з ладу систему, зібрати інформацію або надати віддалений контроль зловмиснику. Сюди входять віруси, трояни, руткіти, шпигунські програми, програми-вимагачі (ransomware), які шифрують дані й вимагають викуп за їх розблокування. Одним із класичних векторів зараження є SQL-ін'єкція – вид атаки на вебдодатки, коли у поле введення користувача зловмисник вводить шкідливий SQL-запит, що дозволяє маніпулювати базою даних: переглядати, змінювати або видаляти інформацію, а іноді й отримувати повний доступ до серверної частини.

Окрему категорію загроз становлять APT-атаки – це цілеспрямовані, добре сплановані й довготривалі операції, що зазвичай ініціюються висококваліфікованими групами, зокрема державними. Такі атаки мають на меті проникнення в мережу, закріплення у ній і непомітне переміщення задля збору розвідданих або втручання в критичні процеси. APT-загрози складно виявити, оскільки вони маскуються під легітимні процеси, використовують унікальні шкідливі компоненти і поетапну тактику з ретельним уникненням виявлення.

Виявлення кібератак – це ключовий етап у системі захисту, який базується на застосуванні різних алгоритмів і методик. Сигнатурні методи виявлення є найбільш традиційними й полягають у порівнянні поточного трафіку або поведінки об'єктів з базою відомих ознак шкідливої активності – так званих сигнатур. Наприклад, якщо у файлі чи мережевому пакеті виявлено фрагмент коду, який відповідає відомому вірусу, система одразу повідомляє про

загрозу. Цей підхід ефективний для вже відомих атак, але є неефективним щодо нових, ще не задокументованих загроз – атак нульового дня. Саме тому на зміну сигнатурному аналізу приходять евристичні методи, які ґрунтуються на пошуку підозрілої, але не обов'язково вже ідентифікованої поведінки. Наприклад, якщо програма виконує нехарактерні для себе дії – змінює системні файли, створює копії себе у різних директоріях або підключається до зовнішніх серверів без очевидної причини – система може позначити її як потенційно небезпечну.

Ще ефективнішими є поведінкові алгоритми, які аналізують активність об'єктів у системі в динаміці, створюючи профілі «нормальної» поведінки користувачів, програм і пристроїв. Відхилення від цієї норми – наприклад, раптовий сплеск мережевої активності вночі, доступ до критичних файлів зі звичайного користувацького облікового запису або зміни в структурі прав доступу – розцінюються як сигнали можливої атаки. Поведінковий аналіз часто інтегрується зі штучним інтелектом і машинним навчанням, що дозволяє системам самостійно адаптуватися до нових типів загроз, аналізуючи великий обсяг даних у реальному часі. Такий підхід значно підвищує точність виявлення, зменшує кількість хибнопозитивних спрацьовувань і забезпечує проактивне реагування на атаки ще до того, як вони заподіють шкоду. Завдяки комбінації сигнатурних, евристичних і поведінкових алгоритмів сучасні системи кіберзахисту стають здатними до комплексного виявлення загроз у складному та швидкозмінному цифровому середовищі.

Використання машинного навчання для ідентифікації кіберзагроз стало одним із найперспективніших напрямів розвитку сучасної кібербезпеки, оскільки традиційні методи дедалі частіше виявляються недостатньо ефективними перед складними та швидко змінюваними атаками. На відміну від сигнатурних систем, що потребують заздалегідь відомого опису загрози, алгоритми машинного навчання здатні виявляти аномальні патерни в даних, які не підпадають під звичайну поведінку системи або користувача. Це дає змогу виявити нові або

модифіковані атаки, зокрема атаки нульового дня, які раніше не були зафіксовані в базах даних. Суть машинного навчання у сфері безпеки полягає в тому, що модель навчається на великій кількості даних про мережевий трафік, поведінку процесів, вхід/вихід користувачів, журнали системних подій тощо, а потім використовує отримані знання для класифікації майбутніх подій як нормальних або підозрілих.

Методи машинного навчання, що застосовуються в цьому контексті, можуть бути як підконтрольними (supervised learning), коли модель тренується на розмічених даних, так і безпідконтрольними (unsupervised learning), які корисні тоді, коли структура даних невідома, і йдеться про виявлення кластерів або аномалій без попереднього маркування. Наприклад, за допомогою алгоритмів класифікації, таких як дерева рішень, логістична регресія, метод опорних векторів або нейронні мережі, можна точно передбачати, чи є певна сесія з'єднання зловмисною. Безпідконтрольні методи, як-от алгоритм k-середніх або кластеризація DBSCAN, дозволяють ідентифікувати відхилення у великих масивах даних, які можуть бути ранніми ознаками несанкціонованої активності. Крім того, використовуються методи навчання з підкріпленням, які дозволяють системам автоматично вдосконалювати свої рішення шляхом аналізу наслідків своїх дій. Усе це робить машинне навчання надзвичайно корисним інструментом для побудови інтелектуальних систем безпеки, які не просто реагують на інциденти, а здатні передбачати та запобігати загрозам до їх реалізації.

Системи IDS/IPS можуть бути побудовані за різними архітектурами: мережева (Network-based IDS/IPS), яка аналізує мережевий трафік у режимі реального часу, або хостова (Host-based IDS/IPS), яка контролює активність окремих пристроїв або серверів, зокрема операційні системи, журнали подій, доступ до файлів та інші критичні параметри. Основу таких систем становлять алгоритми виявлення загроз: сигнатурні, що орієнтуються на відомі шаблони атак, і аномальні, які виявляють відхилення від нормальної поведінки. З розвитком технологій IDS/IPS

інтегруються з SIEM-системами (Security Information and Event Management), що дозволяє аналізувати загрози у ширшому контексті, враховуючи різні джерела даних, історію подій і поведінкові моделі користувачів.

Алгоритми реагування та автоматизація протидії кібератакам є важливим етапом у забезпеченні безперервного та ефективного захисту інформаційних систем.

У сучасних умовах, коли атаки відбуваються зі стрімкою швидкістю та часто охоплюють численні етапи й вектори впливу, людський фактор виявляється недостатньо швидким для ефективного реагування. Саме тому автоматизовані алгоритми, вбудовані у систему захисту, покликані оперативно виявляти загрозу, визначати її природу, локалізувати зону ураження та здійснювати необхідні дії для нейтралізації або мінімізації наслідків. Такі алгоритми можуть бути запрограмовані на виконання широкого спектру заходів: від блокування підозрілого мережевого трафіку до ізоляції вразливих сегментів системи, від скидання сесій користувача до оновлення політик доступу або виклику резервної інфраструктури. Автоматизоване реагування дозволяє реалізувати концепцію SOAR (Security Orchestration, Automation and Response), в межах якої система безпеки функціонує не лише як детектор, а як активний захисник, що приймає рішення на основі заздалегідь прописаних сценаріїв або динамічних оцінок ситуації. Важливо, що ці алгоритми можуть працювати в режимі 24/7 без зниження швидкості чи втоми, що дає змогу ефективно протистояти навіть масованим або складним атакам з мінімальним людським втручанням.

Однією з головних проблем, яка виникає при використанні автоматизованих алгоритмів виявлення і реагування на загрози, є висока ймовірність помилкових спрацьовувань – як false positive, так і false negative. Помилкове позитивне спрацювання (false positive) відбувається тоді, коли система ідентифікує легітимну активність як загрозу.

Це може спричинити небажане блокування критичних сервісів, призупинення доступу для користувачів або запуск непотрібних процедур реагування, що, в свою чергу, може призвести до порушень бізнес-процесів і навіть втрати довіри до системи безпеки.

З іншого боку, помилкове негативне спрацювання (false negative) є ще небезпечнішим, адже у такому випадку реальна атака залишається непоміченою, що дозволяє зловмиснику безперешкодно проникати в систему, розширювати присутність, викрадати дані або руйнувати інфраструктуру. Причинами таких помилок можуть бути як обмеженість алгоритму, недосконалість моделей машинного навчання, неякісно підготовлені дані для навчання, так і динамічна зміна поведінки зловмисників, які цілеспрямовано маскують свої дії під легітимні процеси.

Балансування між чутливістю системи до загроз і її толерантністю до помилкових сигналів є складним завданням, що потребує постійного аналізу, донавчання моделей, уточнення правил та участі фахівців із кібербезпеки, які мають коригувати поведінку системи відповідно до конкретного середовища.

Ще одним критично важливим елементом ефективного використання алгоритмів безпеки є їх інтеграція в комплексні SIEM-системи (Security Information and Event Management). SIEM-платформи об'єднують у собі функціональність збору, агрегації, зберігання, аналізу та візуалізації журналів подій з різних джерел – серверів, мережевих пристроїв, прикладного ПЗ, антивірусів, фаєрволів та інших компонентів IT-інфраструктури.

Інтеграція алгоритмів виявлення загроз у SIEM дозволяє отримати цілісну картину стану безпеки організації, розуміти, як взаємопов'язані різні події, виявляти складні атаки, що складаються з багатьох етапів, і оцінювати ризики в контексті бізнесу. Зокрема, SIEM надає змогу виявляти низькорівневі сигнали, які окремо не є загрозою, але у сукупності свідчать про початок складної атаки. Вбудовані алгоритми в SIEM можуть реалізовувати як

прості правила кореляції подій, так і складні сценарії поведінкового аналізу, засновані на штучному інтелекті.

Більше того, сучасні SIEM-системи все частіше інтегруються з SOAR-рішеннями, що дозволяє поєднувати виявлення загроз із автоматизованими сценаріями реагування: надсилання сповіщень, активація ізоляції, створення інцидентів у системах управління, ініціювання процедур розслідування. Завдяки цьому алгоритми стають не лише інструментом технічного аналізу, а й дієвим компонентом стратегічного управління безпекою, який дозволяє керівництву приймати обґрунтовані рішення в умовах постійного потоку подій та атак. Інтеграція алгоритмів у SIEM – це крок до створення єдиної аналітичної платформи, що забезпечує прозорість, гнучкість та ефективність усієї системи кіберзахисту.

Висновок

У сучасних умовах ефективний захист від кібератак неможливий без використання алгоритмічних засобів, здатних виявляти, локалізовувати й нейтралізовувати загрози у реальному часі. Застосування машинного навчання, поведінкового аналізу та автоматизованого реагування значно підвищує рівень безпеки, дозволяючи системам адаптуватися до нових типів атак.

Проте важливим залишається завдання мінімізації хибних спрацьовувань і забезпечення гнучкості алгоритмів у різних інфраструктурних середовищах. Інтеграція таких алгоритмів у SIEM/SOAR-платформи забезпечує комплексний підхід до управління кіберзагрозами та стратегічне ухвалення рішень у сфері інформаційної безпеки.

Список використаних джерел

1. Algorithmic Detection Approaches in Cybersecurity: Exploring New Innovations and Techniques: https://www.researchgate.net/publication/384367638

2. Methods and Technologies to Detect Cyber Attacks: https://www.eccu.edu/blog/methods-technologies-detect-cyber-attacks/

3. Machine Learning Algorithms for Detecting and Preventing Cyber Threats: https://www.oxjournal.org/machine-learning-algorithms-for-detecting-and-preventing-cyber-threats/

ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В КІБЕРБЕЗПЕЦІ

Оніщенко Є.[1], Мальцев В.[1]

[1]*Харківський національний університет внутрішніх справ*

Сучасний цифровий простір характеризується стрімким зростанням обсягів даних та ускладненням кіберзагроз. Традиційні засоби захисту ІТ-інфраструктури все частіше виявляються недостатніми. У відповідь на це зростає потреба у впровадженні інноваційних підходів до забезпечення кібербезпеки, серед яких провідне місце займає штучний інтелект (ШІ).

ШІ має здатність оперативно аналізувати великі масиви даних, розпізнавати закономірності та аномалії, передбачати загрози і навіть автоматично реагувати на них. У даній роботі розглянуто основні аспекти застосування ШІ в сфері кібербезпеки, його переваги, виклики та перспективи розвитку такі як:

1. Виявлення загроз в режимі реального часу.Одним із ключових напрямів застосування ШІ в кібербезпеці є виявлення загроз у реальному часі. Алгоритми машинного навчання здатні обробляти лог-файли, мережеві пакети та поведінкові патерни користувачів для виявлення аномалій.

2. Прогнозування потенційних атак. Завдяки здатності ШІ до навчання на історичних даних, системи можуть виявляти закономірності, які передують певним типам атак. Це дозволяє розробляти превентивні заходи.

3. Автоматизація реагування на інциденти. ШІ дозволяє автоматизувати процес реагування на інциденти – від блокування IP-адреси до ізоляції зараженого пристрою у мережі.

4. Розпізнавання шкідливого ПЗ. Сучасні антивірусні рішення використовують не лише бази сигнатур, але й поведінковий аналіз, що ґрунтується на машинному навчанні.

Штучний інтелект також використовується у фінансовому секторі, сфері охороні здоров’я, урядових

установах, приватних компаніях, хмарних сервісах, забезпечуючи ефективний захист.

Таким чином, штучний інтелект відкриває нові горизонти в забезпеченні кібербезпеки. Проте ефективне впровадження потребує етичного підходу, якісних даних, прозорості моделей і людського контролю.

Список використаних джерел

1. Колесніков С. В. Штучний інтелект і безпека інформації : монографія / С. В. Колесніков, І. М. Шелестов. – Київ : Видавництво НАУ, 2020. – 236 с.

2. Бондар І. І. Кібербезпека: методи виявлення та реагування на загрози з використанням ШІ / І. І. Бондар // Інформаційна безпека. – 2021. – № 1. – С. 34–41.

3. Russell S. Artificial Intelligence: A Modern Approach / S. Russell, P. Norvig. – 4th ed. – Boston : Pearson, 2020. – 1152 p.

4. Sommer R. Outside the Closed World: On Using Machine Learning For Network Intrusion Detection / R. Sommer, V. Paxson // IEEE Symposium on Security and Privacy. – 2010. – P. 305–316. – DOI: 10.1109/SP.2010.25.

ЗМІСТ

Матеріали III Всеукраїнської
науково-практичної конференції

ТЕОРЕТИЧНА І ПРИКЛАДНА КІБЕРБЕРБЕЗПЕКА

Випуск 3

Всі матеріали надані в авторській редакції