MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE

KHARKIV NATIONAL
UNIVERSITY OF RADIO ELECTRONICS

RADIOTEKHNIKA

All-Ukrainian
interdepartmental scientific and technical collection

ISSN 0485-8972
eISSN 2786-5525

Founded in 1965

I S S U E 2 1 8

Kharkiv
Kharkiv National
University of Radio Electronics
2024

UDC 621.3

The collection is included in the List of scientific professional publications of Ukraine, category «Б», technical and physical-mathematical sciences (approved by orders of the Ministry of Education and Science from 17.03.2020 № 409; from 02.07.2020 № 886; from 24.09.2020 № 1188) by specialties: 105 – Applied Physics and Nanomaterials; 125 – Cybersecurity and information protection; 163 – Biomedical Engineering; 171 – Electronics; 172 – Electronic communications and Radio Engineering; 173 – Avionics; 174 –Automation and Computer-Integrated Technologies and Robotics; 175 – Metrology and information-measuring technique; 176 – Micro- and Nanosystem Technology.

Website: rt.nure.ua

Registration certificate KV № 12098-969 PR dated 14. 12. 2006.

The authors are responsible for the content of the article.

Responsible for the issue: *I.V. Svyd, PhD, Assoc. prof., I.D. Gorbenko, Dr. Sc. (Tech.), prof.*
Technical Secretary: *O.S. Polyakova.*

Recommended by the Scientific and Technical Council of Kharkiv National University of Radio Electronics, protocol № 9 dated 26.09.2024.

Address of the editorial board: Kharkiv National University of Radio Electronics (NURE), ave. Nauky, 14, Kharkiv, 61166, tel. (0572) 7021-397.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ
УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

РАДІОТЕХНІКА

Всеукраїнський
міжвідомчий науково-технічний збірник

ISSN 0485-8972
eISSN 2786-5525

Засновано в 1965 р.

В И П У С К 2 1 8

Харків
Харківський національний
університет радіоелектроніки
2024

УДК 621.3

Збірник включено до Переліку наукових фахових видань України, категорія "Б", технічні та фізико-математичні науки (затверджено наказами МОНУ від 17.03.2020 № 409; від 02.07.2020 № 886; від 24.09.2020 № 1188) за спеціальностями: 105 – Прикладна фізика та наноматеріали; 125 – Кібербезпека та захист інформації; 163 – Біомедична інженерія; 171 – Електроніка; 172 – Електронні комунікації та радіотехніка; 173 – Авіоніка; 174 – Автоматизація, комп'ютерно-інтегровані технології та робототехніка; 175 – Метрологія та інформаційно-вимірювальні технології; 176 – Мікро- та наносистемна техніка.

Сайт: rt.nure.ua

Реєстраційне свідоцтво КВ № 12098-969 ПР від 14. 12. 2006.

За зміст статті відповідальні автори.

Відповідальні за випуск: *І.В. Свид, канд. техн. наук, доц.*, *І.Д. Горбенко, д-р техн. наук, проф.*
Технічний секретар: *О.С. Полякова.*

Рекомендовано Науково-технічною радою Харківського національного університету радіоелектроніки, протокол № 9 від 26.09.2024.

Адреса редакційної колегії: Харківський національний університет радіоелектроніки (ХНУРЕ), просп. Науки, 14, Харків, 61166, тел. (0572) 7021-397.

CONTENT

SYSTEMS AND METHODS OF INFORMATION PROTECTION

MEANS OF TELECOMMUNICATIONS

RADIO ELECTRONIC SYSTEMS

PHYSICS OF DEVICES, ELEMENTS AND SYSTEMS

RELATED PROBLEMS OF RADIO ENGINEERING

ЗМІСТ

СИСТЕМИ І МЕТОДИ ЗАХИСТУ ІНФОРМАЦІЇ

ЗАСОБИ ТЕЛЕКОМУНІКАЦІЙ

РАДІОЕЛЕКТРОННІ СИСТЕМИ

ФІЗИКА ПРИЛАДІВ, ЕЛЕМЕНТІВ І СИСТЕМ

СУМІЖНІ ПРОБЛЕМИ РАДІОТЕХНІКИ

УДК 004.05 **DOI:10.30837/rt.2024.3.218.01**

В.І. ЄСІН, д-р техн. наук, В.В. ВІЛІГУРА, Д.Ю. УЗЛОВ, канд. техн. наук

АРХІТЕКТУРА НУЛЬОВОЇ ДОВІРИ: ПРОБЛЕМИ ТА РЕКОМЕНДАЦІЇ ЩОДО УСПІШНОГО ВПРОВАДЖЕННЯ

Вступ

Протягом останніх десятиліть підприємства активно переходять на цифрові технології. Сьогодні вони підтримують хмарні технології, частіше використовують віддалену роботу, додають рішення на кшталт «як послуга» (as-a-service), а також здійснюють безліч інших важливих перетворень, що призводить до розширення існуючої IT-інфраструктури підприємства. Все більше і більше пристроїв (у тому числі CYOD та BYOD) та сервісів обмінюються інформацією всередині корпоративних мереж, а також за їх межами. Ці зміни призводять до появи нових складних вимог до мережевої безпеки, яким існуючі рішення слабо відповідають [1]. У цих умовах фахівці з безпеки відповідно до нових потреб змушені масштабувати систему мережевої безпеки, часто посилюючи захист шляхом сегментування мережі на дрібніші частини. На жаль, таке рішення створило більше сприятливих можливостей для зловмисників через появу додаткового набору вразливостей, з яким не в змозі впоратися навіть новим, ефективнішим засобам захисту з підвищеним рівнем контролю, через те, що їхнє застосування обмежується межею корпоративної мережі та не дозволяє контролювати те, що відбувається за її межами. Як відомо, традиційна безпека мережі фокусується на захисті периметра. Отримуючи доступ до облікових даних користувача, зловмисники можуть переміщатися по мережі, розповсюджуючи шкідливе програмне забезпечення та додаючи привілеї в міру свого переміщення [2]. Недоліки цього підходу стають очевидними, якщо врахувати, що зловмисники у разі компрометації суб'єктів (кінцевих користувачів, застосунків та інших нефізичних сутностей) можуть отримати доступ до ресурсів усередині або за межами мережі. Більше того, багато підприємств більше не мають чітко визначеного периметра. Периметр втрачає свою актуальність через кілька факторів, включаючи зростання застосування хмарних технологій, мобільність та використання віддалених працівників [3]. Крім того, слід враховувати, що загрозу становлять і внутрішні зловмисники – інсайдери. Таким чином, ідея про те, що жодна мережа (ні внутрішня, ні зовнішня) не заслуговує на довіру, просувається як у наукових колах, так і на практиці [1].

Щоб сьогодні захистити сучасне цифрове підприємство, необхідна комплексна стратегія для безпечного доступу у будь-який час і в будь-якому місці до власних корпоративних ресурсів (застосунків, застарілих / успадкованих систем, даних, пристроїв тощо) незалежно від того, де вони розташовані [4]. Тому підприємства стали переосмислювати традиційний периметр безпеки мережі, схиляючись до нової концепції та архітектури захисту [5].

Такою концепцією є парадигма безпеки, що отримала назву «нульова довіра» (Zero Trust – ZT). Концепція нульової довіри стала дуже популярним підходом до створення захищених систем, що просувається промисловістю та державними органами як новий спосіб створення систем з високим ступенем безпеки [6]. За своєю суттю «нульова довіра» – це філософія, підхід та набір керівних принципів кібербезпеки, що використовуються для створення стратегії, яка фокусується на переміщенні захисту мережі від широких статичних периметрів мережі до вужчого зосередження уваги на суб'єктах, активах підприємства (пристроях, компонентах інфраструктури, застосунках, віртуальних та хмарних компонентах) та окремих або невеликих групах ресурсів [3, 7, 8]. Ідея концепції нульової довіри виникла ще на початку розвитку безпечних обчислень [1]. В її основі лежить застосування принципів безпечного проєктування, взятих із класичної роботи Saltzer та Schroeder [9] и более поздних работ [6, 10, 11], серед яких слід відзначити повне посередництво (перевіряйте кожен доступ на наявність повноважень), відкрите проєктування (розробка не повинна бути закритою; механізми захисту повинні залежати не від безграмотності потенційних зловмисників, а від володіння певними ключами або паролями, що легко захищаються; не повинно бути «безпеки через невідомість»), найменші привілеї (компоненти повинні мати не більше повноважень, ніж їм потрібно), глибокий захист, багаторівневий захист. Особливістю концепції нульової довіри можна вважати додатковий, більш жорсткий принцип «ніколи не довіряти, завжди перевіря-

ти» [12]. При цьому, якщо «*нульова довіра*» являє собою набір концепцій та ідей, розроблених для мінімізації невизначеності в застосуванні точних рішень щодо доступу з найменшими привілеями для кожного запиту в інформаційних системах і службах / сервісах, коли мережу вважають скомпрометованою, то *архітектура нульової довіри* (ZTA – zero trust architecture; іноді в перекладі можна зустріти назву – архітектура з нульовою довірою або архітектура безпеки з нульовою довірою), відповідно до визначення NIST [8] – це план кібербезпеки підприємства, який використовує концепції нульової довіри та охоплює зв'язки компонентів, планування робочого процесу та політики доступу; це архітектура кібербезпеки підприємства, яка базується на принципах нульової довіри та призначена для запобігання витоку даних і обмеження внутрішнього бічного (горизонтального) руху (переміщення). На відміну від архітектури, орієнтованої на захист по периметру, в якій будь-який об'єкт усередині заданого периметра вважається довіреним, ZTA забезпечує обробку будь-якого запиту та надання ресурсу суб'єкту, не покладаючись на неявну довіру [13]. Архітектура нульової довіри враховує нові тенденції, приділяючи особливу увагу захисту ресурсів, а не периметру мережі, оскільки розташування мережі більше не розглядається як основний компонент забезпечення безпеки, необхідної для ресурсу [3, 8]. В рамках архітектури нульової довіри доступ до ресурсів не надається до тих пір, поки суб'єкт, актив або робоче навантаження не будуть верифіковані за допомогою процедур автентифікації та прав / дозволів на виконання певних дій (авторизації) [3]. Тобто в основі архітектури нульової довіри лежить керування ідентифікацією, активами, автентифікація застосунків, сегментація мережі, політики, механізми та фактори аналізу загроз [14].

Одним із факторів, що визначають потребу в ZTA, є складність та гетерогенність сучасних IT-систем [6]. Тому ідея про те, що для забезпечення безпеки системи потрібне повсюдне, детальне та безперервне розгортання багаторівневих засобів контролю безпеки, є цілком очевидною.

Однак незважаючи на популяризацію концепції нульової довіри та очевидні переваги у сфері безпеки від її застосування на підприємствах виникають певні складнощі щодо її реалізації [5, 15, 16-19]. Розгортання архітектури нульової довіри є досить складним як з технічної, так і з організаційної точки зору [18]. Можна відзначити кілька основних перешкод на шляху створення систем нульової довіри в державних і приватних комп'ютерних системах [18, 19]: застарілі системи та інфраструктуру часто неможливо оновити до рівня нульової довіри; навіть якщо оновлення можливе, воно обійдеться недешево; однорангові технології не відповідають концепції нульової довіри, оскільки вони покладаються переважно на паролі, а не на багатофакторну автентифікацію в режимі реального часу; перенесення інформаційних систем організації з власних комп'ютерів на хмарні сервіси може стимулювати розвиток нульової довіри, але тільки в тому випадку, якщо все буде зроблено справді коректно (організації повинні знати, як планувати забезпечення безпеки на основі концепції нульової довіри під час переходу до хмарного середовища, а не простого здійснювати перенесення існуючих додатків у хмару). Крім того, основними серйозними факторами, що перешкоджають впровадженню концепції нульової довіри, за даними звіту компанії Fortinet [15], стали брак інформації для вибору рішення з нульовою довірою та відсутність кваліфікованих розробників/постачальників. У публікаціях галузевих видань технічні аспекти подібних систем описуються нечітко, ігноруються накопичені раніше знання у сфері безпеки, а звіти, що описують реальний досвід створення та використання архітектур нульової довіри, вельми нечисленні. Деякі постачальники IT-послуг стверджують, що вони мають способи реалізації ZTA, але вони не надають подробиць про те, як це зробити. У технічних документах (white papers) цих компаній описуються їх підходи, але вони надають лише високорівневі огляди або списки специфікацій [6]. Ця проблема посилюється ще й тим, що архітектура нульової довіри передбачає використання детального контролю безпеки, тому з великою ймовірністю можна припустити, що доведеться визначати, впроваджувати, розгортати та керувати величезною кількістю політик [18], а знань та досвіду може виявитися недостатньо.

Як видно з викладеного, існує проблема, пов'язана з певним дефіцитом поінформованості про концепцію та архітектуру нульової довіри (про їх теоретичну та практичну значущість) для вибору правильного рішення при побудові системи безпеки корпоративної інфор-

маційної системи підприємства у сучасних умовах. Стаття націлена на вирішення цієї проблеми шляхом узагальнення наявних досліджень та досвіду різних міжнародних компаній, які впроваджують даний підхід на практиці. У ній у стислому викладі розглядаються концептуальна архітектура нульової довіри, її основні логічні компоненти, моделі розгортання ZTA, загрози, пов'язані з архітектурою нульової довіри, а також деякі рекомендації щодо успішного впровадження архітектури нульової довіри на IT-підприємстві, які допоможуть зрозуміти фундаментальні зміни у підході до інформаційної безпеки, кібербезпеки.

1. Архітектура нульової довіри

Архітектура нульової довіри сьогодні це концепція, що розвивається, для якої поки не існує ні сертифікації, ні практичних стандартів [2]. Хоча робота у цьому напрямі активно проводиться. Що стосується нормативних документів, то на сьогодні є основна спеціальна публікація NIST SP 800-207 (в цьому документі дається абстрактне визначення архітектури нульової довіри, наводяться загальні моделі розгортання та варіанти використання, в яких нульова довіра може підвищити загальний рівень безпеки інформаційних технологій підприємства), спеціальні публікації серії NIST SP 1800-35 (у яких коротко описується, як NCCoE та його співробітники використовують комерційно доступні технології для створення сумісних, заснованих на відкритих стандартах прикладів реалізації ZTA, які відповідають концепціям та принципам, викладеним у NIST SP 800-207), документ Національного центру кібербезпеки Великобританії (NCSC – National Cyber Security Centre) [20] та деякі інші. Далі при викладанні матеріалу скористаємося інформацією та висновками з цих документів, а також з деяких інших публікацій різних міжнародних авторитетних видань, присвячених архітектурі нульової довіри.

Якщо підприємство вирішує прийняти нульову довіру як свою основну стратегію і створити ZTA, розроблену з урахуванням принципів нульової довіри, то воно насамперед зосереджується на суті проблеми, викладеної у визначенні «нульова довіра» [8], яка полягає у тому, щоб запобігти несанкціонованому доступ до даних та послуг/сервісів у поєднанні з максимально детальним контролем доступу. Правила доступу максимально деталізуються, щоб забезпечити мінімальні привілеї, необхідні для виконання дій, зазначених у запиті.

При цьому слід відразу усвідомити (про що говорять багато дослідників), що неможливо створити універсальну архітектуру, принаймні через те, що існують різні підходи до нульової довіри, які залежать від базової архітектури підприємства та вибору, зробленого фахівцями з безпеки. А ось представити набір відповідних компонентів та вимог, які можна використовувати для створення актуальної та ефективної архітектури для конкретного підприємства, це безумовно буде корисним внеском та допомогою у застосуванні тієї чи іншої моделі розгортання ZTA (яких, у принципі, може бути кілька для різних бізнес-процесів на одному підприємстві).

Для початку доречно звернутися до абстрактної моделі доступу з нульовою довірою (рис. 1), представленої в документі NIST [8], та визначити її основні компоненти.

Суб'єктом може бути людина (кінцевий користувач – user) або не фізична сутність (non-person / nonhuman entity) така як пристрій, сервер, сервіс, застосунок тощо [21, 22]. Система може являти собою пристрій, такий як ноутбук, мобільний телефон, віртуальна машина, контейнер тощо. Система покладається на *точку застосування політики* (*policy enforcement point* – PEP), щоб дозволити взаємодію з ресурсом. Суб'єкт взаємодіє із системою, і система має перевірити справжність/ідентифікацію (identity) суб'єкта, а також виконати автентифікацію та авторизацію.

Рис. 1. Абстрактна модель доступу із нульовою довірою

Суб’єкту (користувачу, пристрою, застосунку) потрібен доступ до корпоративного ресурсу. Цей ресурс, що знаходиться під контролем підприємства, може являти собою обчислювальні ресурси, дані, застосунки / сервіси (робоче навантаження), розміщені та керовані локально, у хмарі, на межах або в якійсь їх комбінації [3]. Доступ надається через *точку / пункт прийняття рішення про політику* (*policy decision point* – PDP) і відповідну *точку застосування політики* (PEP). Система з нульовою довірою повинна переконатися, що суб’єкт є справжнім (автентичним), а запит дійсним. PDP/PEP приймає належне рішення, щоб дозволити суб’єкту отримати доступ до ресурсу. Це означає, що нульова довіра стосується двох основних сфер: автентифікації та авторизації. Керування доступом залежить від стану безпеки пристрою (механізму, засобу) та потенційного розгляду інших ситуативних факторів (наприклад, часу та місцезнаходження, попередньої поведінки при доступі тощо), які можуть вплинути на рівень довіри до того, як доступ до ресурсу буде наданий відповідно до певних політик.

Загалом, підприємствам необхідно розробити та підтримувати динамічну політику доступу до ресурсів, що базується на оцінці ризиків, і налаштувати систему, яка гарантуватиме правильне та послідовне застосування цих політик для окремих запитів на доступ до ресурсів. Підприємству не слід покладатися на передбачувану надійність, коли суб’єкт відповідає базовому рівню автентифікації (наприклад, при вході до системи), а усі наступні запити ресурсів вважаються однаково дійсними.

1.1. Основні логічні компоненти архітектури кібербезпеки підприємства, яка базується на принципах нульової довіри

Існує безліч логічних компонентів, з яких може складатися впроваджена на підприємстві ZTA. Ці компоненти можуть працювати як у локальній мережі, так і у хмарі.

Представимо концептуальну архітектуру нульової довіри, яка ґрунтується на ідеї та принципах, викладених у документі NIST [8], одночасно уточнюючи її. Основні логічні компоненти архітектури нульової довіри, а також взаємозв'язок між ними наведено рис. 2.

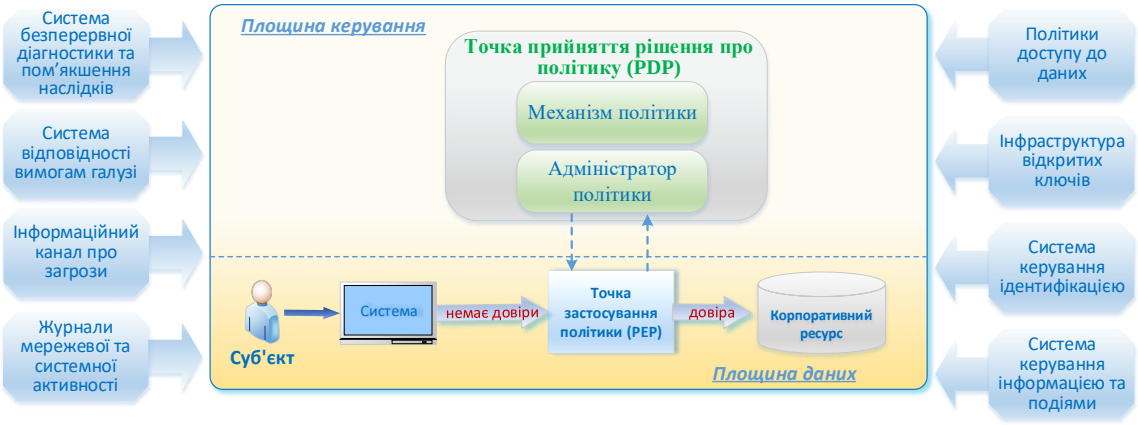


Рис. 2. Основні логічні компоненти архітектури нульової довіри

Передбачається, що суб'єкт функціонує в *недовіреному середовищі* та *недовіреній мережі*, і йому дозволено доступ до *Ресурсу* тільки *через точку застосування політики* (PEP). PEP контролює доступ суб'єкта до ресурсу через те, що NIST називає *зоною неявної довіри* (*implicit trust zone* – представляє собою область, де всі об'єкти є довіреними) [8]. PEP не зберігає і не визначає політику (докладніше призначення PEP розглядається нижче). Цю функцію виконує *точка прийняття рішення про політику* (PDP). Слід також звернути увагу, що *Суб'єкт* взаємодіє з корпоративним ресурсом через так звану *площину даних* (*data plane*), яка відрізняється від *площини керування* (*control plane*), яка використовується логічними компонентами ZTA для зв'язку.

Механізм політики (*policy engine* – PE). Цей компонент відповідає за остаточне рішення щодо надання доступу до ресурсу для даного суб'єкта. PE використовує політику підприємс-

тва, а також вхідні дані із зовнішніх джерел (наприклад, системи безперервної діагностики та пом'якшення наслідків (Continuous Diagnostics and Mitigation – CDM), служби аналізу загроз тощо, які описуються далі) як вхідні дані для алгоритму довіри, щоб надавати, забороняти або скасовувати доступ до ресурсу.

Алгоритм довіри (ТА – *trust algorithm*) – це процес, який використовується РЕ для ухвалення рішення з урахуванням таких вихідних даних, як записи в базі даних політик, роль користувача, відомості про поведінку, інформація про загрози (наприклад, сигнатури атак, що діють в Інтернеті, шкідливі програми тощо) і т. д. відповідно до потреб конкретного розгортання. Кожного разу, коли користувач робить запит на доступ, ТА використовує основну інформацію про ресурс і сторону, що запитує (наприклад, операційну систему (ОС), рівень виправлення / оновлення (patch), використовуваний застосунок). Інформація про користувача, автентифікація, що виконується РЕ, та інші атрибути, такі як час та місцезнаходження, також можуть бути використані ТА для обчислення рівня довіри. Щодо вимог доступу до ресурсу, то сюди можна віднести політики, що ґрунтуються на мінімальних вимогах для доступу, встановлених власником / адміністратором, наприклад, вимога багатофакторної автентифікації з нового місцезнаходження.

Алгоритм довіри може бути реалізований у різний спосіб [8, 23]:

– *ТА на основі критеріїв та оцінок*. ТА на основі критеріїв (Criteria-based ТА) вимагає, щоб сукупність певних атрибутів була обов'язково врахована, перш ніж дозволити виконання певної дії (наприклад, читання/запис). Ці критерії налаштовуються підприємством і мають бути налаштовані окремо для кожного ресурсу. Доступ надається або дія застосовується до ресурсу, лише якщо виконано всі критерії. *ТА на основі оцінок* (*Score-based TA*) обчислює рівень довіри на основі значень для кожного джерела даних і вагових коефіцієнтів, налаштованих підприємством. Якщо оцінка перевищує налаштоване порогове значення для ресурсу, доступ надається або виконується дія. В іншому випадку запит відхиляється або привілеї доступу зменшуються (наприклад, надається доступ для читання, але не для запису до файлу, або повна заборона доступу).

– *Сингулярний* (*Singular*) та *контекстний / контекстуальний* (*Contextual*) ТА. При ухваленні рішень у *сингулярному* ТА не враховується історична інформація користувача, що може прискорити процес ухвалення рішення, але існує ризик того, що деякі атаки можуть залишитись непоміченими, якщо вони залишаються в межах дозволеної ролі користувача (суб'єкта). Навпаки, *контекстуальний ТА* використовує історичні моделі поведінки користувача (суб'єкта) або мережевого агента під час оцінювання запитів на доступ. Це означає, що РЕ повинен підтримувати деяку інформацію про стан усіх суб'єктів і застосунків, але при цьому він може з більшою ймовірністю виявити зловмисника, який використовує підроблені облікові дані для доступу до інформації за шаблоном, який є нетиповим для того, що спостерігає РЕ для даного суб'єкта. Це також означає, що РЕ повинен бути поінформований про поведінку користувача від адміністратора політики (і РЕР), з якими суб'єкти взаємодіють під час комунікації. Аналіз поведінки суб'єкта може бути використаний для створення моделі прийнятного використання, і відхилення від цієї поведінки можуть викликати додаткові перевірки автентифікації або відхилення запитів ресурсів.

Механізм політики працює в парі з іншим компонентом – адміністратором політики (РА). Механізм політики приймає та реєструє рішення (як ухвалене або відхилене), а адміністратор політики виконує це рішення. Як тільки ТА прийняв відповідне рішення, РЕ передає його РА, який налаштовує всі відповідні точки застосування політики, щоб увімкнути або вимкнути з'єднання. Наприклад, він може надіслати налаштування шлюзам та агентам, тим самим вимагаючи повторної автентифікації, повторної авторизації або розриву з'єднання відповідно до певних політик.

Адміністратор політики (*Policy administrator* – РА). Цей компонент відповідає за встановлення та/або відключення шляху зв'язку між суб'єктом і ресурсом (через команди відповідним точкам застосування політики). Він може генерувати будь-які специфічні для сеансу

ідентифікатори та токени автентифікації або облікові дані, що використовуються клієнтом для доступу до корпоративного ресурсу. Він тісно пов'язаний з PE і покладається на його рішення – остаточно дозволити або заборонити сеанс. Якщо сеанс авторизовано, а запит автентифікований, PA налаштовує точку застосування політики (PEP), щоб дозволити розпочати сеанс. Якщо сеанс відхилено (або попереднє схвалення відхилено), PA сигналізує PEP закрити з'єднання. У деяких реалізаціях PE та PA можуть розглядатися як одна служба. В даному випадку вона поділяється на дві логічні складові. PA взаємодіє з PEP під час створення каналу зв'язку. Цей зв'язок здійснюється через площину керування.

Точка застосування політики (*Policy enforcement point* – PEP). Цей компонент відповідає за підключення, моніторинг та завершення з'єднань між суб'єктом та корпоративним ресурсом. PEP зв'язується з PA для надсилання запитів та/або отримання оновлень політики від PA. Це єдиний логічний компонент в архітектурі нульової довіри, але він може бути розбитий на два різні компоненти: клієнт (наприклад, агент на пристрої) та сторона ресурсу (наприклад, компонент шлюзу перед ресурсом, який контролює доступ) або єдиний компонент порталу, що виконує роль сторожа для комунікаційних шляхів. За межами PEP знаходиться довірча зона, де розміщено корпоративний ресурс.

Існує три типи PEP [7]: *PEP агента користувача* (*user agent PEP*), *PEP мережі* та *PEP застосунку*.

PEP мережі вважається найбільш простими з концептуальної точки зору в моделях нульової довіри, оскільки створення мереж нульової довіри зазвичай є найпоширенішою відправною точкою. Мережеві PEP існують у багатьох організаціях. Корпоративні міжмережові екрани (сьогодні це, як правило, брандмауери наступного покоління – NGFW – *Next-Generation Firewalls*) можна розглядати як PEP з нульовою довірою, хоч і з деякими застереженнями. Наприклад, чи можна вважати точкою застосування політики деякий базовий брандмауер п'ятирічної давності? Так, він, звичайно ж, є «точкою контролю мережі» у тому сенсі, що він має правила контролю доступу, які він забезпечує. Однак такий міжмережевий екран навряд чи буде точкою застосування політики (PEP) з нульовою довірою, оскільки він не відповідатиме таким вимогам, як: мати можливість застосовувати модель політики PDP, орієнтовану на ідентифікацію та контекстну залежність; автоматично реагувати на зміни політики, які визначаються PDP; використовувати канал керування для зв'язку з PDP. Адже кожен PEP повинен мати можливість отримувати постійні оновлення від PDP і автоматично коригувати політику, яку він застосовує практично в реальному часі і без втручання людини. Це єдиний спосіб досягти гнучкого та динамічного характеру, властивого підходу нульової довіри, навіть у невеликих масштабах [7].

Оскільки такі точки застосування політики працюють на мережному рівні, вони можуть здійснювати контроль мережного трафіку (як метаданих, так і фактичних даних про трафік), і тому вони є природними точками застосування політики.

PEP застосунків можуть бути зовнішніми по відношенню до застосунків (наприклад, система PAM – *Privileged Access Management* – керування привілейованим доступом) або DLP – *Data Loss Prevention* – запобігання втраті даних) або внутрішніми, наприклад, агент, що працює з *робочим навантаженням* (*workload*). В останньому випадку PEP можна використовувати для локального застосування політик на хості, наприклад, для застосування правил брандмауера локальної ОС. Крім того, точка застосування політики логічно може бути частиною самого застосунку, спираючись на зовнішні атрибути або дії, що впливають на застосунок.

PEP агента користувача – це компоненти, які запускаються на пристрої користувача та надають функції, які часто потрібні для систем з нульовою довірою, такі як встановлення зашифрованого з'єднання через ненадійну мережу. Ці PEP часто використовуються для аналізу пристрою з метою отримання інформації, яка використовується як вхідні дані для політик (наприклад, конфігурації пристрою та стану безпеки). Такий PEP також може взає-

модіяти з суб'єктом (кінцевим користувачем), наприклад, запитуючи додаткову автентифікацію або повідомляючи його.

Слід мати на увазі, що в деяких випадках межа між цими типами точок застосування політики нечітка, і функції, які вони виконують, можуть частково перекриватися. Наприклад, відомо, що IDS/IPS (система виявлення вторгнень / система запобігання вторгненням) можуть бути мережевими або хост-орієнтованими. Аналогічно функції DLP можуть бути реалізовані всередині мережевого пристрою, такого як NGFW, або на хості. В принципі, не так важливо, чи діють конкретні точки застосування політики, такі як DLP та PAM, на мережевому рівні або на рівні застосунків (або на обох). Важливо те, що і DLP, і PAM слід розглядати як частину PEP з нульовою довірою, а їхні політики логічно мають бути частиною моделі нульової довіри. В ідеалі для цього має бути інтеграція між ними та системою нульової довіри. Зрештою, все залежить від конкретної реалізації. Тобто функціональність та поведінка PEP залежатимуть від обраної платформи та від варіанта її розгортання.

Крім цих основних компонентів на підприємстві, що реалізує архітектуру нульової довіри, є ряд зовнішніх компонентів, які сприяють реалізації безпеки з нульовою довірою. А саме існує кілька джерел даних, які надають вхідні дані та правила політик, які використовуються механізмом політик при прийнятті рішень про доступ. До них належать локальні джерела даних, а також зовнішні (тобто не контрольовані або не створені підприємством) джерела даних. Зокрема такі (рис. 2):

– *Система безперервної діагностики та пом'якшення наслідків* (CDM). Вона збирає інформацію про поточний стан корпоративного активу та застосовує оновлення до конфігурації та компонентів програмного забезпечення. Корпоративна система CDM надає механізму політики (PE) інформацію про актив, який надсилає запит на доступ, наприклад, чи працює на ньому відповідна виправлена операційна система, чи забезпечується цілісність дозволених для використання на підприємстві програмних компонентів або присутні недозволені компоненти і чи немає в активу відомих вразливостей. Системи CDM також відповідають за ідентифікацію та потенційне застосування підмножини політик на некорпоративних пристроях, активних в інфраструктурі підприємства.

– *Система відповідності вимогам галузі* (*Industry compliance system*). Гарантується, що підприємство буде відповідати всім нормативним вимогам, під які воно може підпадати (наприклад, вимоги щодо безпеки інформації в галузі охорони здоров'я або фінансової галузі). Сюди входять усі правила політики, які підприємство розробляє задля забезпечення відповідності.

– *Інформаційний(і) канал(и) про загрози (Threat intelligence feed(s)).* Це інформація з внутрішніх або зовнішніх джерел, яка допомагає механізму політики приймати рішення про доступ. Це можуть бути кілька служб, які отримують дані з внутрішніх і/або кількох зовнішніх джерел і надають інформацію про нещодавно виявлені атаки або вразливості, а також нещодавно виявлені недоліки в програмному забезпеченні, нещодавно виявлені шкідливе програмне забезпечення, зареєстровані атаки на інші активи, до яких механізм політики захоче заборонити доступ із корпоративних активів.

– *Журнали мережевої та системної активності (Network and system activity logs).* Ця корпоративна система об'єднує журнали активів, мережевий трафік, дії з доступу до ресурсів та інші події, які дозволяють в режимі реального часу (або майже реального часу) отримувати інформацію про стан безпеки корпоративних інформаційних систем.

– *Політики доступу до даних (Data access policies).* Це атрибути, правила та політики доступу до ресурсів підприємства. Цей набір правил може бути закодований (через інтерфейс керування) або динамічно згенерований механізмом політик. Ці політики є відправною точкою авторизації доступу до ресурсу, оскільки вони надають основні привілеї доступу для облікових записів і застосунків / сервісів на підприємстві. Ці політики мають ґрунтуватися на певних цільових ролях та потребах організації.

– *Інфраструктура відкритих ключів підприємства (Public Key Infrastructure – PKI).* Ця система відповідає за створення та реєстрацію сертифікатів, виданих підприємством ресурсам, суб'єктам, службам та застосункам. Вона також включає глобальну екосистему центрів сертифікації та національну PKI, яка може бути інтегрована або не інтегрована із корпоративною PKI. Це також може бути PKI, що не базується на сертифікатах X.509.

– *Система керування ідентифікацією (ID management system).* Вона відповідає за створення, зберігання та керування обліковими записами корпоративних користувачів та ідентифікаційними записами (наприклад, сервер полегшеного протоколу доступу до каталогів – LDAP). Ця система містить необхідну інформацію/дані про суб'єкт (наприклад, ім'я, адресу електронної пошти, сертифікати) та інші характеристики, такі як роль, атрибути доступу та призначені активи підприємства. Вона часто використовує інші системи (наприклад, PKI) для артефактів (об'єктів), пов'язаних з обліковими записами користувачів. Крім того вона може бути частиною більшої об'єднаної спільноти і може включати некорпоративних співробітників або посилання на некорпоративні активи для спільної роботи.

– *Система керування інформацією та подіями безпеки* (*SIEM* – це область комп'ютерної безпеки, в якій програмні продукти та послуги поєднують у собі управління інформацією про безпеку (*SIM – security information management*) та управління подіями безпеки (*SEM – security event management*)). Тут збирається інформація, пов'язана з безпекою, для подальшого аналізу. Ці дані потім використовуються для уточнення політик та попередження про можливі атаки на активи підприємства.

Усі перелічені вище логічні компоненти не обов'язково повинні бути унікальними системами. Один актив може виконувати функції кількох логічних компонентів, і так само логічний компонент може складатися з кількох апаратних або програмних елементів для виконання завдань. Наприклад, інфраструктура відкритих ключів, керована підприємством, може складатися з одного компонента, що відповідає за видачу сертифікатів для пристроїв, та іншого компонента, що використовується для видачі сертифікатів кінцевим користувачам, але обидва використовують проміжні сертифікати, видані одним і тим самим корпоративним кореневим центром сертифікації підприємства. У деяких продуктах, що підтримують концепцію нульової довіри та представлені сьогодні на ринку, компоненти PE і PA об'єднані в одну службу.

1.2. Підходи до реалізації архітектури нульової довіри

На думку фахівців NIST, архітектура нульової довіри може різнитися залежно від потреб компанії. Для цього вони розглядають можливість застосування кількох різних підходів (шляхів), за допомогою яких підприємство може запровадити ZTA для робочих процесів, зокрема: *вдосконалене/покращене управління ідентифікацією* (*enhanced identity governance*), *логічну мікросегментацію* (*logical micro-segmentation*) та *сегментацію на основі мережі* (*network-based segmentation*) [8]. Ці підходи відрізняються компонентами, що використовуються, і основним джерелом правил політики для організації. При цьому слід зазначити, що одні підходи найбільше підходять для одних випадків, тоді як інші доцільно використовувати в інших ситуаціях. Тому підприємство, яке прагне розробити ZTA, може виявити, що обраний ним варіант використання та існуючі політики виділяють один підхід серед інших існуючих. Однак це не означає, що інші підходи не працюватимуть. Зважаючи на все, це лише вказує на те (і не більше), що інші підходи можуть бути більш важкими для реалізації і можуть вимагати кардинальніших змін в організації розвитку бізнесу.

Архітектура нульової довіри з вдосконаленим управлінням ідентифікацією

Підхід до розробки архітектури нульової довіри, що ґрунтується на *вдосконаленому управлінні ідентифікацією*, використовує ідентичність (*identity*) учасників як ключовий компонент створення політики. У рамках цього підходу політики доступу до ресурсів підприємства ґрунтуються на ідентифікації та встановлених атрибутах. Основна вимога для отримання доступу до ресурсів базується на привілеях доступу, наданих даному суб'єкту.

Інші фактори, такі як використовуваний пристрій, стан активів і фактори середовища, можуть змінити остаточний результат визначення рівня довіри (і остаточний дозвіл на доступ) або будь-яким чином скоригувати результат, наприклад, надати тільки частковий доступ до заданого джерела даних в залежності від розташування в мережі. Окремі ресурси або компоненти PEP, що захищають ресурс, повинні мати можливість перенаправляти запити до служби механізму політики або автентифікувати суб'єкт і схвалити запит перед наданням доступу. Підходи, що ґрунтуються на вдосконаленому управлінні ідентифікацією, для підприємств часто застосовуються з використанням моделі відкритої мережі або корпоративної мережі з доступом сторонніх користувачів («гостей») або частою присутністю в мережі пристроїв, що не належать підприємству (некорпоративних пристроїв). Спочатку доступ до мережі надається всім активам, але доступ до корпоративних ресурсів обмежується обліковими записами з відповідними привілеями доступу. При цьому надання базових можливостей підключення до мережі має також і недолік: зловмисники все одно можуть спробувати провести розвідку мережі та/або використовувати її для атак типу «відмова в обслуговуванні» як усередині компанії, так і проти третьої сторони. Підприємствам, як і раніше, необхідно відстежувати таку поведінку та реагувати на неї перш, ніж вона вплине на робочі процеси.

Архітектура нульової довіри з використанням мікросегментації

Підприємство може вибрати реалізацію архітектури нульової довіри, що базується на розміщенні окремих ресурсів або груп ресурсів в унікальному сегменті мережі, захищеному компонентом безпеки шлюзу. У цьому підході підприємство розміщує інфраструктурні пристрої, такі як інтелектуальні комутатори (або маршрутизатори) або міжмережеві екрани наступного покоління або спеціальні шлюзи, які діють як PEP, захищаючи кожен ресурс або невелику групу пов’язаних ресурсів. Як альтернатива (або додатково) підприємство може вибрати мікросегментацію на основі хоста за допомогою програмних агентів або брандмауерів на активі(ах) кінцевої точки. Ці шлюзові пристрої динамічно надають доступ до окремих запитів від клієнта, активу або сервісу. Залежно від моделі, шлюз може бути єдиним компонентом PEP або частиною багатокомпонентного PEP, що складається зі шлюзу та агента на стороні клієнта. Для повноцінного функціонування цього підходу потрібна програма/система управління ідентифікацією (identity governance program – IGP), але при цьому передбачається, що компоненти шлюзу будуть виступати в ролі PEP, що захищає ресурси від несанкціонованого доступу та/або виявлення. Ключова необхідність цього підходу полягає в тому, щоб компоненти PEP керувалися та мали можливість реагувати та змінювати конфігурацію за потреби для відповіді на загрози або зміни в робочому процесі. Можна реалізувати деякі функції мікросегментованого підприємства за допомогою менш просунутих шлюзів і навіть міжмережевих екранів без збереження стану, але витрати на адміністрування та труднощі швидкої адаптації до змін роблять такий вибір вкрай небажаним.

Архітектура нульової довіри з використанням мережевої інфраструктури та програмно визначених периметрів

Даний підхід використовує мережеву інфраструктуру для реалізації архітектури нульової довіри. Реалізація архітектури нульової довіри може бути досягнута за допомогою оверлейної мережі (overlay network). Подібні підходи іноді називають підходами програмно визначеного периметра (*software defined perimeter* – SDP) [24, 25] і часто включають концепції програмно визначених мереж (*Software Defined Networks* – SDN) [26] і мереж IBN (intent-based networking – іноді перекладаються як мережа на основі намірів, інтенційно-орієнтована мережа) [27]. У цьому підході адміністратор політики виступає у ролі мережевого контролера, який налаштовує та реконфігурує мережу на основі рішень, прийнятих механізмом політики. Клієнти продовжують запитувати доступ через точки застосування політики, якими керує компонент PA. Коли підхід реалізується на мережевому рівні застосунків (тобто на рівні 7 моделі OSI – *Open Systems Interconnection*), найпоширенішою моделлю розгортання є модель розгортання на основі агента пристрою/шлюзу (яка буде розглянута далі). У цій реалізації агент і шлюз ресурсу (діють як єдиний PEP та налаштовані PA) встановлюють за-

хищений канал, що використовується для зв'язку між клієнтом та ресурсом. Можливі інші варіанти цієї моделі, наприклад, для хмарних віртуальних мереж, мереж, не заснованих на IP-протоколах тощо.

З іншого боку, доцільним є розгляд потенційних рішень з нульовою довірою з погляду їх моделей розгортання, які можуть бути корисною основою, за допомогою якої підприємства зможуть оцінити потенційних постачальників відповідних рішень, аналізуючи їх плюси та мінуси. Вважається, що багато з корпоративних моделей нульової довіри, що надаються постачальниками, будуть відповідати одній або декільком моделям розгортання. Такі моделі дозволяють оцінити, як насправді можуть бути розгорнуті системи з нульовою довірою, хоча, звичайно ж, реальна архітектура розгортання залежатиме від можливостей обраної технології.

Існує кілька варіантів розгортання вибраних компонентів архітектури. Залежно від того як налаштовано корпоративну мережу для різних бізнес-процесів на одному підприємстві можуть використовуватися кілька моделей розгортання архітектури нульової довіри. Розглянемо деякі відомі моделі розгортання нульової довіри.

1.3. Моделі розгортання нульової довіри

Насамперед, слід звернути увагу на те, що для простоти (зручності) подальшого викладу матеріалу на всіх рисунках, наведених далі та асоційованих з відповідними моделями розгортання, будуть опущені пов'язані з PDP системи (такі як: CDM, система відповідності вимогам галузі, система керування ідентифікацією тощо), які були представлені раніше на рис. 2. Хоча реально всі ці зв'язки PDP з відповідними логічними компонентами архітектури нульової довіри, як і раніше, залишаються актуальними і присутні незалежно від того, яка модель розгортання з нульовою довірою буде обрана та використовуватися для конкретної реалізації.

Модель розгортання на основі агента пристрою/шлюзу

Модель агента пристрою/шлюзу (*Device Agent/Gateway Model*; іноді цю модель називають моделлю розгортання на основі ресурсів – *resource-based deployment model* [7]) представлена на рис. 3.

Рис. 3. Модель розгортання на основі агента пристрою/шлюзу

У даній моделі розгортання зазвичай у системі суб'єкта є розгорнутий користувальницький агент, що діє як PEP агента користувача. Крім того, існує вбудований PEP (шлюз), який розгортається (відповідно до бачення NIST [8]) на ресурсі або у вигляді компонента безпосередньо перед ресурсом. Наприклад, на кожному корпоративному активі, наданому підприємством, встановлений агент пристрою, який координує з'єднання, а на кожному ресурсі є компонент (шлюз), який розміщується безпосередньо перед ресурсом, так що ресурс взаємодіє тільки зі шлюзом (по суті, виступаючи як проксі для ресурсу). Агент направляє частину або весь потік даних на відповідний PEP обробки запитів. Шлюз відповідає за взаємодію з адміністратором політики та дозволяє лише узгоджені комунікаційні шляхи, налаштовані PA. Наприклад, у типовій ситуації суб'єкт із ноутбуком (як деяким активом виданим підприємст-

вом) хоче підключитися до корпоративного ресурсу (наприклад, до застосунку/бази даних відділу кадрів). Запит на доступ приймається локальним агентом, який потім перенаправляється адміністратору політики. Адміністратор політики та механізм політики можуть бути локальним ресурсом підприємства або хмарною службою. Адміністратор політики передає запит до механізму політики для аналізу. Якщо запит авторизований, адміністратор політики настроює канал зв'язку між агентом пристрою та відповідним шлюзом ресурсів через площину керування. Для цього може використовуватися така інформація, як адреса інтернет-протоколу (IP), інформація про порт, ключ сеансу та інші параметри безпеки. Після чого агент пристрою та шлюз з'єднуються, і починається передача зашифрованих потоків даних застосунків/служб. З'єднання між агентом пристрою та шлюзом ресурсів переривається або після завершення робочого процесу, або з ініціативи адміністратора політики – через порушення безпеки (наприклад, закінчення часу очікування сеансу, неможливості повторної автентифікації тощо). Цю модель доцільніше використовувати для підприємств, які мають ефективну систему керування пристроями, а також окремі ресурси, які можуть взаємодіяти зі шлюзом [8]. Наприклад, для підприємств, які активно використовують хмарні сервіси, – це клієнт-серверна реалізація програмно-визначуваного периметра (SDP) альянсу з безпеки хмарних обчислень (*Cloud Security Alliance* – CSA) [28, 29]. Ця модель також підходить для підприємств, які не хочуть вводити політику BYOD (оскільки в ній доступ можливий лише через агента пристрою, розміщеного на активах, що належать підприємству).

Як зазначається у роботі [7], дана модель має такі переваги: комплексний (наскрізний – end-to-end) контроль доступу до застосунків та мережевого трафіку; дуже компактна неявна зона довіри, яка знаходиться за шлюзом.

У цій же роботі вказується й на недоліки, притаманні цій моделі, зокрема:

– необхідність розгортання PEP як на пристроях, так і на ресурсах;

– існує ймовірність технічних конфліктів між компонентами ресурсів та PEP;

– PEP повинні бути доступні для розгортання на різних, можливо, застарілих / успадкованих ОС;

– існує ймовірність негативного відношення (сприйняття) з боку власників ресурсів застосунків (власники можуть не захотіти розгортати будь-яке додаткове програмне забезпечення у своїх комерційних або критично важливих для бізнесу застосунках);

– потрібні зв'язки один до одного: між PEP і ресурсами (тобто ця модель вимагає розгортання одного PEP для кожного керованого ресурсу, що може накласти значне функціональне навантаження (навантаження з керування) на систему нульової довіри і команду (робочу групу), що відповідає за неї;

– наскрізний захищений тунель може перешкоджати (заважати) нормальному функціонуванню існуючих вбудованих засобів контролю безпеки. У багатьох системах нульової довіри мережевий трафік від PEP агента користувача до PEP ресурсу шифрується. Це безпечно та ефективно, але зазвичай має побічний ефект, який полягає в тому, що весь цей трафік стає непрозорим для будь-якого посередника. Це вигідно, якщо посередником є зловмисник, але шкідливо, якщо це, наприклад, розгорнутий на підприємстві компонент безпеки, такий як мережева IDS/IPS;

– PEP має бути видимим та доступним віддаленим користувачам. Насправді для розгортання систем з нульовою довірою відповідно до цієї моделі буде потрібна окрема можливість безпечного віддаленого доступу – в ідеальному випадку як частина платформи нульової довіри. Зазвичай комерційні платформи з нульовою довірою вирішують це завдання за допомогою комбінації периферійного PEP і відповідного PEP агента користувача.

Модель розгортання з мікросегментацією

Дана модель розгортання (micro-segmentation deployment model) орієнтована на варіант використання між серверами («сервер-сервер») [7]. Ця модель дозволяє підійти до проблеми з точки зору ресурсів, а не користувачів. Ресурси вважаються основними суб'єктами (нефізичними особами/сутностями – non-person entities – NPE), щодо яких мають бути розроблені та

застосовані політики (рис. 4). Дана модель по суті є варіантом попередньої розглянутої моделі (моделі, що базується на ресурсах, – моделі розгортання на основі агента пристрою/шлюзу). Важливою її відмінністю від попередньої моделі є те, що ресурси насправді є суб'єктами (автентифікованими ідентифікаторами).

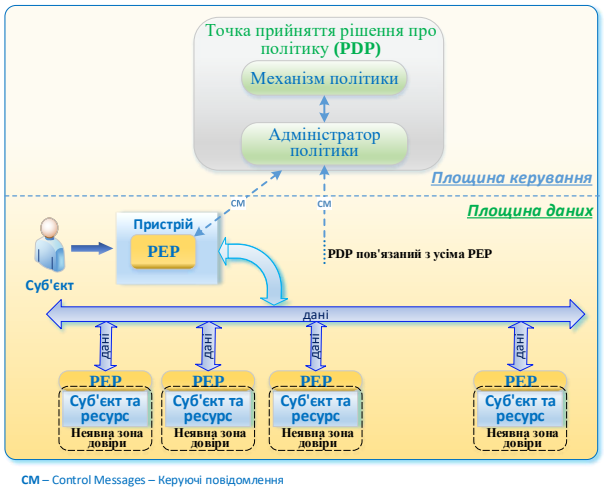


Рис. 4. Модель розгортання із мікросегментацією

Як правило, суб'єкти NPE матимуть більш слабкі форми ідентифікації, ніж люди, а саме, здебільшого на основі сертифікатів і, швидше за все, на основі одного фактора автентифікації. Найчастіше цей сертифікат генерується та керується центром сертифікації підприємства.

Переваги використання цієї моделі [7]: невелика неявна зона довіри (зазвичай обмежена лише самим ресурсом); точний, двонаправлений контроль доступу до ресурсів (для серверів або мікросервісів). Оскільки PEP виконується локально по відношенню до ресурсу, його політики можуть контролювати як вихідні, так і вхідні мережеві з'єднання (ці політики можна застосовувати як до ресурсів на рівні сервера, так і до мікросервісів).

До недоліків цієї моделі можна віднести [7]: необхідність розгортання PEP як на пристроях користувача, так і на ресурсах; наявність ймовірності технічних конфліктів між компонентами ресурсів та PEP; PEP повинні бути доступні для розгортання на множині, можливо, застарілих або успадкованих ОС; наявність можливого негативного відношення з боку власників ресурсів; необхідні взаємозв'язки один до одного між PEP та ресурсами; може не найкращим чином підходити для доступу користувачів до ресурсів; відсутність віддаленого доступу (не передбачено) – потрібний прямий доступ суб'єктів до PEP. Крім того, не слід забувати про потенційний недолік, зумовлений функціональними або архітектурними вадами (недоліками), пов'язаними з реалізацією сценарію «користувач–сервіс» конкретним постачальником або у відкритому вихідному коді.

Модель розгортання на основі анклаву

Ця модель розгортання є також різновидом розглянутої вище моделі агента пристрою/шлюзу. У цій моделі компоненти шлюзу (PEP) знаходяться перед кількома ресурсами, які називаються анклавом ресурсів (resource enclave). Цей набір ресурсів може бути фізично розташований разом (наприклад, у локальному або суміщеному центрі обробки даних) або логічно пов'язаний (наприклад, набір серверів хмари). Як і в попередній моделі, суб'єкт має додатковий локально встановлений PEP агента користувача (рис. 5).

Рис. 5. Модель розгортання на основі анклаву

Важливо розуміти, що у цій моделі неявна зона довіри містить кілька мережевих ресурсів, які, швидше за все, взаємодіють між собою. Тобто дуже важливо, щоб у цій моделі анклав ресурсів працював виключно у логічній приватній мережі, яка перебуває під контролем підприємства. Хоча ресурси всередині анклаву здатні і можуть взаємодіяти один з одним за межами видимості та контролю PEP, єдиний спосіб для суб'єктів поза зоною довіри – взаємодіяти з ним через PEP (який повинен знаходитись під контролем політики). Тобто, використовуючи цю модель, підприємствам необхідно ретельно вивчити дані ресурсів та моделі взаємодії.

Ця модель в цілому простіше в розгортанні, ніж попередня – модель агента пристрою/шлюзу, оскільки в ній задіяно на порядок менше PEP завдяки зв'язку «один до багатьох» між PEP та ресурсами. А відсутність необхідності розгортання додаткового програмного забезпечення на ресурсах не лише спрощує роботу, а й дозволяє уникнути більшості технічних або конфліктів політик із застосунками та їх власниками. Перевага цього рішення полягає також у тому, що PEP розгортаються на межі корпоративної мережі (у DMZ – Demilitarized Zone – демілітаризована зона), тому вони можуть бути природною точкою входу для віддалених користувачів. Зрозуміло, вони також служитимуть точкою застосування політики для локальних користувачів, чий трафік залишиться повністю всередині підприємства. Ідея полягає в тому, щоб PEP могли реагувати на зміни в ресурсах, що захищаються, наприклад, виявляючи поява нових ресурсів і використовуючи атрибути ресурсів (метадані) для застосування до них політик.

Ця модель корисна для підприємств із застарілими / успадкованими (legacy) застосунками або локальними центрами обробки даних, у яких неможливо встановити окремі шлюзи (для встановлення / налаштування агентів пристроїв підприємству необхідна надійна програма (система) управління активами та конфігурацією). Також організації, які працюють у нових середовищах (особливо на базі IaaS – Infrastructure-as-a-Service) або використовують програмно-керовану інфраструктуру (наприклад, DevOps – development and operations), добре підходять для цієї моделі. Організаціям з нижчим рівнем операційної зрілості, меншою прозорістю або складними успадкованими мережами може знадобитися розгорнути більше PEP, щоб зменшити розмір та масштаб кожної неявної зони довіри. Як альтернативу, вони можуть використовувати гібридний підхід, який підтримують деякі постачальники архітектури нульової довіри, поєднуючи цю модель з моделлю мікросегментації.

Недоліком моделі розгортання на основі анклаву є те, що шлюз (PEP) захищає колекцію ресурсів і може бути не в змозі захистити кожен ресурс окремо. Це також може дозволити суб'єктам бачити ресурси, до яких вони не мають прав доступу.

Модель розгортання з використанням хмарної маршрутизації

У наступній моделі (рис. 6), що одержала назву моделі розгортання з використанням хмарної маршрутизації (cloud-routed), весь трафік від суб'єкта проходить через хмарне середовище, перш ніж досягти ресурсу. Ця модель є досить поширеним підходом, який використовується багатьма комерційними виробниками [7].

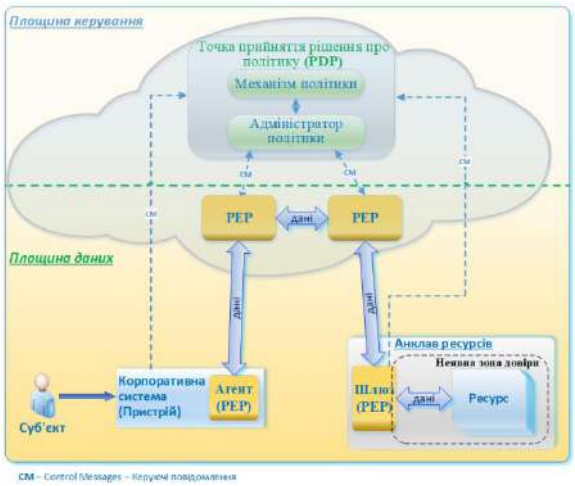


Рис. 6. Модель розгортання з використанням хмарної маршрутизації

Як видно з рис. 6, у цій моделі PEP, що знаходяться перед анклавами ресурсів підприємства, діють аналогічно до PEP (шлюзу) у моделі, показаній вище (рис. 5). Однак у цих PEP є одна важлива відмінність – вони не є точкою входу до корпоративної мережі. Замість цього дана функціональність була логічно перекладена на PEP, розташовані в хмарному середовищі постачальника. У цій моделі PEP, розташовані лише на рівні підприємства (локальні – on-premises), виступають у ролі сполучних ланок, забезпечуючи вихідні з'єднання з хмарним PEP. Оскільки ці внутрішньокорпоративні конектори не вимагають будь-яких вхідних з'єднань, вони часто спрощують розгортання цієї моделі в обмін на деякі обмеження. Коли суб'єкт хоче зв'язатися з ресурсом, він спочатку проходить автентифікацію в PDP, а потім його потік даних спрямовується на один з хмарних PEP, зазвичай, той, який розташований найближче до нього (або має найменшу затримку). Потім відповідний потік даних проходить через PEP у хмарі до PEP, який має з'єднання з цільовим анклавом ресурсів. Локальний (розташований на підприємстві) PEP забезпечує безпеку анклаву ресурсів так само, як і у моделі розгортання на основі анклаву (рис. 5).

Як зазначається у роботі [7], дана модель має певні переваги та недоліки (табл. 1).

Таблиця 1

Переваги та недоліки моделі розгортання з використанням хмарної маршрутизації

Переваги	Недоліки
✓ Більш просте налаштування для підприємств. ✓ Платформа «як послуга» (As-a-Service) знижує операційні витрати підприємства. ✓ Постачальники, які використовують цю модель, також для захисту доступу до загальнодоступних веб-сайтів надають послугу SWG (Secure Web Gateway – безпечний / захищений веб-шлюз).	✓ PEP можуть бути розгорнуті без належного контролю за безпекою, мережею та дотриманням нормативних вимог. ✓ Збільшує затримку трафіку користувача і потенційно знижує пропускну здатність. ✓ Підтримує лише обмежену кількість мережевих протоколів. ✓ Не підходить для локальних користувачів, які отримують доступ до локальних ресурсів. ✓ Потенційно велика, непрозора або шумна неявна зона довіри.

Модель розгортання на основі порталу ресурсів

У цій моделі розгортання (рис. 7) точка застосування політики (PEP) є окремим компонентом, що виконує роль шлюзу для запитів суб'єктів. Портал-шлюз може бути призначений для окремого ресурсу або безпечного анклаву для сукупності ресурсів, що використовуються для виконання однієї бізнес-функції. Як приклад можна навести портал-шлюз у приватній хмарі або центрі обробки даних, що містить успадковані застосунки.

Основна перевага цієї моделі перед іншими полягає в тому, що програмний компонент (PEP) не потрібно встановлювати на всі пристрої клієнта. До того ж, ця модель гнучкіша для політики BYOD та проєктів спільної роботи між організаціями.

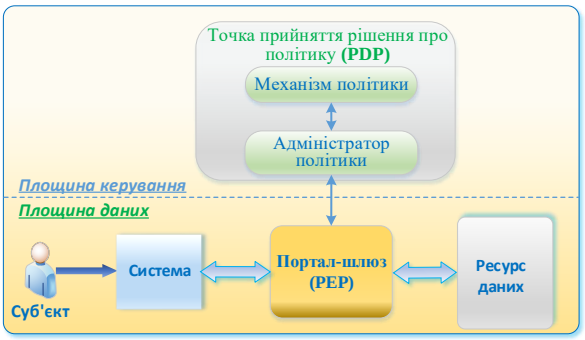


Рис. 7. Модель розгортання на основі порталу ресурсів

Корпоративні адміністратори не повинні стежити за тим, щоб на кожному пристрої перед використанням було встановлено відповідний локальний агент – агент пристрою (PEP агента користувача). Хоча від пристроїв, які вимагають доступ, можна отримати певну інформацію. Крім того, у цій моделі сканування та аналіз активів та пристроїв може здійснюватися тільки після їх підключення до порталу PEP, а постійний моніторинг щодо наявності шкідливих програм, невиправлених вразливостей та відповідної конфігурації може бути неможливим. А оскільки локальний агент, який обробляє запити, відсутній, то на корпоративному рівні може не бути повної прозорості або можливості здійснювати довільний контроль над активами (оскільки їх можна побачити / просканувати лише при підключенні до порталу). Хоча для пом'якшення або компенсації наслідків на підприємстві можуть застосовуватися такі заходи, як ізоляція браузера (browser isolation – це модель кібербезпеки, метою якої є фізична ізоляція активності/дій користувачів Інтернету (і пов'язаних з ними кіберризиків) від їх локальних мереж та інфраструктури).

Як недолік слід зазначити, що використання цього рішення також дозволяє зловмисникам виявляти портал і використовувати його для доступу до нього з метою проведення атаки на кшталт «відмова в обслуговуванні» (DoS – denial-of-service). У зв'язку з чим системи порталу повинні мати достатні ресурси, щоб забезпечити доступність при DoS-атаках або збоях у роботі мережі.

Модель розгортання на основі використання «пісочниці» застосунків

Ще один варіант моделі розгортання на основі агента пристрою/шлюзу полягає в тому, що перевірені / схвалені застосунки або процеси запускаються на активах в ізольованих середовищах/системах (пісочниці – sandboxing). Ці системи можуть бути віртуальними машинами, контейнерами, модулями довіреної платформи та іншими реалізаціями. Вони мають одну мету – захистити застосунок або екземпляри застосунків від можливо скомпрометованого хоста або інших застосунків, що працюють на цьому активі / хості [8]. На рис. 8 пристрій суб'єкта запускає попередньо узгоджені та перевірені застосунки у «пісочниці».

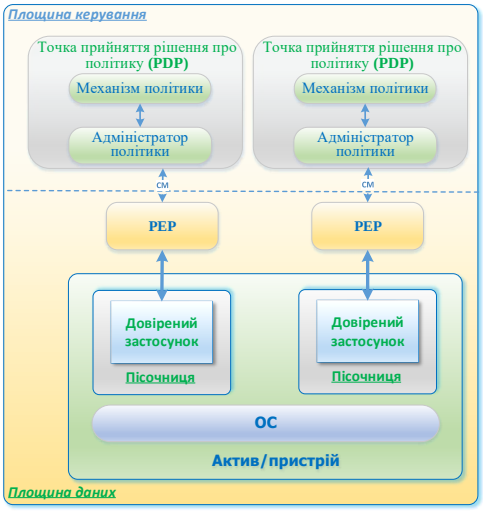


Рис. 8. Модель розгортання на основі використання «пісочниці» застосунків

Застосунки можуть взаємодіяти з PEP для запиту доступу до ресурсів, але PEP відхилятиме запити від інших застосунків активу. Ця модель PEP може бути локальною службою підприємства або хмарною службою.

Основна перевага цього варіанта моделі полягає в тому, що деякі застосунки відокремлені від решти активу. Якщо актив не може бути просканований на предмет вразливостей, ці окремі застосунки, поміщені в «пісочницю» (так звані ізольовані застосунки), можуть бути захищені від потенційного зараження шкідливим програмним забезпеченням (ПЗ) на хості активу.

Одним із недоліків цієї моделі є те, що підприємства повинні підтримувати ці ізольовані застосунки для всіх активів і можуть не мати повної видимості клієнтських активів. Крім того, підприємству необхідно переконатися в безпеці кожного ізольованого застосунку, що може вимагати більше зусиль, ніж простий моніторинг пристроїв.

Таким чином, очевидно, що всі розглянуті моделі можуть дати новий рівень деталізації того, як насправді можуть бути розгорнуті системи з нульовою довірою. Хоча при цьому не слід забувати, що реальна архітектура розгортання, звичайно ж, залежатиме від можливостей обраної технології. Також було б некоректним під час розгляду парадигми нульової довіри не відзначити можливість існування певних ризиків, пов'язаних із використанням архітектури нульової довіри.

1.4. Загрози, пов’язані з архітектурою нульової довіри

Як відомо, жодне підприємство не може усунути ризик кібербезпеки [8]. І це, незважаючи на належно реалізовані та підтримувані архітектурою нульової довіри вказівки щодо кібербезпеки, управління ідентифікацією та доступом, безперервним моніторингом та загальною кібергігієною, що дозволяють зменшити загальний ризик та захистити від поширених загроз. Причому деякі загрози мають унікальні особливості під час впровадження архітектури нульової довіри. Тому далі розглянемо найбільш відомі загрози, пов’язані з архітектурою нульової довіри, та деякі рекомендації щодо їх усунення.

Спотворення процесу прийняття рішень у рамках архітектури нульової довіри

В архітектурі нульової довіри механізм політики (PE) та адміністратор політики (PA) є ключовими компонентами всього підприємства. Жодної взаємодії між ресурсами підприємства не відбувається, якщо вона не схвалена і, можливо, не налаштована механізмом політики та адміністратором політики. Це означає, що ці компоненти необхідно правильно налаштувати, постійно контролювати (будь-які зміни конфігурації повинні реєструватися та підлягати аудиту) та підтримувати у відповідному стані. В іншому випадку будь-який адміністратор підприємства, який має доступ до конфігурації правил PE, може внести несхвалені зміни або зробити помилку, яка може порушити роботу підприємства, або скомпрометований PA може дозволити доступ до ресурсів, які в інших випадках (наприклад, до зламаного особистого пристрою) не були б схвалені.

Відмова в обслуговуванні або порушення роботи мережі

Як було зазначено вище, в архітектурі нульової довіри адміністратор політики є ключовим компонентом забезпечення доступу до ресурсів, тому ресурси підприємства не можуть підключатись один до одного без дозволу PA і, можливо, без дій щодо налаштування. Якщо зловмисник порушує або забороняє доступ до PEP або PE/PA (тобто DoS-атака або захоплення каналу), це може негативно вплинути на роботу підприємства. Підприємства можуть знизити цю загрозу, розташувавши механізм застосування політик у належно захищеному хмарному середовищі або реплікувавши її в кількох місцях відповідно до рекомендацій щодо забезпечення кіберстійкості [30]. Це зменшує ризик, але не усуває його. Наприклад, ботнети, подібні до Mirai (це сімейство шкідливих програм, схожих на хробаків, які заражали IoT-пристрої та об'єднували їх у DDoS-ботнет) [31–33] здійснюють масовані DoS-атаки на основних інтернет-провайдерів і порушують роботу мільйонів інтернет-користувачів. Також існує ймовірність того, що зловмисник може перехопити та заблокувати трафік до PEP або

PA для частини або всіх облікових записів користувачів на підприємстві. Наприклад, у філії або навіть в одного віддаленого співробітника (у таких випадках страждає лише частина суб'єктів підприємства). Хоча, насправді, таке можливе і у застарілих VPN з віддаленим доступом і не є унікальним тільки для архітектури нульової довіри. З іншого боку хостинг-провайдер може випадково перевести хмарний механізм політики або адміністратора політики в режим offline. У цьому випадку механізм політики або компонент адміністратора політики стане недоступним із мережі. І як результат – через неполадки в роботі, може припинитися робота всього підприємства. Крім того, існує також ризик, що ресурси підприємства можуть бути недоступні для адміністратора політики, тому навіть якщо суб'єкту буде надано доступ, адміністратор політики не зможе налаштувати канал обміну даними через мережу. Це може статися в результаті DDoS-атаки або просто через надмірну інтенсивність використання, що раптово виникла. Все це аналогічно до будь-якого іншого порушення роботи мережі, коли деякі або всі суб'єкти підприємства не можуть отримати доступ до певного ресурсу, оскільки він з якоїсь причини недоступний.

Крадіжка облікових даних

Відомо, що розроблена та реалізована архітектура нульової довіри повинна перешкоджати доступу скомпрометованого облікового запису або активу до ресурсів. З іншого боку, правильно реалізовані політики нульової довіри, інформаційної безпеки та стійкості до відмов, а також кращі практики знижують ризик отримання зловмисником широкого доступу до ресурсів підприємства через вкрадені облікові дані або інсайдерську атаку. Тому, виходячи зі всього сказаного, з великою часткою ймовірності, можна припустити, що основною метою для зловмисників будуть відповідні облікові записи з політиками доступу до ресурсів. Оскільки зловмисникам, щоб впровадитись на підприємство, необхідно скомпрометувати існуючий обліковий запис або пристрій. З цією метою зловмисники можуть використовувати фішинг, соціальну інженерію або комбінацію атак, щоб отримати повноваження важливих облікових записів. При цьому застосування MFA (multi-factor authentication – багатофакторна автентифікація) для обробки запитів на доступ може знизити ризик втрати інформації внаслідок компрометації облікового запису. Однак зловмисник з валідними обліковими даними (або зловмисник-інсайдер) все одно зможе отримати доступ до ресурсів, до яких цьому обліковому запису вже було надано права доступу. Наприклад, зловмисник або скомпрометований співробітник, який має облікові дані та активи дійсного співробітника відділу кадрів, що належать підприємству, все одно зможе отримати доступ до бази даних працівників.

У свою чергу, використання архітектури нульової довіри все ж таки дозволяє знизити ризик і запобігати бічному переміщенню будь-яких скомпрометованих облікових записів або активів по мережі. Якщо скомпрометовані облікові дані не авторизовані для доступу до певного ресурсу, їм, як і раніше, буде відмовлено у доступі до цього ресурсу. Крім того, контекстний алгоритм довіри з більшою ймовірністю виявить цю атаку і швидко відреагує на неї, ніж при атаці в застарілій (успадкованій) мережі, що базується на периметрі. Контекстний алгоритм довіри може виявити шаблони доступу, які виходять за рамки звичайної поведінки, та заборонити скомпрометованому обліковому запису або внутрішній загрозі доступ до конфіденційних (чутливих, важливих, критичних – sensitive) ресурсів.

Видимість у мережі

Для виявлення та реагування на потенційні атаки на ресурси підприємства весь трафік у мережі має перевірятися, реєструватися та аналізуватись [8]. Однак частина трафіку в мережі підприємства може бути непрозорою для засобів мережевого аналізу на рівні 3 (мережевому рівні) моделі OSI. Цей трафік може походити від активів, що не належать підприємству (наприклад, від сторонніх служб, які використовують інфраструктуру підприємства для доступу до Інтернету), або від застосунків / сервісів, які не піддаються пасивному моніторингу. Тому підприємство, яке не може виконати ретельну перевірку пакетів або вивчити зашифрований трафік – змушене використовувати інші методи для оцінки можливої присутності зловмисника у мережі. Наприклад, на підприємстві можна збирати метадані (адреси джерела

та призначення тощо) про зашифрований трафік і використовувати їх для виявлення активного зловмисника або можливого шкідливого ПЗ, що працює в мережі. Крім того, для аналізу трафіку, який неможливо розшифрувати та вивчити, можна використовувати методи машинного навчання. Зокрема, автори роботи [34], проаналізувавши шість поширених алгоритмів машинного навчання, показали, як кожен із них вирішує проблему виявлення шкідливих зашифрованих мережевих сеансів. Тому використання машинного навчання потенційно дозволяє підприємству класифікувати трафік як дійсний або, навпаки, шкідливий та той що потребує ліквідації наслідків.

Зберігання системної та мережевої інформації

Наступна загроза для корпоративного моніторингу та аналізу мережного трафіку – це сам компонент аналізу. Якщо результати сканування монітора, мережевий трафік та метадані зберігаються для побудови контекстних політик, проведення експертизи або подальшого аналізу, ці дані стають об'єктом атаки зловмисників. Подібно до мережевих діаграм, конфігураційних файлів та інших документів мережевої архітектури, ці ресурси повинні бути захищені. Якщо зловмиснику вдасться отримати доступ до цієї інформації, він зможе отримати уявлення про архітектуру підприємства та визначити активи для подальшої розвідувальної діяльності та атак.

Ще одним джерелом інформації для зловмисника на підприємстві з нульовою довірою є інструмент керування, який використовується для кодування політик доступу. Як і трафік, що зберігається, цей компонент містить політики доступу до ресурсів і може дати зловмиснику інформацію про те, які облікові записи найбільш цінні для компрометації (наприклад, ті, які мають доступ до ресурсів даних, що цікавлять зловмисника). Що стосується всіх важливих корпоративних даних, то для запобігання несанкціонованому доступу та спробам доступу до них також необхідно забезпечити належний захист. Оскільки ці ресурси є критично важливими для безпеки, вони повинні мати найсуворіші політики доступу та бути доступними лише зі встановлених або спеціальних облікових записів адміністраторів.

Залежність від пропрієтарних форматів даних або рішень

При прийнятті рішень про доступ архітектура нульової довіри спирається на кілька різних джерел даних, включаючи інформацію про запитуючого суб'єкта, використовувані активи, корпоративні та зовнішні джерела інформації, а також дані із аналізу існуючих загроз. Найчастіше активи, що використовуються для зберігання та обробки цієї інформації, не мають загального, відкритого стандарту взаємодії та обміну інформацією. Це може призвести до того, що через проблеми із сумісністю підприємство буде прив'язане до певного кола постачальників. Якщо у одного з постачальників виникають проблеми з безпекою або збої в роботі, підприємство не зможе перейти до нового постачальника без значних витрат (наприклад, заміни кількох активів) або тривалої процедури перетворень (наприклад, перетворення правил політики з одного пропрієтарного формату на інший). Тому щоб знизити пов'язані з цим ризики, підприємствам слід оцінювати постачальників послуг комплексно, враховуючи такі фактори, як контроль безпеки постачальника, витрати підприємства на його переоснащення та управління ризиками в ланцюжку поставок, безумовно, на додаток до більш характерних показників, таких як продуктивність, стабільність тощо.

Використання сутностей, які не є фізичними особами, при адмініструванні архітектури нульової довіри

Зважаючи на те, що штучний інтелект та інші програмні агенти все частіше використовуються для управління безпекою в корпоративних мережах, вони (як відповідні компоненти) повинні взаємодіяти з компонентами управління архітектури нульової довіри (наприклад, механізмом політики, адміністратором політики), причому іноді замість людини-адміністратора. Однак питання про те, як ці компоненти проходять автентифікацію на підприємстві, що запроваджує ZTA, залишається відкритим.

У більшості автоматизованих інформаційних систем передбачається, що при використанні програмного інтерфейсу застосунку (API – application programming interface) для

доступу до компонентів ресурсів використовуватимуться ті чи інші засоби автентифікації. Пов'язаний із цим ризик полягає в тому, що зловмисник зможе спонукати або змусити NPE виконати якесь завдання, на виконання якого він не має привілеїв. Програмний агент може мати більш низькі вимоги до автентифікації (наприклад, ключ API порівняно з MFA) для виконання адміністративних завдань або завдань, пов'язаних із безпекою, порівняно з людиною-користувачем. Якщо зловмисник може взаємодіяти з агентом, він теоретично може обманом змусити агента надати йому ширший доступ або виконати якесь завдання від імені зловмисника. Існує також ризик, що зловмисник може отримати доступ до облікових даних програмного агента та видати себе за нього під час виконання завдань.

При цьому слід мати на увазі, що найбільший ризик при використанні автоматизованих технологій для конфігурування та застосування політик – це можливість помилкових спрацьовувань (нешкідливі дії, прийняті за атаки) та помилкових заперечень (атаки, прийняті за нормальну діяльність), що впливають на рівень безпеки підприємства. Зменшити цей ризик можна за допомогою регулярного проведення додаткового аналізу для виправлення помилкових рішень та покращення процесу прийняття рішень.

Таким чином, можна зробити висновок, що реалізація архітектури нульової довіри – це досить складний процес, причому це швидше за все шлях (рух), ніж повна заміна інфраструктури або технологічних процесів. Організація повинна прагнути до поступового впровадження принципів нульової довіри, змін у процесах та технологічних рішень, які захищають її найцінніші активи даних. Більшість підприємств, швидше за все, продовжуватимуть працювати у комбінованому режимі з використанням нульової довіри та периметра протягом невизначеного періоду часу, продовжуючи вкладати кошти у постійну модернізацію IT [8]. При цьому наявність плану модернізації IT, що включає перехід до архітектури, яка базується на принципах ZT, може допомогти підприємству сформувати дорожні карти для здійснення невеликих переходів на нові робочі процеси. Зрештою, те, як підприємство переходитиме на цю концепцію, залежатиме від його поточного стану кібербезпеки та операційної діяльності. Причому підприємство має досягти базового рівня підготовленості (базовий рівень включає визначення та класифікацію активів, суб'єктів, бізнес-процесів, потоків даних і відображення залежностей для підприємства), перш ніж стане можливим розгорнути масштабну систему, орієнтовану на нульову довіру. Підприємству необхідна ця інформація, щоб визначити список бізнес-процесів-кандидатів та суб'єктів/активів, які будуть залучені до цього процесу. При цьому, найбільшою проблемою, що перешкоджає впровадженню успішних рішень в області нульової довіри, на думку фахівців ACT-IAC (American Council for Technology-Industry Advisory Council – Американська рада з технологій – Консультативна рада з питань промисловості), може бути загальний недостатній рівень кібербезпеки [35]. Ними зазначається, наприклад, що більшість державних установ не мають фундаментальних основ (таких як, політик, процесів та інструментів), необхідних для розгортання систем, що відповідають концепції нульової довіри.

В цілому ж, основними перевагами впровадження концепції нульової довіри, у тому числі виділеними авторами робіт [21, 36], можна вважати такі:

– Поліпшення видимості мережі, виявлення порушень та керування вразливостями (покращення видимості того, які користувачі, коли, як і звідки отримують доступ до тих чи інших ресурсів).

– Перешкода поширенню шкідливих програм (більш ефективне виявлення, реагування та відновлення після інцидентів, що дозволяє мінімізувати наслідки витоків; мінімізація ризику бокового переміщення).

– Скорочення капітальних та експлуатаційних витрат на безпеку.

– Скорочення обсягу та вартості робіт із забезпечення відповідності нормативним вимогам.

– Усунення необхідності пошуку винних (наприклад, при деяких інцидентах мережева команда може звинуватити службу безпеки у зриві робіт, а служба безпеки може звинуватити

мережеву команду; Zero Trust намагається змусити технологічні підрозділи подолати відповідні протиріччя між різними робочими групами).

– Підвищення обізнаності та розуміння даних.

– Захист ресурсів та активів (запобігання витоку важливих даних до зловмисників (обмеження кількості зломів (витоків) за рахунок зниження можливостей зловмисників); зменшення внутрішньої загрози; захист важливих (sensitive) корпоративних даних за допомогою надійного шифрування як під час їх передачі, так і у стані спокою; виконання динамічної оцінки доступу до ресурсів з урахуванням ризиків шляхом безперервної переоцінки всіх операцій та сеансів доступу, збору інформації, отриманої в результаті періодичної повторної автентифікації та повторної авторизації, постійної перевірки стану пристрою, аналізу поведінки, постійної перевірки стану ресурсів, виявлення аномалій та інших аналітичних даних щодо безпеки).

– Забезпечення цифрової трансформації бізнесу (підтримка віддаленої роботи; надання працездатних пристроїв від постачальників; підвищення якості обслуговування кінцевих користувачів).

2. Рекомендації щодо успішного впровадження архітектури нульової довіри

Спираючись на досвід впровадження систем нульової довіри відомими у світі організаціями та компаніями, викладений у різних авторитетних міжнародних виданнях [7, 8, 21], сформулюємо деякі рекомендації для успішного впровадження архітектури нульової довіри на типовому IT-підприємстві у вигляді послідовності певних кроків/етапів. При цьому доречно відразу ж звернути увагу (знову ж таки, виходячи з досвіду впровадження) на те, що керівництво IT-підприємства має прагнути до поступового впровадження принципів нульової довіри, зміни процесів та технологічних рішень, що захищають його найцінніші активи даних. Такий поступовий підхід дозволить знизити ризик відмов і помилок у системі, допоможе зрозуміти подальші процеси розгортання елементів системи, а також полегшить перехід персоналу до нової архітектури.

Перш ніж розпочинати впровадження архітектуру нульової довіри на підприємстві, необхідно провести дослідження активів, суб'єктів, потоків даних та робочих процесів, оскільки підприємство не в змозі визначити, які нові процеси або системи необхідно запровадити, якщо немає уявлення про поточний стан операційної діяльності. Крім того, щоб підприємство «нульової довіри» могло успішно функціонувати, механізм політики повинен мати знання про суб'єктів підприємства. Суб'єкти можуть включати як людей, так і можливі нефізичні особи/сутності, такі як облікові записи сервісів, які взаємодіють із ресурсами. На рис. 9 подано процес поетапного впровадження архітектури нульової довіри на IT-підприємстві.

Рис. 9. Процес поетапного впровадження архітектури нульової довіри

Етап 1. *Виявлення та інвентаризація активів підприємства*. Перший крок, який мають зробити відповідні співробітники підприємства на шляху до «нульової довіри», – це ідентифікувати всі свої активи, визначивши, які ресурси є в існуючому середовищі (обладнання, програмне забезпечення, застосунки, дані та сервіси). Для цього може знадобитися розгортання інструментів для моніторингу трафіку з метою виявлення активних ресурсів, які використовуються та до яких здійснюється доступ. Необхідно

мати повне уявлення про ресурси підприємства (як локальних, так і хмарних) та провести їх облік (визначити їхню кількість, розташування, поточну захищеність, їх важливість та вплив на роботу підприємства), оскільки це саме ті об'єкти, для захисту яких буде розроблено архітектуру нульової довіри. Крім того, передбачається необхідність ідентифікації та моніторингу пристроїв, що не належать підприємству, які можуть перебувати в мережній інфраструктурі, що належить підприємству, або мати доступ до корпоративних ресурсів. Після створення чіткого списку ресурсів, необхідно визначити їх стан: чи потребують вони заміни на більш нові пристрої та чи мають користувачі проблеми із доступом до них на даний момент. Якщо ресурси не враховані, то, швидше за все, вони не будуть належним чином захищені в рамках архітектури нульової довіри. Вони можуть бути вразливими для витоку, модифікації, видалення, відмови в обслуговуванні або інших типів атак.

Етап 2. *Розробка політик доступу для підтримки завдань та корпоративних сценаріїв використання*. Після того, як співробітниками підприємства будуть визначені всі ресурси, які потрібно захистити, та місце їх розташування, необхідно сформулювати політики, які будуть застосовуватись у рамках архітектури нульової довіри, щоб визначити, кому та за яких умов дозволено доступ до кожного ресурсу. Політики доступу повинні бути розроблені таким чином, щоб дозволи та повноваження на доступ до кожного ресурсу відповідали принципам найменших привілеїв та поділу обов'язків. Це вимагає розуміння категорій користувачів, які отримуватимуть доступ до ресурсів, їх вимог до доступу, місць роботи, умов контракту, типів пристроїв та моделей власності (наприклад, BYOD, корпоративна), оскільки все це впливатиме на створення політики. Дозволи на доступ можуть бути обмежені залежно від місцезнаходження особи, яка запитує доступ, часу доби або інших параметрів, які можуть додатково обмежувати доступ без втручання у роботу підприємства. Усі політики доступу повинні ґрунтуватися на важливості ресурсу, що захищається. У тому числі, враховуючи особливості впровадження багаторівневих політик для архітектури нульової довіри в середовищах застосунків Cloud Native (Cloud Native – це програмний підхід до створення, розгортання та управління сучасними застосунками у середовищах хмарних обчислень [37]), описаних у роботі [38]. При цьому слід враховувати, що підприємства під час розробки політик можуть зіткнутися з певними проблемами. Одна з яких полягає в тому, що архітектура нульової довіри підприємства може складатися з багатьох компонентів, кожен з яких може виконувати функції механізму політики та адміністратора політики. В результаті політика доступу не може бути централізована в одному місці. Правила можуть бути розподілені між численними компонентами, наприклад: деякі правила можуть бути задані в компоненті захисту кінцевих точок, деякі – в компонентах керування ідентифікацією, обліковими даними та доступом, інші правила – у компоненті мережевої безпеки, а треті – у компоненті захисту даних або інших компонентах. Відсутність єдиного місця, де могли б централізовано зберігатись усі правила політики, може ускладнити для підприємства підтримку організованого, повного та послідовного розуміння політики доступу. Щоб допомогти підприємствам керувати політикою доступу, необхідно чітко відстежувати не лише правила доступу, а й те, де кожне з них визначене.

Етап 3. *Визначення існуючих можливостей та технологій забезпечення безпеки*. Якщо підприємство планує впровадити архітектуру нульової довіри у нове середовище, тобто у нього немає існуючого IT-оснащення чи засобів безпеки, які воно хотіло б використати або пристосувати, цей етап не потрібен. Проте більшість підприємств, які впроваджують систему нульової довіри, не починатимуть із нуля. Для них важлива існуюча інфраструктура та технологічні системи, які вже виконують функції безпеки. Як правило, підприємства мають як мінімум мережеві брандмауери та системи виявлення вторгнень для забезпечення безпеки периметра, а також системи керування ідентифікацією та доступом до облікових даних, які дозволяють їм автентифікувати користувачів та забезпечувати авторизований доступ на основі ідентифікаційних даних та ролей. На їх ноутбуках та/або мобільних пристроях можуть бути встановлені системи захисту кінцевих точок, що виконують функції брандмау-

ера та забезпечують роботу необхідного антивірусного або іншого програмного забезпечення. Вони можуть мати інструментальні засоби для управління вразливістю та конфігураціями, ведення журналів та інших функцій, пов'язаних з безпекою. Крім того, вони, швидше за все, мають свого роду центр управління безпекою.

Підприємство повинне визначити та скласти опис існуючих компонентів та можливостей технології безпеки, щоб зрозуміти, які засоби захисту вони вже забезпечують, а потім визначити, чи повинні ці компоненти продовжувати забезпечувати захист у межах розгорнутої архітектури нульової довіри чи їх слід модифікувати (замінити). Щоб заощадити гроші, підприємство буде прагнути продовжувати використовувати або модернізувати якомога більше існуючих технологій, не жертвуючи при цьому безпекою. Продовження використання існуючих технологій вимагатиме від підприємства розуміння того, з якими потенційними компонентами та продуктами нульової довіри буде інтегрована існуюча технологія безпеки. Будь-які додаткові компоненти, які придбаються спеціально для розгортання в рамках ZTA, в ідеалі повинні інтегруватися з компонентами технології безпеки, які організація вже має та планує продовжувати використовувати.

Етап 4. *Усунення недоліків у політиках та процесах забезпечення нульової довіри шляхом застосування підходу, заснованого на оцінці ризиків та цінності даних*. Після того, як співробітники підприємства складуть перелік ресурсів, які необхідно захистити, та наявні можливості щодо забезпечення безпеки, можна приступати до планування відповідної технології захисту доступу, визначаючи, чи буде інфраструктура сегментована і на якому рівні захищатиметься кожен ресурс. Технологія доступу має бути розроблена з використанням підходу, що ґрунтується на оцінці ризику, при якому критично важливі ресурси ізолюються у власних зонах довіри, захищених точками застосування політики, але при цьому допускається спільне використання кількох менш важливих ресурсів в одній зоні довіри. Для успішного захисту ІТ-підприємства треба визначити ризики та надати їм рівні. Рівень небезпечності ризику визначається за допомогою розуміння ймовірності його появи та збитків, які принесе підприємству його реалізація. Після документування ризиків буде зрозуміло, які ресурси є найбільш критично важливими для роботи підприємства та які вразливості вони мають. Ці дії можна порівняти з діями у межах системи керування ризиками (RMF – *Risk Management Framework* [39]), оскільки будь-яке впровадження архітектури нульової довіри – це процес зниження ризику для бізнес-функцій підприємства.

При розробці технології захисту доступу організація повинна визначити, яка точка застосування політики відповідає за захист кожного ресурсу, а також які допоміжні технології братимуть участь у ухваленні рішень щодо доступу до ресурсів. Спочатку мережа підприємства може бути зовсім не сегментована. Фактично до впровадження нульової довіри, коли підприємство все ще покладається на захист на основі периметра, таку технологію можна представити як захист усіх ресурсів підприємства за допомогою однієї точки застосування політики, тобто брандмауером на периметрі. У міру впровадження ZTA підприємство повинне сегментувати свою інфраструктуру на дрібніші частини. Така сегментація дозволить йому обмежити потенційний вплив порушень або атак та полегшить моніторинг мережевого трафіку. Під час розробки системи захисту доступу підприємство повинне застосовувати контроль доступу на кількох рівнях, а саме, на рівні застосунків, вузлів та мережі.

Етап 5. *Впровадження компонентів архітектури нульової довіри та поступове використання існуючих безпекових рішень для досягнення кінцевої мети*. Як тільки на підприємстві буде: а) правильне розуміння існуючого оточення з погляду ресурсів, які необхідно захистити, та вже розгорнутих засобів безпеки; б) сформульовані політики доступу, які підходять для підтримки його завдань та бізнес-показників; в) розроблено технологію захисту доступу із зазначенням рівня деталізації, з яким захищатиметься доступ до різних ресурсів, та допоміжних технологій, які будуть використовуватись у точці прийняття рішення про політику (PDP), підприємство може безпосередньо розпочинати поступове впровадження архітектури нульової довіри.

Як було вказано вище, наразі вже існує декілька моделей розгортання нульової довіри, які є доступними для встановлення на підприємстві. Тому важливо зрозуміти конкретні потреби ІТ-підприємства для визначення придатної для нього архітектури. Так, при визначенні придатного для ІТ-підприємства рішення при застосуванні міжмережевих екранів із точкою прийняття рішень про політику необхідно виходити з функціональних ролей брандмауерів, як елементів інфраструктури – шлюзів (PEP), які беруть участь у детальних мережевих політиках та політиках рівня ідентифікації, у забезпеченні безпеки. NIST, наприклад, визначає наступні типи шлюзів [38]:

– *Вхідний шлюз* (*Ingress gateway*). Цей шлюз керує тим, як застосунки в кластері/групі виходять за його/її межі (наприклад, керує тим, які імена, сертифікати, порти, протоколи та кінцеві точки застосунків обслуговуються за межами кластера).

– *Вихідний шлюз* / *шлюз виходу* (*Egress gateway*). Керує взаємодією програм у кластері/групі із зовнішнім світом. Він може використовуватися для традиційної фільтрації та реєстрації вихідних повідомлень, як проксі Squid, але також може реалізовувати політику на основі ідентифікації для того, що дозволено викликати та виконувати обмін обліковими даними або надавати набір облікових даних від імені програми, щоб програмі не потрібно було їх обробляти.

– *Граничний / крайовий шлюз* (*Edge gateway*). Цей шлюз знаходиться на межі між мережею та вхідним шлюзом. Він приймає зовнішній потік / трафік перед вхідним шлюзом та виконує тонке балансування навантаження між групами або вузлами. Використовується для переривання зовнішнього трафіку, забезпечення стійкості до відмови на рівні інфраструктури, розгортання «синьо-зелених» (blue-green) груп і спрощення розгортання вхідного шлюзу для кожної групи користувачів, не вимагаючи від кожної з цих груп наявності вхідних шлюзів, що публічно маршрутизуються. Синьо-зелене або канаркове (canary) розгортання – це методологія, що дозволяє впровадити удосконалення програми для невеликої підмножини кінцевих користувачів, і якщо все йде добре, поступово збільшувати це співвідношення, поки всі користувачі не перейдуть на нове розгортання [40].

– *Sidecar gateway*. Особливість цього шлюзу полягає у його розташуванні поряд з кожним екземпляром застосунку для перехоплення всього трафіку, що входить і виходить із застосунку, та обробки внутрішніх комунікацій між сервісами в інфраструктурі.

Ідентифікація, автентифікація та авторизація мають вирішальне значення для прийняття рішень щодо доступу до ресурсів. Враховуючи, що ухвалення та виконання рішень про доступ – це дві основні сфери відповідальності ZTA, підприємство захоче використовувати існуюче або нове рішення ICAM (*Identity, Credential, Access Management – керування ідентифікацією, обліковими даними та доступом*) як фундаментальний компонент для початкової реалізації архітектури нульової довіри.

Підприємству слід уважно розглянути можливість запровадження багатофакторної автентифікації для всіх користувачів. Захист кінцевих точок або аналогічне рішення, що може оцінювати стан пристрою та інтегрується з рішенням ICAM, також може стати ще одним основним компонентом початкового розгортання ZTA. Початкова архітектура нульової довіри, що базується на цих основних компонентах, зможе використовувати ідентифікацію та авторизацію суб'єктів, а також стан і відповідність вимогам кінцевих пристроїв, що запитують, як основу для прийняття рішень про доступ. Потім можуть бути розгорнуті додаткові допоміжні компоненти та функції для задоволення більшої кількості вимог ZTA. Які типи компонентів та в якому порядку будуть розгорнуті, залежить від завдань підприємства та конкретних умов використання. Якщо важлива безпека даних, пріоритетними будуть компоненти безпеки даних. Якщо важливо виявлення аномалій з урахуванням поведінки, можна встановити моніторинг і систему аналізу на основі штучного інтелекту (AI – *artificial intelligence*). Архітектуру нульової довіри доцільно будувати поступово, додаючи та інтегруючи все більше допоміжних компонентів, функцій та можливостей, щоб крок за кроком перетворити її на більш повнофункціональну.

При цьому слід розуміти, що при розгортанні рішень, що задовольняють концепцію нульової довіри, може виникнути проблема, пов'язана з так званим людським фактором. На жаль, керівники служб безпеки на деяких підприємствах стикаються із протидією змінам із боку деяких осіб. Це може бути обумовлено культурою, технічними упередженнями або емоційною орієнтацією на існуючі інструменти або архітектуру безпеки [7]. Ця проблема може здаватися незначною у порівнянні з фінансовими та апаратними складовими, однак це може викликати серйозні проблеми із впровадженням. Є кілька способів протистояти цьому. Насамперед, це навчання (підвищення рівня обізнаності, кваліфікації). Тому дуже важливо належним чином підійти до питання створення майбутніх користувачів архітектури нульової довіри. Успіх впровадження архітектури нульової довіри багато в чому залежатиме від знання основних принципів нульової довіри, які застосовуються до ситуації, а також методів, необхідних для забезпечення дотримання цих принципів. У узагальненому вигляді дана інформація міститься у табл. 2 [23].

Таблиця 2

Принципи (основи) нульової довіри, завдання та методи, пов'язані з реалізацією принципів

Принципи (основи)	**Ціль (завдання)**	**Методи реалізації принципу**
Ресурси.	Ідентифікація та класифікація ресурсів (попередня умова для всіх інших принципів).	Керування ідентифікацією (користувачі, пристрої).
Усі комунікаційні зв'язки захищені незалежно від розташування мережі.	Застосування однієї і тієї ж політики безпеки для внутрішніх та зовнішніх запитів доступу.	Сегментація (сегментація мережі та застосунків для застосування політик ближче до даних або ресурсів), автентифікація та авторизація всіх запитів як внутрішніх, так і зовнішніх, шифрування всіх комунікацій як внутрішніх, так і зовнішніх.
Доступ до ресурсів надається на основі кожного з'єднання.	Права доступу до попереднього з'єднання або сеансу не впливають на права наступного сеансу; доступ суворо обмежений ресурсом, що запитується.	Автентифікація сеансів, детальний контроль доступу на рівні ресурсів.
Доступ до ресурсів надається на основі динамічних політик, які враховують стан ідентифікації користувача/пристрою та можуть включати інші поведінкові атрибути.	Необхідно реалізувати динамічні політики доступу, які враховують стан ідентифікатора користувача, стан пристрою користувача та його атрибути поведінки.	Автентифікація та авторизація на основі контексту, ризик-орієнтована (розрахунок оцінки ризику на основі контексту та історії), адаптивні та динамічні методи контролю доступу.
Усі належні ресурси/пристрої знаходяться у найбільш безпечному стані.	Постійна діагностика та пом'якшення наслідків для оцінки стану пристрою; скомпрометовані пристрої повинні бути позбавлені доступу.	Моніторинг стану пристрою, зв'язку. Автентифікація на основі поведінки для заборони доступу до скомпрометованих пристроїв.
Усі автентифікації та авторизації ресурсів є динамічними та підлягають неухильному виконанню.	Автентифікація та авторизація є безперервними та автоматичними процесами, що переглядають та адаптують довіру до нових та поточних комунікацій.	Безперервна автентифікація та авторизація, адаптивні політики доступу, повторна автентифікація та повторна авторизація, автоматизація автентифікації та авторизації.
Збір інформації для коригування та підвищення рівня безпеки.	Безперервний збір даних для виявлення загроз у системі з автоматичним застосуванням необхідних заходів безпеки.	Реєстрація дій, моніторинг мережі, виявлення та аналіз загроз; оперативна реконфігурація.

Етап 6. *Перевірка реалізації для підтвердження підсумкових досягнень під час розгортання архітектури нульової довіри*. Вибрані для тестування сценарії використання повинні відповідати тим, які найбільше точно відображають повсякденний доступ користувачів

підприємства до його ресурсів. В ідеалі підприємство може створити набір тестів, які можна використовувати для перевірки можливостей у рамках ZTA не лише перед розгортанням кожної нової можливості в процесі поступового розгортання архітектури нульової довіри, а й на основі регулярного тестування, коли розгортання ZTA вважається завершеним. Наприклад, проведення тестування на проникнення та можливості викрадення даних, протистояння фішинговим атакам, фальсифікації тощо.

Етап 7. *Постійне вдосконалення та розвиток відповідно до змін характеру загроз, завдань, технологій та нормативних документів*. Після того як архітектура нульової довіри буде розгорнута і буде вважатися завершеною, вона повинна продовжувати адаптуватися до умов, що змінюються. Якщо технологічні компоненти, що використовуються в ZTA, модернізуються або застарівають, їх слід замінити. Якщо з'являються нові інноваційні технології, підприємство повинне розглянути можливість їх інтеграції в існуючу архітектуру нульової довіри, щоб скористатися перевагами нових засобів захисту, методів та механізмів, які можуть підвищити рівень безпеки підприємства. Якщо цілі підприємства в сфері безпеки змінюються, або внаслідок зміни завдань, або внаслідок змін у нормативних документах, може знадобитися зміна політик та самої архітектури нульової довіри, щоб найкраще відповідати цим новим цілям. Тобто в рамках цього безперервного процесу валідації та вдосконалення підприємства повинні постійно контролювати свою мережу та іншу інфраструктуру, а також оновлювати політики, технології та топології сегментації мережі, щоб переконатися, що вони залишаються ефективними.

Наприкінці хочеться, по перше, ще раз відзначити, що розгортання та впровадження архітектури нульової довіри – це складний, тривалий, багатоетапний і безперервний процес, який передбачає виконання кропіткої роботи для досягнення якісніших рішень у забезпеченні безпеки ІТ-підприємства. По-друге, зупинитися на існуючих сьогодні проблемах і тенденціях, пов'язаних із розвитком та впровадженням архітектури нульової довіри на підприємствах. А саме, сьогодні, в принципі, визначено основи архітектур нульової довіри, але як зробити так, щоб різні технології відповідали вимогам до архітектури нульової довіри, як і раніше, залишається непростим завданням [41]. В даний час контроль доступу, автентифікація особистості та оцінка довіри в архітектурі нульової довіри все ще знаходяться на стадії дослідження, тому питання про те, як використовувати зазначені напрацювання для підвищення рівня захисту безпеки та практичного застосування архітектури нульової довіри, залишається актуальною темою, яка заслуговує на окремий розгляд.

Тим не менш, існуючі вже сьогодні рішення побудови систем на основі архітектур нульової довіри та досвід їх використання свідчать про істотні переваги нової концепції перед традиційними архітектурами, орієнтованими на захист по периметру. А матеріал, представлений у цій роботі, на нашу думку, допоможе відповідним спеціалістам розібратися з фундаментальними змінами у підході до інформаційної безпеки, кібербезпеки та використати відповідні рекомендації щодо успішного впровадження архітектури нульової довіри на своїх ІТ-підприємствах.

Висновки

1. Архітектура нульової довіри – це план кібербезпеки підприємства, який використовує концепції нульової довіри та охоплює зв’язки компонентів, планування робочого процесу та політики доступу. А її реалізація – це шлях, який потрібно пройти, а не просто повна заміна інфраструктури або технологічних процесів.

2. Модель розгортання на основі агента пристрою/шлюзу доцільніше використовувати для підприємств, які мають ефективну систему управління пристроями, а також окремі ресурси, які можуть взаємодіяти зі шлюзом. Ця модель також підходить для підприємств, які не хочуть запроваджувати політику BYOD.

3. Модель розгортання з мікросегментацією орієнтована на варіант використання між серверами. Переваги використання даної моделі полягають у невеликій неявній зоні довіри

та точному двонаправленому контролі доступу до ресурсів (для серверів або мікросервісів). Серед недоліків даної моделі можна виділити: необхідність розгортання точок застосування політики як на пристроях, так і на ресурсах; наявність ймовірності технічних конфліктів між компонентами ресурсів та PEP; точки застосування політики повинні бути доступні для розгортання на множині, можливо, застарілих або успадкованих ОС; наявність можливих негативних відносин з боку власників ресурсів застосунків; необхідність взаємозв'язків один до одного між PEP та ресурсами; може не підходити для доступу користувачів до ресурсів; потрібний прямий доступ суб'єктів до точок застосування політики.

4. Модель розгортання на основі анклаву є корисною для підприємств із успадкованими застосунками або локальними центрами обробки даних, у яких неможливо встановити окремі шлюзи. Також цю модель доречно використовувати організаціям, які працюють у нових середовищах (особливо на базі IaaS) або використовують програмно-керовану інфраструктуру. Недоліком цієї моделі розгортання є те, що шлюз захищає колекцію ресурсів і може бути не в змозі захистити кожен ресурс окремо. Це може дозволити суб'єктам бачити ресурси, до яких вони не мають прав доступу.

5. Модель розгортання з використанням хмарної маршрутизації більш проста та зручна в установці система для підприємств. Проте слід пам'ятати, що простота розгортання не може бути виправданням використання недостатньо ефективного механізму управління безпекою. Дана модель зазвичай добре підходить тільки для віддалених користувачів, оскільки весь трафік повинен проходити через хмару постачальника, а якщо користувачі знаходяться в локальній мережі і отримують доступ до локальних ресурсів, їх дані недоцільно переправляти через хмару постачальника, так як це збільшує затримки, знижує пропускну здатність та збільшує витрати підприємства.

6. Основною перевагою моделі розгортання на основі порталу ресурсів перед іншими моделями є те, що програмний компонент (PEP) не потрібно встановлювати на усі клієнтські пристрої. До того ж, ця модель гнучкіша для політики BYOD та проєктів спільної роботи між організаціями. Як недолік слід зазначити, що використання цього рішення може дозволити зловмисникам виявляти портал та використовувати його для доступу до нього з метою проведення атаки на кшталт «відмова в обслуговуванні». У зв'язку з чим системи порталу повинні мати достатні ресурси, щоб забезпечити доступність при DoS-атаках або збоях у роботі мережі.

7. Основна перевага моделі розгортання на основі використання «пісочниці» застосунків полягає в тому, що певні програми відокремлені від решти активу. При цьому одним із недоліків цієї моделі є те, що підприємства повинні підтримувати такі ізольовані застосунки для всіх активів і можуть не мати повної видимості клієнтських активів. Крім того, підприємству необхідно переконатися в безпеці кожного ізольованого застосунку, що може вимагати більше зусиль, ніж простий моніторинг пристроїв.

8. Говорячи про можливість і доцільність застосування тієї чи іншої моделі розгортання з нульовою довірою, слід відзначити те, що на одному підприємстві для різних бізнес-процесів можуть використовуватися кілька моделей розгортання ZTA, все залежатиме від того, як організована корпоративна мережа на конкретному підприємстві.

9. Під час впровадження архітектури нульової довіри можуть виникати деякі специфічні загрози, які мають унікальні особливості (наприклад, спотворення процесу прийняття рішень у рамках архітектури нульової довіри; відмова в обслуговуванні або порушення роботи мережі; крадіжка облікових даних (інсайдерська загроза); видимість у мережі; зберігання системної та мережевої інформації; залежність від пропрієтарних / власних форматів даних або рішень; використання сутностей, які не є фізичними особами при адмініструванні ZTA), для яких також мають місце певні рішення у межах реалізації певної ZTA.

10. Представлені дослідження покликані допомогти спеціалістам з безпеки розібратися в новій парадигмі забезпечення кібербезпеки інформаційних систем сучасних цифрових

підприємств та використати надані рекомендації щодо успішного впровадження архітектури нульової довіри на своїх IT-підприємствах.

Список літератури:

1. Buck C., Olenberger C., Schweizer A., Völter F., Eymann, T. Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust // Computers & Security. 2021. 110. 102436.
2. Trend Micro Incorporated. What Is Zero Trust? URL: https://www.trendmicro.com/en_us/what-is/what-is-zero-trust.html. (дата звернення: 10.07.2024).
3. Kerman A., Borchert O., Rose S., Division E., Tan A. Implementing a zero trust architecture // National Institute of Standards and Technology, 2020. 17 p. URL: https://www.nccoe.nist.gov/sites/default/files/legacy-files/zta-project-description-final.pdf. (дата звернення: 10.07.2024).
4. National Cybersecurity Center of Excellence (NCCoE). Implementing a Zero Trust Architecture. URL: https://www.nccoe.nist.gov/projects/implementing-zero-trust-architecture. (дата звернення: 10.07.2024).
5. Єсін В. І., Вілігура В. В., Узлов Д. Ю. Огляд існуючих моделей та основних принципів нульової довіри // Радіотехніка. 2024. Вип. 217. С. 39–54.
6. Fernandez E. B., Brazhuk A. A critical analysis of Zero Trust Architecture (ZTA) // Computer Standards & Interfaces. 2024. Vol. 89. 103832. https://doi.org/10.1016/j.csi.2024.103832.
7. Garbis J., Chapman J. W. Zero Trust Security: An Enterprise Guide. Berkeley, CA: Apress, 2021. 300 p.
8. Rose S., Borchert O., Mitchell S., Connelly S. Zero Trust Architecture. NIST Special Publication 800-207. 2020. https://doi.org/10.6028/NIST.SP.800-207.
9. Saltzer J. H., Schroeder M. D. The protection of information in computer systems // Proceedings of the IEEE. 1975. 63(9). P. 1278–1308.
10. Shapiro J. S., Hardy N. EROS: A principle-driven operating system from the ground up // IEEE software. 2002. 19(1). P. 26–33.
11. Bishop M. Introduction to computer security. Addison-Wesley Professional. 2004. 747 p.
12. Samaniego M., Deters R. Zero-trust hierarchical management in IoT // 2018 IEEE international congress on Internet of Things (ICIOT). IEEE, 2018. P. 88–95.
13. Teerakanok S., Uehara T., Inomata A. Migrating to zero trust architecture: Reviews and challenges // Security and Communication Networks. 2021. 2021(1). 9947347. https://doi.org/10.1155/2021/9947347.
14. Adahman Z., Malik A. W., Anwar Z. An analysis of zero-trust architecture and its cost-effectiveness for organizational security // Computers & Security. 2022. Vol. 122. 102911. https://doi.org/10.1016/j.cose.2022.102911
15. Fortinet. The State of Zero Trust. Report. 2023. URL: https://www.fortinet.com/content/dam/fortinet/assets/reports/report-state-of-zero-trust.pdf. (дата звернення: 10.07.2024).
16. Martinez J. Zero Trust Architecture: 2024 Complete Guide. URL: https://www.strongdm.com/zero-trust. (дата звернення: 10.07.2024).
17. Shore M., Zeadally S., Keshariya A. Zero trust: the what, how, why, and when // Computer. 2021. 54(11). P. 26–35. https://doi.org/10.1109/MC.2021.3090018.
18. Bertino E. Zero Trust Architecture: Does It Help? // IEEE Security & Privacy. 19(05). P. 95-96, 2021. https://doi.org/10.1109/MSEC.2021.3091195.
19. Shackelford S. Zero-trust security: Assume that everyone and everything on the internet is out to get you – and maybe already has. The Conversation. URL: https://theconversation.com/zero-trust-security-assume-that-everyone-and-everything-on-the-internet-is-out-to-get-you-and-maybe-already-has-160969. (дата звернення: 10.07.2024).
20. The National Cyber Security Centre. Zero trust architecture design principles. Guidance. Version 1.0. 2021. URL: https://www.ncsc.gov.uk/collection/zero-trust-architecture. (дата звернення: 10.07.2024).
21. NIST Special publication 1800-35B. Implementing a Zero Trust Architecture. Vol. B: Approach, Architecture, and Security Characteristics. 2023. 264 p. URL: https://www.nccoe.nist.gov/projects/implementing-zero-trust-architecture. (дата звернення: 10.07.2024).
22. Rais R., Morillo C., Gilman E., Barth D. Zero Trust Networks. Building Secure Systems in Untrusted Networks. 2nd ed. O'Reilly Media, 2024. 332 p.
23. Syed N. F., Shah S. W., Shaghaghi A., Anwar A., Baig Z., Doss R. Zero Trust Architecture (ZTA): A Comprehensive Survey // IEEE Access. 2022. Vol. 10. P. 57143–57179. https://doi.org/10.1109/ACCESS.2022.3174679.
24. SDP Specification 1.0. Cloud Security Alliance (CSA). 2014. URL: https://cloudsecurityalliance.org/artifacts/sdp-specification-v1-0. (дата звернення: 10.07.2024).
25. Software-Defined Perimeter (SDP) Specification v2.0. Cloud Security Alliance (CSA). 2022. URL: https://cloudsecurityalliance.org/artifacts/software-defined-perimeter-zero-trust-specification-v2. (дата звернення: 10.07.2024).
26. Nadeau T. D., Gray K. SDN: Software Defined Networks: An authoritative review of network programmability technologies. O'Reilly Media, Inc., 2013. 382 p.

27. Cohen R., Barabash K., Rochwerger B., Schour L., Crisan D., Birke R., ... & Jain V. An intent-based approach for network virtualization // Proc. 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013). IEEE. 2013. P. 42–50.

28. Bilger B., Boehme A., Flores B., Guterman Z., Hoover M., Iorga M., Islam J., Kolenko M., Koilpilla J., Lengyel G., Ludlow G., Schroeder T., Schweitzer J. Software defined perimeter working group. SDP specification 1.0. Cloud Security Alliance, Tech. Rep. 2014. URL: https://cloudsecurityalliance.org/artifacts/sdp-specification-v1-0/. (дата звернення: 10.07.2024).

29. Koilpillai J., Garbis J., Islam J., Flores B., Bailey D., Chen B., Bremler E., Roza M., Mahmud S. Software-Defined Perimeter (SDP) Specification 2.0. Cloud Security Alliance, Tech. Rep. 2022. URL: https://cloudsecurityalliance.org/artifacts/software-defined-perimeter-zero-trust-specification-v2. (дата звернення: 10.07.2024).

30. Ross R., Pillitteri V., Graubart, R., Bodeau D., McQuaid R. Developing Cyber-Resilient Systems: A Systems Security Engineering Approach. NIST Special Publication 800–160. Vol. 2. Revision 1. 2021. 310 p.

31. Antonakakis M., April T., Bailey M., Bernhard M., Bursztein E., Cochran J., ... & Zhou Y. Understanding the Mirai Botnet // 26th USENIX security symposium (USENIX Security 17). 2017. P. 1093–1110.

32. Mirai Botnet. URL: https://web.archive.org/web/20161212084605/https://www.cyber.nj.gov/threat-profiles/botnet-variants/mirai-botnet. (дата звернення: 10.07.2024).

33. Bursztein E. Inside the infamous Mirai IoT Botnet: A Retrospective Analysis. URL: https://blog.cloudflare.com/inside-mirai-the-infamous-iot-botnet-a-retrospective-analysis/. (дата звернення: 10.07.2024).

34. Anderson B., McGrew D. Machine learning for encrypted malware traffic classification: accounting for noisy labels and non-stationarity // Proceedings of the 23rd ACM SIGKDD International Conference on knowledge discovery and data mining. 2017. P. 1723–1732. https://doi.org/10.1145/3097983.3098163.

35. American Council for Technology and Industry Advisory Council (ACT-IAC). Zero Trust Cybersecurity Current Trends. 2019. 29 p. URL: https://www.actiac.org/documents/zero-trust-cybersecurity-current-trends. (дата звернення: 10.07.2024).

36. Cunningham C., Holmes D., Pollard J. The eight business and security benefits of zero trust. Forrester Research, Inc., 2019 URL: https://www.forrester.com/report/the-eight-business-and-security-benefits-of-zero-trust/RES134863. (дата звернення: 10.07.2024).

37. What is Cloud Native? URL: https://aws.amazon.com/what-is/cloud-native/?nc1=h_ls. (дата звернення: 10.07.2024).

38. Chandramouli R., Butcher Z. NIST Special Publication 800-207A. A Zero Trust Architecture Model for Access Control in Cloud-Native Applications in Multi-Location Environments. 2023. https://doi.org/10.6028/NIST.SP.800-207A.

39. NIST Special Publication 800-37 Revision 2. Risk Management Framework for Information Systems and Organizations. A System Life Cycle Approach for Security and Privacy. https://doi.org/10.6028/NIST.SP.800-37r2.

40. Mullen-Schultz G. Blue/Green Deployment with Azure Front Door. URL: https://techcommunity.microsoft.com/t5/azure-architecture-blog/blue-green-deployment-with-azure-front-door/ba-p/1609178. (дата звернення: 10.07.2024).

41. He Y., Huang D., Chen L., Ni Y., Ma X. A survey on zero trust architecture: Challenges and future trends // Wireless Communications and Mobile Computing. 2022. 2022(1). 6476274. https://doi.org/10.1155/2022/6476274.

Надійшла до редколегії 15.09.2024

Відомості про авторів:

Єсін Віталій Іванович – д-р техн. наук, професор, Харківський національний університет імені В.Н. Каразіна, професор кафедри кібербезпеки інформаційних систем, мереж і технологій, навчально-науковий інститут комп'ютерних наук та штучного інтелекту; Україна; e-mail: v.i.yesin@karazin.ua; ORCID: https://orcid.org/0000-0003-1977-7269

Вілігура Владислав Вікторович – Харківський національний університет імені В.Н. Каразіна, викладач кафедри кібербезпеки інформаційних систем, мереж і технологій, навчально-науковий інститут комп'ютерних наук та штучного інтелекту; Україна; e-mail: viligura93@gmail.com; ORCID: https://orcid.org/0000-0002-1137-2382

Узлов Дмитро Юрійович – канд. техн. наук, Харківський національний університет імені В.Н. Каразіна, в.о. директора навчально-наукового інституту комп'ютерних наук та штучного інтелекту; Україна; e-mail: dmytro.uzlov@karazin.ua; ORCID: https://orcid.org/0000-0003-3308-424X

УДК 004.056.5 **DOI:10.30837/rt.2024.3.218.02**

В.О. ПОДДУБНИЙ, Р.Ю. ГВОЗДЬОВ, О.В. СЄВЄРІНОВ, канд. техн. наук

ВИКОРИСТАННЯ НУЛЬОВИХ ВОДЯНИХ ЗНАКІВ ДЛЯ ПІДТВЕРДЖЕННЯ АВТОРСТВА ЗОБРАЖЕНЬ ТА БАГАТОФАКТОРНОЇ АВТЕНТИФІКАЦІЇ

Вступ

З розвитком інформаційних систем, з інтеграцією та цифровізацією держави, цифровою трансформацією суспільства, реалізацією сервісів, послуг в електронному форматі питання захисту систем, що обробляють інформацію, стає все більш актуальним. Реалізація даних послуг та процесів потребує безліч адміністраторів, операторів, обслуговуючого персоналу, які поділяються за функціями, можливостями, обов'язками тощо. Проте, такий ріст цифровізації сприяє збільшенню кількості атак на інформаційно-комунікаційні системи (далі – ІКС). Однією з найрозповсюдженіших атак є фішинг [1], який направлений на кінцевого користувача чи оператора, тобто найпростіший метод злому – це не намагатися обійти засоби захисту ІКС, знайти вразливості, заразити ІКС вірусом, а отримати облікові дані адміністратора чи користувача. Одним із методів захисту від крадіжки акаунтів та облікових засобів в ІКС є використання багатофакторної автентифікації. Нульові водяні знаки досить новий метод, що з'явився на заміну «звичайним» цифровим водяним знакам. Нульові водяні знаки першочергово повинні забезпечувати перевірку авторства, проте все частіше з'являються ідеї їх використання для автентифікації [2]. Даний метод підтвердження авторства та автентифікації досить гнучкий та універсальний, існують різні нульові водяні знаки для зображень, тексту чи програмного коду [3], засновані на різних математичних моделях.

В даній роботі висуваються наступні цілі:

– здійснення огляду нульових водяних знаків;

– модифікація існуючого алгоритму нульового водяного знаку для його тестування;

– визначення можливості використання нульового водяного знаку для підтвердження авторства зображень;

– пропозиція використання нульового водяного знаку в схемі багатофакторної автентифікації.

Водяні знаки, розповсюдження та використання для підтвердження авторства

Цифровий водяний знак – технологія, створена для захисту авторських прав мультимедійних файлів. Зазвичай цифрові водяні знаки невидимі. Однак ЦВЗ можуть бути видимими на зображенні або відео [4]. Зазвичай ця інформація являє собою текст або логотип, який ідентифікує автора.

Цифрові водяні знаки доцільно використовувати в багатьох сферах діяльності, де корисно зв'язати деяку додаткову інформацію (метадані) з контейнером даних (об'єктом, в який вбудовується цифровий водяний знак). Ці метадані можуть бути вбудованим як водяний знак. Є й інші способи асоціювати інформацію з об'єктом, наприклад, розміщення його в заголовку цифрового файлу, кодування у видимому штрих-коді на зображенні, QR-коди, цифрові підписи та печатки та ін.

В подальшому при розгляді цифрового водяного знаку буде матися на увазі цифровий водяний знак для зображення.

Цифрові водяні знаки мають дві основні характеристики для порівняння: стійкість і непомітність. Стійкість – це характеристика, яка визначає можливість вилучення водяного знака після додавання до контейнера шуму, спотворення або стиснення. Непомітність є характеристикою ефекту вбудованого знака на контейнер. Оптимальне значення невидимості – це коли водяний знак не впливає безпосередньо на контейнер. На рис. 1 показано приклад цифрового водяного знаку для перевірки авторства зображення. В такому випадку в контей-

нер (зображення) вбудовується додаткова інформація про власника, а до контролюючого органу передається кінцева інформація про зображення.

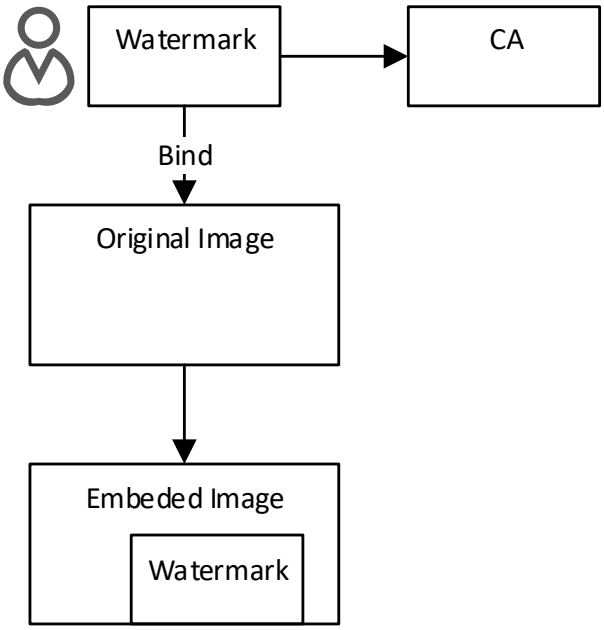


Рис. 1. Спрощена схема ідентифікації володільця за допомогою цифрового водяного знаку

Проте «класичні» водяні знаки мають такий істотний недолік, як спотворення оригінального зображення при вбудові водяного знаку. Це може бути важливо, наприклад, для медичних даних [5, 6]. Також такі водяні знаки сильно спотворюються при зміні форматів зображення, стисканні, форматуванні.

Нульові водяні знаки, їх використання для підтвердження авторства

Zero digital watermarking (нульовий водяний знак) – це метод створення водяного знаку для зображень, який має на меті максимально зменшити вплив водяного знаку на візуальну якість зображення. У цьому методі використовується підхід "zero visibility" (нульова видимість), який означає, що водяний знак невидимий для людського ока.

Цей підхід корисний в різних областях, таких як захист авторських прав на зображення, відстеження поширення зображень в Інтернеті, інформаційна безпека та інші. Zero digital watermarking дозволяє додавати ідентифікуючу інформацію до зображень, не руйнуючи їх зовнішній вигляд, і може бути корисним інструментом для вирішення проблем з підробкою та незаконним використанням зображень.

Основна перевага методу нульового водяного знаку полягає в тому, що водяний знак не вбудовано в саме зображення, тому, на відміну від традиційних методів водяних знаків, не вносить жодних змін до зображення, таким чином уникаючи будь-яких спотворень зображення. Приховані функції витягуються з головного зображення та поєднуються з водяним знаком (якась прихована інформація, наприклад логотип), після чого шифруються та створюється ключ. Потім секретний ресурс має зберігатися в довіреному органі для майбутнього вилучення водяних знаків. Таким чином, вилучення внутрішньої репрезентативної інформації про ознаки з даних зображення є основним підходом нульових водяних знаків [7].

Графічне зображення нульового цифрового знаку показано на рис. 2.

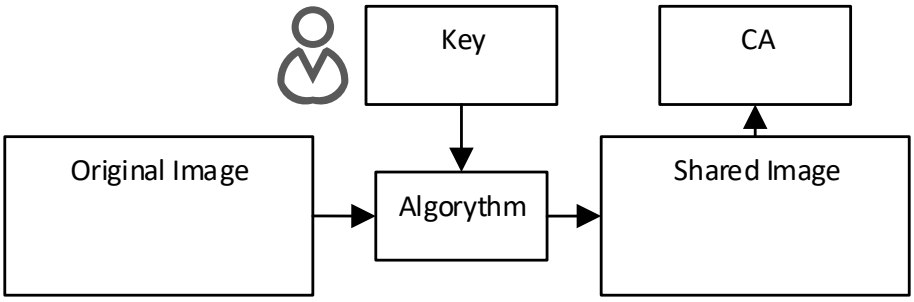


Рис. 2. Спрощена схема ідентифікації володільця за допомогою цифрового водяного знаку

Алгоритм нульового водяного знаку

В роботі [7] запропоновано дві схеми нульового водяного знаку для зображень. Для тестування та демонстрації алгоритму модифіковано першу схему. Схему даного алгоритму зображено на рис. 3.

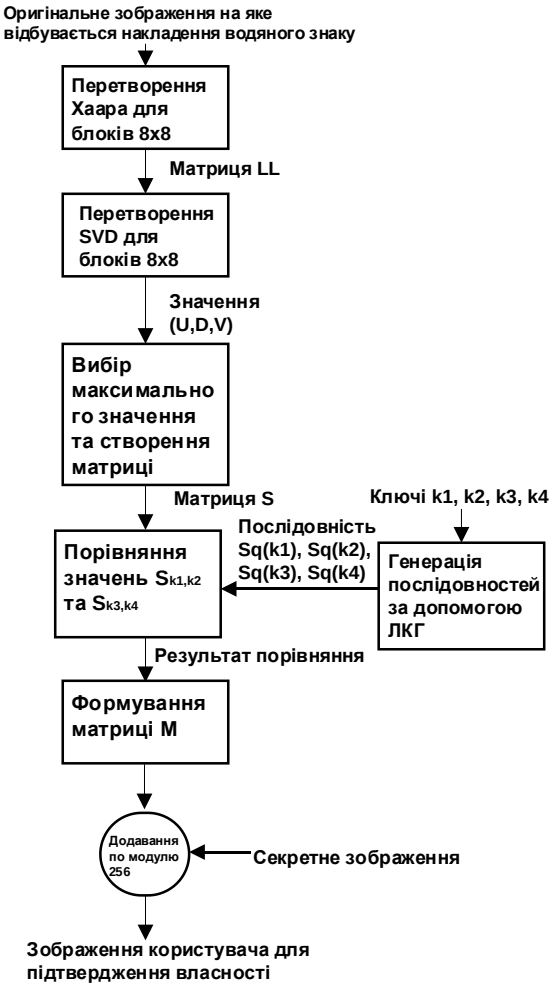


Рис. 3. Запропонована схема алгоритму

Таким чином, алгоритм складається з наступних кроків:

1) Оригінальне зображення (O) розбивається на блоки 8x8.

2) Для кожного блоку 8x8 виконується однорівневе перетворення Хаара [8], з якого в подальшому використовується високочастотний спектр LL. Спектр LL має розмір *m/2*n/2*.

3) Для кожного блоку 8x8 виконується CVD перетворення [9].

4) З отриманих коефіцієнтів перетворення CVD: U,D,V^T вибирається максимальне значення, з якого створюється матриця S.

5) Генеруються ключі, які задають стан ЛКГ key1, key2, key3, key4.

6) Відбувається генерація чотирьох послідовностей k1, k2, k3, k4. Оскільки ці коефіцієнти будуть встановлювати значення кожного пікселю, то довжина послідовностей повинна складати *m/2*n/2*. Послідовність k1, k3 складається з діапазону *[0;m/2]*. Послідовність k2, k4 складається з діапазону *[0;n/2]*, генерація відбувається за допомогою алгоритму $k_{i+1}=(a* k_i+1) \bmod C$.

7) Відбувається порівняння двох вибраних значень D_i та B_i з матриці M. Значення вибираються як $D_i = S[k1_i,k2_i]$; $B_i=S[k3_i,k4_i]$.

8) На підставі вибраних коефіцієнтів складається Master Share M. Пікселі нумеруються зліва направо, зверху вниз, кожен піксель має номер в і *[0;(m/2)*(n/2)]* Якщо значення D_i не дорівнює значенню B_i, то значення пікселя в Master Share обчислюється як

$M_i=(k1_i + k2_i) \bmod (256)$

в іншому випадку

$M_i=(k4_i + k4_i) \bmod (256)$

в кінцевому випадку отримується майстер-зображення для кожного спектру M.

9) Генерується секретне зображення (S) так, що $M+S \bmod (256) = Sig$, де *Sig* – зображення користувача, що зберігається в засвідчуваному центрі.

Використання нульових водяних знаків для підтвердження авторства

Нульові водяні знаки з'явилися як засіб підтвердження авторства, проте з часом були проведені дослідження, які показують складність використання нульових знаків для цієї мети. В роботі [10] проведено дослідження можливості використання нульових водяних знаків для захисту авторства медичних даних.

Після тестування запропонованого вище алгоритму було встановлено наступні результати при вилученні водяного знаку, що наведені в табл. 1

Таблиця 1

№	1		2		3	
	PSTR	NCC	PSTR	NCC	PSTR	NCC
1	-	1	-	1	-	1
2	23,60	0,92	20,36	0,76	24,95	0,85
3	16,05	0,90	14,14	0,66	19,05	0,82
4	14,89	0,90	13,15	0,66	18,21	0,81

Для перевірки водяного знаку використовувалося дві метрики:

– NCC;

– PSTR.

Значення NCC визначається як

$$NCC = \frac{\sum_{i=1}^{n}(A_i-\bar{A})(B_i-\bar{B})}{\sqrt{\sum_{i=1}^{n}(A_i-\bar{A})^2 \sum_{i=1}^{n}(B_i-\bar{B})^2}} \ . \quad (1)$$

Значення NCC визначає нормальну кореляцію між зображанням *A* та *B*.

Значення $\bar{A}$ та $\bar{B}$ становлять середні значення пікселів зображення *A* та *B*.

Математичне представлення *PSNR* виглядає наступним чином (1):

$$PSNR = 20log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right), \quad (2)$$

де MAX_f – максимальне значення сигналу, яке існує в вихідному зображенні; *MSE* – (середня квадратична помилка) [11].

Отримавши аналогічні результати, що і автори [10], можна зробити висновки що нульові водяні знаки мають наступні особливості:

– складність забезпечення авторства в подібних зображеннях;

– складність забезпечення авторства різних зображень одного автора.

Ці особливості, на нашу думку, не дозволяють використовувати нульові водяні знаки для підтвердження авторства.

Схема багатофакторної автентифікації на основі нульового водяного знаку

Як описано вище, нульові водяні знаки не підходять для підтвердження авторства, проте можуть бути корисними для схем авторизації.

Схема ґрунтується на двох ознаках:

1. Знання чогось: пароль.

2. Володіння чимось: зображення.

Додатковою відмінністю алгоритму є ще один фактор, що включає ознаку

1. Знання чогось: чи є фото користувача ключем.

Дана ознака включає в себе знання того, яке саме з фото є ключем. Користувач може зберігати безліч фото на своєму носії інформації, проте зловмисник не має інформації про те, чи є якесь фото на носії користувача ключем. Отримавши доступ до носія, зловмисник побачить лише фото/набір фото без додаткової інформації.

Наприклад, на телефоні користувача збережений набір фото, як на рис. 4 [12]:

Рис. 4. Приклад використання зображення в якості ключових даних

Одне з цих фото може бути ключем для водяного знаку, але й може не бути, зловмисник не знає, яке саме фото є ключем, та чи є взагалі.

Ця ознака перетворює даний алгоритм на трьохфакторний та дозволяє розмити інформацію на носії користувача.

Проте це дійсно лише для унікальних фотографій, оскільки якщо фото буде неунікальне та неконфіденційне, то ознака володіння користувачем втрачається. В такому випадку користувачу слід самому вибрати фото, що послаблює стійкість алгоритму (оскільки користувач може вибрати неунікальне фото). Теоретичним рішенням є генерація фото за допомогою штучного інтелекту, це також теоретично може включати процес відновлення фото, оскільки ШІ генеруватиме фото на основі ключових фраз та солі (наприклад ID користувача в системі або інша сіль КЗЗ), що слугуватимуть для відновлення ключа.

Запропонована схема складається з наступних компонентів:

– алгоритм розширення ключа;

– генератор геш-значень;

– алгоритм нульового цифрового знаку;

– центр сертифікації (опціонально);

– засоби зчитування зображення (опціонально).

– засіб збереження фото (опціонально).

– модуль автентифікації.

Розглянемо кожний модуль окремо.

Алгоритм розширення ключа розширює ключ на основі пароля користувача. Пароль користувача є вектором ініціації для генерації псевдовипадкової послідовності, під час тестування було використано лінійно конгруентний генератор випадкових послідовностей

проте було встановлено залежність запропонованого алгоритму від ключів згенерованих за допомогою ЛКГ. Вхідні дані: пароль користувача. Вихідні дані: розгорнутий ключ на основі паролю користувача, що відповідає вимогам алгоритму нульового водяного знаку.

Пароль користувача розподіляється на чотири частини, кожна з частин проходить через генератор геш-значень, та слугує вектором ініціації для ЛКГ, що генерує ключі key1, key2, key3, key4. Такий спосіб генерації ключових послідовностей дозволяє забезпечити стійкість від зворотної атаки на пароль. Спрощена схема розгортання ключів зображена на рис. 5.

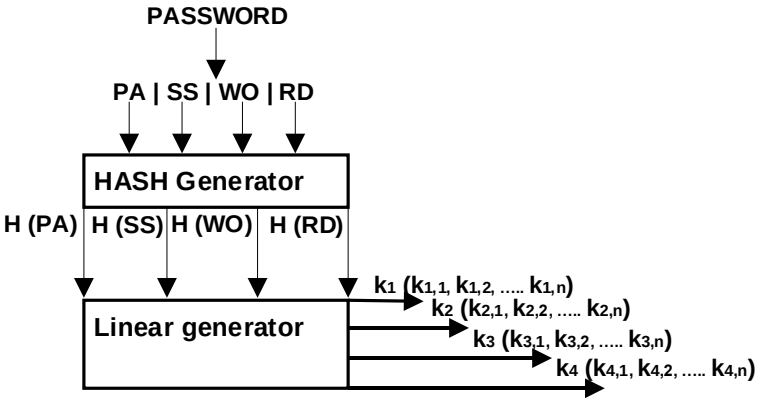


Рис. 5. Схема розгортання ключів для запропонованого алгоритму

Може бути використаний інший алогічний алгоритм. Вхідні дані для алгоритму: ключ користувача, ключове зображення користувача. Вихідні дані: зображення отримане в результаті вилучення водяного знаку.

Генератор геш-значень – використовує пароль користувача для генерації геш-послідовності, що перевіряється КЗЗ. В якості генератора можна використовувати будь-яку стійку геш-функцію, таку як SHA-2, SHA-3, Купина тощо. Вхідні дані: пароль користувача. Вихідні дані: геш-сума паролю користувача.

Центр сертифікації (опціонально). Центр сертифікації здійснює порівняння результатів вилучення водяного знаку з власними даними для кожного користувача. Центр сертифікації використовується опціонально, так як його функції може виконувати локальний модуль в ІКС, проте він може використовуватися, коли в ІКС необхідно здійснювати зовнішню ідентифікацію або коли необхідно використання довіреної третьої сторони. Центр сертифікації використовує власний ключ для генерації кінцевого водяного знаку, який порівнюється з еталонною копією водяного знаку. Після порівняння центр сертифікації повертає Модулю доступу числове значення достовірності (на скільки відсотків вірне зображення користувача). На основі цього коефіцієнту Модуль доступу приймає рішення про правильну/хибну ідентифікацію. Вхідні дані: вилучений водяний знак користувача. Вихідні дані: коефіцієнт достовірності.

Модуль доступу – є модулем, що обробляє та приймає рішення про доступ користувача до системи. Модуль доступу зберігає (або має доступ) до геш-значень паролів користувача. Модуль доступу перевіряє геш-значення паролю користувача та отриманий в результаті вилучення водяного знаку коефіцієнт схожості. Оскільки оригінальне зображення може (в залежності від типу його передачі в ІКС) містити помилки, то можливі погрішності в коефіцієнті схожості. Необхідний поріг налаштовується в самому модулю доступу, це може бути як повна схожість (значення = 1), так і якийсь коефіцієнт (наприклад, 0.95). Вхідні дані: геш-значення пароля користувача, коефіцієнт достовірності водяного знаку.

На рис. 6 зображена спрощена схема автентифікації з використанням нульового водяного знаку.

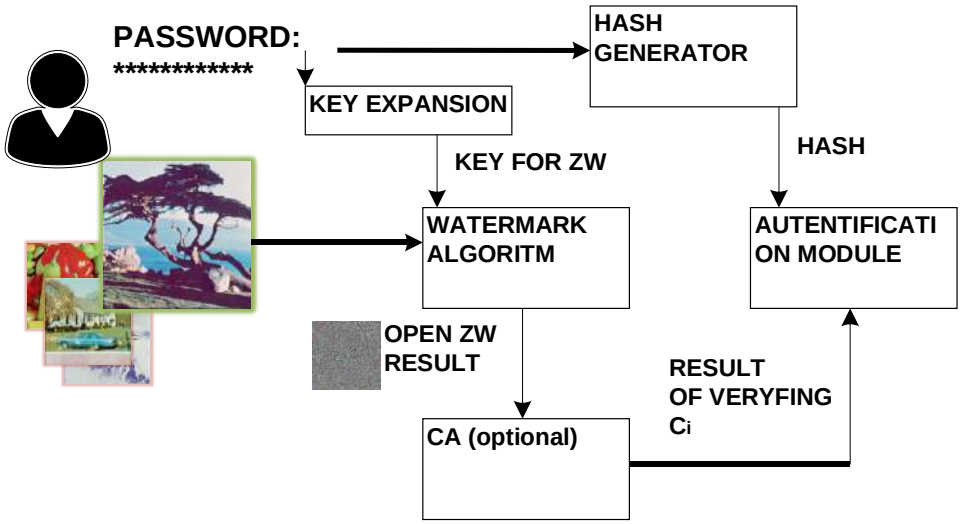


Рис. 6. Запропонована схема автентифікації (спрощена)

Схема автентифікації складається з наступних етапів:

1) Користувач має у володінні секретне зображення на будь-якому носії інформації, що має здатність відображати фото (телефон, планшет, паперова версія фото). Користувач знає, яке саме фото є ключем, пароль.

2) Користувач ініціює початок автентифікації.

3) Користувач демонструє фото засобам візуального зчитування та вводить пароль.

4) Засоби візуального зчитування цифровізують (якщо необхідно) зображення та передають до КЗЗ, в якому це зображення вноситься до вхідних даних алгоритму водяного знаку.

5) Пароль передається до КЗЗ. КЗЗ здійснює порівняння паролю з геш-значенням в БД. Пароль передається до вхідних даних алгоритму розширення ключа.

6) Алгоритм розширення ключа здійснює розгортання ключа за наявним паролем, пароль видаляється з оперативної пам'яті КЗЗ, надалі пароль не використовується, використовується лише розгорнений ключ.

7) Алгоритм розширення ключа передає розгорнутий ключ до алгоритму водяного знаку.

8) Алгоритм водяного знаку отримує розгорнутий ключ та відцифроване зображення. Алгоритм за наявними даними здійснює розгортання водяного знаку, в результаті розгортання отримується майстер зображення користувача.

9) КЗЗ передає майстер зображення до СА, СА здійснює порівняння майстер-зображення з еталонним значенням, повертаючи коефіцієнт схожості до КЗЗ. За необхідності СА може здійснювати накладання майстер-зображення з відповідним зображенням, що зберігається для даного користувача в СА.

10) КЗЗ приймає рішення на основі коефіцієнту, отриманого від СА [13].

Висновки

Дослідження та тестування алгоритму нульового водяного знаку показали, що нульові водяні знаки не підходять для підтвердження авторства, оскільки не можуть чітко «відрізнити» схоже зображення користувача, або стороннє схоже зображення при правильності введення ключових даних. Це актуально для медичних зображень [14], оскільки вони мають високу подібність. Проте нульові водяні знаки можуть бути використані в схемах автентифікації.

Запропоновано схему ідентифікації та автентифікації користувача за допомогою нульових водяних знаків для зображень. Алгоритм має як і суттєві плюси, так і декілька мінусів. До концептуальних плюсів відноситься:

– наявність мінімум двох ознак ідентифікації;

– наявність трьох ознак ідентифікації при умові унікальності зображень;

– можливість використання центрів сертифікації;

– можливість використання зовнішніх сенсорів для передачі інформації в ікс (фото камери, сканери тощо);

– можливість збереження ключа на будь-якому носії інформації, у випадку використання зовнішніх сенсорів дозволяється використання навіть паперових фото.

До концептуальних мінусів належить:

– наявність помилок першого та другого роду при використанні зовнішніх сенсорів;

– використання паролю, що не є стійким методом ідентифікації.

До проблем, що потребують вирішення, відносяться:

– розробка стійкого алгоритму водяного знаку для зображень;

– тестування запропонованого алгоритму на стійкість до атак (людина посередні, маскарад тощо)

Також можна виділити наступні вектори розвитку нульових водяних знаків:

– використання штучного інтелекту для розпізнавання зображень, їх характеристик та ключових параметрів;

– знаходження балансу між стійкістю та чутливістю алгоритму водяного знаку, алгоритм повинен бути достатньо стійким, щоб ігнорувати шуми та пошкодження зображення, та достатньо чутливим, щоб реагувати на зміну ключових параметрів;

– комбінація алгоритмів перетворення зображення для досягнення згаданих вище характеристик та швидкодії; в якості алгоритмів можуть бути використані: DWT, FWT, SVD, WFT, MWT, PCA, LDA, Fourier-Mellin, Radon-перетворення.

Підсумовуючи, можна зробити припущення, що запропонована модель є перспективною, проте потребує досліджень з реалізації, тестування й доведення ефективності. Також ускладнюючим фактором є необхідність розробки й тестування багатьох комбінацій компонентів моделі.

Список літератури:

1. Марія Огнівчук. Прогноз кіберзагроз 2024 // H-X Technologies [Електронний ресурс] https://www.h-x.technology/ua/blog-ua/cyber-threats-forecast-2024-ua
2. X. Qi and Y. Liu. Cloud Model Based Zero-Watermarking Algorithm for Authentication of Text Document // 2013 Ninth International Conference on Computational Intelligence and Security, Emeishan, China. 2013. P. 712–715. doi: 10.1109/CIS.2013.155.
3. Iwendi Celestine, Srivastava Gautam, Jo Ohyun, Javed, Abdul Rehman. KeySplitWatermark: Zero Watermarking Algorithm for Software Protection Against Cyber-Attacks // 2020/04/15, IEEE Access, 10.1109/ACCESS.2020.2988160
4. Cox I., Miller M., Bloom J., Fridrich J., Kalker T. Digital Watermarking and Steganography. 2nd Edition, 2007.
5. A. Zulfiqar and M. H. Fazal-e-Amin. A Novel Fragile Zero Watermarking Algorithm for Digital Medical Images // Electronics. 2022. Vol. 11. P.710. doi: https://doi.org/10.3390/electronics1105 0710
6. Wu D., Wang M. and Zhao J. Color Zero-Watermarking Algorithm for Medical Images Basedon BEMD-Schur Decomposition and Color Visual Cryptography // Hindawi Security and Communication Networks, 2021, doi: https://doi.org/10.1155/2021/7081194
7. Asha Rani, Amandeep K. Bhullar, Deepak Dangwal, Sanjeev Kumar. A Zero-Watermarking Scheme using Discrete Wavelet Transform // Procedia Computer Science. 2015. Vol. 70. C. 603–609.
8. Nidhi Sethi, Ram KrishnaImage, Prof R.P. Arora. Image Compression Using Haar Wavelet Transform // Computer Engineering and Intelligent Systems. Vol. 2, No.3
9. Elizabeth A. Compton and Stacey L. Ernstberger Singular Value Decomposition: Applications to Image Processing // Citations Journal of Undergraduate Research. 2020. Vol. 17.

10. Roček A., Javorník M., Slavíček K. et al. Zero Watermarking: Critical Analysis of Its Role in Current Medical Imaging // Digit Imaging. 2021. Vol. 34. P. 204–211.

11. Peak Signal-to-Noise Ratio as an Image Quality Metric [Електронний ресурс]. Режим доступу: http://www.ni.com/white-paper/13306/en/

12. The USC-SIPI Image Database [Електронний ресурс]. Режим доступу: URL: https://sipi.usc.edu/database/

13. R. Gvozdov, V. Poddubnyi, O. Sieverinov, A. Buhantsov, A. Vlasov and V. Sukhoteplyi. Method of Biometric Authentication with Digital Watermarks // PIC S&T, IEEE. 2021. P.569–571. https://doi.org/10.1109/PICST54195.2021.9772134

14. Hongyan X. Digital media zero watermark copyright protection algorithm based on embedded intelligent edge computing detection // Mathematical Biosciences and Engineering. 2021. Вип. 18(5). С. 6771–6789.

Надійшла до редколегії 03.09.2024

Відомості про авторів:

Поддубний Вадим Олександрович – Харківський національний університет радіоелектроніки, аспірант кафедри безпеки інформаційних технологій; Україна; e-mail: vadym.poddubnyi@nure.ua; ORCID: https://orcid.org/0000-0002-4380-491X

Гвоздьов Роман Юрійович – Харківський національний університет радіоелектроніки, аспірант кафедри безпеки інформаційних технологій; Україна; e-mail: roman.hvozdov@nure.ua; ORCID: https://orcid.org/0000-0002-5408-943X

Сєвєрінов Олександр Васильович – канд. техн. наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри безпеки інформаційних технологій; Україна; e-mail: oleksandr.sievierinov@nure.ua; ORCID: https://orcid.org/0000-0002-6327-6405

УДК 004.056.5 **DOI:10.30837/rt.2024.3.218.03**

О.І. ФЕДЮШИН, канд. техн. наук, Є.В. ГОЛОВКО, А.О. СМІРНОВ, канд. техн. наук, В.М. СУХОТЕПЛИЙ, О.В. ЧЕЧУЙ, канд. техн. наук

МЕТОДИ ЗАХИСТУ ІНФОРМАЦІЇ НА ОСНОВІ КВАНТОВОЇ СТЕГАНОГРАФІЇ ЗОБРАЖЕНЬ

Вступ

Стеганографія – це процес, в якому конфіденційні дані в тій чи іншій формі ховаються в зображенні, яке може бути будь-яким звичайним зображенням, наприклад, кота або дерева. Зображення, яке використовується в цьому процесі, називається стего-зображенням. Цей процес не змінює жодної видимої риси вихідного контейнера. Таким чином, будь-який зловмисник, який хоче отримати доступ до даних, навіть не знає, що в ньому приховані дані. Доступ до даних може отримати лише особа, яка має спеціальний ключ, в той час як для будь-якої іншої особи він навіть не існує.

Безпека даних дуже важлива, коли мова йде про обмін інформацією між користувачами. Наш мотив полягає в тому, щоб заховати дані в зображенні та отримати їх за допомогою алгоритму вилучення, і зробити це за допомогою квантових обчислень. У квантовій обробці зображень [1] квантове представлення зображень відіграє ключову роль, яка визначає типи завдань обробки і те, наскільки добре вони можуть бути виконані.

Існують різні методи представлення зображень: Qubit Lattice [2], Entangled image, Real Ket [3], Flexible Representation of Quantum Images (FRQI) [1], Novel Enhanced Quantum Image Representation (NEQR) [4].

FRQI використовує нормалізовану суперпозицію для зберігання всіх пікселів зображення, однакові операції можна виконувати одночасно над усіма пікселями, і тому FRQI може полегшити обчислювальну проблему обробки зображень. Обмеження FRQI полягає в тому, що він використовує лише один кубіт для зберігання інформації про відтінки сірого для кожного пікселя зображення, тому деякі операції цифрової обробки зображень, такі як складні операції з кольором, не можуть бути виконані на основі FRQI.

Модель NEQR використовує лінійний незалежний базовий стан кубітової послідовності для зберігання значення відтінку сірого для кожного пікселя. Таким чином, для зберігання цифрового зображення з використанням квантової механіки в NEQR використовуються дві переплетені кубітні послідовності, які представляють інформацію про відтінки сірого та положення всіх пікселів на зображенні. У представленні FRQI інформація про відтінки сірого зображення кодується за допомогою одного кубіта, тоді як у NEQR інформація про відтінки сірого кодується в базисних кубітних станах. Оскільки кожен базисний кубітний стан є лінійно незалежним, завдання обробки зображення стає набагато простішим, ніж у FRQI. Крім того, часова складність підготовки квантового зображення NEQR зменшується приблизно в квадратичному відношенні порівняно з FRQI. Використовуючи особливості зберігання даних в різних форматах можна організувати стеганографічне вбудовування даних. Актуальним завданням роботи є проведення наукового пошуку ефективних методів квантової стеганографії. В подальшому планується провести якісний аналіз переваг та недоліків, перспектив та труднощів їх практичного впровадження та розглянути ефективні методи для моделювання подібних систем.

Представлення зображень у квантових станах

Квантова обробка зображень займається представленням зображень і зберіганням даних про зображення у квантових станах, а також перетворенням цих станів для досягнення поставленої мети. Першим кроком при обробці зображень у квантових комп'ютерах є переведення пікселів зображення у квантові стани. Розглянемо різні підходи до цього процесу.

Квантовий комп'ютер оперує так званими квантовими бітами. Можна визначити квантовий біт, або скорочено q-біт (кубіт), як квантово-механічну систему, що має два стани, які позначаються відповідно, як $|0\rangle$ і $|1\rangle$.

Однак на відміну від класичного випадку, у квантовій механіці ці два стани можуть перебувати у стані суперпозиції, тобто загальний стан квантового біта може бути записаний як:

$$|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{1}$$

де α і β – комплексні коефіцієнти.

Іншими словами, можна сказати, що закони квантової механіки допускають інші значення кубіта, які називаються станами суперпозиції. Таким чином стани суперпозиції являють собою значення між екстремумами 0 і 1, а квантовий біт може приймати нескінченно багато значень. Кубіт можна визначити, як вектор одиничної довжини у двовимірному гільбертовому просторі над полем комплексних чисел.

Стани $|0\rangle$ і $|1\rangle$ разом являють собою базисні вектори. Як і всі вектори, вони вказують напрямок і мають величину.

Для запису двох станів кубітів можна використовувати позначення бра ($\langle\ |$) і кет ($|\ \rangle$) – позначення Дірака. Вектори виду $|\ \rangle$ називаються кет-векторами, а виду $|\ \rangle$ бра-векторами.

Формула (1) для хвильової функції $|\varphi\rangle$ описує, в якій пропорції нескінченна множина всіх варіантів значень квантового стану містить варіанти базисних станів $|0\rangle$ і $|1\rangle$.

Візуалізація стану кубіту можлива за допомогою спеціального інструменту, названого сферою Блоха. Сфера Блоха – це сфера з одиничним радіусом, при цьому точка на її поверхні відповідає стану кубіта.

На рис. 1 зображено побудову Сфери Блоха з використанням бібліотек Qiskit від IBM та Python [5].

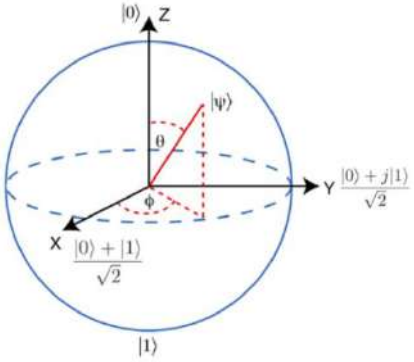


Рис. 1. Модель Сфера Блоха. Стан у верхній частині сфери представляє $|0\rangle$,
а стан у нижній частині сфери представляє $|1\rangle$

Квантові решітки (Quantum Lattice).

Група пікселів збирається разом, щоб сформувати зображення. Ці пікселі містять такі властивості, як положення та інтенсивність кольору. Першим методом, який дозволив перетворити значення пікселів у квантовий стан, була модель "Квантової решітки". Дослідження забезпечило аналогове перетворення у квантові стани. Для зображення розміром $N \times N$ використовується $N \times N$ кубітів, які розташовані у матричному форматі, де кожен кубіт еквівалентний кожному пікселю. Дослідження було зосереджено на представленні пікселя у вигляді частот, а не лінійної комбінації колірної моделі RGB. Частоти, представлені пікселем, обчислюються за допомогою гіпотетичної машини, яка перетворює електромагнітні хвилі світла в ініціалізовані кубіти.

Всю систему можна представити у вигляді

$$A: F \to \varphi, \tag{2}$$

де φ представляє кубіт, який ініціалізується частотою F, зчитаною машиною A.

Процес перетворення показаний на рис. 2.

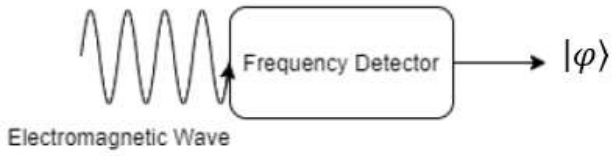


Рис. 2. Модель Qubit Lattice

Структура "решітки кубітів" формується шляхом розміщення ідентичних $K-1$ кубітів за кожним кубітом, які представляють частоту пікселя. Ці частоти кодуються в θ – кут відхилення кубіта в суперпозиції $|0\rangle$ та $|1\rangle$ (див. рис.1). Основним завданням зберігання частоти в ідентичному кубіті була можливість повернення до значень пікселів, щоб реконструювати зображення з квантового стану. Кількість часу, необхідного для отримання даних, залежить від кількості пікселів/кубітів, задіяних у процесі. Кількість кубітів прямо пропорційна точності вилучення зображень з його квантового стану. Нехай кубіт має стан α і β.

Позначимо кількість вимірювань для цих станів через M_α і M_β:

$$\cos^2\frac{\theta}{2} = M_\alpha \Big/ \left(M_\alpha + M_\beta\right) \tag{3}$$

Тоді частоту можна знайти, якщо розв'язати рівняння для тета-квантів.

Модель Real Ket.

З введенням кубітної решітки [2], Латорре у 2005 р. представив заплутану модель представлення зображень [3], де кожне зображення було розділене на 4 квадранти, де кожен квадрант був пронумерований, починаючи зліва направо з верхнього рядка. Потім ці квадранти були знову поділені на ще 4 квадранти з такими ж номерами. Цей поділ продовжувався до тих пір, поки ми не отримали один піксель. Модель ділить зображення на менші зображення (розбиваючи зображення на 4 частини) і створює структуру Quadtree з коефіцієнтами, як на зображенні у відтінках сірого на рис. 3 (тут реалізовано стиснення та заплутування зображення).

І Real Ket, і квантова решітка надавали способи представлення зображення, але вони мали свої обмеження.

Решітчаста модель вимагала більше кубітів для представлення зображення, чого практично неможливо досягти. Модель Real Ket забезпечувала краще стиснення зображення, і виявилася швидшою за модель квантової решітки, але була обмежена випадковістю, яка була необхідна пікселям для ефективної роботи коли ми розглядаємо реальне зображення, оскільки пікселі пов'язані між собою.

Модель FRQI.

У 2010 р. було запроваджено модель FRQI (Flexible Representation of Quantum Images – гнучкого представлення квантових зображень), яка мала перевагу над іншими розглянутими моделями, оскільки вимагає меншу кількість кубітів для представлення зображення. У FRQI потрібно лише $2n+1$ кубіт для зображення розміром $2^n \times 2^n$.

Таке представлення використовує переваги суперпозиції між кубітами і дозволяє кодувати інтенсивність кольору та положення пікселів у нормалізованому стані кубіта [1].

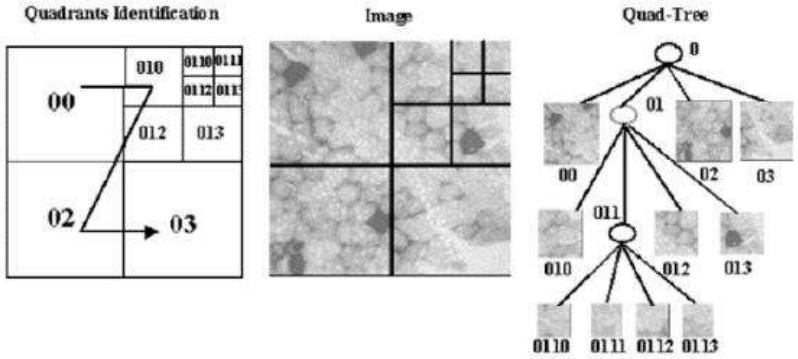


Рис. 3. Модель Real Ket

Зображення представлено в залежності від θ у рівняннях:

$$\left|I(\theta)\right\rangle = \frac{1}{2^n}\sum_{i=0}^{2^{2n-1}}\left(\cos\theta\left|0\right\rangle + e^{i\phi}\sin\theta\left|1\right\rangle\right)\otimes\left|i\right\rangle, \tag{4}$$

де $\theta \in \left[0, \frac{\pi}{2}\right], i = 0,1,2,3,\ldots,2^n - 1$.

Коефіцієнт тета дозволяє нам кодувати інтенсивність кольору, в той час як положення пікселя представлено через $\left|I\right\rangle$.

Зображення перетворюються з початкового стану ($\left|0\right\rangle^{\otimes 2n}$) кубіта у стан FRQI за допомогою унітарного перетворення (позначається P), яке включає в себе гейт Адамара (позначається H) для створення суперпозицій початкового стану кубіта з наступними контрольованими обертаннями (позначаються R) для створення FRQI. Використовуючи гейт H на керовані обертання навколо осей X та Y, результатом є стан FRQI, представлений як $P = HR$. Приклад представлення показано на рис. 4.

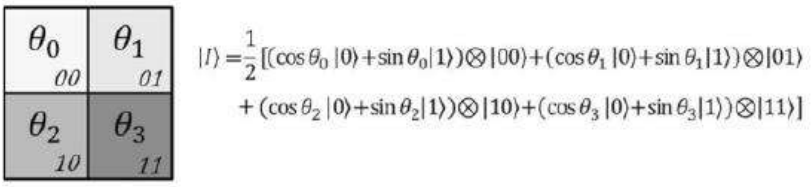


Рис. 4. Приклад представлення зображення в моделі FRQI

Перевагою FRQI є суперпозиція послідовності кубітів, що дозволяє нам трансформувати всі пікселі, змінивши лише один кубіт. Недоліком цього методу представлення зображень є кількість гейтів, необхідних для підготовки зображення у квантовому форматі FRQI. Пряма реалізація цього методу вимагає 2^{2n} керованих обертань і $2n$ гейтів Адамара. Загальна обчислювальна складність зростає до квадратичної (2^{4n}).

Іншим недоліком, який пов'язаний зі станом FRQI, є те, що його можна застосовувати лише до квадратних зображень.

Модель NEQR.

Нова модель NEQR (Novel Enhanced Quantum Image Representation – Нове покращене квантове представлення зображень) спирається на переваги, які надає FRQI, і використовує 2

заплутаних кубіти для зберігання пікселя і властивостей кольору пікселя. Метод зменшує загальну обчислювальну складність з (2^{4n}) до (2^{2n}) [4]. Він зосереджується на недоліках моделі FRQI і зберігає інформацію на основі кубітової послідовності, що дозволяє вдвічі зменшити обчислювальну складність та покращити коефіцієнт стиснення в 1,5 рази. Існує набагато більше методів, які дослідники опублікували для представлення зображення у квантових станах. Але FRQI і NEQR є найпоширенішими, з яких NEQR має більшу застосовність в поточному сценарії роботи для задач стеганографії.

Колірна схема зображення складається з трьох значень інтенсивності, відомих як RGB-значення зображення, інтенсивність кожного кольору може змінюватися від 0, де 0 означає чорний, а 255 – білий. Для кодування кожної інтенсивності ($2^q = 255$), де q – кількість кубітів, необхідних для кодування різних інтенсивностей певного кольору, а для кодування позиції нам потрібен інший набір кубітів. Оскільки ми будемо представляти двовимірне (2x2) піксельне зображення, ми будемо визначати позицію зображення через його рядок і стовпець, Y та X відповідно, а колір – через формулу

$$f(Y,X) = C_{YX}^{q-1} C_{YX}^{q-2} \ldots C_{YX}^{1} C_{YX}^{0}, C_{YX}^{q} \in [0,1], f(Y,X) \in \left[0, 2^q - 1\right] \qquad (5)$$

Промоделюємо представлення зображення в моделі NEQR за допомогою середовища Qiskit Jupyter [5].

Qiskit – фреймворк Python з відкритим вихідним кодом, наданий IBM, який використовується для маніпулювання, написання квантових програм, а потім реалізації їх на реальному квантовому комп'ютері або симуляторах, наданих на сайті IBM з квантових досліджень. Дані необхідні для побудови зображення наведено в табл. 1.

Таблиця 1

Положення пікселя	Бінарне представлення	Інтенсивність в градаціях сірого
$\|00\rangle$	$\|00000000\rangle$	0 – Black
$\|01\rangle$	$\|01100100\rangle$	100 – Darkshade
$\|10\rangle$	$\|11001000\rangle$	200 – Lightshade
$\|11\rangle$	$\|11111111\rangle$	255 – White

Етапи обробки зображень за допомогою NEQR є наступними:

Крок 1: Генеруємо класичне зображення у відтінках сірого (2×2) за допомогою python.

Крок 2: Квантовий ланцюг зображення формується на основі налаштування кольору зображення за допомогою NEQR-коду. Цей крок передбачає зберігання класичних даних у вигляді квантового стану, як показано на рис. 5.

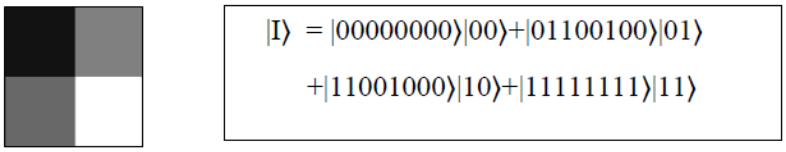


Рис. 5. Приклад представлення зображення в моделі NEQR

Код виконується в блокноті Jupyter на мові Python, а потім на симуляторах Quantum, наданих IBM. Кодується класичне зображення розміром 2×2 пікселі, як показано на рис. 5, зі значенням шкали сірого в діапазоні 0-255 у схему квантового зображення за допомогою коду NEQR. Схема, яка перетворює дані зображення в дані Quantum, показана на рис. 6. Значення всіх чотирьох пікселів розділені за допомогою бар'єрів (показані пунктирними вертикальними лініями).

На рис. 6 перші вісім кубітів (0-7) використовуються для кодування інтенсивності пікселів, а інші два кубіти (8 і 9) використовуються для кодування інформації про положення пікселів. Можливі положення пікселів отримуються за допомогою перетворення Адамара двопозиційних кубітів, яке дає нам всі чотири можливі стани положення.

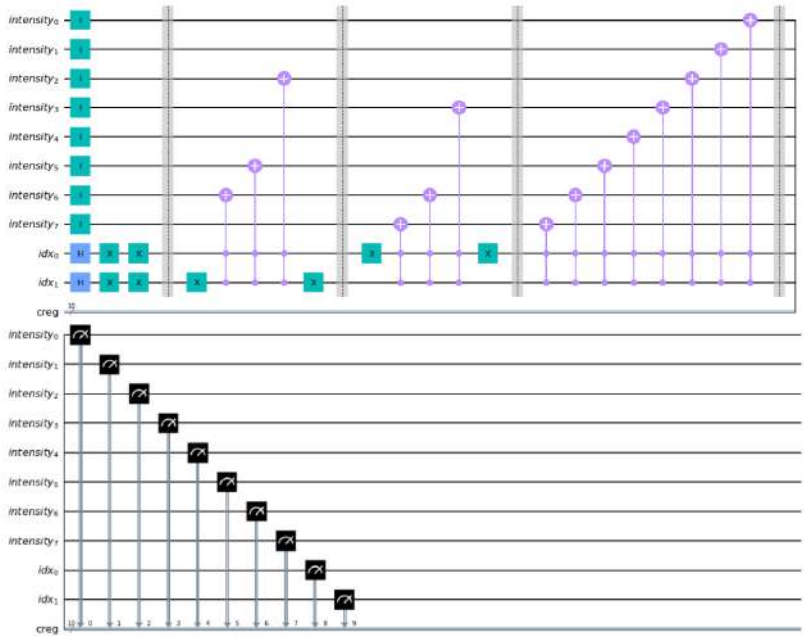


Рис. 6. Квантова схема для перетворення зображення в модель NEQR

Перший піксель на рис. 6 має інтенсивність 0, яка в NEQR представлена бітовим рядком |00000000⟩. Це кодується за допомогою восьми кубітів. Аналогічно, значення інших пікселів задаються за допомогою гейтів CC-NOT у різних позиціях схеми залежно від значення інтенсивності пікселів. В кінці схеми десять класичних регістрів використовуються для вимірювання вихідного сигналу. Всі пікселі кодуються відповідними значеннями інтенсивності. Вихід схеми – десятикубітний бітовий рядок, в якому перші два біти представляють інформацію про положення, а решта вісім бітів – інформацію про колір пікселів.

Технології безпеки на основі квантової обробки зображень

Технологію квантової обробки зображень побудовано на розширенні цифрової обробки зображень до області квантових обчислень, що призводить до реалізації безпечних, ефективних і передових технологій для криптографії та приховування інформації.

На рис. 7 представлено загальну схему квантових технологій захисту зображень в рамках цих двох широких областей.

Рис. 7. Технології квантового захисту зображень

В науці криптографії шифрування розглядається як процес прямого приховування інформації, щоб зробити її нечитабельною без спеціальних знань. Зазвичай це робиться для

збереження таємниці і, як правило, для конфіденційних комунікацій. Криптографія спрямована на захист змісту повідомлень, тоді як приховування інформації фокусується на приховуванні самого факту їхнього існування. Приховування інформації за допомогою таких стратегій, як стеганографія та водяні знаки, видається більш безпечною, оскільки такі методи не так легко помітити зловмисникам. Серед основних обмежень цих методів є лімітований об'єм інформації для передачі. Адже кількість інформації залежить від розмірів контейнера-носія і алгоритму вбудовування, також зображення-носій після додавання прихованого повідомлення не повинно мати видимих спотворень.

Алгоритми для квантової стеганографії зображень

Як було сказано раніше, стеганографія зображень – це метод приховування інформації орієнтований на приховування секретного повідомлення в зображенні-носії [6]. На рис. 8 наведено загальну схему протоколів квантової стеганографії зображень, а решта цього розділу висвітлює деякі досягнення на їх основі.

Рис. 8. Схема квантової стеганографії зображень

У 2014 р. Цзян та ін. запропонували стратегію стеганографії зображень NEQR на основі муарових моделей [7]. Стратегія була розроблена як стеганографічний алгоритм з відповідними квантовими схемами для приховування двійкового зображення у відтінках сірого.

Алгоритм вбудовування починається з вибору початкової решітки муару, тобто стохастичного зображення, як зображення прикриття. Потім початкова муарова решітка модифікується відповідно до секретного зображення, і деформована розглядається як муаровий шаблон. Зрештою, зображення муару змінюється для отримання стего-зображення.

Після цього дослідження була розроблена вдосконалена версія з використанням двох сліпих алгоритмів стеганографії з використанням найменшого значущого біта (LSB) на основі NEQR представлення [7]. Перший алгоритм базується на стандартному (або простому) LSB, який використовує кубіти повідомлень для безпосередньої заміни LSB пікселів.

Хоча стандартна система стеганографії LSB є простою, її стійкість є низькою. Інший алгоритм – блоковий LSB, вбудовує кубіт повідомлення в декілька пікселів, які належать до одного блоку зображення. Стеганографічна схема блочного LSB має на меті покращити стійкість стандартної схеми LSB. Це досягається шляхом розбиття зображення обкладинки на блоки, кожен з яких приховує один кубіт повідомлення замість пікселя. Експериментальні результати, представлені в цьому дослідженні, демонструють що невидимість є хорошою, а баланс між пропускною здатністю і надійністю можна регулювати відповідно до потреб додатків.

Алгоритм простого LSB вбудовування

Припустимо, що зображення контейнеру є $2^n \times 2^n$ квантовим зображенням $|I\rangle$ з діапазоном сірого 2^q (як визначено у рівнянні (5)), а повідомлення – $2^n \times 2^n$ двійкове квантове зображення $|M\rangle$, як показано нижче:

$$|M\rangle = \frac{1}{2^n} \sum_{i=0}^{2^{2n}-1} |m_i\rangle \otimes |i\rangle, \tag{6}$$

де $m_i \in \{0,1\}, i = 0,1,\ldots,2^{2n}-1$.

Схема вбудовування простого алгоритму LSB представлена на рис. 9, в якій $2n$ вентилів CNOT використовуються для перевірки того, чи збігається інформація про положення $|I\rangle$ та $|M\rangle$. Якщо інформація про позицію ідентична, то інформація про позицію $|M\rangle$ змінюється на $|00\ldots0\rangle$. Таким чином, під їхнім контролем LSB для $|I\rangle$ (тобто, $|C_i^0\rangle$) міняється місцями з кубітом повідомленням $|m_i\rangle$ для отримання стего-зображення $|I'\rangle$.

Схема вилучення показана на рис. 9, *б*, де $2n$ вентилів Адамара використовуються для перетворення початкового стану (тобто послідовності $|0\rangle$) у порожнє зображення. Аналогічно зі схемою вбудовування, коли інформація про положення $|I\rangle$ та $|M\rangle$ однакова, LSB для $|I\rangle$ міняється місцями з повідомленням кубіта $|m_i\rangle$ для отримання повідомлення $|M\rangle$.

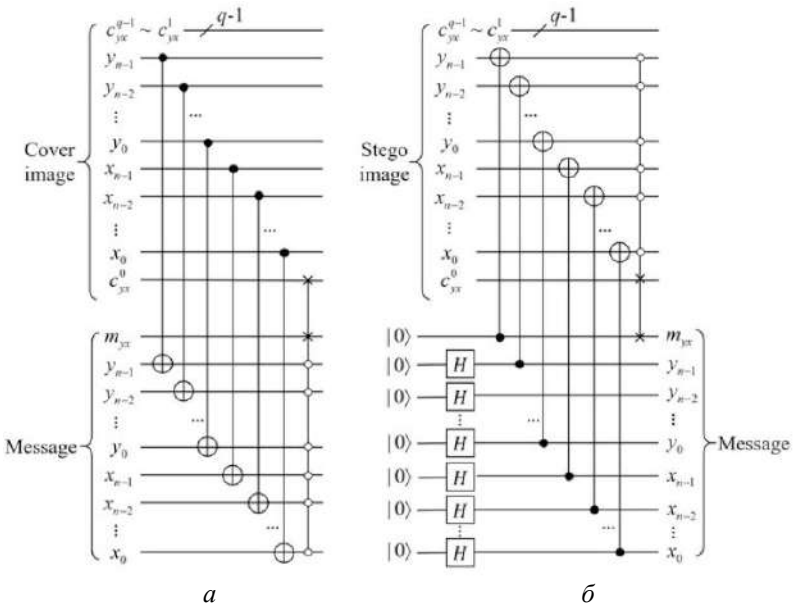


Рис. 9. Квантові схеми реалізації LSB стеганографії (*а* – вбудовування, *б* – вилучення)

Алгоритм блочної стеганографії LSB

Хоча алгоритм простої LSB-стеганографії простий, він має низьку стійкість [8–10]. Для покращення стійкості та непомітності схеми LSB, алгоритм блокової стеганографії розбиває зображення контейнера на блоки, в кожному з яких (замість кожного пікселя) ховається одне повідомлення довжиною в кубіт. Насправді, звичайну LSB-стеганографію можна розглядати як окремий випадок блокової LSB, в якій кожен блок вміщує лише один піксель.

Для реалізації схеми вбудовування та вилучення даних потрібно використовувати додатково квантовий лічильник та компаратор.

Квантовий лічильник.

Схему квантового лічильника [8] показано на рис. 10, де $|b\rangle$ – кубіт на вході і $b \in \{0,1\}$. $|a_{n-1}\ldots a_1 a_0\rangle$ – лічильник з початковим значенням $|00\ldots0\rangle$. Якщо вхідний кубіт $|b\rangle$ дорівнює $|1\rangle$, то $|a_{n-1}\ldots a_1 a_0\rangle$ збільшується на 1, інакше $|a_{n-1}\ldots a_1 a_0\rangle$ залишається без змін.

Квантовий компаратор.

Схему квантового компаратора [9] показано на рис. 11. Компаратор порівнює a та b, де $|a\rangle = |a_{n-1}\ldots a_1 a_0\rangle$ та $|b\rangle = |b_{n-1}\ldots b_1 b_0\rangle$, $a_i, b_i \in \{0,1\}$, $i = 0,1,\ldots,n-1$. Кубіти $|e_1\rangle$ та $|e_0\rangle$ є вихідними даними. Якщо $e_1 e_0 = 10$, то a > b; якщо $e_1 e_0 = 01$, то $a < b$; і якщо $e_1 e_0 = 00$, то $a = b$.

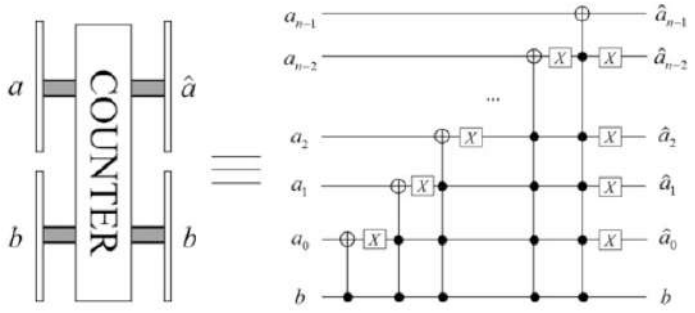


Рис. 10. Квантова схема реалізації лічильника

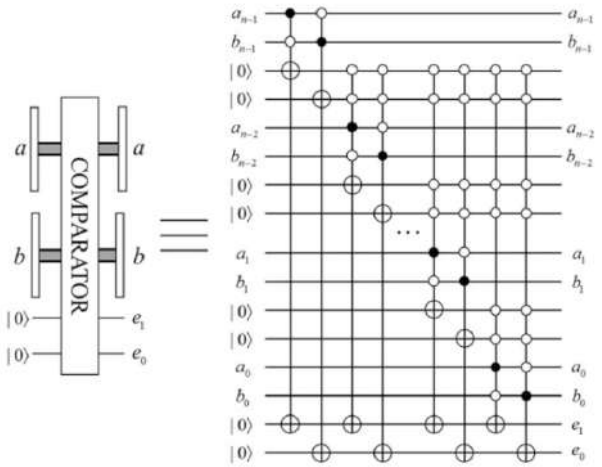


Рис. 11. Квантова схема реалізації компаратора

Процедура вбудовування блоків.

У блоковій схемі LSB зображення контейнеру $|I\rangle$ розміром $2^n \times 2^n$ має бути розбите на блоки $2^{n-p_1} \times 2^{n-p_2}$, де кожен блок має розмір $2^{p_1} \times 2^{p_2}$, де $p_1, p_2 \in \{0,1\}$. Блок зображення $|B\rangle$ можна визначити наступним чином:

$$\left|B_{k,l}\right\rangle = \frac{1}{2^n} \sum_{k=0}^{2^n-1} \sum_{l=0}^{2^n-1} \left|b_{k,l}\right\rangle \otimes |kl\rangle, \tag{7}$$

де $|k\rangle = \left|y_{n-1}, y_{n-2}, \ldots, y_{p_1}\right\rangle$, $|l\rangle = \left|x_{n-1}, x_{n-2}, \ldots, x_{p_2}\right\rangle$.

Припустимо, що повідомлення є двійковим квантовим зображенням, як показано в рівнянні (6). Його розмір $2^{n-p_1} \times 2^{n-p_2}$, а інформація про колір $p_1, m_{k,l} \in \{0,1\}$, де

$|k\rangle = |y_{n-1}, y_{n-2}, \ldots, y_{p_1}\rangle$ та $|I\rangle = |x_{n-1}, x_{n-2}, \ldots, x_{p_2}\rangle$. Процедура вбудовування полягає у наступному [7]:

К р о к 1 : Зображення контейнеру $|I\rangle$ зашифровано, щоб підвищити його непомітність у схемі. Для цього використовується квантовий алгоритм скремблювання зображень Гільберта.

К р о к 2 : Якщо інформація про положення $|y_{n-1}, y_{n-2}, \ldots, y_{p_1}\rangle$ $|x_{n-1}, x_{n-2}, \ldots, x_{p_2}\rangle$ в $|I\rangle$ дорівнює інформації $|M\rangle$, то операція вбудовування міняє місцями LSB $|I\rangle$ (тобто, $|C^0_{yx}\rangle$ і кубіт повідомлення $|m_{k,l}\rangle$.

К р о к 3 : Для відновлення зашифрованого зображення використовується обернене гільбертово скремблювання зображення.

Процедура вилучення блоків.

Після вбудовування кожного кубіта повідомлення 2^p разів (де $p = p_1 + p_2$) стегозображення може бути атаковано зловмисниками, що може змінити деякі LSB значення. Це призведе до того, що сума LSB всіх пікселів, що належать до одного блоку буде дорівнювати не 0 або 2^p, а якомусь значенню між ними.

Визначення того, чи дорівнює витягнутий кубіт повідомлення 0 або 1 відповідно до значення суми полегшується встановленням порогового значення. Якщо сума більша або дорівнює порогу, то кубіт повідомлення дорівнює 1, інакше він дорівнює 0. Процедура вилучення відбувається наступним чином [7]:

К р о к 1 : Повторює Крок 1 процедури вбудовування.

К р о к 2 : Схема керування використовується для розділення стегозображення на $2^{n-p_1} \times 2^{n-p_2} = 2^{n-p_1-p_2}$ блоків. Крім того, схема включає $2^{n-p_1-p_2}$ керуючих шарів, кожен з яких відповідає одному блоку зображення.

К р о к 3 : Квантові лічильники використовуються для підсумовування всіх LSB пікселів, які належать до одного блоку. Блок містить $2^{n-p_1-p_2}$ лічильників, а числа підрахунку представлені у вигляді $a_{y_{n-1}, y_{n-2}, \ldots, y_{p_1}\ x_{n-1}, x_{n-2}, \ldots, x_{p_1}}$.

К р о к 4 : Оскільки кожен блок складається з 2^p пікселів, число, отримане на кроці 3, слід порівняти з 2^{p-1}, що є порогом T, який встановлюється за допомогою квантового компаратора. Якщо $a_{y_{n-1}, y_{n-2}, \ldots, y_{p_1}\ x_{n-1}, x_{n-2}, \ldots, x_{p_1}} \geq 2^{p-1}$, то вилучене повідомлення дорівнює 1, інакше витягнуте повідомлення дорівнює 0.

П р и к л а д

Розглянемо просте зображення контейнеру розміром 4×4 та 8-бітне повідомлення 00110110, яке є ASCII-кодом символу "6" (див. рис. 12), зображення контейнеру розбивається на вісім блоків розміром 1×2 (у цьому випадку у рівнянні (7) $n = 2$, $p_1 = 0$, $p2 = 1$), і повідомлення перебудовується у двійкове зображення 4×2, як показано на рис. 12.

На рис. 13 показано схему вбудовування блоку LSB, яка складається з трьох частин, що відповідають трьом етапам, описаним вище. Перша та третя частини виконують гільбертове скремблювання зображення та його зворотну операцію.

Блок-схема вилучення LSB (на рис. 14) складається з чотирьох частин, які відповідають чотирьом етапам, описаним вище. Частина гільбертового скремблювання є такою ж, як і в операції вбудовування. Модуль розділення – це схема керування, яка визначає, який лічильник $C^0_{y_1 y_0 x_1 x_0}$ використовується. Наприклад, якщо керуючим значенням є 000, то $C^0_{000 x_0}$ міняється місцями з першим допоміжним кубітом $|0\rangle$, тобто входить у перший лічильник.

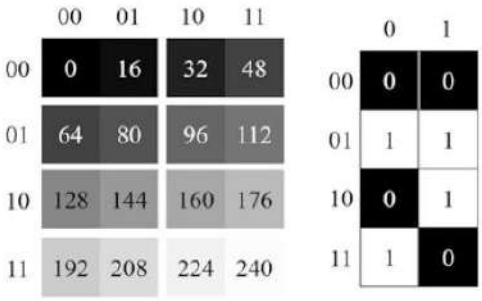


Рис. 12. Зображення-контейнер та повідомлення

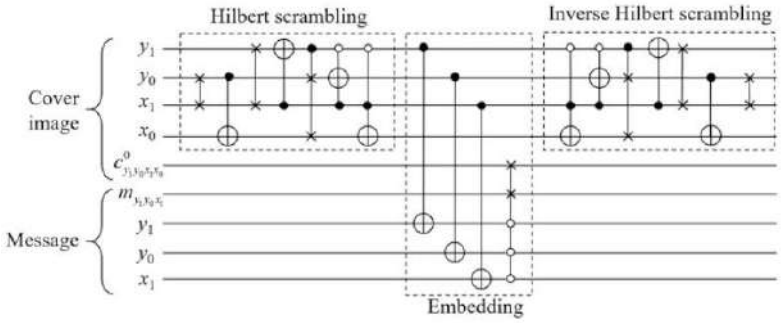


Рис. 13. Приклад блочного LSB вбудовування

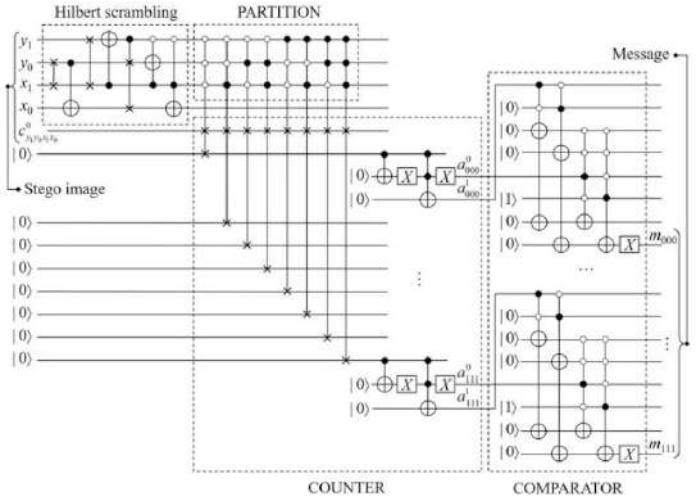


Рис. 14. Приклад вилучення даних для блочного LSB

Лічильний модуль складається з $2^{n-p_1} \times 2^{n-p_2} = 8$ лічильників, які відповідають восьми блокам. Кожен лічильник $a_{y_1y_0x_1}$ підсумовує LSB пікселів блоку $B_{y_1y_0x_1}$. Оскільки кожен блок має два пікселі, максимальне значення $a_{y_1y_0x_1}$ дорівнює 2, тому достатньо двох кубітів, тобто $\left|a_{y_1y_0x_1}\right\rangle = \left|a^1_{y_1y_0x_1} a^0_{y_1y_0x_1}\right\rangle$.

Крім того, частина порівняння містить $2^{n-p_1} \times 2^{n-p_2} = 8$ компараторів, які порівнюють $a_{y_1y_0x_1}$ з порогом $2^{p-1} = (01)_2$. Як показано вище, якщо $a_{y_1y_0x_1} > 01$, то молодші два кубіти кожного компаратора дорівнюють 10, і так далі. Отже, потрібно лише інвертувати нижній кубіт, щоб отримати кубіт повідомлення $\left|m_{y_1y_0x_1}\right\rangle$.

Висновки

Квантові водяні знаки – це область квантового приховування інформації, що швидко розвивається. Квантові зображення забезпечують міцну основу для цієї галузі. Різні моделі представлення квантових зображень мають різні переваги. Вибір відповідної моделі представлення квантових зображень для конкретних процесів квантової обробки зображень може суттєво вплинути на ефективність та результативність. В роботі ми представили декілька основних моделей квантових зображень, розглянули їх переваги та недоліки відносно використання в задачах стеганографії.

Детально зупинились на двох алгоритмах стеганографії на основі LSB для квантових зображень. Вони відрізняються тим, чи вбудовується кубіт повідомлення в піксель або блок контейнера-зображення. Обидва алгоритми є сліпими, тобто процедура вилучення не потребує оригінального зображення або оригінального повідомлення. Аналіз і моделювання на основі експериментальних результатів показують, що невидимість алгоритмів є гарною, і існує природний компроміс між їхньою пропускною здатністю та стійкістю.

Список літератури:

1. Le P. Q., Dong F. and Hirota K. A flexible representation of quantum images for polynomial preparation, image compression, and processing operations // Quantum Information Processing. 2010. Vol. 10. P. 63–84.
2. Venegas-Andraca S. and Bose S. Storing, processing, and retrieving an image using quantum mechanics // Proc. SPIE Conf. Quantum Information and Computation. 2003. P. 134–147.
3. Latorre J. I. Image Compression and Entanglement. [Електронний ресурс] Режим доступу: arXiv: quant-ph/0510031, 2005.
4. Zhang Y., Lu K., Gao Y. and Wang M. Neqr: a novel enhanced quantum representation of digital images // Quantum Information Processing. 2013. Vol. 12. P. 2833–2860.
5. Zulehner A., Wille R. Simulation and Design of Quantum Circuits // Ulidowski I., Lanese I., Schultz U.P., Ferreira C. (eds) Reversible Computation: Extending Horizons of Computing. Lecture Notes in Computer Science. 2020. Vol 12070. Springer, Cham. https://doi.org/10.1007/978-3-030-47361-7_3.
6. Gill S.S., Kumar A., Singh H., et al. Quantum computing: A taxonomy, systematic review and future directions // Softw: Pract Exper. 2022; 52(1):66-114. doi:10.1002/spe.3039.
7. Jiang N., Zhao N., Wang L. LSB based quantum image steganography algorithm // Theoret. Phys. 2016. 55(1). P.107–123.
8. Ma L., Lu J. Construction of controlled quantum counter // Chin. J. Quantum Electr. 2003. 20(1). P. 47–50.
9. Wang D., Liu Z., Zhu W., Li S. Design of quantum comparator based on extended general Toffoli gates with multiple targets // Comput. Sci. 2012. 39(9). P. 302–306.
10. Zhou RG., Luo J., Liu X. et al. A Novel Quantum Image Steganography Scheme Based on LSB // Theor Phys. 2018. 57. P.1848–1863. https://doi.org/10.1007/s10773-018-3710-x.

Надійшла до редколегії 10.09.2024

Відомості про авторів:

Федюшин Олександр Іванович – канд. техн. наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри безпеки інформаційних технологій, факультет комп'ютерної інженерії та управління; Україна; e-mail: oleksandr.fediushyn@nure.ua; ORCID: http://orcid.org/0000-0002-3600-405X

Головко Євген Вікторович – Харківський національний університет радіоелектроніки, аспірант кафедри безпеки інформаційних технологій; Україна; e-mail: yevhen.holovko1@nure.ua; ORCID: https://orcid.org/0009-0000-9684-7369

Смірнов Антон Олександрович – канд. техн. наук, Харківський національний університет радіоелектроніки, доцент кафедри безпеки інформаційних технологій, факультет комп'ютерної інженерії та управління; Україна; e-mail: anton.smirnov@nure.ua, ORCID: https://orcid.org/0000-0003-4121-3902

Сухотеплий Владислав Миколайович – Харківський національний університет Повітряних Сил імені Івана Кожедуба, старший викладач кафедри радіоелектронних систем пунктів управління Повітряних Сил, Україна; e-mail: vladislav181168@gmail.com; ORCID: https://orcid.org/0000-0002-2366-4167

Чечуй Олександр Вікторович – канд. техн. наук, доцент, Харківський національний університет Повітряних Сил імені Івана Кожедуба, доцент кафедри радіоелектронних систем пунктів управління Повітряних Сил; Україна; e-mail: alche1972@ukr.net; ORCID: https://orcid.org/0000-0002-7584-4457

УДК 621.3.095.221 **DOI:10.30837/rt.2024.3.218.04**

В.І. ЗАБОЛОТНИЙ, канд. техн. наук, А.М. ОЛЕЙНІКОВ, канд. техн. наук,
Д.М. ЗАБОЛОТНИЙ, А.К. КУСТОВ

ТЕХНІЧНИЙ КАНАЛ ВИТОКУ ІНФОРМАЦІЇ ПОБІЧНИМИ ЕЛЕКТРОМАГНІТНИМИ ПЕРЕВИПРОМІНЮВАННЯМИ ДОПОМІЖНИХ ТЕХНІЧНИХ ЗАСОБІВ І СИСТЕМ

Вступ

Проєктування та впровадження комплексу технічного захисту інформації (ТЗІ) на об'єкті електронно-обчислювальної техніки (ЕОТ) потребує виявлення технічних каналів витоку інформації (ТКВІ), що підлягають захисту. При проведенні передпроєктних робіт на об'єкті ЕОТ визначаються якісні моделі всіх ТКВІ, які пізніше, за потребою, досліджуються як кількісні. Зміст і параметри моделей ТКВІ залежать від: конкретного складу основних технічних засобів (ОТЗ), призначених для обробки інформації з обмеженим доступом (ІзОД); переліку допоміжних засобів і систем (ДТЗС), що знаходяться поряд з ОТЗ у виділеному приміщенні (ВП); архітектурно-будівельних особливостей споруди з об'єктом ЕОТ; контрольованої зони (КЗ), елементів місцевості, що оточує об'єкт ЕОТ, тощо.

Більшість типових якісних і кількісних моделей ТКВІ широко відомі [1, 2]. Існують також ТКВІ, які не часто зустрічаються у практиці діяльності фахівців з комплексів ТЗІ. Наприклад, канал побічних електромагнітних випромінювань (ПЕМВ) ДТЗС, що утворюється шляхом перехоплення приймачами засобів технічної розвідки (ТР) за межами КЗ небезпечних сигналів у вигляді побічних електромагнітних полів ОТЗ, які перевипромінюються ДТЗС [1]. В ряді випадків дослідження небезпеки таких ТКВІ є актуальною задачею [3]. Її рішення може проводитися в ході інженерного аналізу ДТЗС або їх спеціального дослідження [1].

Особливістю різновиду ТКВІ, що розглядається в статті, є його початок формування в зоні 1 ОТЗ. Зона 1 – територія (сфера) навколо основних технічних засобів, в межах якої здійснюється наведення небезпечних сигналів на інші технічні засоби, системи та їх комунікації. Зона 1 характеризується радіусом R_1, що визначає граничну відстань від ОТЗ до межі, за якою вважається неможливим наведення небезпечних сигналів на технічні засоби [1]. Даний випадок взаємного розташування ОТЗ, ДТЗС та засобу ТР не передбачає наявності гальванічного зв'язку між ДТЗС з засобом ТР. Засіб ТР реалізується як радіоприймальний пристрій, що сприймає перевипромінювання ДТЗС. Розташування засобу ТР при цьому – Зона 2.

Зона 2 – територія (сфера) навколо технічних засобів обробки інформації, за межами якої вважається неможливим перехоплення небезпечного сигналу з метою відтворення інформації [1]. Характеризується радіусом R_2, що визначає найбільшу відстань від технічних засобів обробки інформації до межі, за якою напруженості електричного та магнітного полів небезпечного сигналу відносно шумових завад не перевищують нормованого значення. В Зоні 2 можливе перехоплення інформації, а за її межами – ні.

Стаття призначена для розробки підходів до методичного забезпечення процесу створення і застосування кількісної моделі ТКВІ, що утворюється шляхом перехоплення приймачами засобів ТР за межами КЗ небезпечних сигналів у вигляді побічних електромагнітних полів зони 1 ОТЗ, які перевипромінюються ДТЗС.

Якісна модель ТКВІ, що утворюється шляхом перехоплення приймачами засобів ТР за межами КЗ небезпечних сигналів у вигляді побічних електромагнітних полів ОТЗ, які перевипромінюються ДТЗС

В літературі ТКВІ, що досліджується в цій роботі, описаний як різновид каналів ПЕМВ, який утворюється шляхом перехоплення за межами КЗ приймачами засобів ТР небезпечних

сигналів у вигляді побічних електромагнітних полів ОТЗ, які перевипромінюються ДТЗС (рис. 1) [1]. Причому ці ДТЗС повинні знаходитися в зоні 1 ОТЗ.

ДТЗС, якщо вони знаходяться в зоні 1 ОТЗ, є випадковими антенами (ВА) і можуть призводити до витоку інформації небезпечними сигналами, наведеними на них ПЕМВ. На основі цієї якісної моделі ТКВІ розробляється кількісна модель, що дозволяє оцінювати рівень загроз витоку ІзОД.

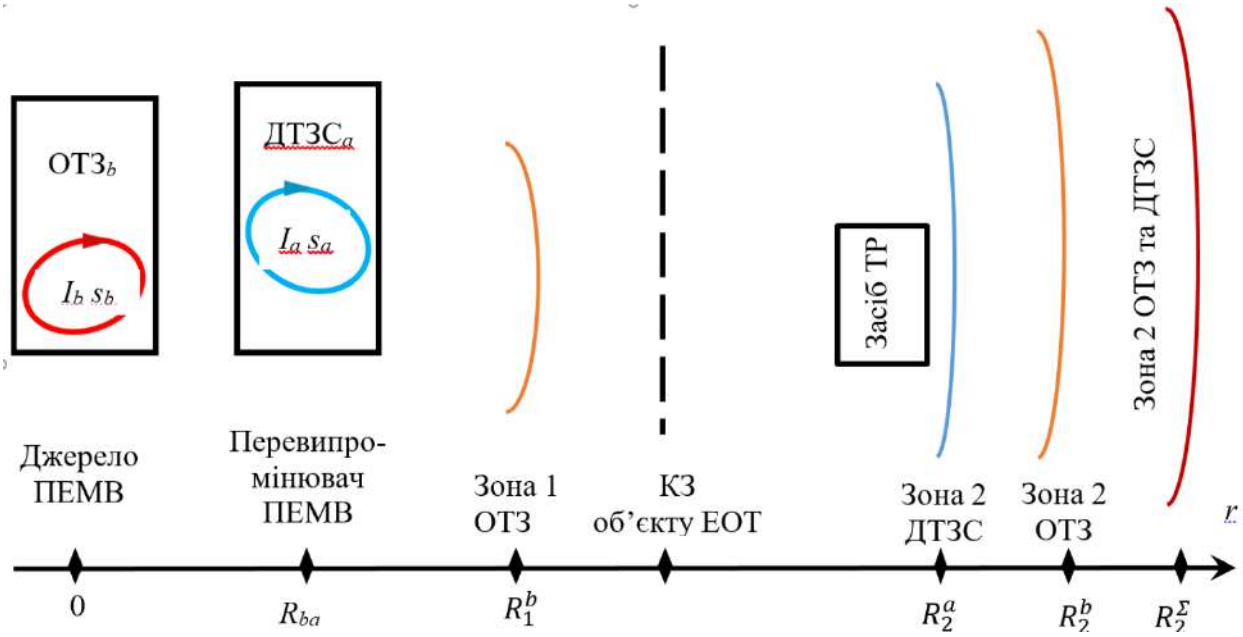


Рис. 1. Ситуаційна схема дослідження ТКВІ перевипромінення

Кількісна модель ТКВІ, досліджувана в статті

Кількісна модель ТКВІ призначена для оцінки можливості витоку інформації, яку створює конкретна ситуація застосування ОТЗ та ДТЗС об'єкту ЕОТ. Математичний апарат моделі базується на теорії електромагнітного поля, випромінювання, розповсюдження і прийому радіосигналів. Крім того, в формуванні моделі ТКВІ використовуються прийняті в практиці ТЗІ визначення, підходи, теоретичні положення розповсюдження електромагнітних випромінювань, показники можливого витоку інформації тощо.

Показники основних параметрів можливого витоку інформації каналами ПЕМВ [4]:

E – напруженість електричного поля інформативного (небезпечного) сигналу;

H – напруженість магнітного поля інформативного (небезпечного) сигналу.

В ході проведення досліджень за відповідними методиками визначаються радіуси R_1 та R_2, за межами яких показники менше гранично допустимих величини – норм захисту E_N та H_N.

Така кількісна модель ТКВІ складається з окремих моделей, застосування яких має певні загальні і окремі особливості.

Потенційні можливості ТКВІ

В практиці ТЗІ принципово недопустимо зменшувати оцінку можливості апаратури розвідки по одержанню інформації. Тому при виборі вихідних даних для розрахунків необхідно, по перше, визначати їх точно (при інструментальних вимірюваннях або теоретичних обґрунтуваннях). По друге, у випадках невизначеності брати для підстановки граничні значення, найбільш несприятливі для сторони захисту ІзОД. Такий підхід до вибору даних доцільно віднести до застосування принципу потенційно-можливих ТКВІ (далі – ПМТКВІ). Наприклад, через невизначеність поляризаційних характеристик ПЕМВ ОТЗ і антени технічного засобу їх радіорозвідки признати, що вони співпадають. Тоді поляризаційний коефіцієнт прийому в радіоканалі розвідки взяти максимально можливим ($\gamma = 1$). Такі припущення

будуть позначатися при застосуванні відміткою ПМТКВІ. Це дасть можливість акцентувати увагу на більш прискіпливе спостереження за поведінкою цих параметрів при експлуатації ОТЗ та ДТЗС, кореигувати заходи захисту від витоку інформації, що передбачено оперативним вирішенням задач ТЗІ, викладених у ДСТУ 3396.1-96.

Модель джерела ТКВІ

Теорія електромагнітного поля розглядає декілька класичних моделей випромінювачів електромагнітних полів [5], але в реальних системах, які описують канал ПЕМВ, здебільшого зустрічаються лише деякі з них – рамка та диполь. Причому диполь, як джерело ПЕМВ ОТЗ, характерний для варіанту нештатної роботи апаратури (обрив інформаційного провідника) і для подальшого аналізу неактуальний на відміну від перевипромінювача ДТЗС рамки.

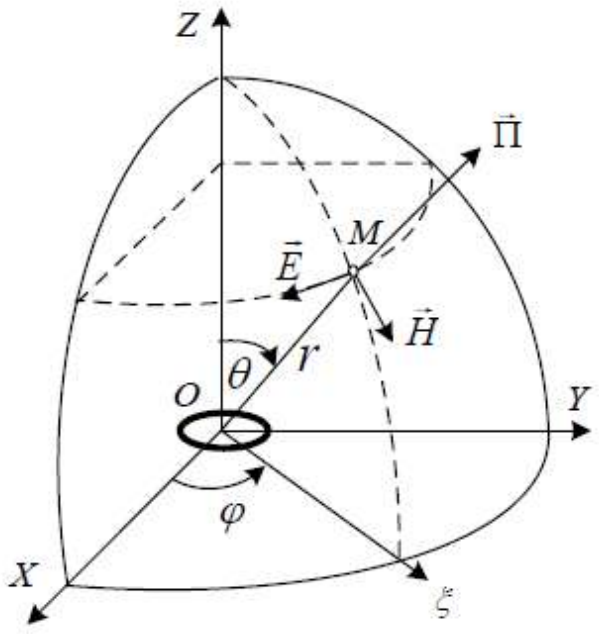


Рис. 2. Схема орієнтації складових поля випромінювання рамки зі струмом

Рамка (рис. 2) – модель реальної електричної схеми ОТЗ кола перетворювання (підсилення) небезпечного сигналу, яка має певні фізичні розміри та орієнтацію у просторі. Навколо рамки за законами електродинаміки створюється електромагнітне поле [5].

Рішення рівнянь Максвела для елементарного рамкового випромінювача відомі для гармонійного струму частоти ω й амплітуди струму *I*. При цьому повинна виконуватись умова квазістаціонарності – фаза гармонійного коливання в рамці скрізь однакова. Це еквівалентно співвідношенню: загальна довжина провідників рамки – не більше, принаймні, половини довжини хвилі λ коливання в рамці. У найбільш небезпечних ПЕМВ відеотракту ПЕОМ тестові сигнали виду «піксель чорний – піксель білий» мають нижні частоти гармонік ПЕМВ порядку 10 – 60 МГц (довжина хвилі λ = 5 – 30 м). При необхідності частоти гармонік дослідження можна кратно знижувати у *n* разів використанням тестового сигналу «*n*-пікселів чорних – *n*-пікселів білих». На практиці для найбільшої досліджуваної частоти *f*=1ГГц загальна довжина провідників рамки не повинна перевищувати 15 см. Якщо рамка перевищує зазначені розміри, то розрахунки випромінювань необхідно виконувати по напівемпіричним формулам теорії антен. Цей варіант лежить за межами даного дослідження.

Модель розповсюдження ПЕМВ у зонах 1 та 2

Рішення рівнянь Максвелла [5] у комплексній формі для векторної складової поля показує зв'язок між амплітудою, частотою та фазою гармонійного струму в рамці з аналогічними параметрами складових електромагнітного поля в обраній точці простору. Ілюстрація наведена в полярній системі координат (рис. 2).

З метою наступного аналізу, спрощення й наочності одержання результатів варто систему рівнянь Максвела записати у виді

$$\begin{cases} E_\varphi = \omega^2 \frac{IsW}{4\pi r c^2}\left[1 - i\frac{1}{kr}\right], \\ H_r = i\omega \frac{Is}{2\pi r^2 c}\left[1 - i\frac{1}{kr}\right], \\ H_\theta = \frac{Is}{4\pi r^2 c}\left[1 + i\frac{1}{kr} - \frac{1}{(kr)^2}\right]. \end{cases} \quad (1)$$

де ω – кутова частота; W – хвильовий опір простору (W=120π); r – відстань до точки спостереження; ϑ – меридіональний кут на точку спостереження; s – площа рамки; c – швидкість світла; I – комплексна амплітуда струму в рамці.

При виведенні (1) використано співвідношення для хвильового числа $k = 2\pi/\lambda = \omega/c$ та хвильового опору простору $W=\mu c$, де μ – магнітна стала простору.

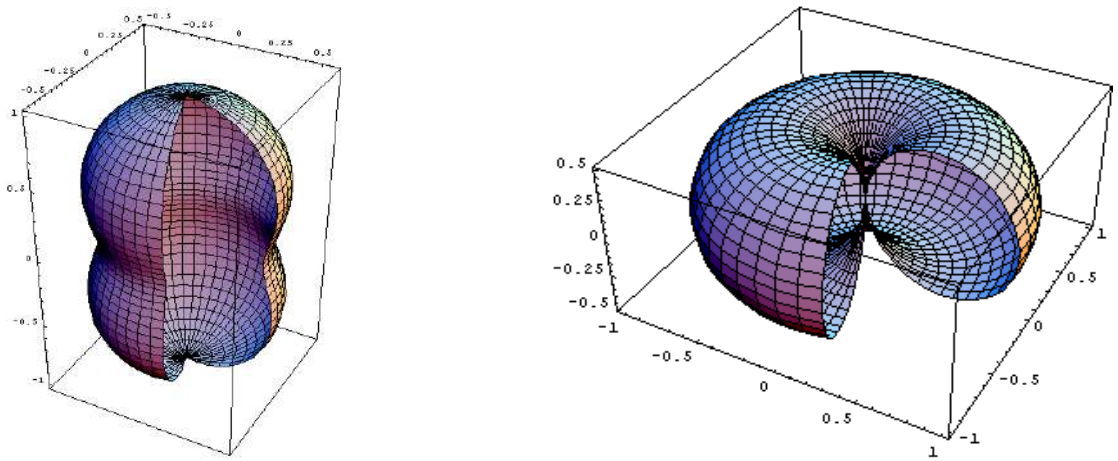
Крім того, на рис. 1 вектор H_r співпадає з напрямом вектору Пойнтінга $П$. Хвильовий множник $e^{i(\omega t - kr)}$опущений. Оскільки φ азимутальний кут ніде не фігурує, це означає, що поле симетричне щодо осі OZ.

Після нескладних перетворень при $k = 2\pi/\lambda = \omega/c$ можна одержати систему рівнянь для зони 1, з урахуванням об'єднання радіальної та тангенціальної складових магнітного поля:

$$H \approx \frac{Is}{4\pi r^3 c}\sqrt{1 + 3cos^2\theta} \quad (2)$$

$$E \approx \frac{IsW}{4\pi r^2 c}\sin\theta \quad (3)$$

В рівняннях (2), (3) множник з кутом θ характеризує спрямованість ПЕМВ у просторі (рис. 3).



а – магнітна складова ПЕМВ

б – електрична складова ПЕМВ

Рис. 3. Діаграма спрямованості ПЕМВ у ближній зоні.
Положення рамки – горизонтальне, відповідає рис. 2

Область зони 2 відповідає дальній зони рівнянь Максвела (kr >>1).

Вираз для зони 2 можна одержати з (1) аналогічно визначенню складових поля ближньої зони.

Після спрощення виразів (1) за умовою (kr >>1) можна одержати:

$$H_\alpha \approx \frac{Is}{4\pi r c^2} sin\theta. \quad (4)$$

У дальній зоні – зоні 2 ЕМП має дві синфазні складові – електричну E_α та магнітну H_θ, величини яких співвідносяться, як $E_\alpha = H_\theta * W$.

Модель випромінювача ОТЗ

Для обраного варіанту досліджень магнітного поля випромінювань ОТЗ[1] приймається рамка зі струмом I_b еквівалентною площею s_b. Частота, на якій проводиться інструментальна оцінка зон 1 та 2, – ω. Як правило, конкретні значення I_b та s_b безпосередньо не оцінюються. Радіуси ПЕМВ зон 1 та 2 (R_1 , R_2) визначаються аналітично за відповідними методиками після проведення інструментальних вимірювань. Зазначені величини R_1 і R_2 відповідної частоти ω фіксуються в проєктній та експлуатаційній документації на об'єкт ЕОТ.

Модель перевипромінювання ПЕМВ

Як визначено в описі досліджуваного ТКВІ, в зоні 1 ПЕМВ ОТЗ знаходиться ДТЗС певної конструкції (являє собою сукупність електричних провідників власне ДТЗС, меблів, будівельних конструкцій, предметів інтер'єру, зовсім необов'язково електричних схем). Ці провідники можуть займати певний простір в конструкції ДТЗС. По формі провідники можуть бути замкненими у формі кільця еквівалентної площі s_a. Провідники у формі кільця реагують на магнітну складову ПЕМВ, в них створюється відповідний струм I_a. Оці два фактори s_a та I_a , в свою чергу, згідно з принципом двоїстості, формують магнітне поле перевипромінення, яке за умовами, визначеними у задачі дослідження, слід оцінити для дальньої зони 2.

Розімкнені провідники довжиною l_a реагують на електричну складову поля E і у даній роботі не досліджуються за визначеними у Моделі джерела ТКВІ причинам. В перспективі на даний випадок теж можна буде звернути увагу.

Кількісну модель зазначеного перевипромінювача доцільно скласти на основі повітряного трансформатора (без осердя) [6], первинна обмотка якого складається з одного витка проводу електричної схеми ОТЗ. Площа витка s_b, струм небезпечного сигналу I_b. Частина силових ліній магнітного поля первинного витка створює потокозчеплення з вторинним витком. Це потокозчеплення називають взаємною індуктивністю первинного та вторинного витків.

Особливістю такого повітряного трансформатора є значне поле розсіювання, що обумовлено достатньо довільним взаємним розташуванням осей витків та певної відстані R_{ba} між ними. Розсіювання поля первинної обмотки приводить до зниження коефіцієнту зв'язку між витками, зменшення сили наведеного струму I_a. Визначення частини магнітного поля, що проходить крізь площу витка s_a, дозволяє визначити зону 2 ДТЗС (радіус R_2^a).

Магнітний потік зони R_1^b, як було зазначено вище, вважається полем індукції. В місці розташування випадкової антени на відстані R_{ba}. Виходячи з принципу ПМТКВІ, магнітне поле індукції вважається радіальним. При наявності даних щодо орієнтації вторинного витка береться до уваги кут ζ_{ba} між взаємним положенням відповідної лінії магнітного поля зони 1 і вісі зазначеного центру вторинного витка. В цьому випадку ефективна площа вторинного витка визначається як $S_a cos\zeta_{ba}$. При відсутності інформації щодо орієнтації – застосовується принцип ПМТКВІ, тобто $cos\zeta_{ba}$ =1.

Зазначене дає змогу оцінити частину потоку магнітного поля ПЕМВ, що передається з ОТЗ на ВА ДТЗС. Це поле далі буде використовуватися на перевипромінювання і формувати відстань зони 2 ДТЗС. Вираз для цієї оцінки:

$$K = S_a cos\zeta_{ba} / 4\pi (R_{ba})^2 . \quad (5)$$

З виразу (2) витікає, що поле магнітної індукції H в зоні 1 по інтенсивності змінюється за законом $1/r^3$. Прирівнюючи значення поля магнітної індукції H нормі H_N, можна записати межу зони 1 R_1^b. В подальшому доцільно використовувати формулу про значення поля магнітної індукції H ОТЗ ближньої зони 1 через зазначені вище параметри:

$$H = \frac{B_1 \, I_b s_b}{r^3}, \quad (6)$$

[1] Надалі в позначках величин будуть використовуватися індекси: для ОТЗ – *b*, а для ДТЗС – *a*.

де B_1 – параметр, який враховує усі константи у виразі (2) та умови впливу середовища, що оточує ОТЗ на розповсюдження ЕМП у ближній зоні.

Визначення значення параметра B_1 можна зробити по одержаним даним щодо розміру зони 1 R_1^b, на межі якої виконується норма безпечного поля з її показником H_N. Після нескладних перетворювань можна скласти вираз через параметри R_1^b та H_N:

$$H = H_N \frac{\left(R_1^b\right)^3}{r^3}. \quad (7)$$

Таким чином, якщо ОТЗ створює магнітну складову ПЕМВ (7) в місці розташування рамки ВА ДТЗС площею s_a, то в ній буде індуктуватися струм I_a . Це і породжує поле перевипромінювання зони 2 радіус R_2^a, якої і потрібно оцінити.

Визначення параметрів джерела поля перевипромінення

У рамці ВА ДТЗС під дією поля випромінення ОТЗ наводиться е.р.с., величина якої складає, згідно з законом індукції Фарадея [7] в позначеннях даної статті:

$$\varepsilon = \mu H s_a cos\zeta. \quad (8)$$

Доречно застосувати принцип ПМТКВІ і позбутися множника $cos\zeta$, що не зменшує значення е.р.с. і спрощує запис для подальшого використання. Для визначення струму I_a, що діє в рамці, спочатку потрібно оцінити комплексне значення опору в ній, а потім одержати потрібний вираз:

$$I_a = \frac{\varepsilon}{\sqrt{(R_a+R_\Sigma)^2+(X)^2}}, \quad (9)$$

де R_a – тут «омічний» опір матеріалу витка ВА рамки ДТЗС з урахуванням «скін ефекту» на частоті ω (розраховується по [8]); R_Σ – опір випромінювання ВА рамки ДТЗС на частоті ω (розраховується по [7]); X – індуктивний опір рамки ВА рамки ДТЗС на частоті ω. X= ωL_a; L_a – індуктивність ВА рамки ДТЗС (розраховується по [8])

Як і раніше, до оцінки зазначених вище параметрів слід застосовувати принцип ПМТКВІ. При цьому слід брати до уваги, що знаменник (9) має параметри, значення яких достатньо складно точно оцінювати на реальних об'єктах ЕОТ при проведенні обстежень. Тому принцип ПМТКВІ, орієнтований на їх менше значення, може приводити до збільшення оцінки I_a, що, в свою чергу, приведе до деякого збільшення оцінки зони 2 ДТЗС. А це цілком вкладається в принципи застосування ПМТКВІ.

Узагальнена зона 2 ПЕМВ ОТЗ та ДТЗС

ПЕМВ в дальній зоні – зоні 2 складається з двох сигналів – прямого від випромінювача ОТЗ та того, що одночасно перевипромінюється ВА, сформованою конструкцією ДТЗС. Слід очікувати, що ці сигнали будуть складатися, від цього розмір зони 2 очікувано буде зростати. Оцінка збільшеного розміру зони 2 природньо являє собою важливу практичну задачу.

Визначити сумарну зону 2 ОТЗ та ДТЗС слід з урахуванням принципу ПМТКВІ. В його основу слід закласти синфазне, когерентне складання двох сигналів. Межу сумарної зони 2 визначити рівнем двох когерентно складених сигналів тоді, коли будуть мати загальний рівень H_N, який є нормою захисту по полю H. Позначення цьому – R_2^Σ .

Тут можливі два крайні випадки. По-перше, взаємне розташування ОТЗ, ДТЗС і точки оцінки складення сигналів на стороні ДТЗС (рис. 1) (позитивній напрям). По-друге, (від'ємний напрям) – точка оцінки складення сигналів, ОТЗ і ДТЗС при від'ємному значенню r (рис. 1).

Обидва сигнали монотонно зменшуються зворотньо-пропорційно відстані r від джерел. Для першого випадку це дає вираз

$$\frac{I_b s_b}{4\pi r c^2} + \frac{I_a s_a}{4\pi (r-R_{ba}) c^2} = H_N. \quad (10)$$

Можна спростити вираз (10) врахуванням визначення значень R_2 обох джерел на зразок формули (7), а саме:

$$H_N \frac{R_2^b}{r} + H_N \frac{R_2^a}{r - R^{ba}} = H_N \tag{11}$$

Даний вираз, після нескладних перетворень, приводить до рішення квадратного рівняння, що (після позбавлення від фізично неспроможного значення) має вигляд

$$R_2^{\Sigma+} = \frac{(R_2^b + R_2^a + R^{ba}) + \sqrt{(R_2^b + R_2^a + R^{ba})^2 - 4R_2^b R^{ba}}}{2}. \tag{12}$$

Якщо оцінювати випадок розташування ДТЗС за ОТЗ, слід змінити знак у (12) перед R^{ba} на протилежний і тоді

$$R_2^{\Sigma-} = -\frac{(R_2^b + R_2^a - R^{ba}) + \sqrt{(R_2^b + R_2^a - R^{ba})^2 + 4R_2^b R^{ba}}}{2}. \tag{13}$$

Тут знак мінус вказує на від'ємну координату на осі лінії розташування ДТЗС та ОТЗ.

Слід зауважити, що про форму зони 2 у вигляді сфери вже не йдеться. Але значення (12) та (13) обмежать форму фігури сумарної межі зони 2 у просторі.

Висновки

Проведені дослідження, що засновані на використанні відповідних розділів теорії електромагнітного поля, чинних організаційних документів в області ТЗІ, дозволили запропонувати обґрунтування для побудови окремих кількісних моделей ТКВІ, що описують процес можливого витоку інформації шляхом перевипромінення ПЕМВ ОТЗ магнітного поля, розташованими в зоні індукції ДТЗС.

Запропоновано використовувати принцип потенційних можливостей ТКВІ при складанні моделей ТКВІ, що дозволяє гарантовано визначати найперші умови витоку інформації.

Одержані результати доцільно використовувати в ході проєктування, впровадження і супроводу комплексів ТЗІ для розробки методик оцінки можливих ТКВІ.

Подальші дослідження

Доцільно дослідити кількісні моделі:

- витоку інформації за рахунок явища електричної індукції електричної складової поля ОТЗ на випадкову електричну антену ДТЗС;
- витоку інформації за рахунок перевипромінення електричною антеною ДТЗС, створеною магнітною антеною ОТЗ складової електричного поля, і навпаки.

Список літератури:

1. Технічні канали витоку інформації. Порядок створення комплексів технічного захисту інформації : навч. посіб. / С.О. Іванченко, О.В. Гавриленко, О.А. Липський, А.С. Шевцов. Київ : ІСЗЗІ НТУУ «КПІ», 2016. 104 с.

2. Засоби та системи технічного захисту інформації / І.Є. Антіпов, А.М. Олейніков, Ю.В. Ликов, В.Д. Кукуш, І.О. Милютченко. Харків : ХНУРЕ, 2019. 216 с.

3. Заболотний В.І., Ткач О.О. Аналіз впливу під'єднання допоміжних технічних засобів на характеристики побічних електромагнітних випромінювань основних технічних засобів // Дванадцята міжнар. наук.-техн. конф. «Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління», 27 – 28 квітня 2022 р. Баку – Харків – Жиліна, с. 144.

4. Тимчасові рекомендації з технічного захисту інформації від витоку каналами побічних електромагнітних випромінювань та наводок ТР ТЗІ – ПЕМВН-95.

5. Шокало В.М., Правда В.І., Усін В.А., Вунтесмері В.С., Грецьких Д.В. Електродинаміка та поширення радіохвиль. Ч. 1 ; за заг. ред. В.М. Шокало та В.І. Правди : підручник для студентів ВНЗ. Харків : ХНУРЕ ; Колегіум, 2009. 286 с.

6. Теорія електричних і магнітних кіл : підручник / С. В. Панченко, О. М. Ананьєва, М. М. Бабаєв та ін. ; 2-ге вид., випр. та допов. Харків : УкрДУЗТ, 2020. 246 с.

7. Драбкин А.Л., Зузенко В.Л., Кислов А.Г. Антенно-фидерные устройства ; изд. 2-е доп и перераб. Москва : Сов. радио, 1974. 536 с.
8 Немцов М.И., Шамаев Ю.М. Справочник по расчету параметров катушек индуктивности. Москва : Энергоиздат, 1981. 136 с.

Надійшла до редколегії 10.09.2024

Відомості про авторів:
Заболотний Володимир Ілліч – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри безпеки інформаційних технологій, Україна; email: volodymyr.zabolotnyi@nure.ua , ORCID: https://orcid.org/0000-0003-3258-8489
Олейніков Анатолій Миколайович – кандидат технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри комп'ютерної радіоінженерії та систем технічного захисту інформації, Україна; email: anatoly.oleynikov@nure.ua; ORCID: https://orcid.org/0000-0002-4458-8833
Заболотний Дмитро Миколайович – АТ «ІІТ», начальник відділу комплексів ТЗІ, Україна; email: dmytro.zabolotnyi@nure.ua , ORCID: https://orcid.org/0009-0008-5891-4426
Кустов Андрій Костянтинович – Харківський національний університет радіоелектроніки, аспірант кафедри безпеки інформаційних технологій, Україна; email: andrii.kustov@nure.ua, ORCID: https://orcid.org/0009-0003-2919-0768

УДК 621.372(075) **DOI:10.30837/rt.2024.3.218.05**

М.С. КАВЕЦЬКИЙ, В.І. РУЖЕНЦЕВ, д-р техн. наук

ВИЯВЛЕННЯ ВЕБ-АТАК ПО HTTP ЗАПИТАМ З ВИКОРИСТАННЯМ ТЕХНІК NLP

Вступ

Зростання кількості веб-додатків і їх значущість у сучасному інтернет-просторі супроводжуються збільшенням загроз безпеці. Веб-атаки, що спрямовані на вразливості в мережевих протоколах, стають дедалі складнішими, використовуючи різноманітні техніки для обходу захисних заходів.

Основним видом атак є зловживання HTTP-трафіком, що вимагає нових та ефективних методів виявлення, оскільки старі методи мають свої слабкі місця, що призводить до хибних результатів і неправильного блокування трафіку.

Робота зосереджена на вдосконаленні методів виявлення веб-атак через аналіз HTTP-трафіку з використанням технік обробки природної мови (NLP) та моделей на базі трансформерів, зокрема BERT.

Метою дослідження є розробка високоефективного класифікатора, здатного точно ідентифікувати аномальну активність у HTTP-трафіку.

Гіпотеза, яка підтверджується в ході дослідження, полягає в тому, що рішення на основі машинного навчання можуть класифікувати зловмисний трафік з більшою точністю, ніж традиційні системи на основі правил або сигнатур, та мають нижчий рівень хибних спрацювань.

Машинне навчання. Основні поняття

Машинне навчання (ML) та обробка природної мови (NLP) є двома взаємопов'язаними областями досліджень, які значно впливають на розвиток сучасних технологій обробки даних. NLP займається взаємодією між комп'ютерами та людською мовою, тоді як ML забезпечує алгоритми та моделі, що дозволяють автоматично виявляти патерни в даних і робити прогнози на їх основі. У цьому розділі розглядаються основні поняття машинного навчання в контексті NLP, що є важливими для розуміння методів, використаних у цьому дослідженні.

Методи NLP та моделі на основі трансформерів, зокрема BERT, були використані для аналізу HTTP-трафіку та виявлення веб-атак. Трансформери, з їхньою здатністю до контекстного розуміння тексту, показали високу ефективність у задачах класифікації, що підтверджується результатами проведених експериментів. Впровадження цих методів дозволяє значно підвищити точність ідентифікації зловмисного трафіку порівняно з традиційними методами, що базуються на правилах або сигнатурах.

Токенізація є першим і основним етапом в обробці природної мови, що включає поділ тексту на окремі компоненти, звані токенами. Токени можуть бути словами, фразами або навіть окремими символами. У контексті HTTP-запитів токенізація дозволяє розділити запит на окремі частини, які можуть бути проаналізовані окремо.

Лематизація і стемінг – це техніки нормалізації тексту. Лематизація приводить слова до їх базової або словникової форми (леми), а стемінг обрізає слова до їх кореневої форми. Ці процеси допомагають зменшити кількість різновидів слів, що полегшує подальший аналіз та підвищує ефективність моделей машинного навчання.

Після токенізації та нормалізації тексту наступним кроком є векторизація, яка перетворює текстові дані в числові вектори. Одним із популярних методів векторизації є метод "мішок слів" (Bag of Words), де текст представляється як вектор кількостей слів. Більш просунуті методи включають TF-IDF (Term Frequency-Inverse Document Frequency) та векторизацію за допомогою попередньо навчених моделей, таких як Word2Vec або GloVe.

Трансформери є архітектурою глибинних нейронних мереж, що призначені для обробки послідовностей даних, таких як текст. Вони були вперше представлені в 2017 р. у статті «Attention is All You Need» [1] Вони використовують механізм самоуваги (self-attention), що дозволяє моделі фокусуватися на різних частинах вхідної послідовності для більш точного контекстного розуміння. Модель BERT (Bidirectional Encoder Representations from Transformers) є однією з найпопулярніших моделей на основі трансформерів. Вона тренується на завданнях маскування слів і прогнозування наступних речень, що дозволяє їй отримувати глибоке контекстне розуміння тексту.

Класифікація тексту є одним з основних завдань у NLP, що включає категоризацію текстових даних в один або кілька класів. У контексті виявлення веб-атак, класифікація тексту дозволяє ідентифікувати зловмисні HTTP-запити серед нормального трафіку. Для цього використовуються алгоритми машинного навчання, такі як логістична регресія, дерева рішень, випадкові ліси, градієнтний бустинг та нейронні мережі.

Отже, в цьому розділі наведено основні відомості щодо теоретичної частини, необхідної для розуміння цієї роботи. Далі буде наведено ключові переваги моделі BERT та використаної в ній архітектурі трансформерів.

Архітектура трансформерів та модель BERT

В даній роботі зосереджено увагу на використанні сучасних моделей на основі трансформерів для аналізу HTTP трафіку з метою виявлення веб-атак. Однією з ключових моделей, які розглядаються, є BERT (Bidirectional Encoder Representations from Transformers). Ця модель, створена компанією Google, стала проривом в обробці природної мови (NLP) завдяки своїй здатності розуміти контекст слів в обох напрямках, що дозволяє ефективно вирішувати різноманітні задачі, включаючи класифікацію тексту, машинний переклад та інші.

Переваги цієї моделі для задачі аналізу HTTP трафіку наступні: розуміння контексту, адаптивність до розміру вхідних даних, попереднє навчання.

Щодо першого, то вона має двонаправлений контекст: BERT аналізує текст як зліва направо, так і справа наліво, що дозволяє краще розуміти контекст і семантичні зв'язки між словами (в даному випадку слова – це текст з HTTP запиту).

Щодо адаптивності до довжини послідовностей: HTTP запити можуть варіюватися за довжиною, і BERT здатна ефективно працювати з послідовностями різної довжини.

Попереднє навчання даної моделі допомагає моделі мати велику кількість фонових знань, що допомагає їй виявляти складні загрози, які можуть бути важкими для моделей, навчальних лише на вузькому наборі даних. Також дана перевага дозволяє тренувати модель трішки швидше, бо певні базові «знання» у моделі вже є.

Дана модель має переваги з трансформерів у вигляді механізму уваги (attention mechanism), який дозволяє моделі ефективно визначати та зважувати важливість різних частин тексту відносно один одного. Це означає, що модель може зосередитися на релевантних словах і фразах, навіть якщо вони розташовані далеко один від одного в тексті, що забезпечує точне розуміння контексту. Механізм уваги працює паралельно, що дозволяє обробляти довгі послідовності тексту швидше та ефективніше, ніж традиційні рекурентні нейронні мережі. Таким чином, трансформерна архітектура робить можливим захоплення довгострокових залежностей та складних взаємозв'язків у тексті, що у випадку аналізу HTTP трафіку доволі гарно допомогає, як показують результати дослідження у наступних підрозділах.

Отже, в цьому розділі описано основні переваги щодо архітектури трансформерів, а також відомості щодо моделі BERT. У наступному розділі буде проаналізовано схожі роботи, описано основні відомості щодо реалізації детекторів атак в цих роботах, а також наведено недоліки та порівняння з запропонованим підходом.

Огляд літератури

У цій роботі була використана модель BERT зі своїми перевагами, детально описаними у попередньому розділі.

В роботі [2] також використовують датасет CSIC2010 (про датасети детально описано у наступних розділах). Але в [2] автори, на жаль, описують свої експерименти по-різному, оскільки використовують аномальний трафік в процесі генерації векторів.

Робота, яка базується на схожій ідеї, що і в цьому дослідженні, заснована на Doc2Vec методі векторизації [3]. Запити HTTP трансформуються через Doc2Vec модель в векторну форму, яка потім використовується для того, щоб передбачити чи нормальний, чи аномальний цей трафік. Ця модель натренована на повній колекції датасету CSIC2010. Вони створили ніби «документи» з 10 нормальними чи аномальними запитами, а класифікація відбувається ансамблем класифікаторів натренованих на тренувальній вибірці, яка містить 70 % всієї вибірки. І хоча стверджується, що точність класифікації наближається до 99 %, але на нашу думку, головним недоліком цієї роботи є те, що вони перевіряють свої результати на доволі обмеженому наборі даних, який може не мати багато узагальнюючих властивостей і обмежуватися тільки певними, які є в датасеті. Під час виконання цієї роботи, було зібрано великий датасет та проведено більше досліджень з-поміж моделей, створених нами, а також класифікаторів, які вже існують на ринку.

Наступна робота [4] представляє HTTP запити як біграми в словнику з 80 символів ASCII. Алгоритм Isolation Forest був використаний для визначення належності даного HTTP-запиту до нормального чи аномального трафіку в отриманому векторному просторі. Незважаючи на проведення експериментів на відомому датасеті CSIC2010, спостерігається помітна різниця в кількості даних, що використовуються, тому можна припустити, що було використано меншу підмножину даних.

Робота [5] представляє повністю керований підхід. Для класифікації HTTP-трафіку використовується нейронна архітектура LSTM-CNN. Спочатку рекурентна мережа LSTM обробляє HTTP-запит на основі блочних характеристик, потім обрані стани мережі LSTM передаються в згорткову мережу, яка після обробки векторів передає їх на вихід у вигляді MLP мережі, що класифікує HTTP-запит в один з двох класів. Автори методу повідомляють про дуже хороші результати не тільки для колекції CSIC2010, але й для колекцій CICIDS 2017 та ISCX 2012, що містять різні типи атак.

У роботі [6] описано систему виявлення веб-атак, яка базується на ансамблі класифікаторів та методах векторного подання з області обробки природної мови (NLP). Спочатку система токенізує текст на основі вручну підготовленого словника, що містить токени, характерні для мережевого трафіку. Отримані текстові представлення векторизуються паралельно за допомогою нейронних моделей, що базуються на рекурентних та згорткових мережах. Після цього проводиться комплексна перевірка, яка повертає оціночний вектор, що визначає впевненість у взаємній різниці векторів. Отриманий вектор разом з векторами з нейронних моделей передається до ансамблю класифікаторів, який оцінює, чи є HTTP-запит нормальним трафіком чи атакою. Метод був протестований на колекції CSIC2010 та власних колекціях, але, на жаль, немає інформації про те, як проводилось навчання моделі.

У роботі [7] автори представляють нову систему виявлення Locate-Then-Detect (LTD), яка може точно виявляти веб-загрози в реальному часі, використовуючи глибокі нейронні мережі з механізмом уваги. Спочатку LTD витягує підозрілі сегменти з довгих запитів, потім запропоновані області додатково перевіряються класифікатором. Завдяки ефективному фільтруванню нерелевантних рядків фону, LTD може збільшити точність на 22,3 % і зменшити 82,6 % обчислювальних витрат. Експерименти на реальному веб-трафіку показують, що LTD перевершує кілька провідних методів і може ефективно виявляти атаки із середнім часом відгуку 5 мс, що робить його добре придатним для реальних додатків. Незважаючи на численні переваги LTD, на даний момент він може обробляти лише атаки типу SQLi та XSS.

Поточна робота над LTD охоплює виявлення більшої кількості веб-атак, таких як File Inclusion і Code Execution.

У порівнянні з іншими дослідженнями, ця робота має кілька ключових вдосконалень, які роблять модель більш ефективною у виявленні веб-атак. Перша відмінність полягає у використанні сучасної архітектури трансформерів, зокрема моделі BERT, яка показала значну ефективність у задачах обробки природної мови (NLP). В певних роботах [2 – 6] не використовуються трансформери з усіма їх перевагами (описані в відповідному розділі). І хоча ці роботи використовують векторний простір для представлення датасету, проте вони базуються кожна на різних методах машинного навчання, не маючи переваг архітектури трансформерів. Робота [7], яка хоч і використовує архітектуру трансформерів, має описані вище недоліки (значна обмеженість атак, які модель може виявляти). Також серед усіх описаних в цьому розділі робіт, дана робота має ключову відмінність, яка дає більшу перевагу для моделі в області машинного навчання – більш широкий датасет. Ключова перевага побудованого нами рішення полягає в тому, що було зібрано та скомбіновано кілька датасетів, включаючи CSIC2010, PositiveTechnologies, openappsec, а також власноруч зібрані дані (про це більш детально в розділі з датасетом) для створення одного великого та збалансованого набору даних (в проаналізованих роботах датасет або використовувався менше, що дає меншу узагальнюючу спроможність для моделі машинного навчання). Датасет з цієї роботи містить понад 195 тисяч записів, що значно перевищує обсяги даних, використаних у попередніх дослідженнях. Завдяки цьому модель здатна краще узагальнювати та виявляти складні патерни в HTTP-трафіку, що забезпечує високу точність і надійність класифікації. Крім того, методи попередньої обробки даних, такі як видалення дублікативних записів і непотрібної інформації, сприяли покращенню якості моделі. В результаті, створена модель демонструє високу збалансовану точність та здатність розпізнавати складні веб-атаки з мінімальною кількістю хибних спрацювань, що робить її конкурентоспроможною серед існуючих рішень у цій сфері.

Отже, в цьому розділі було описано схожі роботи, описано їх підходи та недоліки. Певні роботи самі по собі мають недолік через використання архітектур, які не мають певних механізмів (наприклад механізму уваги), що значно послаблює їх узагальнюючі властивості під час навчання. Деякі роботи хоча і мають схожу архітектуру, проте вони будуються на інших моделях, а також на значно меншому датасету. Використання малого датасету веде до того, що модель може передбачувати менше атак (що і було продемонстровано у роботі [7]), а також ця модель сама по собі буде гірше виявляти певні атаки, бо вона не має таких переваг в узагальнюючих властивостях перед моделлю, яка навчалася на датасеті з більшою кількістю даних. В наступному розділі розглянемо особливості запропонованого підходу, а також основні відмінності між цим підходом та підходами зі схожих робіт. Також в інших розділах далі буде описано важливість вибору більш широкого датасету.

Огляд запропонованого підходу

Відомо, що методи на основі нейронних мереж вже давно застосовуються у галузі виявлення кібератак. Натомість, основна пропозиція цієї роботи полягає в удосконаленні існуючих підходів шляхом використання сучасних моделей на базі трансформерів, зокрема BERT, для аналізу HTTP-трафіку. Запропонований підхід має кілька ключових відмінностей від традиційних методів, що дозволяє досягти більш високої точності та зменшити кількість хибних спрацювань.

На рис. 1 наведено загальний підхід до класифікації. Основні відмінності між запропонованим в роботі та класичними підходами – це наявність іншого токенізатору, а також моделі (та архітектури) для класифікації.

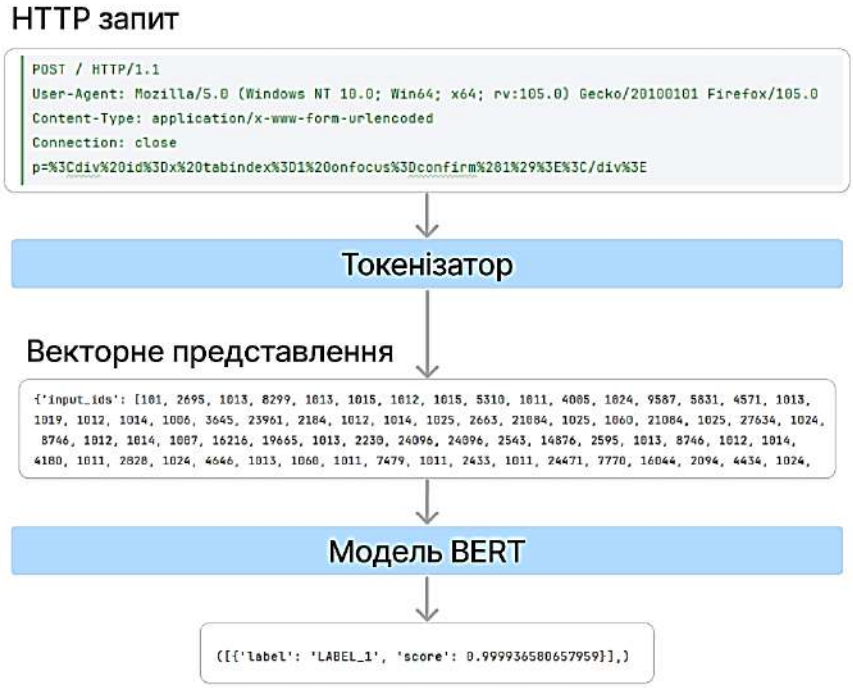


Рис. 1. Пояснення підходу

Щодо елементу токенізатору з рис. 1, то BERT токенізатор використовує WordPiece токенізацію, яка дозволяє ефективно розбивати слова на менші підслова чи символи. Це підвищує здатність моделі працювати з рідкісними або незнайомими словами, покращуючи узагальнення та розуміння тексту. Також відмінність – це подвійний процес токенізації: WordPiece токенізатор спочатку розбиває текст на токени, використовуючи регулярні вирази, а потім розбиває їх на менші підслова. Це знижує ймовірність утворення невідомих токенів (UNK) порівняно з іншими методами. BERT токенізатор є унікальним, бо враховує контекст, що дозволяє моделі краще розуміти багатозначні слова, порівняно з простішими методами токенізації, які можуть втратити контекст. Наприклад, у роботах, де використовується Doc2Vec (як у [3]), запити HTTP перетворюються в вектори без врахування глибокого контексту, що забезпечується механізмом уваги в BERT. Це може призвести до втрати важливої інформації про контекст, яку BERT зберігає. Крім того, Doc2Vec токенізатор не має тих переваг, які є у WordPiece токенізатора BERT, що дозволяє краще працювати з рідкісними або незнайомими словами. Алгоритм Isolation Forest (як у [4]) використовується для класифікації на основі просторового розподілу даних, що відрізняється від текстової обробки BERT. Він не використовує глибоке контекстне розуміння, яке забезпечує BERT, а токенізація в такій роботі зазвичай не враховує поділ на підслова, що знижує ефективність обробки рідкісних слів.

Щодо моделі (а також її архітектури) з рис. 1 можна виділити відмінності: BERT базується на архітектурі трансформерів, яка використовує механізм уваги (Attention Mechanism) для обробки всього тексту одразу, на відміну від рекурентних нейронних мереж (RNN) та LSTM, які обробляють текст послідовно. BERT навчається двонаправлено, що означає одночасне врахування контексту зліва направо і справа наліво. Це значно покращує розуміння контексту порівняно з однонаправленими моделями. А от, наприклад, в роботі [4] алгоритм Isolation Forest використовується для класифікації на основі просторового розподілу даних, що відрізняється від текстової обробки BERT. Він не використовує глибоке контекстне розуміння, яке забезпечує BERT, а токенізація в такій роботі зазвичай не враховує поділ на підслова, що знижує ефективність обробки рідкісних слів. У роботах з використанням LSTM-CNN (як у [5]), моделі обробляють послідовні дані, але можуть втратити контекст на великих відстанях у тексті, чого уникає BERT завдяки своєму глобальному механізму уваги. Токенізатори, що використовуються в цих роботах, зазвичай менш ефективні у порівнянні з

WordPiece токенізатором BERT. Роботи, які використовують ансамблі з методами NLP (як у [6]), можуть мати обмеження в узагальненні також через те, що вони не використовують трансформери, що забезпечують глибше контекстне розуміння та узагальнюючі властивості. Токенізація в таких системах часто базується на простих методах розбиття на слова або символи, без переваг WordPiece токенізації BERT.

На рис. 1 наведено пояснення щодо підходу. Можна бачити HTTP трафік, який є певним текстом і подається в токенізатор. Суть токенізатору зробити з цього тексту певні токени, які модель може розуміти. Перевага моделі BERT в тому, що для неї є натренований токенізатор, який має певний набір токенів найбільш частовикористовуваних, регістр слів не важливий, а також кожне слово – це не токен, бо одне слово може включати декілька токенів. Таким чином, після розбиття токенізатором отримуємо векторне представлення з певною довжиною. Ця довжина фіксована і під час роботи розрахована на 400 токенів, чого цілком вистачає, як буде показано далі під час проведення аналізу датасету. За потреби кількість токенів можна збільшити, проте це збільшить час, який потрібен для передбачення та тренування моделі. Далі це векторне представлення подається моделі на класифікацію, де вона видає «label» – зі значеннями «label_0», «label_1» це просто мітки класу «аномальний» або «нормальний» трафік, а також score, який показує наскільки модель впевнена в своєму передбаченні від 0 до 1. Де 1 – 100 % впевненості в тому, що саме певний label є правильним.

В машинному навчанні основні відмінності відбуваються через зміну архітектури, а також використаного датасету. Якщо в певних традиційних архітектурах був відсутній механізм уваги, то тут він є. Це дає змогу цьому методу робити більш релевантні узагальнюючі висновки щодо датасету в процесі навчання (дозволяє фокусуватися на різних частинах вхідних даних залежно від їх важливості). Також не малу роль грає сам датасет, який був значно розширений якраз з задумом, щоб модель BERT, використовуючи перевагу трансформерів, більш гарно зробила узагальнюючі висновки щодо всього спектру атак, які можуть бути в датасеті з розширеними характеристиками (тобто більш глибокий аналіз до розуміння). В цьому криється основна відмінність. І хоча робота [7] також використовує механізм уваги, але все одно використовується інша модель (не BERT), а також менший датасет, і в результаті може передбачуватись зовсім малий набір атак та зі зменшеною точністю (про це – в розділі з описом схожих робіт).

Насамперед, цей метод аналізує HTTP трафік, який можна представити у вигляді тексту (звичайний текст запиту), а далі даний текст буде класифікуватися моделлю, яка натренована на великій кількості прикладів як поганого (є атака), так і легітимного трафіку. Це дає змогу автоматично виявляти патерни в даних та будувати залежності між даними. А з використанням моделі BERT можна отримати перевагу в контекстному розумінні тексту. Тобто, залежності між даними будуть гарно знаходитися завдяки механізму уваги від трансформерів. Це основна перевага використання трансформерів, яка була описана у минулому підрозділі.

Далі буде наведено відмінності між реалізацією традиційних методів та реалізованого у ході дослідження. У традиційних методах дані зазвичай збираються з журналів серверів, мережевих пакетів та інших джерел. Вони часто представляються у вигляді рядків тексту або простих векторів ознак. В даному ж підході робиться все те саме, але відбувається й додаткова підготовка у вигляді токенізації, лематизації та видалення стоп-слів для забезпечення максимально точного представлення текстових даних.

Щодо векторизації слів, то в традиційних методах зазвичай це відбувається через так званий «мішок слів» (bag of words) або TF-IDF (term frequency-inverse document frequency), але такі методи не враховують контекстуальних зв’язків між словами, що може призводити до втрати важливої інформації. В запропонованому методі відбувається токенізації через трансформери, зокрема моделі BERT. Це дозволяє враховувати контекстуальні зв’язки між словами в запитах, що значно покращує точність представлення даних.

Навчання моделі в традиційних методах – зазвичай використовуються моделі типу логістичної регресії, випадкові ліси. Ці моделі добре працюють на простих задачах, але їх ефек-

тивність може знижуватися на складних даних через обмежені можливості врахування контексту. В цьому дослідженні використовується модель BERT, яка була донавчена на специфічних даних HTTP-запитів. Використання трансформерів дозволяє моделі краще розуміти контекст і взаємозв'язки між словами, що покращує її здатність ідентифікувати зловмисні запити.

Загалом, наведений підхід використовує контекстуальне розуміння тексту (тобто він розуміє контекст, що у звичайній розмові грає велику роль). Це дозволяє більш точно ідентифікувати аномалії та зловмисні дії. Завдяки цьому досягається високий рівень точності моделі, що було продемонстровано в ході експериментів. Про це – в наступному підрозділі.

Отже, в цьому розділі описано основні відмінності щодо запропонованого підходу з цієї роботи, а також підходів, які використовувалися в інших роботах. А також розглянуто перевагу цієї роботи над іншими, які могли так чи інакше також використовувати схожі архітектури, але різні моделі. Описано важливість розширення датасету.

Опис існуючих датасетів. Процес розширення датасету

Потрібно було зібрати та підготовити набір даних, що включав HTTP-запити як нормальні, так і зловмисні. Для цього спочатку було проаналізовано наявні датасети, а також можливість їх застосування.

Датасет від CSIC2010 [2] включає тисячі автоматично згенерованих веб-запитів та призначений для тестування систем захисту від веб-атак. Наявні в датасеті дані включають 36,000 нормальних та понад 25,000 аномальних запитів. Але в цьому датасеті була проблема, що він мав багато однакових даних, які відрізнялися тільки параметром сесії, що, як показала практика, ніяк не допомагає в класифікації трафіку (бо цей параметр сесії – це рандомне число і не більше). Тому з цього датасету було взято всі дані після видалення повторів.

Другий датасет, наданий PositiveTechnologies через GitHub [8], містить дані з 21,991 безпечних та 1,097 аномальних HTTP-запитів із банківської програми. Цей датасет є цінним ресурсом для вивчення та виявлення потенційних загроз у банківському середовищі. Він мав також свої дублюючі дані. Ще в нього була особливість, що перший рядок починався з дати та часу – відмітки часу, коли він був зібраний. Ці дані абсолютно були непотрібні, тому датасет треба було обробити та прибрати непотрібне.

Загалом, на основі цих двох датасетів спочатку було сформовано першу версію датасету з 58 тисяч даних. Було поділено на тренувальну (40 тис.), тестувальну (13 тис.) та валідаційну (5 тис.) вибірки. Це робиться для того, щоб тренувати модель на тренувальній, а також паралельно перевіряти на тестувальній. І вже після тренування провести фінальне випробування на валідаційній вибірці, яку модель «не бачила» під час тренування, а отже не могла брати з неї ніякі залежності, не вивчала ці дані і тд. Це дає більш об’єктивний спосіб перевірити результати.

Під час тренування стало очевидно, що сформованого датасету недостатньо для створення гарної узагальнюючої моделі (в порівнянні моделей це описано), тож, було взято третій датасет від openappsec [9], який є доволі великим та включає в себе 973,964 легітимних HTTP-запитів з 185 реальних веб-сайтів у 12 категоріях, а також 73,924 зловмисних завантажень. В ньому також видалені дублікати (майже кожен п’ятий запис), але їх було видалено і сформовано другу версію датасету зі 195 тис. записів. Де тренувальна включала 117 тис., тестувальна – 51 тис., і валідаційна – 27 тис. записів.

Також в ході формування датасету було використано багато різних маленьких джерел в інтернеті, з яких бралася певна кількість даних для доповнення фінального датасету.

Отже, далі всі датасети було скомбіновано, видалено повторення, а також певні дані, які містили помилки. В результаті вийшло, що всі датасети збалансовані, мають середнє відхилення 0,5 та середнє значення 0,5 за колонкою «benign», яка містить значення «1» або «0» в залежності від того, є аномалія, чи її немає. Наступним етапом треба було тренувати моделі, процес чого описано в наступному розділі.

Хід проведення експериментів

Експерименти проводилися з метою розробки та оцінки ефективності класифікатора для виявлення веб-атак через аналіз HTTP-трафіку з використанням методів обробки природної мови (NLP) та моделей на базі трансформерів, зокрема BERT. Для цього використовувався датасет, методику збору якого описано у попередньому розділі.

Для підготовки даних використовувалися такі методи: токенізація, лематизація та видалення стоп-слів. Після цього HTTP-запити були представлені у вигляді векторів, придатних для обробки моделлю BERT. Модель BERT була обрана за її здатність до контекстного розуміння тексту, що є критичним для ідентифікації складних шаблонів у HTTP-запитах.

Експерименти проводилися в кілька етапів. На першому етапі модель BERT тренувалася на підготовленому наборі даних з метою класифікації запитів як нормальних або зловмисних. Для тренування використовувалися стандартні методи оптимізації, такі як градієнтний спуск, та функції втрат, що підходять для задач бінарної класифікації.

Результати тренування чотирьох моделей на підготовлених датасетах можна побачити на рис. 2 (графіки двох моделей «bert-tiny-4.39m-sql-e70» та «bert-tiny-4.39m-sql-e50» майже повністю співпадають). На графіку відображено як змінювалась похибка моделі (loss, вертикальна вісь) в залежності від кроків навчання (train step), які змінюються по горизонтальній осі. Як видно, модель «bert-tiny-4.39m-nosql-e1-extended» стала найкращою серед усіх (має найменшу похибку). Вона навчалася на розширеному датасеті, який включав більше даних, ніж попередні.

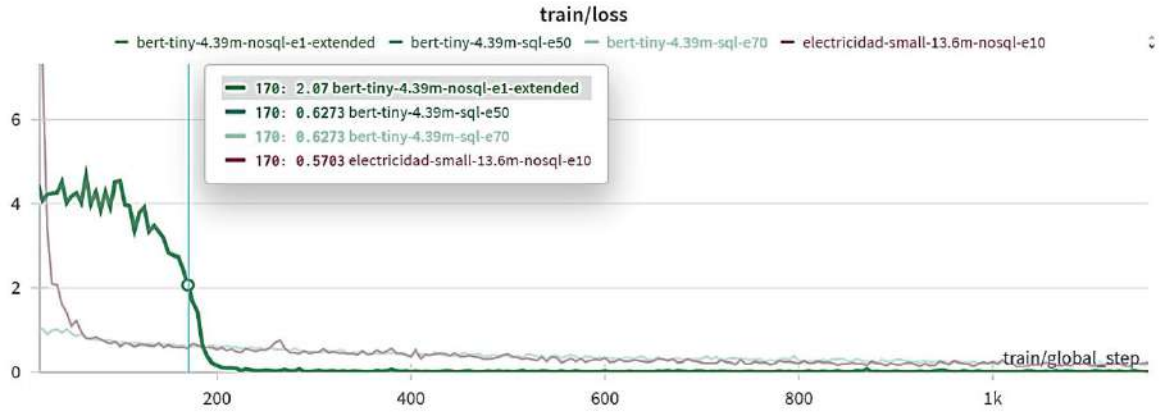


Рис. 2. Результати тренування моделей

На рис. 2 відображено графік, який автоматично сформовано під час навчання з використанням бібліотеки wandb мови Python. Ця бібліотека дозволяє відслідковувати параметри під час навчання, дивитися проміжні та фінальні результати. Для повторення експерименту необхідно використовувати саме цю бібліотеку, яка допоможе автоматично сформувати даний графік після навчання моделі. По вертикальній осі відображена похибка (це просто певне число) у вигляді функції втрат (різниця між кінцевим результатом передбачення моделі та справжнім результатом. Це просто число, і в ході навчання воно повинне зменшуватися, що і видно на рисунку), а горизонтально наведено шаг (або ітерацію) тренування моделі. Різні початкові точки графіків для різних моделей можна пояснити різними датасетами, що використовувались при навчанні, а також різними внутрішніми параметрами для різних моделей.

Вертикальна блакитна лінія – це просто графічний елемент, який було отримано наведенням курсора миші на довільну ділянку графіку. Вона допомагає відображати певний момент на графіку, щоб можна було побачити результати моделі в цій точці, а також виділити «товстим» лінію з найкращою моделлю у відповідному редакторі для наочності.

Самі по собі моделі відрізняються лише даними, на яких вони тренувалися. Це потрібно, щоб знайти баланс між необхідною кількістю даних для тренування, а також отриманою

помилкою. Назви моделей «bert-tiny-4.39m-sql-e70» та «bert-tiny-4.39m-nosql-e50», а також модель «electricidad-small-13.6m-nosql-e10» мають певний сенс в плані найменування, бо відображають дані, на яких тренувалися, та інші параметри. В їх назві є кількість параметрів (4,39 та 13,6 мільйонів), кількість епох «e70», а також примітка «sql» або «nosql». Ця примітка показує, чи були в тренувальному датасеті моделі дані нормального трафіку, який містив нормальний SQL запит до бази даних. Фінальний датасет для тренування налічував 195 тис. запитів та був збалансований (тобто, там була майже однакова кількість прикладів поганого та хорошого трафіку). Цей датасет розбивався на 117 тис. тренувальних даних, а також на 51 тис. тестувальних і 27 тис. – валідаційних. На тренувальних відбувалося тренування і використовувалися тестувальні дані для корекції під час тренування, а далі після тренування на готовій моделі валідаційний датасет використовувався як фінальна перевірка результату моделі (але це не той датасет зі статті [9], на якій перевірялася остаточна точність, бо це лише датасет для перевірки проміжних результатів для відбору кращих серед побудованих моделей). Тренування відбувалося завдяки можливостям мови програмування Python. Воно брало певну порцію даних, навчалося на ній, а далі брався тестувальний датасет і проганявся по моделі, видаючи похибку, яку можна бачити на рис. 2. Таким чином відбувалося до тих пір, поки похибка не буде зведена до мінімуму.

Також з рис. 2 можна бачити, що фінальна модель «bert-tiny-4.39m-nosql-e1-extended» (навчалася на розширеному датасеті) на початку свого тренування мала більшу похибку, аніж інші моделі, такі як «bert-tiny-4.39m-sql-e70» та «bert-tiny-4.39m-sql-e50» (навчалися на вузькому датасеті). Цей момент показує важливість навчання на більшому датасеті, бо моделі, які навчалися на зменшеному дата сеті, мають меншу узагальнену властивість. Тому що, коли датасет розширили та почали навчати іншу модель (відбувалося «донавчання»), то така модель почала показувати багато помилок (бо взяла менше характеристик з малого датасету).

Отже, в цьому розділі описано процес, в ході якого проводилися основні експерименти, а також результати щодо побудованих моделей. Обрано кращу модель. Наведено відомості щодо важливості використання збільшеного датасету і як це відображається на загальній картині тренування та якості моделі. В наступному розділі наведено процес оцінки та порівняння моделей між собою, що передбачає визначення основних метрик для оцінки, а також порівняння на базі цих метрик побудованих моделей з іншими фаєрволами, які використовуються в промисловості. Наведено числові результати ефективності побудованого класифікатору для аналізу веб-трафіку та проаналізовано ці результати з результатами інших методів.

Процес оцінки та порівняння моделей між собою

Після тренування модель тестували на незалежному наборі даних, щоб оцінити її ефективність. Далі необхідно порівняти побудовану модель з класичними методами, які вже зарекомендували себе та використовуються на ринку, таким чином підтвердити або спростувати поставлену на початку роботи гіпотезу. Тому для порівняння ефективності побудованої моделі було також використано результати тестування з роботи [9], де порівняно між собою дев'ять провідних рішень систем захисту веб-додатків (WAF). Результати кожної системи порівнювалися з результатами запропонованої моделі, щоб визначити її переваги та недоліки. На завершальному етапі експериментів проведено аналіз помилок, щоб ідентифікувати випадки, коли модель помилялася, та визначити можливі напрямки для подальшого вдосконалення. Цей етап включав детальний розгляд неправильно класифікованих запитів та аналіз їхніх особливостей. Таким чином, проведені експерименти дозволили не тільки підтвердити гіпотезу про перевагу машинного навчання над традиційними методами, але й надати конкретні рекомендації щодо подальшого розвитку методів виявлення веб-атак.

Ключовою метрикою є збалансована точність. Вона отримується як середнє між метриками TNR (true negative rate), або її ще називають Specificity, та метрикою TPR (true positive rate), або Recall.

Метрика TNR, також відома як Specificity, є однією з метрик, яка використовується для оцінювання ефективності класифікаційних моделей, особливо в задачах двокласової класифікації. Вона вимірює здатність моделі правильно ідентифікувати всі дійсні негативні приклади з усіх доступних негативних прикладів у наборі даних. Іншими словами, TNR показує, яку частку реальних негативів (об'єктів, що не належать до класу, який нас цікавить) модель змогла правильно передбачити.

Метрика TPR (або recall) вимірює здатність моделі правильно ідентифікувати всі дійсні позитивні приклади з усіх доступних позитивних прикладів у наборі даних. Простіше кажучи, recall показує, яку частку реальних позитивів (об'єктів, що належать до класу, який нас цікавить) модель змогла коректно передбачити. Формально, recall визначається як відношення кількості правильно передбачених позитивних випадків (True Positives, TP) до суми дійсних позитивних випадків, що складається з правильно передбачених позитивів (TP) та неправильно передбачених негативів (False Negatives, FN).

Загалом, їх потрібно отримувати, щоб вирахувати збалансовану точність, яка всього лише є середнім значенням між Specificity та Recall. Ця метрика дозволяє комплексно оцінити здатність моделі правильно класифікувати зловмисні запити, мінімізуючи хибні спрацювання.

Для порівняння з існуючими фаєрволами було обрано кращу навчену модель та порівняно за основними параметрами з метою зрозуміти, чи буде така модель краще працювати, чи ні. Для цієї мети підходить стаття [9], яка досліджує якість сучасних фаєрволів на конкретному датасеті, який було використано і в цій роботі. У рамках дослідження автори вирішили провести глибокий, але простий експеримент, спрямований на виклик як шкідливих, так і легітимних веб-запитів у різних веб-захистах та вимірювання їх результатів. Для проведення тесту використовувався дуже великий набір даних: 973,964 легітимних HTTP-запитів з 185 реальних веб-сайтів у 12 категоріях, а також 73,924 зловмисних запитів з широкого спектру загроз, які зазвичай зустрічаються в атаках.

Експеримент, проведений авторами статті [9] у липні 2023 р., порівнював кілька популярних рішень WAF, серед яких були Microsoft Azure WAFv2 з правилами OWASP CRS 3.2, AWS WAF з керованим набором правил AWS, AWS WAF з керованим набором правил AWS та F5 Ruleset, CloudFlare WAF з керуванням та OWASP Core Rulesets, F5 NGINX App Protect WAF з профілем за замовчуванням, F5 NGINX App Protect WAF зі строгим профілем, NGINX ModSecurity з правилами OWASP CRS 3.3.4, open-appsec / CloudGuard AppSec з конфігурацією за замовчуванням (High Confidence), а також open-appsec / CloudGuard AppSec з критичною конфігурацією впевненості.

Результати тесту підкреслюють значущі відмінності у продуктивності різних захисних рішень. Наприклад, CloudFlare WAF продемонстрував майже ідеальну якість виявлення (99,945 %), але найнижчу якість безпеки (67,297 %) серед усіх перевірених продуктів. Azure WAF забезпечує дуже високу якість безпеки (98,547 %), але також має надзвичайно високий рівень хибнопозитивних результатів (38,346 %). Такі результати вказують на значний ризик безпеки або необхідність ретельного налаштування як спочатку, так і в подальшому для ефективного використання продукту в реальних умовах.

Для врахування дисбалансу в кількості екземплярів у різних класах у цьому порівнянні також використали метрику Balanced Accuracy, яка враховує середню точність по кожному класу. Результати порівняння по збалансованій точності (ЗТ) можна побачити на наступному рисунку. Автори зазначають, що open-appsec / CloudGuard AppSec зі стандартною конфігурацією забезпечує найвищий показник збалансованої точності (97,28 %), за ним іде цей же продукт із конфігурацією Critical Profile (ЗТ – 96,8 %), а також NGINX AppProtect зі строгим профілем (ЗТ – 92,52 %). На рис. 3 [9] можна побачити кількісне значення збалансованої точності, яке було отримано для того чи іншого фаерволу в ході оцінки його точності на перевірочному датасеті.

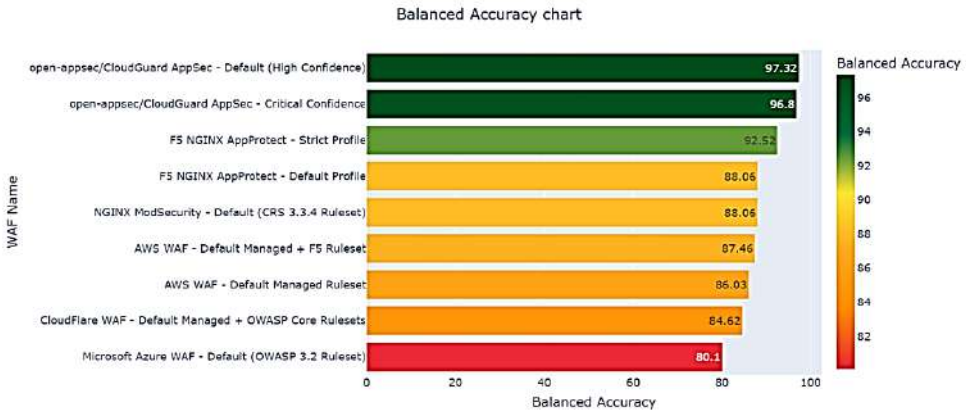


Рис. 3. Наявні метрики звичайних фаєрволів для порівння [9]

Експеримент в рамках цього дослідження проводився на тому ж самому датасеті, що й від авторів статті [9] (використовувався вкрай обширний набір даних: 973,964 легітимних HTTP-запитів з 185 реальних веб-сайтів у 12 категоріях, а також 73,924 зловмисних запитів з широкого спектру загроз, які зазвичай зустрічаються в атаках). А саме було взято найкращу натреновану модель під назвою «bert-tiny-4.39m-nosql-e1-extended» та зроблено спроби передбачити по всьому датасету чи то нормальний трафік, чи ні.

В результаті отримано метрики, з яких можна отримати збалансовану точність (основная метрика для порівняння). Створена модель демонструє Recall (Sensitivity) на рівні 0,9997 та Specificity на рівні 1,0. Тоді середнє значення цих показників, яке є збалансованою точністю, дорівнює 0,9998. Це значення значно перевищує найкраще рішення з порівняння зі статті [9], яке, за словами авторів, має збалансовану точність на рівні 0,9732. Тобто, наведений метод з використанням моделі транформерів BERT можна цілком назвати конкурентоспроможним завдяки його здатності розуміти контекст та автоматично вивчати зв'язки в даних під час тренування. До недоліків можна віднести необхідність мати датасет для навчання моделі, а також обчислювальні потужності, бо тренування займає певний час і може виникнуть необхідність мати потужну GPU для цієї задачі.

Висновки

Проведене дослідження демонструє значні можливості використання методів машинного навчання та обробки природної мови (NLP) для підвищення ефективності виявлення веб-атак. Використання моделі BERT, заснованої на трансформерах, дозволило створити високоефективний класифікатор для аналізу HTTP-трафіку.

Результати експериментів показали, що запропонована модель здатна з більшою точністю ідентифікувати зловмисні запити порівняно з традиційними методами, що базуються на правилах і сигнатурах.

Результати дослідження можуть бути використані для вдосконалення існуючих систем кібербезпеки, сприяючи розробці більш надійних і ефективних рішень для захисту від веб-атак. Найкраща модель, натренована в рамках цього дослідження під назвою bert-tiny-4.39m-nosql-e1-extended, демонструє чутливість (Recall) на рівні 0,9997 та специфічність (Specificity) на рівні 1,0. Середнє значення цих показників, яке є збалансованою точністю, становить 0,9998. Це значення значно перевищує найкраще рішення, описане у статті [9], де автори зазначають збалансовану точність на рівні 0,9732.

Порівняння з іншими роботами показало, що наша модель має кілька ключових переваг.

По-перше, модель BERT використовує сучасний підхід до токенізації, зокрема WordPiece токенізатор, який дозволяє краще обробляти рідкісні та незнайомі слова. На відміну від токенізаторів у інших роботах, WordPiece забезпечує розбиття слів на підслова, що підвищує ефективність обробки тексту та узагальнюючу здатність моделі.

По-друге, BERT забезпечує глибоке контекстне розуміння завдяки своєму глобальному механізму уваги, що дозволяє моделі враховувати взаємозв'язки між словами на всіх рівнях тексту. Це є значною перевагою над методами, які використовують Doc2Vec, LSTM-CNN, Isolation Forest або ансамблі з методами NLP, що не враховують глибокий контекст або обмежуються локальними зв'язками в тексті.

По-третє, наша модель була навчена на великому та збалансованому наборі даних, який включає кілька датасетів, таких як CSIC2010, PositiveTechnologies, openappsec та власноруч зібрані дані. Це значно покращило узагальнюючу здатність моделі та її точність у виявленні складних веб-атак.

Порівняння з дев'ятьма провідними системами захисту веб-додатків (WAF) підтвердило конкурентоспроможність та переваги запропонованого методу. Гіпотеза підтвердилася.

Отже, створена модель демонструє високу збалансовану точність та здатність розпізнавати складні веб-атаки з мінімальною кількістю хибних спрацювань, що робить її конкурентоспроможною серед існуючих рішень у цій сфері. Враховуючи ці переваги, модель BERT може бути рекомендована для використання у системах виявлення веб-атак для забезпечення високого рівня безпеки та надійності.

Дослідження робить важливий внесок у розвиток методів виявлення веб-атак, демонструючи, що інтеграція передових технологій NLP та моделей машинного навчання може суттєво підвищити рівень кібербезпеки в умовах постійно зростаючих загроз.

Список літератури:

1. Attention Is All You Need [Електронний ресурс]. Режим доступу до ресурсу: https://arxiv.org/pdf/1706.03762.pdf
2. HTTP DATASET CSIC 2010 [Електронний ресурс]. Режим доступу до ресурсу: https://www.isi.csic.es/dataset/
3. Saikat Das, Mohammad Ashrafuzzaman, Frederick T Sheldon, and Sajjan Shiva. Network intrusion detection using natural language processing and ensemble machine learning // 2020 IEEE Symposium Series on Computational Intelligence (SSCI), pages 829–835. IEEE, 2020.
4. Ali Moradi Vartouni, Saeed Sedighian Kashi, and Mohammad Teshnehlab. An anomaly detection method to detect web attacks using stacked auto-encoder // 2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), pages 131–134. IEEE, 2018.
5. Jiaxin Liu, Xucheng Song, Yingjie Zhou, Xi Peng, Yanru Zhang, Pei Liu, and Dapeng Wu. Deep anomaly detection in packet payload. arXiv preprint arXiv:1912.02549, 2019.
6. Chaochao Luo, Zhiyuan Tan, Geyong Min, Jie Gan, Wei Shi, and Zhihong Tian. A novel web attack detection system for Internet of things via ensemble classification // IEEE on Industrial Informatics, 2020
7. Tianlong Liu, Yu Qi 2, Liang Shi, Jianan Yan. Locate-Then-Detect: Real-time Web Attack Detection via Attention-based Deep Neural Networks
8. mrm8488/bert-tiny-finetuned-sms-spam-detection [Електронний ресурс]. Режим доступу до ресурсу: https://huggingface.co/mrm8488/bert-tiny-finetuned-sms-spam-detection
9. Best WAF solutions in 2023 - real-world comparison [Електронний ресурс]. Режим доступу до ресурсу: https://www.openappsec.io/post/best-waf-solutions-in-2023-real-world-comparison

Надійшла до редколегії 05.09.2024

Відомості про авторів:

Кавецький Максим Сергійович – Харківський національний університет радіоелектроніки, магістр кафедри безпеки інформаційних технологій факультет комп'ютерної інженерії та управління; Україна; e-mail: maksym.kavetskyi@nure.ua; ORCID: https://orcid.org/0009-0008-7419-1029

Руженцев Віктор Ігорович – д-р техн. наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри безпеки інформаційних технологій; Україна; e-mail: viktor.ruzhentsev@nure.ua, ORCID: http://orcid.org/0000-0002-1007-6530

УДК 004.056.55 **DOI:10.30837/rt.2024.3.218.06**

С.О. КАНДІЙ, І.Д. ГОРБЕНКО, д-р техн. наук

УТОЧНЕННЯ ОЦІНОК БЕЗПЕКИ КВАНТОВО-СТІЙКИХ СТАНДАРТІВ АСИМЕТРИЧНОГО ШИФРУВАННЯ З ВРАХУВАННЯМ СТРУКТУРИ Q-АРНИХ РЕШІТОК

Вступ

Квантово-стійка криптографія з кожним роком використовується все більше для вирішення практичних задач. Такі квантово-стійкі стандарти як ДСТУ 8961:2019 («Скеля») [1], FIPS 203 (CRYSTALS-Kyber) [2] та FIPS 204 (CRYSTALS-Dilithium) [3] ґрунтуються на складних проблемах з теорії решіток, що природнім чином призводить до підвищення інтересу до криптоаналізу проблем з теорії решіток.

У останні роки спостерігається значний прогрес у моделях редукції решіток. У роботі [5] запропоновано модель безпеки, що враховує q-арної структуру решіток у криптографічних проблемах, на яких ґрунтується безпека сучасних квантово-стійких стандартів на решітках. Проте, у роботі [5] дослідження проводилося на абстрактних наборах параметрів, а дослідження конкретних криптографічних схем не проводилося.

Ця робота є продовженням роботи [5] і присвячена застосуванню розробленої методики на стандартизованих криптографічних перетвореннях на решітках. Для кожного стандарту спочатку надаються існуючі оцінки у формальних моделях безпеки IND-CCA [6] або EUF-CMA [7], після чого надаються оцінки для атак вкладення та декодування, відповідно до роботи [5].

1. Моделі безпеки

Відомі атаки на проблеми LWE, SIS та NTRU можливо поділити [5] на наступні класи:

- Комбінаторні атаки;
- Алгебраїчні атаки;
- Атаки декодування;
- Атаки розпізнавання;
- Атаки вкладення;
- Гібридні атаки.

У межах цієї роботи представляють інтерес атаки вкладення, атаки декодування та гібридні атаки.

Атаки вкладення є одними з найбільш ефективних атак на LWE та NTRU. Їх сутність полягає у побудові решіток спеціального вигляду, найменший вектор яких містить шуканий секрет. Такі атаки ще називають первинними (англ. Primal) атаками. Також їх можливо використовувати для вирішення проблеми SIS.

Сутність атак декодування полягає у зведенні проблеми LWE або NTRU до проблеми CVP. Атаки такого роду вимагають побудови та редукції базису решітки таким чином, щоб можливо було вирішити проблему CVP для шуканого таємного вектору. Проведення атак декодування є технічно складнішим за атаки вкладення через двоетапну структуру атаки. Атаки декодування, як і атаки вкладення, також іноді називають первинними атаками (англ. BDD Primal attacks).

Гібридні атаки поєднують комбінаторні методи криптоаналізу з атаками вкладення або атаками розпізнавання (гібридні дуальні атаки). Такі атаки при використанні розріджених секретів часто є найкращими для багатьох криптографічних систем. Це є особливо актуальним для ДСТУ 8961:2019.

У сучасній криптографії однією з обов’язкових вимог до будь-якої криптографічної системи є наявність доказової безпеки. Тобто, має існувати математичний доказ безпеки, який

гарантує відсутність атак у межах обраної формальної моделі, за умови виконання невеликої кількості модельних припущень.

Модель безпеки IND-CCA ґрунтується на ідеї нерозрізнювальності: якщо супротивник не може розрізнити шифротекст повідомлення m_0 від шифротекста повідомлення m_1, то він не може отримати жодної інформації про зашифровані повідомлення.

Для побудови доказу безпеки шифру *E* для супротивника *A* вводяться дві гри (експерименти): $Exp_{A,E}^{IND-CCA-0}(\lambda)$ та $Exp_{A,E}^{IND-CCA-1}(\lambda)$ для параметра безпеки λ [6]. У кожній грі іспитувач генерує випадкову ключову пару $(pk, sk) \leftarrow Gen(1^{\lambda})$ та передає відкритий ключ супротивнику *A*. Супротивник *A* обирає два повідомлення m_0, m_1 однакової довжини та надсилає їх іспитувачу. Іспитувач генерує випадковий біт $b \in \{0,1\}$, чим обирає гру. Якщо $b = 0$, то іспитувач зашифровує повідомлення m_0 та надсилає шифротекст $c^* = Enc(sk, m_0)$ у якості завдання (гра $Exp_{A,E}^{IND-CCA-0}(\lambda)$). Якщо біт $b = 1$, то, іспитувач зашифровує повідомлення m_1 та надсилає шифротекст $c^* = Enc(sk, m_1)$ у якості завдання (гра $Exp_{A,E}^{IND-CCA-1}(\lambda)$). Супротивник має визначити у яку гру він грає (яке повідомлення було зашифровано) та повернути біт b_A. Результатом ігор $Exp_{A,E}^{IND-CCA-0}(\lambda)$ та $Exp_{A,E}^{IND-CCA-1}(\lambda)$ є значення предиката $b == b_A$. Супротивник *A* може робити запити до оракула дешифрування O_{Dec}, який може розшифрувати будь-який шифротекст окрім шифротексту завдання. Розрізняють IND-CCA1 безпеку, де супротивник може робити запити тільки до моменту отримання шифротекста та IND-CCA2 безпеку, де запити можливо робити і після отримання завдання. У межах цього дослідження IND-CCA2 буде вважатися синонімом IND-CCA.

Перевага супротивника у розрізненні ігор визначає безпеку в моделі IND-CCA. Якщо перевага є незначною у теоретико-числовому сенсі, то схема асиметричного шифрування вважається безпечною в моделі IND-CCA:

$$Adv_{A,E}^{IND-CCA}(\lambda) = \Pr[Exp_{A,E}^{IND-CCA-0}(\lambda) - Exp_{A,E}^{IND-CCA-1}(\lambda)] = negl(\lambda). \quad (1)$$

Від схем асиметричного шифрування при побудові механізмів інкапсуляції ключів вимагається безпека у моделі IND-CPA (Indistinguishability under Chosen-Plaintext Attacks), або у моделі OW-CPA (One-Wayness under Chosen-Plaintext Attacks).

Перевагу супротивника *A* у іграх IND-CPA та OW-CPA для схеми асиметричного шифрування PKE позначимо як $Adv_{PKE}^{OW-CPA}(A)$ та $Adv_{PKE}^{IND-CPA}(A)$ відповідно. Стандартним визначенням для переваги супротивника є:

$$Adv_{PKE}^{OW-CPA}(A) = \Pr[OW - CPA(A) == 1], \quad (2)$$

$$Adv_{PKE}^{IND-CPA}(A) = |\Pr[IND - CPA(A) == 1] - 1/2|. \quad (3)$$

Схема асиметричного шифрування у загальному випадку може мати помилки дешифрування, тобто для деяких правильно обчислених шифротекстів розшифрування може давати не правильний результат. Існують різні підходи до врахування помилок дешифрування. У межах цього дослідження будемо слідувати роботі [8]. Для оцінки ймовірності виникнення помилок дешифрування введемо наступну величину:

$$\delta_{wc} = E_{(pk,sk)}[\max_m \Pr[Dec(sk, c) \neq m]]. \quad (4)$$

У моделі EUF-CMA супротивник може звертатися до оракула підпису $Sign(sk, \cdot)$ для отримання підписів довільно обраних повідомлень. Схема підпису вважається безпечною, якщо ймовірність того, що супротивник зможе підробити підпис для будь-якого повідомлен-

ня є не значною. Так само, як і IND-CCA, доказова безпека у моделі EUF-CMA формується через ігри (експерименти). Позначимо відповідний експеримент $Exp_{A,S}^{EUF-CMA}(1^{\lambda})$ для схеми підпису S та супротивника A. У цьому експерименті випробовувач генерує ключову пару (sk, pk) та надає супротивнику відкритий ключ pk. Супротивник може роботи запити $m_1,...,m_q$ до оракула $Sign$. Усі запити до оракула зберігаються у списку Q. Після чого супротивник має повернути пару (m^*,σ^*). Якщо $S.Verify(pk,m^*,\sigma^*)==1$ і $m^* \notin Q$, то супротивник перемагає.

Перевага супротивника визначається як

$$Adv_{A,S}^{EUF-CMA}(1^{\lambda}) = \Pr[Exp_{A,S}^{EUF-CMA}(1^{\lambda}) == 1] \ . \tag{5}$$

Якщо $Adv_{A,S}^{EUF-CMA}(1^{\lambda}) = negl(\lambda)$, то схема підпису вважається безпечною у моделі EUF-CMA.

Посиленим варіантом моделі безпеки EUF-CMA є модель SUF-CMA. Якщо у моделі EUF-CMA вимагається створити підпис для повідомлення, що раніше не було підписано, то у моделі SUF-CMA вимагається створити підпис для будь-якого повідомлення, навіть якщо воно було вже підписано. Відповідний формальний експеримент $Exp_{A,S}^{SUF-CMA}(1^{\lambda})$ відрізняється від $Exp_{A,S}^{EUF-CMA}(1^{\lambda})$ лише тим, що список Q містить не тільки запити до оракула підпису, а й відповіді. І у кінці перевіряється, що $(m^*,\sigma^*) \notin Q$, як наведено у псевдокоді нижче.

Перевага супротивника аналогічно до EUF-CMA:

$$Adv_{A,S}^{SUF-CMA}(1^{\lambda}) = \Pr[Exp_{A,S}^{SUF-CMA}(1^{\lambda})] . \tag{6}$$

Якщо $Adv_{A,S}^{SUF-CMA}(1^{\lambda}) = negl(\lambda)$, то схема підпису вважається безпечною у моделі SUF-CMA.

2. Уточнення оцінок ДСТУ 8961:2019

ДСТУ 8961:2019 [1] використовує перетворення у полі $R_q = Z_q[X]/(X^n - X - 1)$ і грунтується на проблемі NTRU. У табл. 1 перелічено загальносистемні параметри ДСТУ 8961:2019. Параметри N, q, p визначають поле (і ідеал у цьому полі), у якому будуть виконуватися перетворення, параметри t, d_g, d_f задають кількість ненульових коефіцієнтів у поліномах.

Таблиця 1

Основні загальносистемні параметри ДСТУ 8961:2019

Параметр	Значення
N	Параметр поля. Визначає степінь поліномів.
q	Параметр поля. Визначає максимальні значення коефіцієнтів поліномів.
p	«Малий модуль». Визначає структуру таємного ключа. Для всіх наборів параметрів має фіксоване значення – 3.
t	Визначає кількість коефіцієнтів в таємному поліномі
d_g	$d_g = \lfloor 2n/3+1 \rfloor$
d_f	$d_f = 2t$

ДСТУ 8961:2019 підтримує три набори загальносистемних параметрів. Набори загальносистемних параметрів зведено в табл. 2.

Таблиця 2

Загальносистемні параметри ДСТУ 8961:2019

Набір параметрів	N	q	p	t	d_g	d_f
Skelya256	881	7673	3	159	588	318
Skelya384	1201	9221	3	192	801	384
Skelya512	1471	12251	3	255	981	510

Доказ безпеки перетворення SkelyaTransform[PKE] сформульовано в роботі [9].

Т е о р е м а 1 [9]. Нехай PKE є OW-CPA безпечною та δ_{wc}-коректною схемою асиметричного шифрування з властивістю однозначного відновлення, тоді SkelyaTransfom[PKE] є IND-CCA безпечним механізмом інкапсуляції ключів. Більш формально – для кожного квантового алгоритму A у грі IND-CCA проти KEM=SkelyaTransform[PKE], що робить $q_H, q_{BPGM}, q_{KDF}, q_D$ запитів до оракулів H,BPGM,KDF та оракула дешифрування, існує квантовий алгоритм B у грі OW-CPA проти схеми асиметричного шифрування PKE, для якого виконується нерівність

$$\begin{aligned} &Adv_{KEM}^{IND-CCA}(A) \le (2\cdot q_H + 2\cdot q_D + q_{KDF})\cdot \\ &\cdot\sqrt{Adv_{PKE}^{OW-CPA}(B)} + 8\cdot(q_{BPGM} + q_D + 1)^2\cdot\delta_{wc} \end{aligned} . \qquad (7)$$

З теореми 1 випливає, що оцінка безпеки ДСТУ 8961:2019 може бути доказово зведена до проблеми NTRU. Поліноми f та g мають коефіцієнти у множині {-1,0,1}, проте кількість ненульових елементів сильно відрізняються. Для полінома f маємо $\|f\|_\infty = 2t$, де t – загальносистемний параметр, який для усіх наборів параметрів дає кількість ненульових елементів $d_f \approx n/3$. У той же час $\|g\| = 2n/3+1$, що дає близький до рівномірного розподіл на множині {-1,0,1} для полінома g. Тож, можливо вважати, що поліном g має рівномірний розподіл і використовувати апроксимовані параметри розподілів, що отримані в роботі [5]. Для полінома f експерименти показали, що центрований нормальний розподіл з параметром $\sigma_f = 0.6$ достатньо добре апроксимує розподіл ймовірностей, що зображено на рис. 1.

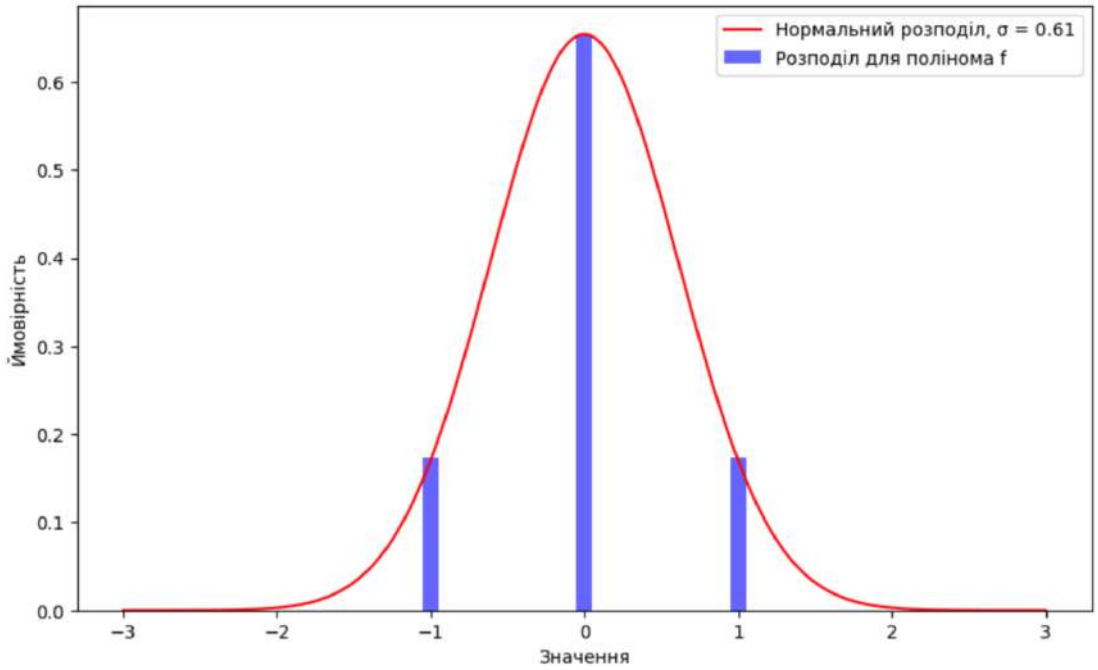


Рис. 1. Апроксимація розподілу ймовірностей для полінома f

У табл. 3 наведено оцінки захищеності від атаки вкладення з використанням запропонованої в розд. 3 моделі.

Таблиця 3

Оцінки атаки вкладення для ДСТУ 8961:2019

Набір параметрів	**Вартість атаки (біт, GSA)**	**Розмір блоку редукції**	**Вартість атаки (біт, симулятор)**	**Розмір блоку редукції**
Скеля 256	178	611	206	706
Скеля 384	252	865	288	988
Скеля 512	312	1071	355	1219

З табл. 3 видно, що врахування q-арної структури решіток дає різницю порядка 30 біт безпеки для усіх наборів параметрів.

В табл. 4 наведено оцінки складності атак декодування для запропонованої в розд. 3 моделі.

Таблиця 4

Оцінки атаки декодування для ДСТУ 8961:2019

Набір параметрів	**Вартість атаки (біт, GSA)**	**Розмір блоку редукції**	**Вартість атаки (біт, симулятор)**	**Розмір блоку редукції**
Скеля256	171	558	209	688
Скеля384	249	815	318	1050
Скеля512	312	1022	408	1349

Як видно з табл. 3, атака декодування при використанні моделі GSA для набору параметрів Скеля256 дає кращі результати, проте при врахуванні алгебраїчної структури q-арних решіток ця перевага нівелюється. На рис. 2 наведено порівняння атак вкладення та декодування.

З рис. 2 видно, що зі збільшенням розмірності q-арна структура решітки все більше впливає на оцінку складності атаки. Якщо для набору параметрів Скеля256 різниця є незначною, то для Скеля384 та Скеля512 вартість атаки декодування стрімко збільшується, у той час як у моделі GSA такого не відбувається.

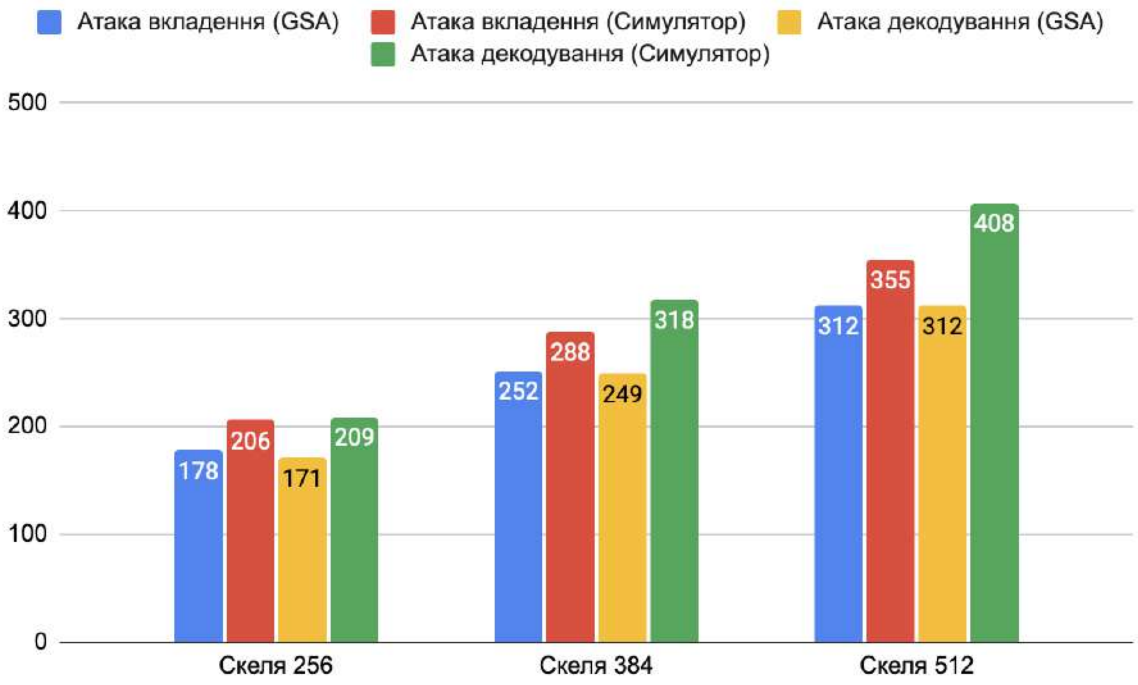


Рис. 2. Порівняння атак вкладення та декодування

Тож, атака вкладення показує себе краще для всіх наборів параметрів. Для ДСТУ8961:2019 також додатково були розраховані оцінки безпеки для гібридних атак. У табл. 5 занесено відповідні оцінки.

Таблиця 5

Оцінки гібридної атаки для ДСТУ 8961:2019

Набір параметрів	**Вартість атаки (біт, GSA)**	**Розмір блоку редукції**	**Вартість атаки (біт, симулятор)**	**Розмір блоку редукції**
Скеля 256	154	504	179	590
Скеля 384	221	723	265	873
Скеля 512	276	903	335	1105

На рис. 3 наведено порівняння оцінок гібридної атаки та атаки вкладення для ДСТУ 8961:2019.

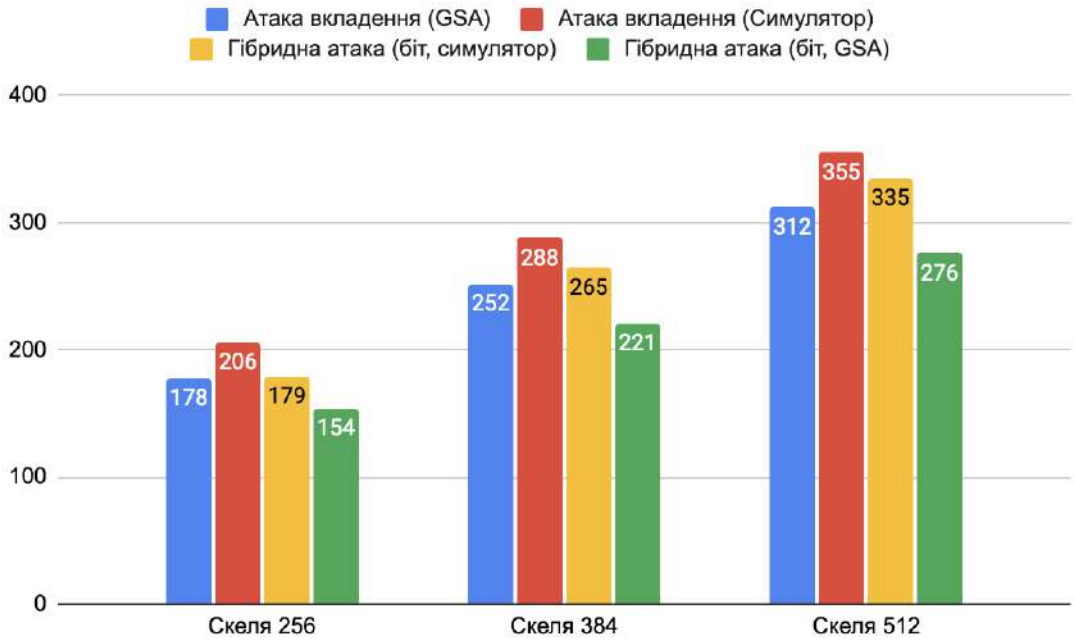


Рис. 3. Порівняння гібридної атаки та атаки вкладення для ДСТУ 8961:2019

З рис. 3 видно, що гібридна атака є найефективнішою для ДСТУ 8961:2019.

3. Уточнення оцінок fips 203 (CRYSTALS-Kyber)

Протокол інкапсуляції ключів CRYSTALS-Kyber [2] використовує перетворення у полі $R_q = Z_q[X]/(X^n+1)$ і грунтується на проблемі Module-LWE. Для отримання протокола інкапсуляції ключів використовується варіант перетворення Фуджісакі–Окамото з неявним відхиленням [10].

Згідно зі специфікацією CRYSTALS-Kyber підтримує три набори загальносистемних параметрів. Параметри зведеніо у табл. 6. Параметри n та q визначають поле, параметр k задає розмірність векторів, параметри η_1 та η_2 є параметрами біноміального розподілу для векторів s та e відповідно. Параметри d_u, d_v використовуються під час кодування поліномів у бітову строку, параметр δ є ймовірністю помилки декапсуляції. З табл. 6 добре видна суто практична перевага проблеми Module-LWE: для всіх рівнів безпеки використовується одне поле. Такий підхід дає змогу значно спростити реалізацію та масштабувати систему для довільного рівня безпеки. Окрім того, використання відносно малого значення для параметра q дозволяє ефективно використовувати векторизацію обчислень.

Таблиця 6

Загальносистемні параметри CRYSTALS-Kyber

Набір параметрів	***n***	***k***	***q***	η_1	η_2	(d_u, d_v)	δ
Kyber512	256	2	3329	3	2	(10,4)	2^{-139}
Kyber768	256	3	3329	2	2	(10,4)	2^{-164}
Kyber1024	256	4	3329	2	2	(11,5)	2^{-174}

В роботі [2] авторами Crystals-Kyber проведений детальний аналіз безпеки у моделі квантового випадкового оракула. Цими результатами можна скористатися для подальшого аналізу.

Т е о р е м а 2 [2]. Нехай XOF, H та G є випадковими оракулами. Тоді для будь-якого класичного супротивника A, що робить не більше q_{RO} запитів до випадкових оракулів XOF, H та G, існують класичні супротивники B та C, для яких

$$Adv_{Kyber}^{IND-CCA}(A) \le 2Adv_{k+1,k,\eta}^{MLWE}(B) + Adv_{PRF}^{prf}(C) + 4q_{RO}\delta. \quad (8)$$

Т е о р е м а 3 [2]. Нехай XOF, H та G є квантовими випадковими оракулами. Тоді для будь-якого квантового супротивника A, що робить не більше q_{RO} запитів до випадкових оракулів XOF, H та G, існують квантові супротивники B та C, для яких

$$Adv_{Kyber}^{IND-CCA}(A) \le 4q_{RO} \cdot \sqrt{Adv_{k+1,k,\eta}^{MLWE}(B)} + Adv_{PRF}^{prf}(C) + 8q_{RO}\delta. \quad (9)$$

З теорем 2, 3 видно, що формальні докази безпеки для Crystals-Kyber мають таку ж структуру, що і отримані докази для ДСТУ 8961:2019.

З теорем 2, 3 випливає, якщо вважати симетричні криптопримітиви безпечними, то безпека Crystals-Kyber цілком зводиться до проблеми MLWE. Оскільки для криптографічних наборів параметрів не відомо як використовувати алгебраїчну структуру MLWE, то можна вважати, що безпека Crystals-Kyber зводиться до LWE. Оскільки Crystals-Kyber використовує біноміальний розподіл, який є дискретним аналогом нормального розподілу, то аналіз полегшується. У табл. 7 наведеніо оцінки атак вкладення на проблему MLWE, що асоційована з кожним набором загальносистемних параметрів.

Таблиця 7

Оцінки атаки вкладення для CRYSTALS-Kyber

Набір параметрів	**Вартість атаки (біт, GSA)**	**Розмір блоку редукції**	**Вартість атаки (біт, симулятор)**	**Розмір блоку редукції**
Kyber512	118	406	131	449
Kyber786	182	625	200	687
Kyber1024	256	878	277.	950

З табл. 7 видно, що врахування q-арної структури решіток дає різницю порядка 20 біт безпеки для усіх наборів параметрів.

В табл. 8 наведено оцінки складності атак декодування.

Таблиця 8

Складність атак декодування для CRYSTALS-Kyber

Набір параметрів	**Вартість атаки (біт, GSA)**	**Розмір блоку редукції**	**Вартість атаки (біт, симулятор)**	**Розмір блоку редукції**
Kyber512	114	372	137	450
Kyber786	182	596	232	764
Kyber1024	263	860	354	1169

З табл. 7, 8 випливає така ж картина, як і для ДСТУ 8961:2019. Зі зростанням розмірності структура q-арних решіток все більше впливає на складність атаки. На рис. 4 наведено порівняння атак вкладення та декодування для Crystals-Kyber.

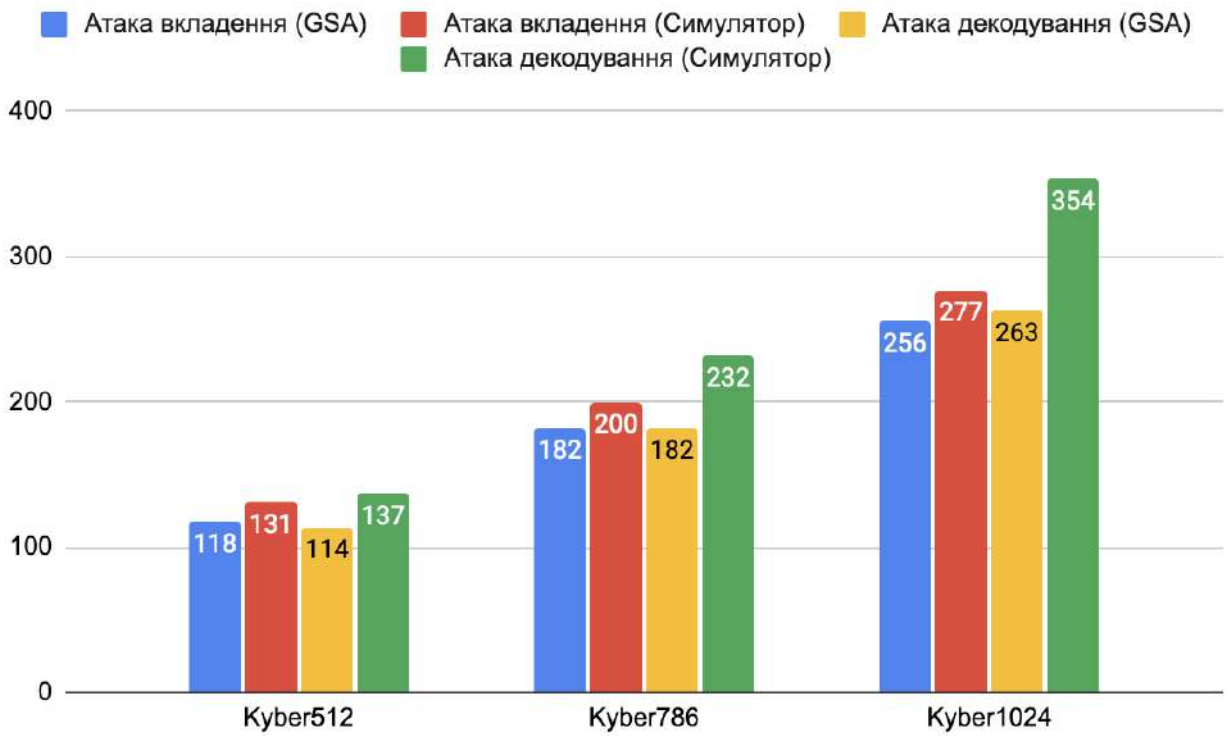


Рис. 4. Порівняння атак вкладення та декодування

4. Уточнення оцінок Falcon

Схема електронного підпису Falcon має в своїй основі фреймворк GPV, що був вперше запропонований в роботі [4] для побудови квантово-стійких електронних підписів на решітках. Сутність фреймворку GPV полягає у наступному:

Відкритий ключ задається матрицею $A \in Z_q^{n \times m}$ (де $m > n$). Ця матриця задає базис q-арної решітки Λ.

Таємний ключ задається матрицею $B \in Z_q^{m \times m}$. Ця матриця задає базис дуальної решітки $\Lambda_q^{\perp}$, яка, згідно з визначенням, є ортогональною до Λ за модулем q. Тобто, для будь-яких векторів $x \in \Lambda$ та $y \in \Lambda_q^{\perp}$ виконується $x \cdot y = 0 \bmod q$, де $\cdot$ – операція скалярного добутку.

Для заданого повідомлення m підписом є малий (у сенсі евклідової норми) вектор $s \in Z_q^m$, для якого виконується $sA^T = H(m)$, де $H : \{0,1\}^* \to Z_q^n$ – стійка до колізій геш функція. Для перевірки підпису достатньо перевірити, що виконується рівняння $sA^T = H(m)$.

Для обчислення підпису спочатку обчислюється довільний випадковий вектор $c_0 \in Z_q^m$, для якого виконується $c_0 A^T = H(m)$. Оскільки до вектора c_0 не накладається вимог щодо значень його евклідової норми, то його знайти можливо стандартними засобами лінійної алгебри за поліноміальний час. Далі використовується таємний базис B для обчислення вектора $z \in \Lambda_q^{\perp}$, який є близьким до вектора c_0. Різниця векторів $s = c_0 - z$ є коректним підписом, оскільки $sA^T = c_0 A^T - zA^T = c - 0 = H(m)$. Якщо c_0 та v є достатньо близькими, то s буде малим.

Електронний підпис Falcon використовує у якості решітки Λ NTRU решітку [4]. Застосовуючи NTRU решітки до фреймворку GPV, Falcon вносить наступні зміни до GPV:

Відкритим ключем є поліном *h*, який використовується для обчислення публічного базису NTRU решітки Λ.

Таємним ключем є поліноми $f, g, F, G \in Z[x]/(\phi)$, які використовуються для обчислення базису дуальної решітки $\Lambda_q^{\perp}$.

Підпис для повідомлення *m* складається з пари поліномів (s_1, s_2), для яких виконується $s_1 + s_2 h = H(r \| m)$, де *r* – сіль (salt). При обчисленні підпису використовується таємний ключ для обчислення вектора $z = (z_0, z_1) \in \Lambda_q^{\perp}$, який є близьким до вектора $t = (H(m \| r), 0)$. Різниця векторів *t* та *z* є коректним підписом.

Для обчислення вектору $z = (z_0, z_1)$ використовується алгоритм семплування (вибірки), що повертає вектор з нормального розподілу. Особливістю алгоритму семплування Falcon [4] є використання для пришвидшення операцій алгебраїчної структури циклотомічного поля та перетворення Фур'є. Falcon також використовує деревовидні структури даних – LDL дерева. Деталі можливо знайти в специфікації [4].

Falcon використовує наступні загальносистемні параметри:

- Параметри поля *(n,q)*
- Параметр розподілу таємних ключів $\sigma_{\{f,g\}} = 1.17\sqrt{q/2n}$
- Параметр розподілу підписів σ
- Обмеження на максимальний розмір підписів *B*

Загальносистемні параметр Falcon зведені в табл. 9.

Таблиця 9

Загальносистемні параметри Falcon

Параметр	**Falcon512**	**Falcon1024**
(n,q)	(512,12289)	(1024,12289)
$\sigma_{\{f,g\}}$	4.0531638033	2.86601961058
σ	165.736 617 183	168.388 571 447
B	5833.92886484	8382.43651929

Не зважаючи на те, що Falcon є фіналістом конкурсу NIST, безпосередньо його аналізу у моделі EUF-CMA присвячено не так багато робіт. Оскільки схема підпису ґрунтується на фрейворку GPV, то можливо адаптувати докази з оригінальної роботи.

Проте, можливо довести безпеку іншим шляхом. Фреймворк GPV є частковим випадком парадигми Hash-and-Sign. В останні роки для парадигми Hash-and-Sign з'явилося багато робіт щодо безпеки EUF-CMA у моделі квантового випадкового оракула. Кожен результат ґрунтується на певних модельних припущеннях. Результат у роботі [11] зручно використовувати, оскільки він грунтується на тих самих припущеннях, що і докази безпеки фреймворку GPV.

У загальному випадку підпис Hash-and-Sign параметризується стійкою до колізій геш функцією *H* та односторонньою функцією з лазівкою *T*, що є стійкою до знаходження прообразу (англ. Preimage-resistant trapdoor function). У випадку Falcon геш-функція *H* реалізується через shake256, тож будемо вважати, що вона є криптографічною. Одностороння функція *T* в Falcon є функцією з фреймворку GPV, до якої додана структура NTRU решітки.

Адаптуємо основний результат роботи [11] до схеми підпису Falcon наступним чином:

Т е о р е м а 4 [11]. Для будь-якого квантового супротивника *A* у грі EUF-CMA для схеми підпису Falcon, що робить не більше q_{sign} класичних запитів до оракулу підпису та q_{qro} квантових запитів до квантового оракулу *H*, існує супротивник *B*, що може

інвертувати односторонню функцію *T*, та супротивник *D*, що може знайти прообраз для *T*, використовуючи q_{sign} запитів до оракулу підпису. При цьому перевага супротивника *A* становить

$$Adv^{EUF-CMA}_{A,Falcon}(1^{\lambda}) \le (2q_{ro}+1)^2 Adv^{INV}_{T}(B) + Adv^{PS}_{T}(D) + 3/2q'_{sign}\sqrt{\frac{q'_{sign}+q_{qro}+1}{|R|}} + 2(q_{ro}+2)\sqrt{\frac{q'_{sign}-q_{sign}}{|R|}}, \quad (10)$$

де $|R|$ – розмір простору бітових строк, що використовуються у якості випадкових значень; q'_{sign} – максимальна загальна кількість запитів до оракула *H* в усіх запитах на підпис; $Adv^{INV}_{T}(B)$ – перевага супротивника *B* в інвертуванні *T*; $Adv^{PS}_{T}(D)$ – перевага супротивника *D* в знаходженні прообразу *T*.

Якщо *A* робить тільки класичні запити до оракула *H*, то

$$Adv^{EUF-CMA}_{A,Falcon}(1^{\lambda}) \le (2q_{ro}+1)^2 Adv^{INV}_{T}(B) + Adv^{PS}_{T}(D) + q'_{sign}\frac{q'_{sign}+q_{qro}+1}{|R|} + (q_{ro}+1)\frac{q'_{sign}-q_{sign}}{|R|}. \quad (11)$$

Для Falcon $|R| = |\{0,1\}^{384}|$. Тож, доказ безпеки у моделі EUF-CMA зводить безпеку Falcon до безпеки односторонньої функції з лазівкою *T*: до складності інвертування та складності пошуку прообразу.

Інвертування *T* означало б вирішення проблеми NTRU, тож

$$Adv^{INV}_{T}(B) \le Adv^{NTRU}_{n,q,\sigma}. \quad (12)$$

Знаходження прообразу *T* є рішенням $s=(s_1,s_2)$ рівняння $s_1 + s_2 h = H(r \| m)$. Знаходження рішення рівняння є в точності проблемою ISIS з параметром *B*, тому

$$Adv^{PS}_{T}(D) \le Adv^{ISIS}_{n,q,B} \le Adv^{SIS}_{n,q,B}. \quad (13)$$

Тож, безпеку Falcon можливо звести до проблем NTRU та SIS на NTRU решітках.

Оскільки задача інвертування односторонньої функції в електронному підписі Falcon зводиться до проблеми NTRU, то конкретні оцінки складності зводяться до оцінки складності атак вкладення та декодування.

У табл. 10 наведені оцінки безпеки екземплярів проблеми NTRU, на яку спирається Falcon, від атак вкладення.

Таблиця 10

Атаки вкладення для проблеми NTRU

Falcon	**Вартість атаки (біт, GSA)**	**Розмір блоку редукції**	**Вартість атаки (біт, симулятор)**	**Розмір блоку редукції**
Falcon512	141	483	147	505
Falcon1024	268	918	285	979

У табл. 11 наведено аналогічні оцінки для атак декодування на проблему NTRU.

Таблиця 11

Атаки декодування для проблеми NTRU

Falcon	Вартість атаки (біт, GSA)	Розмір блоку редукції	Вартість атаки (біт, симулятор)	Розмір блоку редукції
Falcon512	134	439	164	538
Falcon1024	277	907	370	1222

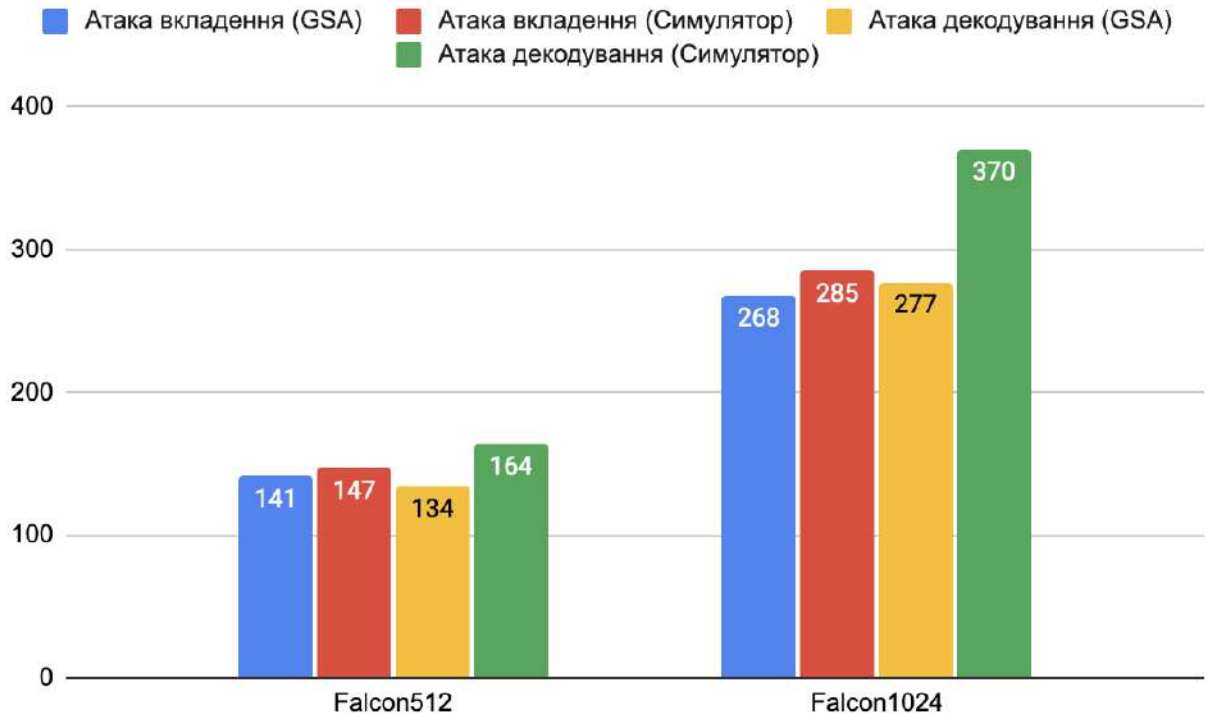


Рис. 5. Порівняння атак вкладення і декодування для Falcon

З табл. 10, 11 та рис. 5 видно, що для Falcon512 складність атак декодування та вкладення не сильно відрізняється, проте при врахуванні алгебраїчної структури решіток для Falcon1024 атаки декодування стають значно гіршими.

У табл.12 зведено результати оцінки безпеки Falcon для підробки підпису (задача SIS) для наборів параметрів.

Таблиця 12

Результати оцінки безпеки Falcon (SIS)

Falcon	Вартість атаки (біт, GSA)	Розмір блоку редукції	Вартість атаки (біт, симулятор)	Розмір блоку редукції
Falcon512	114	373	136	446
Falcon1024	268	878	355	1169

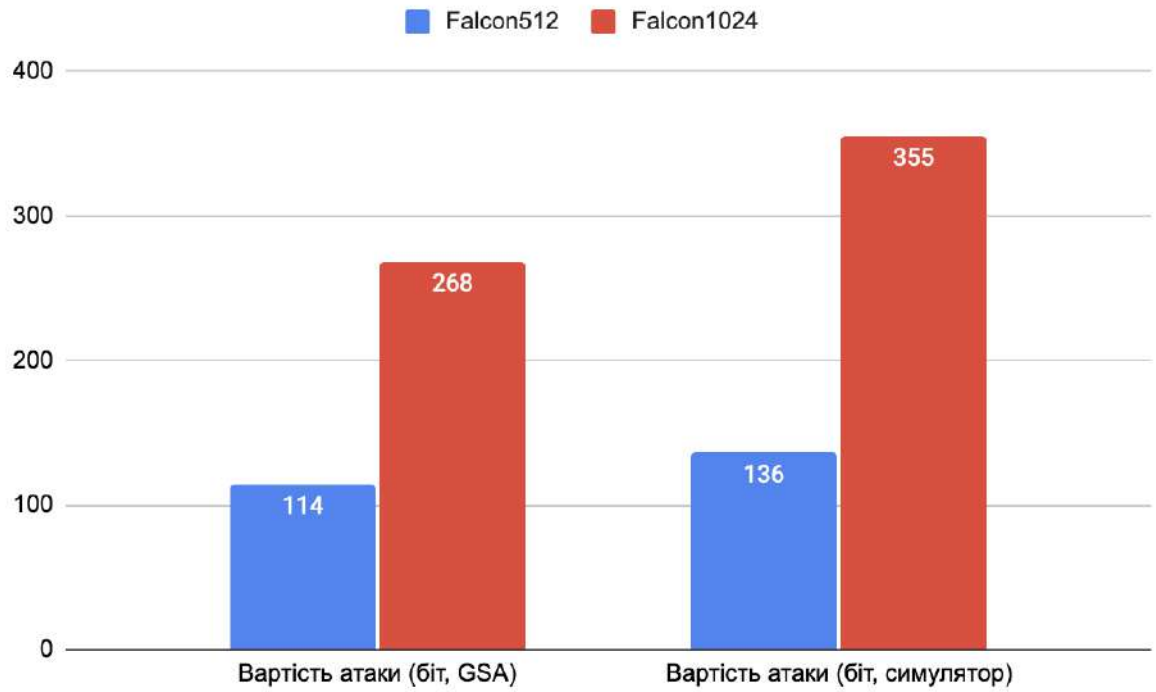


Рис. 6. Вартість атаки на SIS для різних моделей безпеки.

З рис. 6 видно, що попередні спостереження зберігаються.

5. Уточнення оцінок fips 204 (CRYSTALS-Dilithium)

Схема електронного підпису CRYSTALS-Dilithium [3] має в основі перетворення Фіата–Шаміра з перериваннями (*англ*. Fiat–Shamir with Aborts).

У цій схемі параметрами безпеки є

- Параметри поля *(n,q)*, які визначають поле $R_q = Z_q[X]/(X^n+1)$;
- Параметр *d*, що визначає кількість біт, які будуть відкинути з коефіцієнтів вектору *t*;
- Параметр τ, що визначає кількість ± 1 у поліномі *c*;
- Параметр γ_1, що визначає діапазон коефіцієнтів у векторі *y*;
- Параметр γ_2, що визначає параметри округлення;
- Параметри *(k,l)*, що визначають розмірність матриці *A*;
- Параметр η, що визначає діапазон коефіцієнтів в таємному ключі.

Параметри Crystals-Dilithium зведені у табл. 13.

Таблиця 13

Параметри Crystals-Dilithium

Рівень безпеки	**2**	**3**	**5**
(q,n)	(8380417, 512)	(8380417, 512)	(8380417, 512)
d	13	13	13
τ	39	49	60
γ_1	2^{17}	2^{19}	2^{19}
γ_2	(q-1)/88	(q-1)/32	*(q-1)/32*
(k,l)	(4,4)	(6,5)	(8,7)
η	2	4	2

Оскільки CRYSTALS-Dilithium був у центрі уваги конкурсу NIST, то різними авторами для нього був проведений аналіз безпеки у моделі квантового випадкового оракула.

Найбільш детальний аналіз проведений у роботі [12]. Безпека CRYSTALS-Dilithium окрім стандартних проблем Module-SIS та Module-LWE також ґрунтується на нестандартній проблемі SelfTargetMSIS.

Сутність проблеми SelfTragetMSIS полягає у тому, щоб знайти вектор *y*, для якого виконується

$\|y\|_\infty \le \gamma$ для заданного γ

$H([I \mid A]\cdot y \,\|\, M) = c$ для заданих $A \in R_q^{m\times k}, M, c$.

Якщо деякий супротивник *A* має перевагу $Adv_{m,k,2\gamma}^{MSIS}$ у вирішенні проблеми MSIS, то у супротивника *B* перевага у вирішенні проблеми SelfTargetMSIS буде

$$Adv_{H,m,k,\gamma}^{SelfTargetMSIS}(B) \approx \sqrt{Adv_{m,k,2\gamma}^{MSIS}(A)/Q_H}\,, \tag{14}$$

де Q_H – кількість запитів до квантового випадкового оракула *H*.

Для CRYSTALS-Dilithium перевага супротивника *A* у грі SUF-CMA складає

$$Adv_{Dilithium}^{SUF-CMA} \le Adv_{k,l,D}^{MLWE} + Adv_{H,k,l+1,\zeta_1}^{SelfTargetMSIS} + Adv_{k,l,\zeta_2}^{MSIS} + 2^{-\alpha+1} \tag{15}$$

Де α є мінімальною ентропією схеми,

$$\begin{aligned} \zeta_1 &= \max\{\gamma_1 - \beta, 2\gamma_2 + 1 + 2^{d-1}\cdot\rho\},\\ \zeta_2 &= \max\{2(\gamma_2 - \beta), 4\gamma_1 + 2\} \end{aligned} \tag{16}$$

Мінімальну ентропію α для CRYSTALS-Dilithium можливо розрахувати як

$$\alpha > nl\cdot\log(\min(q/((4\gamma_1+1)(4\gamma_2+1)), 2\gamma_2 - 1)\,, \tag{17}$$

якщо $2\gamma_1, 2\gamma_2 < \sqrt{q/2}$ та $l \le k$.

З формул вище випливає, що для оцінки безпеки схеми необхідно оцінити складність вирішення проблеми MLWE з параметрами *k,l,D*, проблем MSIS з параметрами k, l, ζ_1 та $k, l+1, \zeta_2$. І мінімальна ентропія схеми повинна перевищувати цільовий рівень безпеки.

Оскільки алгебраїчну структуру проблем MLWE та MSIS для криптографічних наборів параметрів невідомо як використовувати, то можливо розглядати відповідні проблеми LWE та SIS.

У табл. 14 зведено результати оцінки безпеки CRYSTALS-Dilithium від атак вкладення для моделі безпеки.

Таблиця 14

Оцінка складності атаки вкладення для Crystals-Dilithium

Набір параметрів	**Складність атаки (біт, GSA)**	**Розмір блоку редукції**	**Складність атаки (біт, Симулятор)**	**Розмір блоку редукції**
Dilithium2	123	424	170	583
Dilithium3	182	625	236	811
Dilithium5	252	864	331	1134

У табл. 15 зведено результати оцінки безпеки CRYSTALS-Dilithium від атак вкладення для моделі безпеки.

Таблиця 15

Оцінка складності атаки декодування для Crystals-Dilithium

Набір параметрів	Складність атаки (біт, GSA)	Розмір блоку редукції	Складність атаки (біт, Симулятор)	Розмір блоку редукції
Dilithium2	124	406	171	565
Dilithium3	186	609	247	234
Dilithium5	261	246	352	1164

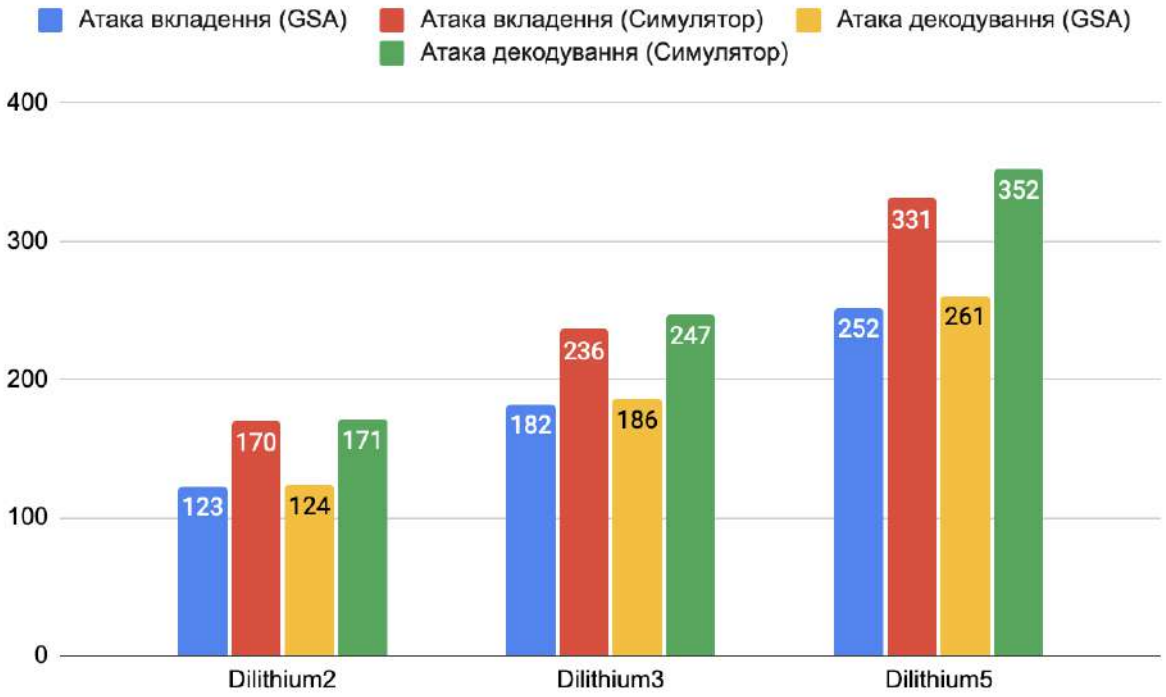


Рис. 7. Порівняння атак вкладення та декодування

У табл. 16 зведено результати оцінки безпеки CRYSTALS-Dilithium (SIS) для наборів параметрів, що представлені в табл. 1.

Таблиця 16

Оцінка складності атаки декодування для Crystals-Dilithium

Dilithium	SIS (автори)	SIS (наша)	SelfTargetSIS (автори)	SelfTargetSIS (наша)
Dilithium2	123	113	121	111
Dilithium3	186	135	175	127
Dilithium5	265	208	253	199

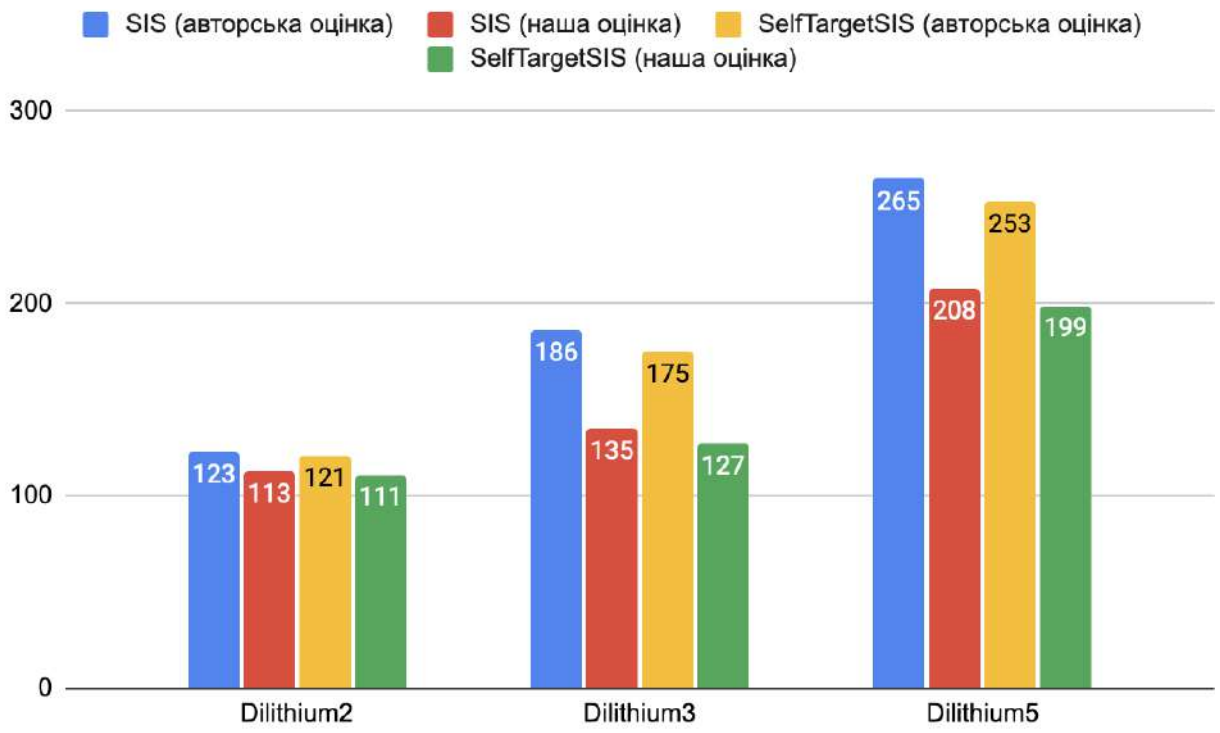


Рис. 8. Порівняння атак на SIS

Висновки

1. Уточнено оцінки безпеки механізмів інкапсуляції ключів ДСТУ 8961:2019 та Crystals-Kyber. В залежності від набору параметрів різниця між оцінками у моделі GSA та моделі, що враховує алгебраїчну структуру q-арних решіток, складає 20–30 біт безпеки. Причому, уточнені оцінки показують, що існуючі атаки є менш ефективними, ніж припускалося при використанні моделі GSA.

2. Атака декодування для 1 рівню безпеки NIST показує кращі результати за атаку вкладення. Для третього рівня безпеки NIST атака декодування має приблизно однакову складність з атаками вкладення. Проте, для п'ятого рівня безпеки атаки декодування значно програють атакам вкладення. Це пояснюється тим, що у атаках декодування форма базису сильніше впливає на параметри атаки.

3. Уточнено оцінки безпеки електронних підписів Falcon та Crystals-Dilithium. В залежності від набору параметрів різниця між оцінками у моделі GSA та моделі, що враховує алгебраїчну структуру q-арних решіток, також складає 20-30 біт безпеки. Причому, уточнені оцінки показують, що існуючі атаки є менш ефективними, ніж припускалося при використанні моделі GSA. Для атак декодування ефективність атаки стрімко падає з ростом розмірності решіток, у той час як для атак вкладення різниця не є такою сильною.

Список літератури:

1. ДСТУ 8961:2019. Інформаційні технології. Криптографічний захист інформації. Алгоритм асиметричного шифрування та інкапсуляції ключів. Чинний від 21.12.2019. Вид. офіц. Київ : УкрНДНЦ, 2019. 72 с.

2. National Institute of Standards and Technology (2024) Module-lattice-based key-encapsulation mechanism standard, CSRC. Available at: https://csrc.nist.gov/pubs/fips/203/final (Accessed: 13 October 2024).

3. National Institute of Standards and Technology (2024a) Module-lattice-based digital signature standard, CSRC. Available at: https://csrc.nist.gov/pubs/fips/204/final (Accessed: 13 October 2024).

4. [PDF] falcon: Fast-fourier lattice-based compact signatures over NTRU | Semantic scholar. Available at: https://www.semanticscholar.org/paper/Falcon:-Fast-Fourier-Lattice-based-Compact-over-Fouque-Hoffstein/423e31b1b96ffa0559078961963baeeb98f01e19 (Accessed: 13 October 2024).

5. Kandii S.O. and Gorbenko, I.D. Assessing the influence of the algebraic structure of q-ary lattices on the complexity of cryptanalysis of problems on lattices // Radiotekhnika. 2024. No217. P. 79–99. doi:10.30837/rt.2024.2.217.07.
6. Bellare M. et al. (1998) Relations among notions of security for public-key encryption schemes // Lecture Notes in Computer Science. 1998. P. 26–45. doi:10.1007/bfb0055718.
7. Goldwasser S., Micali S. and Rivest R.L. A digital signature scheme secure against adaptive chosen-message attacks // SIAM Journal on Computing. 1988. 17(2). P. 281–308. doi:10.1137/0217017.
8. Hövelmanns K., Hülsing A. and Majenz C. Decryption failures and the Fujisaki-Okamoto Transform, Cryptology ePrint Archive. Available at: https://eprint.iacr.org/2022/365.pdf (Accessed: 13 October 2024).
9. Kandii S.O. and Gorbenko I.D. Analysis of DSTU 8961:2019 in the quantum random Oracle Model // Radiotekhnika. 2023. No214. P. 7–16. doi:10.30837/rt.2023.3.214.01.
10. Lyubashevsky V. Fiat-Shamir with aborts: Applications to lattice and factoring-based signatures // Lecture Notes in Computer Science. 2009. P. 598–616. doi:10.1007/978-3-642-10366-7_35.
11. Kosuge H. and Xagawa K. Probabilistic hash-and-sign with retry in the quantum random Oracle Model // Lecture Notes in Computer Science. 2024. P. 259–288. doi:10.1007/978-3-031-57718-5_9.
12. Kiltz E., Lyubashevsky V. and Schaffner C. A concrete treatment of Fiat-Shamir signatures in the quantum random-oracle model // Lecture Notes in Computer Science. 2018. P. 552–586. doi:10.1007/978-3-319-78372-7_18.

Надійшла до редколегії 07.09.2024

Відомості про авторів:
Кандій Сергій Олегович – Харківський національний університет імені В. Н. Каразіна, аспірант кафедри кібербезпеки інформаційних систем, мереж і технологій, навчально-науковий інститут комп'ютерних наук та штучного інтелекту, АТ «Інститут Інформаційних технологій», науковий консультант; Україна; e-mail: sergeykandy@gmail.com; ORCID: https://orcid.org/0000-0003-0552-8341
Горбенко Іван Дмитрович – д-р техн. наук, професор, Харківський національний університет імені В.Н. Каразіна, професор кафедри кібербезпеки інформаційних систем, мереж і технологій, навчально-науковий інститут комп'ютерних наук та штучного інтелекту, АТ «Інститут інформаційних технологій», головний конструктор; Україна; e-mail: i.d.gorbenko@karazin.ua; ORCID: https://orcid.org/0000-0003-4616-3449

УДК 004.056:519.2 **DOI:10.30837/rt.2024.3.218.07**

А.М. ОЛЕКСІЙЧУК, д-р техн. наук, І.В. САМОЙЛОВ, канд. техн. наук

ЙМОВІРНІСНІ ВЛАСТИВОСТІ РОЗВ'ЯЗКІВ СИСТЕМ РІВНЯНЬ ГАМОУТВОРЕННЯ ГЕНЕРАТОРІВ ГАМИ З НЕРІВНОМІРНИМ РУХОМ

Вступ

Традиційною основою для побудови сучасних потокових шифрів є генератори гами, які базуються на лінійних регістрах зсуву (ЛРЗ) та нелінійних елементах ускладнення. Одним з відомих методів підвищення стійкості таких генераторів до алгебраїчних та кореляційних атак, є введення нерівномірності в процес руху ЛРЗ. Зазвичай нерівномірність руху забезпечується одним із двох способів: шляхом зовнішнього управління рухом або шляхом самоуправління, тобто встановлення детермінованої залежності кількості зсувів ЛЗР генератора в кожному такті від його поточного стану (див., наприклад [1, 2]).

Найвідоміші на сьогодні генератори гами з нерівномірним рухом, які застосовуються у потокових шифрах A5/1 [3], Alpha1 [4], LILI-128 [5] та деяких інших, ретельно досліджено ще у дев'яності та нульові роки. Проте інтерес фахівців до таких генераторів зберігається і сьогодні, про що свідчать нещодавні публікації [6, 7], присвячені новим атакам на шифр A5/1 та деякі інші потокові шифри, побудовані на базі генераторів гами з нерівномірним рухом.

Відомо, що за певних загальних умов нерівномірність руху ЛРЗ покращує криптографічні властивості генератора, збільшує значення періодів та еквівалентних лінійних складностей його вихідних послідовностей, підвищує його стійкість відносно кореляційних атак [8, 9]. Проте відомі методи оцінювання стійкості генераторів гами з нерівномірним рухом, розвинуті переважно в дев'яності роки [8 – 13], базуються на спрощеному описі їх функціонування. З одного боку, це надає змогу охопити широкий клас різноманітних генераторів гами, але, з іншого – знижує точність та адекватність висновків про стійкість окремих з них. Як приклад, відзначимо генератори з векторним зовнішнім управлінням рухом, які являють собою багатовимірне узагальнення найбільш дослідженого класу ЛРЗ з нерівномірним рухом, до якого відносяться генератори гами потокових шифрів A5/1 та Alpha1.

Мета статті – дослідити ймовірнісні властивості розв'язків системи рівнянь (СР) гамоутворення довільних та комбінувальних, генераторів гами з нерівномірним рухом.

В п. 1 наведено загальну ймовірнісну модель, яка описує функціонування таких генераторів [2].

В п. 2 за допомогою теоретико-автоматних методів отримано матричне представлення для середнього числа розв'язків СР гамоутворення генератора гами з нерівномірним рухом. Встановлено умови, за яких комбінувальний генератор із зовнішнім управлінням є необоротним за Гаффманом, а також достатні умови експоненційного росту середнього числа розв'язків системи гамоутворення цього генератора від довжини його вихідної послідовності.

В п. 3 отримано аналітичні вирази та оцінки розподілів ймовірностей сум випадкових векторів, які виробляються блоками управління рухом певного класу комбінувальних генераторів гами.

1. Ймовірнісна модель генератора гами з нерівномірним рухом

За означенням генератор гами являє собою скінченний автономний автомат $\Im = (S, Y, h, f)$, де S та Y позначають відповідно внутрішній та вихідний алфавіти автомата $\Im$, $h: S \to S$ і $f: S \to Y$ є відповідно функціями переходів та виходів цього автомата (див., наприклад, [1, 2]).

Позначимо $x(0) \in S$ початковий стан автомата $\Im$, $\bar{x} = \{x(i) : i = 0,1, ...\}$ – його внутрішню послідовність,

$$x(i) = h^i(x(0)),\ i = 0,1, \ldots . \tag{1}$$

Нехай, далі, (A, p_A) – дискретне джерело без пам'яті, де $A \subseteq \mathbf{N}_0$, p_A – розподіл ймовірностей на множині A. Джерело виробляє послідовність незалежних випадкових величин (ВВ) $\varepsilon(0), \varepsilon(1), \ldots$, кожна з яких розподілена на множині A за законом p_A:

$$\mathbf{P}(\varepsilon(i) = a) = p_A(a),\ a \in A,\ i = 0,1, \ldots .$$

Розглянемо ВВ $\delta(0) \equiv 0$, $\delta(i) = \varepsilon(0) + \ldots + \varepsilon(i-1)$, $i = 0,1, \ldots$, та визначимо таку послідовність знаків алфавіту *Y*, що відповідає послідовностям (1) і $\bar{\delta} = \{\delta(i) : i = 0,1,...\}$:

$$y(i) = f(x(\delta(i))),\ i = 0,1,\ldots . \tag{2}$$

Відзначимо, що $\bar{y} = \{y(i) : i = 0,1,\ldots\}$ є випадковою послідовністю, яка залежить від $\bar{\delta}$ та початкового стану $x(0)$ автомата $\Im$. Надалі вважатимемо, що $x(0)$ є випадковим елементом, який не залежить від $\bar{\delta}$ та має рівномірний розподіл ймовірностей на множині S.

Співвідношення (1), (2) (разом із зазначеними вище обмеженнями щодо законів розподілу ВВ $x(0)$ та $\delta(i)$, $i = 0,1, \ldots$) являють собою загальну ймовірнісну модель генератору гами з нерівномірним рухом [2].

Зазвичай на практиці в ролі джерела (A, p_A) використовується деякий автономний автомат, що виробляє послідовність $\varepsilon(0), \varepsilon(1), \ldots$ невід'ємних цілих чисел. Такий автомат називають блоком управління рухом генератора гами $\Im$. Як додаткове припущення часто приймають таку умову: $x(0), x(1), \ldots$ є послідовністю незалежних ВВ, що рівномірно розподілені на множині S. Іншими словами, замість псевдовипадкової послідовності (1) розглядають випадкову послідовність $\{x(i) : i = 0,1, ...\}$, знаки якої є незалежними рівномірно розподіленими на множині S випадковими величинами.

Надалі вважатимемо, що $p_A(a) > 0$ для кожного $a \in A$. В цьому випадку генератор гами, функціонування якого описується співвідношеннями (1) і (2), називається генератором гами з A-рухом. Говорять про обмежений A-рух, якщо $|A| < \infty$, та необмежений A-рух – у протилежному випадку [9, 11].

Приклад 1. Нехай $\Im = (S, Y, h, f)$ – регістр зсуву довжини n з лінійним зворотним зв'язком та функцією ускладнення $f = f(x_1, ..., x_n)$, де $S = V_n = \{0,1\}^n$, $Y = \{0,1\}$. Позначимо $m(x) = x^n \oplus c_{n-1}x^{n-1} \oplus ... \oplus c_0$, де $c_i \in Y$, $i \in \overline{0,\ n-1}$, поліном зворотного зв'язку регістра, U – його супровідну матрицю. Тоді стан $x(i) = (x_0(i), \ldots, x_{n-1}(i))$ регістра $\Im$ в i-му такті визначається за формулою $x(i) = x(0)U^i$, $i = 0,1, \ldots$.

Нехай, далі, $i_1, ..., i_k$ – номери точок знімання інформації з накопичувача регістра зсуву на входи функції f, $0 \le i_1 < ... < i_k \le n-1$, $\prod$ – $(n \times k)$-матриця, j-й стовпець якої має єдину одиницю в j-му рядку $(j \in \overline{1,k})$ та нулі в інших рядках. Тоді знак вихідної послідовності регістра $\Im$ в i-му такті дорівнює $f(x(0)U^i \prod)$, $i = 0,1, \ldots$. Якщо (A, p_A) – блок управління рухом регістра $\Im$, то він виробляє знаки двійкової послідовності $y(i) = f(x(0)U^{\delta(i)} \prod)$, $i = 0,1,\ldots$, де $\delta(0) \equiv 0$, $\delta(i) = \varepsilon(0) + ... + \varepsilon(i-1)$, $\varepsilon(i)$ – незалежні випадкові величини, що розподілені на множині A за законом p_A, $i = 0,1, \ldots$.

Відзначимо, що у більшості публікацій досліджуються властивості найпростішого класу лінійних регістрів зсуву з нерівномірним рухом, які характеризуються умовами $k = 1$, $f(x_1) = x_1$, $x_1 \in \{0,1\}$.

Приклад 2. Розглянемо більш загальний варіант управління рухом автономного автомата $\Im$, що є каскадом паралельного з'єднання n автономних автоматів без виходу та автомата без пам'яті [2].

Нехай $S = V_{L_1} \times ... \times V_{L_n}$, $\Im = (S, Y, h, f)$ – автономний автомат з функцією переходів

$$h(z_1, ..., z_n) = (h_1(z_1), ..., h_n(z_n)), \quad z_j \in V_{L_j}, \ j \in \overline{1, n}, \tag{3}$$

де $h_j : V_{L_j} \to V_{L_j}$, $j \in \overline{1, n}$. Нехай, далі,

$$(x_1(i), ..., x_n(i)) = (h_1^i(x_1(0)), ..., h_n^i(x_n(0))), \ i = 0, 1, ... \tag{4}$$

внутрішня послідовність автомата $\Im$, що відповідає його початковому стану $(x_1(0), \ ..., x_n(0))$.

Припустимо, що $A \subseteq \mathbf{N}_0^n$, а $\varepsilon(i) = (\varepsilon_1(i), ..., \varepsilon_n(i))$ є n-вимірним випадковим вектором, розподіленим на множині A за законом p_A, $i = 0, 1, \ ...$. Позначимо $\delta(0) \equiv 0$, $\delta(i) = (\delta_1(i), ..., \delta_n(i)) = \varepsilon(0) + ... + \varepsilon(i-1)$, $i = 0, 1, \ ...$ та задамо випадкову послідовність

$$y(i) = f(x_1(\delta_1(i)), ..., x_n(\delta_n(i))), \ i = 0, 1, ... \tag{5}$$

Співвідношення (4), (5) являють собою n-вимірне узагальнення співвідношень (1), (2). Назвемо генератор, який функціонує за законами (4), (5), генератором гами з векторним зовнішнім управлінням. Конкретним прикладом такого генератора є комбінувальний генератор гами з нерівномірним рухом [1, 2].

2. Теоретико-автоматний підхід до аналізу криптографічних властивостей генераторів гами з нерівномірним рухом

Нехай $\Im = (S, Y, h, f)$ – довільний автономний автомат, (A, p_A) – блок управління його рухом, де $A \subseteq \mathbf{N}_0$, $|A| < \infty$, p_A – рівномірний розподіл ймовірностей на множині A.

Задамо новий автомат $\Im_A = (A, S, Y, h_A, f_A)$ із вхідним алфавітом A та функціями переходів і виходів

$$h_A(x, \varepsilon) = h^{\varepsilon}(x), \ x \in S, \ \varepsilon \in A, \tag{6}$$

$$f_A(x, \varepsilon) = f(x), \ x \in S, \ \varepsilon \in A \tag{7}$$

відповідно. Зрозуміло, що, за умови фіксації початкового стану $x(0)$ автомата $\Im_A$, він переробляє довільну вхідну послідовність $\varepsilon(0), \varepsilon(1), ..., \varepsilon(t-1)$, $t \in \mathbf{N}$, у вихідну послідовність $f(x(0)), f(x(\delta(1))), ..., f(x(\delta(t)))$, яка співпадає з початковим фрагментом вихідної послідовності генератора гами з нерівномірним рухом, що функціонує за законами (1), (2). Отже, цей генератор можна розглядати як скінченний автомат $\Im_A$, який перетворює вхідні послідовності, що виробляються блоком управління рухом автомата $\Im$, у вихідні послідовності вигляду (2).

Зазначена інтерпретація процесу функціонування генератора гами з нерівномірним рухом надає змогу безпосередньо застосовувати до аналізу його криптографічних властивостей відомі теоретико-автоматні методи. Задача відновлення початкового стану генератора гами з нерівномірним рухом за його вихідною послідовністю зводиться до однієї зі стандартних загальних задач теорії автоматів: відновлення початкового стану скінченного автомату за відомим виходом при невідомому вході (але відомому розподілі ймовірностей на вхідному алфавіті) [14].

Зауважимо, що система рівнянь гамоутворення генератора $\Im_A$ у тактах з номерами $1, 2, ..., t$ має такий вигляд:

$$
\begin{aligned}
f(h^{\varepsilon(0)}(x(0))) &= y(1),\\
\dots \quad \dots \quad \dots \quad &\dots\\
f(h^{\varepsilon(0)+\dots+\varepsilon(i-1)}(x(0))) &= y(i),\\
\dots \quad \dots \quad \dots \quad &\dots\\
f(h^{\varepsilon(0)+\dots+\varepsilon(t-1)}(x(0))) &= y(t),
\end{aligned}
\tag{8}
$$

де $x(0)$ та $(\varepsilon(0),\dots,\varepsilon(t-1))$ є відповідно невідомий початковий стан та невідома вхідна послідовність автомата $\Im_A$, $\bar{y}=(y(1),\dots,y(t))$ – його відома вихідна послідовність.

Позначимо $\eta_t(\bar{y})$ число розв'язків $(x(0),\varepsilon(0),\dots,\varepsilon(t-1))$ системи рівнянь (8). Зауважимо, що, згідно з припущеннями з п. 1 $\eta_t(\bar{y})$ є випадковою величиною, розподіл якої характеризує, зокрема, обчислювальну складність відновлення початкового стану $x(0)$ автомата $\Im_A$ за фрагментом його вихідної послідовності $\bar{y}$ шляхом повного перебору всіх можливих розв'язків системи (8). Це відноситься також і до більш загального варіанту зовнішнього управління рухом автомата $\Im$, коли $A\subseteq \mathbf{N}_0^n$ (див. приклад 2).

Відомо [14], що характер асимптотичної поведінки розподілу ВВ $\eta_t(\bar{y})$ при $t\to\infty$ залежить від того, чи володіє автомат $\Im_A$ властивістю оборотності за Гаффманом. Нагадаємо (див., наприклад, [2]), що автомат $\Im_A$ називається необоротним за Гаффманом, якщо існують стан $x_1\in S$ та вхідні послідовності $\varepsilon_1,\dots,\varepsilon_k$ і $\varepsilon_1',\dots,\varepsilon_k'$, де $k\ge 3$, такі, що

(*а*) $(x_1,\varepsilon_1)=(x_1,\varepsilon_1')$, $(x_k,\varepsilon_k)=(x_k',\varepsilon_k')$;

(*б*) $(\varepsilon_2,\dots,\varepsilon_{k-1})\ne(\varepsilon_2',\dots,\varepsilon_{k-1}')$;

(*в*) $f(x_i,\varepsilon_i)=f(x_i',\varepsilon_i')$, $i\in\overline{1,k}$,

де x_i, x_i' визначаються рекурентно за формулами $x_{i+1}=h^{\varepsilon_i}(x_i)$, $x_{i+1}'=h^{\varepsilon_i'}(x_i')$, $i\in\overline{1,k-1}$.

З результатів [14] випливає, що у випадку, коли автомат $\Im_A$ є оборотним за Гаффманом, значення випадкової величини $\eta_t(\bar{y})$ обмежені зверху певною константою, яка не залежить від $t\in\mathbf{N}$. Навпаки, для необоротного за Гаффманом автомата $\Im_A$ (за певних загальних умов) значення $\eta_t(\bar{y})$ експоненційно зростає при $t\to\infty$ для більшості вихідних послідовностей $\bar{y}=(y(1),\dots,y(t))$.

Як приклад, розглянемо комбінувальний генератор $\Im$, який складається з n лінійних регістрів зсуву та функції ускладнення $f:V_n\to\{0,1\}$. Нехай (A,p_A) – блок управління рухом цього генератора, де $A\subseteq\mathbf{N}_0^n$, $\Im_A$ – скінченний автомат, який йому відповідає.

Твердження 1. Нехай існують вектори $b=(b_1,\dots,b_n)$, $c=(c_1,\dots,c_n)\in A$, $b\ne c$, такі, що всі координати вектора $b+c$ не перевищують довжину найкоротшого регістра зсуву комбінувального генератора $\Im$. Тоді автомат $\Im_A$ є необоротним за Гаффманом.

Доведення. Позначимо $x_j=(x_j(0),\dots,x_j(L_j-1))$ початковий стан j-го регістра генератора $\Im$, $j\in\overline{1,n}$. Розглянемо вхідні послідовності

$$\varepsilon_1=a,\ \varepsilon_2=b,\ \varepsilon_3=c, \tag{9}$$

$$\varepsilon_1'=a,\ \varepsilon_2'=c,\ \varepsilon_3'=b, \tag{10}$$

де $a=(a_1,...,\ a_n)$ – довільний елемент множини *A*. Знаки вихідних послідовностей автомата $\Im_A$ у тактах 0, 1, 2, 3, що відповідають послідовностям (9) та (10), дорівнюють відповідно

$$f(x_1(0),...,x_n(0)),\ f(x_1(a_1),...,x_n(a_n)),\ f(x_1(a_1+b_1),...,x_n(a_n+b_n)),$$
$$f(x_1(a_1+b_1+c_1),...,x_n(a_n+b_n+c_n)) \tag{11}$$

та

$$f(x_1(0),...,x_n(0)),\ f(x_1(a_1),...,x_n(a_n)),\ f(x_1(a_1+c_1),...,x_n(a_n+c_n)),$$
$$f(x_1(a_1+b_1+c_1),...,x_n(a_n+b_n+c_n)). \tag{12}$$

Нехай $(\alpha_1,...,\ \alpha_n)\in V_n$ – довільний вектор такий, що $f(\alpha_1,...,\ \alpha_n)=0$. Тоді за умови твердження можна вибрати початковий стан автомата $\Im_A$ таким чином, щоб виконувалися рівності

$$x_1(a_1+b_1)=x_1(a_1+c_1)=\alpha_1,\ \ldots,\ x_n(a_n+b_n)=x_n(a_n+c_n)=\alpha_n. \tag{13}$$

Безпосередньо з формул (11) – (13) випливає, що вхідні послідовності (9), (10) та вибраний початковий стан автомата $\Im_A$ задовольняють умови (*а*) – (*в*). Отже, автомат $\Im_A$ є необоротним за Гаффманом, що й треба було довести.

Кажучи неформально, отримане твердження показує, що з ростом довжини вихідної послідовності комбінувального генератора гами з нерівномірним рухом число її прообразів, тобто відповідних розв'язків системи рівнянь (8), майже завжди зростає експоненційно швидко.

Більш точні результати про асимптотичну поведінку числа прообразів вихідних послідовностей скінченних автоматів викладено в [14]. Як правило, спроби застосування зазначених результатів до конкретних генераторів з нерівномірним рухом, які використовуються на практиці, пов'язані зі значними аналітичними труднощами. Відмітимо, що навіть для "простішого" випадку ЛРЗ з *A*-рухом, де $|A|=2$, задача отримання аналітичного виразу закону розподілу ВВ $\eta_t(\bar{y})$ залишається дуже складною та є далекою від повного вирішення [9, 11].

Певну (але неповну) інформацію про характер росту типових значень $\eta_t(\bar{y})$ як функції параметру *t* надає залежність від *t* середнього числа розв'язків СР (8), тобто математичного сподівання $\mathbf{E}\eta_t(\bar{y})$ [14].

Наступне твердження показує, що для широкого класу генераторів гами з нерівномірним рухом асимптотична поведінка параметра $\mathbf{E}\eta_t(\bar{y})$ при $t\to\infty$ визначається властивостями послідовності степенів деякої $(0,1)$-матриці, яка відповідає автомату $\Im_A$.

Твердження 2. Нехай автомат $\Im_A = (A,S,Y,h_A,f_A)$ задовольняє таку властивість: для довільних $x,\ x'\in S$ існує не більш одного $\varepsilon\in A$ такого, що $h_A(x,\varepsilon)=x'$. Задамо орієнтований граф $\Delta(\Im_A)$ з множиною вершин $S\times S$, в якому для кожної пари вершин (x_1,x_1'), (x_2,x_2') дуга, що спрямована з (x_1,x_1') до (x_2,x_2'), існує тоді й тільки тоді, коли для деяких $\varepsilon,\varepsilon'\in A$ виконуються умови

$$x_2=h^{\varepsilon}(x_1),\ x_2'=h^{\varepsilon'}(x_1'),\ f(x_2)=f(x_2'). \tag{14}$$

Позначимо $D(\Im_A)$ матрицю суміжності орграфу $\Delta(\Im_A)$. Тоді для кожного $t\in\mathbf{N}$ має місце рівність

$$\mathbf{E}\eta_t(\bar{y})=\frac{1}{|A|^t|S|}\omega(D(\Im_A)^t), \tag{15}$$

де $\omega(\cdot)$ позначає вагу (суму всіх елементів) відповідної матриці.

Доведення. Для будь-яких $t \in \mathbf{N}$, $x,\ x' \in S$ та $\bar{y} \in Y^t$ позначимо $C_{\bar{y}}(x, x')$ число розв'язків $(x(0), \varepsilon(0), \ldots, \varepsilon(t-1))$ системи рівнянь (8) таких, що $x(0) = x$, $h^{\varepsilon(0)+\ldots+\varepsilon(t-1)}(x(0)) = x'$.

Розглянемо матриці $C_{\bar{y}}$ та $D_{\bar{y}}$ порядків $|S|$ та $|S|^2$ з елементами

$$C_{\bar{y}}(x, x'),\ x,\ x' \in S$$

та

$$D_{\bar{y}}((x_1, x_1'), (x_2, x_2')) = C_{\bar{y}}(x_1, x_2) C_{\bar{y}}(x_1', x_2'),\ (x_1, x_1'),\ (x_2, x_2') \in S^2 \tag{16}$$

відповідно. За умови твердження при $t = 1$ виконуються такі умови:

$C_y(x, x') = 1$, якщо існує (єдиний) $\varepsilon \in A$ такий, що $h_A(x, \varepsilon) = x'$, $f(x') = y$;

$C_y(x, x') = 0$ – у протилежному випадку,

де $x,\ x' \in S$, $y \in Y$. Отже, згідно з формулами (14), (16) маємо

$$D(\Im_A) = \sum_{y \in Y} D_y . \tag{17}$$

Далі, для кожного $\bar{y} = (y(1), \ldots, y(t)) \in Y^t$, де $t \in \mathbf{N}$, виконується рівність

$$C_{\bar{y}} = C_{y(1)} \cdots C_{y(t)}, \tag{18}$$

яка, у свою чергу, тягне рівність

$$D_{\bar{y}} = D_{y(1)} \cdots D_{y(t)} . \tag{19}$$

Таким чином, внаслідок формули (18) маємо

$$\eta_t(\bar{y}) = \sum_{(x,x') \in S^2} C_{\bar{y}}(x, x') = \omega(C_{y(1)} \cdots C_{y(t)}),\ \bar{y} = (y(1), \ldots, y(t)) \in Y^t . \tag{20}$$

Для доведення формули (15), скористуємося рівністю [14]

$$\mathbf{E}\eta_t(\bar{y}) = \frac{1}{|A|^t |S|} \sum_{\bar{y} \in Y^t} \eta_t(\bar{y})^2 . \tag{21}$$

Підставляючи вираз (20) у формулу (21) та послідовно використовуючи рівності (16), (19), (17), отримаємо такі співвідношення:

$$\mathbf{E}\eta_t(\bar{y}) = \frac{1}{|A|^t |S|} \sum_{\bar{y} \in Y^t} (\omega(C_{y(1)} \cdots C_{y(t)}))^2 = \frac{1}{|A|^t |S|} \sum_{\substack{(x_1, x_2) \in S^2,\\ (x_1', x_2') \in S^2}} \sum_{\bar{y} \in Y^t} C_{\bar{y}}(x_1, x_2) C_{\bar{y}}(x_1', x_2') =$$

$$= \frac{1}{|A|^t |S|} \sum_{\substack{(x_1, x_2) \in S^2,\\ (x_1', x_2') \in S^2}} \sum_{\bar{y} \in Y^t} D_{\bar{y}}((x_1, x_1'), (x_2, x_2')) = \frac{1}{|A|^t |S|} \omega\left(\sum_{\bar{y} = (y(1), \ldots, y(t)) \in Y^t} D_{y(1)} \cdots D_{y(t)} \right) =$$

$$= \frac{1}{|A|^t |S|} \omega\left(\left(\sum_{y \in Y} D_y \right)^t \right) = \frac{1}{|A|^t |S|} \omega(D(\Im_A)^t) .$$

Твердження доведено.

Зауважимо, що формули (18), (20) та (21) надають змогу встановити прості аналітичні вирази меж параметра $\mathbf{E}\eta_t(\bar{y})$ для різноманітних генераторів гами з нерівномірним рухом. У певних випадках такі межі містять явну інформацію про експоненційну швидкість зростання середнього числа розв'язків СР (8) від параметра t, тобто довжини вихідної послідовності генератора гами.

Як приклад, наведемо твердження, яке є безпосереднім наслідком попереднього та встановлює аналітичні межі середнього числа розв'язків СР гамоутворення комбінувального генератора гами з нерівномірним рухом.

Твердження 3. Нехай $\Im_A$ – комбінувальний генератор з множиною станів S, яка складається з n ЛРЗ максимального періоду та зрівноваженої комбінувальної функції $f: V_n \to \{0, 1\}$, (A, p_A) – блок управління рухом цього генератора, де $|A| < \infty$, p_A – рівномірний розподіл ймовірностей на множині A.

Тоді для кожного $t \in \mathbf{N}$ виконується нерівності

$$1/2^t \,|S|\,|A|^t \le \mathbf{E}\eta_t(\bar{y}) \le 1/2\,|S|\,|A|^t. \tag{22}$$

Зауважимо, що згідно з оцінками (22) величина $|S|^{-1}\mathbf{E}\eta_t(\bar{y})$ експоненційно швидко прямує до нескінченності при $t \to \infty$ у випадку, коли $|A| > 2$.

3. Розподіли ймовірностей сум незалежних випадкових векторів, що виробляються блоками управління рухом генераторів гами

Важливою окремою задачею ймовірнісного аналізу системи рівнянь (8) є дослідження розподілів ймовірностей випадкових векторів $\delta(i) = \varepsilon(0) + ... + \varepsilon(i-1)$, які визначають сумарні величини зсувів регістрів даного генератора гами протягом i тактів, $i = 1, 2, ...$ (див. приклади 1, 2). До цієї задачі приводить, зокрема, аналіз ефективності кореляційних атак на різноманітні генератори гами з нерівномірним рухом [13, 15, 16].

Зауважимо, що, оскільки випадкові вектори $\varepsilon(i)$, $i = 1, 2, ...$, є незалежними у сукупності та однаково розподілені на множині $A \subseteq \mathbf{N}_0^n$ (яка звичайно є скінченною), то, за умови оборотності коваріаційної матриці K випадкового вектора $\varepsilon(0)$, граничний розподіл послідовності $\{\frac{1}{\sqrt{i}}(\delta(i) - i\mathbf{E}\varepsilon(0)) : i = 1, 2, ...\}$ є нормальним з параметрами $(0, K)$ (див., наприклад, [17, с. 174]). Проте, на практиці аналіз криптографічних властивостей генераторів гами з нерівномірним рухом, як правило, потребує неасимптотичних виразів або оцінок ймовірностей значень випадкових величин $\delta(i)$.

Нижче (за певних обмежень відносно блоку управління рухом заданого генератора гами) наведені такі вирази та оцінки.

Нехай $A = \{a_1, ..., a_m\} \subseteq \mathbf{N}_0^n$, $\{\varepsilon(i) : i = 0, 1, ...\}$ – послідовність незалежних випадкових векторів, що розподілені за законом $\mathbf{P}(\varepsilon(i) = a) = m^{-1}$, $a \in A$. Позначимо $\delta(0) \equiv 0$, $\delta(i) = \varepsilon(0) + ... + \varepsilon(i-1)$, $i = 1, 2, \ldots$, та отримаємо вираз розподілу ймовірностей випадкового вектора $\delta(i)$.

Зрозуміло, що

$$\mathbf{P}(\delta(i) = a) = m^{-i} \sum_{\substack{(\alpha_0, ..., \alpha_{i-1}) \in A^i: \\ \alpha_0 + ... + \alpha_{i-1} = a}} 1, \quad i = 1, 2, \ldots, \; a \in \mathbf{N}_0^n. \tag{23}$$

Розіб'ємо набори, за якими ведеться підсумування у формулі (23), на класи, що попарно не перетинаються, відносячи до того ж самого класу такі набори $(\alpha_0, ..., \alpha_{i-1})$, що мають одинакові вектори частот зустрічаємості $\mu_1, ..., \mu_m$ елементів $a_1, ..., a_m$ відповідно.

Помітимо, що для будь-яких $\mu_1, ..., \mu_m \in \mathbf{N}_0$, де $\mu_1 + ... + \mu_m = i$, існує точно $\frac{i!}{\mu_1! ... \mu_m!}$ наборів $(\alpha_0, ..., \alpha_{i-1}) \in A^i$ таких, що частота зустрічаємості елемента a_j в наборі $(\alpha_0, ..., \alpha_{i-1})$ дорівнює μ_j, $j \in \overline{1, m}$. Звідси внаслідок рівності (23) отримаємо, що

$$\mathbf{P}(\delta(i) = a) = m^{-i} \sum_{\substack{(\mu_1, ..., \mu_m) \in \mathbf{N}_0^m: \\ a_1\mu_1 + ... + a_m\mu_m = a, \\ \mu_1 + ... + \mu_m = i}} \frac{i!}{\mu_1! ... \mu_m!}, \quad i = 1, 2, ..., \ a \in \mathbf{N}_0^n. \tag{24}$$

Розглянемо окремий випадок, в якому елементи $a_1, ..., a_m$ множини A є лінійно незалежними (над полем $\mathbf{R}$) n-вимірними векторами. В цьому випадку для будь-якого $a \in \mathbf{N}_0^n$ існує єдиний набір чисел $\mu_1, ..., \mu_m$ такий, що $a = a_1\mu_1 + ... + a_m\mu_m$. Звідси внаслідок формули (24) випливає такий результат.

Твердження 4. Нехай $A = \{a_1, ..., a_m\} \subseteq \mathbf{N}_0^n$, де вектори $a_1, ..., a_m$ є лінійно незалежними над полем $\mathbf{R}$. Тоді для довільних $a \in \mathbf{N}_0^n$, $i = 1, 2, ...$ таких, що

$$a = a_1\mu_1 + ... + a_m\mu_m, \ i = \mu_1 + ... + \mu_m, \ \mu_j \in \mathbf{N}_0, \ j \in \overline{1, m}, \tag{25}$$

виконується рівність

$$\mathbf{P}(\delta(i) = a) = m^{-i} \frac{i!}{\mu_1! ... \mu_m!}. \tag{26}$$

Якщо ж a та i не можуть бути представлені у вигляді (25), то $\mathbf{P}(\delta(i) = a) = 0$.

Розглянемо зараз випадок, у якому вектори $a_1, ..., a_{m-1}$ є лінійно незалежними над полем $\mathbf{R}$, а вектор a_m дорівнює їхній лінійній комбінації з раціональними коефіцієнтами:

$$a_m = a_1c_1 + ... + a_{m-1}c_{m-1}, \ c_j \in \mathbf{Q}, \ j \in \overline{1, m-1}. \tag{27}$$

Покажемо, що за умови

$$c_1 + ... + c_{m-1} \neq 1 \tag{28}$$

сума (24) має не більше одного ненульового доданка, внаслідок чого виконується рівність (26).

Дійсно, припустимо, що існує два різних набори $(\mu_1, ..., \mu_m)$, $(\nu_1, ..., \nu_m) \in \mathbf{N}_0^m$ таких, що

$$\sum_{j=1}^{m} \mu_j a_j = \sum_{j=1}^{m} \nu_j a_j = a, \ \sum_{j=1}^{m} \mu_j = \sum_{j=1}^{m} \nu_j = i. \tag{29}$$

З першої рівності (29) маємо $\sum_{j=1}^{m-1} (\mu_j - \nu_j) a_j = (\nu_m - \mu_m) a_m$, звідки в силу лінійної незалежності векторів $a_1, ..., a_{m-1}$ випливає, що $\nu_m \neq \mu_m$ та

$$c_j = \frac{\mu_j - \nu_j}{\nu_m - \mu_m}, \ j \in \overline{1, m-1}. \tag{30}$$

Підсумовуючи рівності (30) за всіма $j \in \overline{1, m-1}$, знайдемо, що

$$\sum_{j=1}^{m-1} c_j = \frac{1}{\nu_m - \mu_m}\left(\sum_{j=1}^{m-1}\mu_j - \sum_{j=1}^{m-1}\nu_j\right) = \frac{i-\mu_m - i+\nu_m}{\nu_m-\mu_m} = 1.$$

Але це суперечить умові (28). Отже, сума (24) має не більше одного ненульового доданка, що й треба було довести.

Таким чином, отримано наступний результат.

Твердження 5. Нехай множина $A = \{a_1, ..., a_m\}$ задовольняє умови (27), (28), де вектори $a_1, \ldots, a_{m-1}$ є лінійно незалежними над полем $\mathbf{R}$. Тоді справджується висновок твердження 4.

Приклад 3. Нехай $m = n+1$, $A = \{a_1, ..., a_{n+1}\} \subseteq \mathbf{N}_0^n$,

$$a_i = (\beta, ..., \beta, \alpha, \beta, ..., \beta),\ i \in \overline{1, n},\ a_{n+1} = (\beta, ..., \beta), \tag{31}$$

де $\alpha, \beta \in \mathbf{N}_0, \alpha \neq \beta$. Неважко перевірити, що вектори $a_1, ..., a_{n+1}$ задовольняють умову твердження 5, де $a_{n+1} = \dfrac{\beta}{\alpha + (n-1)\beta}\sum_{j=1}^{n} a_j$, тобто $c_j = \dfrac{\beta}{\alpha+(n-1)\beta}$, $j \in \overline{1, n}$.

Отже, згідно з твердженням 5 для довільних $\mu_1, ..., \mu_m \in \mathbf{N}_0$ та a, i, що задовольняють умову (25), виконується рівність (26).

Зауважимо, що в окремому випадку $n = 3$, $\alpha = 0$, $\beta = 1$ співвідношення (31) описують блок (само)управління рухом генератора гами шифру A5/1.

Наведемо оцінки двійкового логарифму ймовірності (26) за умови (25).

Скористаємося відомими нерівностями (див., наприклад, [18]):

$$m^{-i}\frac{2^{iH(q_1,...,q_m)}}{\sqrt{(2\pi i)^{m-1} q_1 ... q_m}} \exp\left\{-\sum_{j=1}^{m}\frac{1}{12 i q_j}\right\} < \mathbf{P}(\delta(i) = a) < m^{-i}\frac{2^{iH(q_1,...,q_m)}}{\sqrt{(2\pi i)^{m-1} q_1 ... q_m}}, \tag{32}$$

де $q_j = \dfrac{\mu_j}{i}$, $j \in \overline{1, m}$, $H(q_1, ..., q_m) = -\sum_{j=1}^{m} q_j \log q_j$.

Нехай $\mu_j = t \in \mathbf{N}$ для всіх $j \in \overline{1, m}$, $a = t(a_1 + ... + a_m)$, $i = tm$. Тоді, на підставі формул (26), (32), отримаємо, що

$$\log \mathbf{P}(\delta(i) = a) < -\frac{1}{2}((m-1)\log(2\pi i) - m\log m),$$

$$\log \mathbf{P}(\delta(i) = a) > -\frac{1}{2}((m-1)\log(2\pi i) - m\log m) - \frac{1}{12i}\log e.$$

Отже, для будь-яких

$$a = t(a_1 + ... + a_m),\ i = tm,\ \ t = 1, 2, ... \tag{33}$$

виконується рівність

$$-\log \mathbf{P}(\delta(i) = a) = \frac{1}{2}(m-1)(\log t + \log(2\pi)) - \frac{1}{2}\log m + \Delta_{t,m}, \tag{34}$$

де $0 \le \Delta_{t,m} \le \dfrac{\log e}{12tm}$. Як видно з формул (33), (34), значення $\mathbf{P}(\delta(i) = a)$ прямують до нуля зі швидкістю порядку $O(t^{-1})$ при $t \to \infty$.

Висновки

Основними результатами статті є ймовірнісні властивості розв'язків систем рівнянь гамоутворення генераторів гами з нерівномірним рухом (див. вище твердження 1 – 5). Отримано матричне представлення для середнього числа розв'язків зазначених систем рівнянь та встановлено умови, за яких комбінувальний генератор із зовнішнім управлінням є необоротним за Гаффманом. Отримано достатні умови експоненційного росту середнього числа розв'язків системи гамоутворення цього генератора від довжини його вихідної послідовності, аналітичні вирази та оцінки розподілів ймовірностей сум випадкових векторів, які виробляються блоками управління рухом певного класу комбінувальних генераторів гами.

Результати можуть бути застосовані при розв'язанні задач оцінювання стійкості генераторів гами із зовнішнім управлінням рухом та обґрунтування вимог до криптографічних параметрів вузлів ускладнення таких генераторів гами, що визначають їхню стійкість відносно кореляційних атак.

Список літератури:

1. Katz J., Lindell Y. Introduction to Modern Cryptography. Taylor & Francis Group, 2021. 628 p.
2. Олексійчук А. М., Курінний О. В. Методи криптоаналізу потокових шифрів : навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2023. 172 с.
3. Anderson R., Roe M. A5. [Електронний ресурс] : http://jya.com/crack-a5.htm.
4. Komninos N., Honary B., Darnell M. An efficient stream cipher Alpha1 for mobile and wireless devices // Proceedings of the 8-th IMA International Conference on Cryptography and Coding. 2001. P. 294–300.
5. Simpson L.R. LILI Keystream Generator / L.R. Simpson, E. Dawson, J.D. Golić, W.L. Millan // Selected Are-as in Cryptography. SAC 2000. Lecture Notes in Computer Science, vol. 2012. Springer, Berlin, Heidelberg. P. 248–261.
6. Sadkhan S.B. A proposed Development of Clock Control Stream Cipher based on Suitable Attack // 2020 1st. Information Technology To Enhance e-learning and Other Application. IT-ELA, 2020, P. 165–170. doi: 10.1109/IT-ELA50150.2020.9253074.
7. Xu Y., Hao Y., Wang M. Revisit two memoryless state-recovery cryptanalysis methods on A5/1 // http://eprint.iacr.org /2023/1557.
8. Meneses A., van Oorschot P., Vanderstone S. Handbook of applied cryptography. CRC Press, 1997.
9. Kholosha A.A. Clock-controlled shift registers for key-stream generation // http://eprint.iacr.org /2001/061.
10. Gollman D., Chambers W.G. Clock-controlled shift registers: a review // IEEE J. on Selected Areas in Communication. 1989. V. 7. № 4. P. 525–533.
11. Golić J., O'Connor L. Embedding and probabilistic correlation attacks on clock-controlled shift registers // Advances in Cryptology – EUROCRYPT'94, Proceedings. Springer Verlag. 1995. P. 230 243.
12. Golić J., Petrovic M.V. A generalized correlation attacks with a probabilistic constrained edit distance // Advances in Cryptology – EUROCRYPT'92, Proceedings. Springer Verlag. 1992. P. 472–476.
13. Johansson T. Redused complexity correlation attacks on two clock-controlled generators // ASIACRYPT'98, Proceedings. Springer Verlag. 1998. P. 342–356.
14. Mikhailov G.V., Chistyakov V.P. On the problems of finite automatons theory related to the number of preimages of the output sequences // Review of Applied and Industrial Mathematics. 1994. Vol. 1, Iss. 1. P. 108–117.
15. Ekdahl P., Johansson T. Another attack on A5/1 // IEEE Trans. on Inform. Theory. 2003. Vol. 49. P. 1–7.
16. Ekdahl P. On LFSR-based stream cipher: analysis and design. Ph. D. Th., 2003.
17. Коваленко І.М., Гнеденко Б.В. Теорія ймовірностей. Київ : Вища шк., 1990. 328 с.
18. Feller W. An introduction to probability theory and its application. Willey. N.-Y., 1950. 420 p.

Надійшла до редколегії 17.09.2024

Відомості про авторів:

Олексійчук Антон Миколайович – доктор технічних наук, доцент, Інститут спеціального зв'язку та захисту інформації Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського", професор спеціальної кафедри № 1; Україна; e-mail: alex-dtn@ukr.net

Самойлов Ігор Володимирович – кандидат технічних наук, доцент, Інститут спеціального зв'язку та захисту інформації Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського", доцент спеціальної кафедри № 1; Україна; e-mail: samoilov1966igor@gmail.com

УДК 621.391:519.2 **DOI:10.30837/rt.2024.3.218.08**

Є.В. КОТУХ, канд. техн. наук, Г.З. ХАЛІМОВ, д-р техн. наук, І.Є. ДЖУРА

ПРОБЛЕМА ЗНАХОДЖЕННЯ ПЕРІОДИЧНОСТІ В КВАНТОВОМУ КРИПТОАНАЛІЗІ АЛГОРИТМІВ ГРУПОВОЇ КРИПТОГРАФІЇ

Вступ

Алгоритми пошуку квантового періоду є центральним компонентом квантових обчислень, особливо в задачах, де періодичність відіграє ключову роль, наприклад алгоритм Шора (для цілочисельної факторизації) та інші програми, що включають періодичні функції над групами. Ці алгоритми використовують принципи квантової суперпозиції та інтерференції, щоб знайти період даної функції експоненціально швидше, ніж класичні алгоритми. Нижче наведено огляд алгоритму визначення квантового періоду, його теоретичні основи та порівняння з існуючими квантовими та класичними методами. Щоб забезпечити повний аналіз алгоритмів пошуку періоду для різних типів груп, включаючи групи Сузукі, Ерміта та Рі, потрібно спочатку зрозуміти загальну структуру та властивості цих груп у зв'язку з квантовими алгоритмами, зокрема з проблемою знаходження періодичності.

Аналіз літератури

Квантові алгоритми визначення періоду відіграють основоположну роль у квантових обчисленнях і криптоаналізі, зокрема завдяки їх застосуванню в алгоритмі Шора, який дає змогу ефективно розкладати великі цілі числа на множники та обчислювати дискретні логарифми. Ці завдання мають вирішальне значення для безпеки багатьох широко використовуваних криптографічних систем, таких як шифрування RSA. Новаторська робота Пітера Шора в 1994 р. продемонструвала, як квантовий комп'ютер може вирішувати ці проблеми експоненціально швидше, ніж класичні методи, створюючи серйозну проблему для класичної криптографії [1]. Алгоритм Шора використовує квантове перетворення Фур'є (QFT) для ідентифікації періоду даної функції, підхід, який став основою для багатьох квантових алгоритмів, що вирішують криптографічні проблеми. Алгоритм пошуку періоду є ключовим для ефективного вирішення таких проблем, як цілочисельна факторізація, яка є центральною для зламу RSA. Це викликало інтерес до квантово-стійких криптографічних систем [2, 3].

Класичні алгоритми для пошуку періоду, такі як алгоритм Ро Полларда, мають значні обмеження, коли йдеться про обробку великих вхідних даних. Алгоритм Полларда Ро, хоч і ефективний для певних циклічних групових структур, працює з експоненціальною складністю часу, що робить його непрактичним для більших екземплярів [4]. Класичні методи грубої сили також неможливі для великих періодів, оскільки вони вимагають перевірки кожного можливого введення, доки не буде знайдено повторення, зі складністю O(T), де T - період.

З іншого боку, квантові алгоритми, такі як Шор, працюють у поліноміальному часі та пропонують експоненціальне прискорення порівняно з класичними методами. Алгоритм Шора, зокрема, має часову складність O((log N)^3), де N є цілим числом, яке розкладається на множники, порівняно з класичною експоненційною складністю O(exp(log N)^c) [1]. Це різке прискорення робить визначення квантового періоду ключовим інструментом у квантовому криптоаналізі, де його можна застосовувати для ефективного зламу класичних криптографічних систем [5].

Квантове перетворення Фур'є (QFT) є основним для квантових алгоритмів пошуку періоду, включаючи алгоритм Шора. QFT ефективно обчислює частотні компоненти періодичної функції, дозволяючи визначити період у логарифмічному часі відносно розміру вхідних даних. Алгоритм Шора починається з ініціалізації суперпозиції станів, застосовує квантовий оракул для обчислення періодичної функції, а потім використовує QFT для

виділення періоду Nielsen2002. Однак застосування КТП до неабелевих груп значно складніше. У випадках, пов'язаних із неабелевими структурами, такими як групи Сузукі та Рі, квантове перетворення Фур'є є багатовимірним і менш простим, що робить пошук періоду в цих групах відкритою дослідницькою проблемою [6]. Ці групи мають вирішальне значення для вивчення проблем прихованих підгруп (HSP) у квантовій криптографії, де поточні квантові алгоритми не можуть ефективно виділити періодичність [7].

Хоча квантові алгоритми показали значний успіх з абелевими групами, неабелевий випадок залишається набагато складнішим. Неабелеві групи, такі як групи Сузукі, Ерміта та Рі, є більш складними через їх багатовимірні представлення та некомутативний характер їхніх елементів. Ці властивості роблять застосування квантових алгоритмів, особливо алгоритмів пошуку періоду, експоненціально складнішим [3].

Наприклад, група Сузукі є неабелевою простою кінцевою групою типу Лі зі скрученою структурою Шевалле. Квантові алгоритми для знаходження періоду борються з цими групами, тому що їхня теорія представлення є набагато більш залученою, і для таких груп не існує ефективної КТП [6]. Подібні проблеми спостерігаються з ермітовими (унітарними) і групами Рі, де періодичність прив'язана до власних значень матриці або скручених автоморфізмів, а поточні квантові методи не пропонують ефективних рішень [8].

Квантовий пошук періоду лежить в основі багатьох криптографічних атак, особливо тих, які загрожують безпеці RSA та криптографії на основі еліптичної кривої. Алгоритм Шора, який використовує пошук періоду для ефективного розкладання цілих чисел, безпосередньо підриває схему шифрування RSA, оскільки безпека RSA залежить від складності розкладання великих цілих чисел [1]. Окрім RSA, визначення квантового періоду також можна застосовувати до таких проблем, як проблема дискретного логарифмування як у скінченних полях, так і в групах еліптичних кривих. Якби були розроблені ефективні квантові алгоритми для неабелевих груп, це могло б призвести до зламу криптографічних систем, які покладаються на жорсткість цих проблем [9]. Незважаючи на прориви, які забезпечують квантові алгоритми визначення періоду, залишаються значні проблеми у застосуванні цих методів до структур неабелевих груп. Продовжуються дослідження щодо розробки ефективних квантових алгоритмів для проблеми прихованих підгруп (HSP) у неабелевих групах, що дозволить квантовим комп'ютерам вирішувати більш широкий спектр криптографічних задач [7]. Розвиток постквантової криптографії, яка спрямована на розробку криптографічних алгоритмів, стійких до квантових атак, є ще одним важливим напрямком поточних досліджень [10, 11].

Метою цієї статті є визначення квантової проблеми знаходження періоду, вивчення її поточного стану щодо неабелевих груп та аналіз критеріїв складності, пов'язаних з найбільш помітними групами, що використовуються в криптографічних програмах.

Результати досліджень

Групи Сузукі, Ерміта та Рі є конкретними прикладами неабелевих груп, що додає значного рівня складності квантовим алгоритмам. Хоча ефективні алгоритми пошуку періоду існують для абелевих груп і деяких неабелевих випадків (такі, як двогранні групи), проблема знаходження періоду залишається складною в цих більш складних групах типу Лі.

Група Сузукі неабелева, проста, скінченна, типу Лі, скручена група Шевалле. Їх позначають як Sz (q), та вона є частиною більшого класу скручених груп Шевалле, що визначаються для полів характеристики 2, де $q = 2^{2n+1}$. Групи Сузукі демонструють високосиметричні, некомутаційні структури. Вони існують для непарних ступенів числа 2 і класифікуються як прості групи (групи без нетривіальних нормальних підгруп). Група має складну внутрішню структуру, яка включає польові автоморфізми та вимагає передових методів теорії Лі та алгебраїчної геометрії для опису. Порядок групи Сузукі дорівнює $Sz(q):\left|Sz(q)\right| = q^2\left(q-1\right)\left(q^2+1\right)$. Періодичність у групах Сузукі надзвичайно важко проаналізувати. Оскільки групи Сузукі є неабелевими, їх теорія представлення набагато складніша, ніж теорія абелевих груп. Квантові алгоритми, які покладаються на перетворення

Фур'є, такі як алгоритм Шора, погано працюють для груп Сузукі, оскільки їх структура призводить до багатовимірних представлень. Квантове перетворення Фур'є (QFT) не є простим, і для груп Сузукі невідомо ефективних алгоритмів пошуку періоду. Отримання інформації про підгрупи в групах Сузукі залишається обчислювально складним. Групи Сузукі вивчаються в контексті скінченних простих груп і груп типу Лі, і розв'язання проблеми HSP для цих груп дасть суттєве розуміння ширшого класу квантових проблем. Відсутність ефективних алгоритмів відображає загальну проблему вирішення проблем знаходження періоду для неабелевих груп.

Ермітові (унітарні) групи є неабелевими, класичними групами. Ермітові групи, також відомі як унітарні групи, складаються з матриць, які зберігають ермітову форму (внутрішній добуток на складні векторні простори). Унітарна група U(n, q) складається з n помножених на n матриць над полем q, де кожна матриця задовольняє умову $U^{\dagger}U = I$ (зберігає ермітову форму). Ермітові групи є неабелевими, коли n > 1, що робить їх частиною класичної групи груп, яка зберігає певні симетрії під час перетворень. Ці групи відіграють важливу роль у квантовій механіці та квантових обчисленнях, оскільки унітарні перетворення керують квантовою еволюцією. Ермітові групи мають складні структури власних значень, періодичність яких пов'язана з властивостями власних значень. Періодичність у ермітових групах пов'язана з поведінкою власних значень. Наприклад, періодичність унітарних матриць включає обертальну симетрію в комплексних векторних просторах. З квантовим перетворенням Фур'є над унітарними групами стає важче працювати через багатовимірну природу представлення, особливо зі збільшенням розміру матриці n. Квантові алгоритми, які включають унітарні матриці (такі як алгоритми quantum walk або алгоритми HSP), повинні мати справу з періодичністю, яка виникає внаслідок складних обертальних симетрій. Для алгоритмів пошуку періоду завдання полягає в ефективному виявленні повторюваних власних значень або шаблонів у перетвореннях матриці, що потребує інтенсивних обчислень і значної постобробки. Ермітові групи тісно пов'язані з проблемами квантової криптографії та квантової корекції помилок, де унітарні операції є фундаментальними. HSP для унітарних груп ще не є ефективно розв'язаним, що відображає ширшу складність вирішення квантових проблем для неабелевих груп. Для виділення періодичності в таких групах часто потрібні методи з теорії представлень і алгебр Лі.

Групи Рі є неабелевими, простими, скрученими групами Шевалле типу Лі. Їх позначають $G_2(q)$ або $F_4(q)$, це кінцеві прості групи, визначені над полями характеристики 3 замість 2 (як це визначено для групи Сузукі). Як і групи Сузукі, вони належать до класу скручених груп Шевалле і виникають із специфічних автоморфізмів алгебраїчних груп. Порядок груп Рі відповідає структурі основної алгебраїчної групи (наприклад, $G_2(q)$ і $F_4(q)$), звичайно може бути описаний за допомогою автоморфізмів скручених полів. Порядок груп Рі дорівнює $G_2(q) = q^3\left(q^3+1\right)\left(q-1\right)$, де $q = 3^n$. Періодичність у групах Рі важко проаналізувати через їхню високосиметричну структуру та складність спотворених автоморфізмів, які їх визначають. Квантові алгоритми борються з неабелевою природою груп Рі. Немає відомого ефективного алгоритму для вирішення проблем пошуку періодів або прихованих підгруп у цих групах. Багатовимірні представлення, які не піддаються ефективному перетворенню Фур'є або алгоритмам quantum walk, ще більше ускладнюють вилучення періодичної інформації. Як і групи Сузукі, групи Рі є частиною класифікації скінченних простих групп. Розуміння того, як вирішити HSP для цих групп, є вирішальним для вдосконалення методів квантових обчислень, що відображає ширшу складність неабелевих груп у квантових алгоритмах. Вирішення проблем знаходження періоду для груп Рі, ймовірно, вимагатиме прориву в теорії квантового представлення та квантової інформаційної науки.

Визначення проблеми

Проблему знаходження квантового періоду можна описати так:

Дано функцію $f:\mathbf{Z}\to G$, де G – деяка група, яка є періодичною з періодом r (тобто $f(x)=f(x+r)$ для всіх $x\in\mathbf{Z}$), завдання полягає в тому, щоб визначити період r.

Кроки в алгоритмі пошуку квантового періоду

К р о к 1 *Суперпозиція.* Ініціалізувати квантовий регістр у суперпозиції всіх можливих входів $|x\rangle$, де $x\in\mathbf{Z}$:

$$\frac{1}{\sqrt{N}}\sum_{x=0}^{N-1}|x\rangle.$$

К р о к 2. *Оцінка функції.* Застосуйте квантовий оракул для обчислення функції $f(x)$, поєднуючи результат із вхідними даними:

$$\frac{1}{\sqrt{N}}\sum_{x=0}^{N-1}|x\rangle|f(x)\rangle.$$

Мета полягає в тому, щоб виміряти r період $f(x)$.

К р о к 3. *Квантове перетворення Фур'є.* Застосуйте QFT до першого регістру (який містить суперпозицію вхідних даних):

$$\frac{1}{\sqrt{r}}\sum_{k=0}^{r-1}e^{2\pi ikx/r}|k\rangle.$$

QFT виявляє частотні компоненти функції, надаючи інформацію про періодичність.

К р о к 4. *Вимірювання.* Після застосування QFT виміряємо стан системи. З високою ймовірністю результат дасть кратне $1/r$, що дозволить визначити період r.

Квантовий алгоритм пошуку періоду працює за поліноміальний час, пропонуючи експоненціальне прискорення порівняно з класичними алгоритмами, які вимагають експоненціального часу в найгіршому випадку для визначення періоду. Це пов'язано з тим, що QFT можна ефективно обчислити в $O(n^2)$ часі, де n – кількість кубітів, які використовуються для представлення вхідного простору. Прикладом вирішення проблеми є існування квантового комп'ютера IBM на 127 кубітів. Проблема знаходження періоду є ядром алгоритму Шора для розкладання великих цілих чисел на множники. Період відповідає порядку числа за модулем N, і знаходження цього періоду дозволяє ефективно розкласти на множники. Пошук періоду тісно пов'язаний з HSP в абелевих групах, де ідентифікація прихованої підгрупи еквівалентна ідентифікації періоду функції.

Існують деякі класичні підходи. Класичні підходи грубої сили для пошуку періоду вимагають багаторазового оцінювання функції для різних вхідних даних, доки не буде знайдено повторення. Складність $O(r)$, де r – період. Цей спосіб стає неможливим протягом тривалого часу. Хоча алгоритм Ро Полларда пропонує швидший підхід для знаходження періодів у певних структурах циклічної групи (наприклад, для дискретних логарифмів або цілочисельної факторизації), він все ще працює в експоненціальному часі відносно розміру вхідних даних.

Як згадувалося раніше, алгоритм Шора використовує пошук квантового періоду як свою основну підпрограму. Він знаходить період модульної функції піднесення до степеня, що призводить до ефективної цілочисельної розкладки. Складність дорівнює $O\left((\log N)^2\right)$ для розкладання N-розрядного цілого числа.

Алгоритм Саймона знаходить період (або приховану маску XOR) функції, яка є періодичною щодо операції XOR. Він працює за поліноміальний час, але вирішує інший тип про-

блеми періодичності порівняно з алгоритмом Шора. Складність дорівнює $O\left(n^2\right)$, де n – кількість бітів у вхідних даних.

Багато квантових алгоритмів для HSP спираються на принципи визначення періоду. Для абелевих груп складність залишається поліноміальною, але для неабелевих груп складність значно зростає (часто стає експоненціальною), оскільки квантове перетворення Фур'є стає важчим для інтерпретації. Хоча алгоритми Шора та Саймона пропонують ефективний пошук періоду для певних типів періодичностей (модульних та XOR відповідно), їхня складність залишається поліноміальною. Однак для неабелевих груп або інших складних структур квантові алгоритми можуть не мати такої переваги.

Таблиця 1

Порівняльний аналіз

Тип групи	**Періодичність**	**Застосування QFT**	**Інші квантові алгоритми**	**Складність**	**Квантова складність**	**Зауваження**
Абелеві	Однозначно існує	Існує ефективна одновимірна QFT	Шор, Саймон, Знахідка періоду	$O(r)$ для фінансування періоду грубої сили	$O\left((\log N)^2\right)$ для алгоритму Шора	Жодних проблем: проста структура, чітка періодичність, легкий QFT.
Циклічні	Чітко визначена періодичність як порядок групи	Ефективний, працює подібно до абелевого випадку	Шор, Саймон	$O(N)$	$O\left((\log N)^2\right)$ для алгоритму Шора	Жодних проблем: більшість проблем можна ефективно вирішити за допомогою QFT
Двогранні	Періодичність включає симетрію як обертання, так і відображення.	Виклик QFT через неабелеву структуру	Субекспоненціальні алгоритми для HSP	$O(r)$	Субекспоненціальний	Проблема: неабелева природа ускладнює пошук підгрупи.
Симетричні	Періодичність на основі перестановок у структурах циклу	Експоненціально складна КТП	Ефективних алгоритмів невідомо	$O(n!)$	Експоненціальний	Завдання: багатовимірна QFT потрібна і залишається відкритою проблемою.
Неабелеві	Складна періодичність, яку часто важко визначити	Багатовимірна QFT, дуже складна	Ефективних алгоритмів невідомо	Експоненціальний	Експоненціальний	Завдання: некомутативна природа груп
Сузукі	Високосиметрична, неабелева, скручена періодичність	Дуже складно, без ефективного QFT	Ефективних алгоритмів невідомо	Експоненціальний	Експоненціальний	Виклик: належить до спотворених груп типу Брехні, важко аналізувати.

Рі	Високо-симетрична, неабелева, скручена періодичність	Дуже складно, без ефективного QFT	Ефективних алгоритмів невідомо	Експонен-ціальний	Експонен-ціальний	Виклик: належить до спотворених груп типу Брехні, важко аналізувати
Ерміта	Періодичність прив'язана до структури власних значень матриці	Складно через пред-ставлення на основі матриці	Ефективних алгоритмів невідомо	Експонен-ціальний	Експонен-ціальний	Завдання: періодичність на основі власних значень і QFT
Вінцевого добутку	Періодичність походить від добутку циклічних та інших груп	Багато-вимірна QFT, дуже складна	Ефективних алгоритмів невідомо	Експонен-ціальний	Експонен-ціальний	Завдання: складне поєднання циклічної та двогранної структур.
Кінцеві прості	Складна періодичність за рахунок структури простих груп	QFT взагалі нездійсненна для неабелевих	Немає ефективних алгоритмів	Експонен-ціальний	Експонен-ціальний	Виклик: те саме для Suzuki, Ree та інших груп
Кінцеві поля	Періодичність легко визначити завдяки добре структурованому полю	Ефективна QFT для абелевих підполів	Алгоритм Шора для скінченних полів	$O(q)$	$O\left((\log q)^2\right)$	Завдання: тільки абелеві підгрупи мають ефективне рішення фінансування періодичності

Висновок

Виявлення квантового періоду є одним із найважливіших проривів у квантових обчисленнях, що дозволяє ефективно розв’язувати проблеми, які класично нерозв’язні. Хоча ці алгоритми працюють виключно добре для абелевих груп, неабелеві групи створюють значні проблеми. Потрібні подальші дослідження, щоб розкрити весь потенціал квантових алгоритмів для неабелевих структур, але наразі квантові алгоритми, такі як Шора та Саймона, залишаються найпотужнішими інструментами для пошуку періоду в абелевих налаштуваннях. Результати порівняльного аналізу представлено в табл. 1.

Основна проблема у застосуванні квантових алгоритмів для проблеми прихованих підгруп (HSP) у неабелевих групах пов’язана зі складністю ефективного вилучення інформації про підгрупи за допомогою квантових перетворень Фур’є. У випадку абелевих груп алгоритм Шора та пов’язані з ним методи досягають успіху завдяки здатності виконувати ефективну квантову вибірку за Фур’є, яка фіксує достатньо інформації для ідентифікації прихованої підгрупи. Однак у неабелевих групах квантове перетворення Фур’є стає значно складнішим, оскільки уявлення груп більше не є одновимірними. Ця складність призводить до труднощів з ефективним обчисленням або інтерпретацією квантових зразків Фур’є, які поширені у просторах вищої розмірності. Як наслідок, існуючим квантовим алгоритмам важко визначити приховані підгрупи в неабелевих групах, особливо коли підгрупа не є нормальною або легко виділяється. Більше того, неабелев HSP включає знамениті складні проблеми, такі як ізоморфізм графів, де проблема прихованих підгруп для симетричних груп, як відомо, є важкою. Спроби узагальнити успішні абелеві методи на неабелеві випадки часто призводять до неповних або неоптимальних рішень, що вимагає нових квантових алгоритмічних методів або розуміння теорії представлень. Крім того, неабелеві групи можуть проявляти більш складну та непередбачувану поведінку під час вибірки квантових станів, що ускладнює зусилля з розробки ефективних алгоритмів.

Список літератури

1. 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
2. Nielsen M. A., & Chuang I. L. (2002). Quantum Computation and Quantum Information. Cambridge University Press. https://shorturl.at/09toE
3. Watrous J. Quantum Computational Complexity // Meyers, R. (eds) Encyclopedia of Complexity and Systems Science. Springer, New York, 2009. NY. https://doi.org/10.1007/978-0-387-30440-3_428
4. Pollard J. M. A Monte Carlo method for factorization // BIT Numerical Mathematics. 1975. No 15(3). P. 331–334. https://link.springer.com/article/10.1007/BF01933667
5. Simon D. R. On the Power of Quantum Computation // Proceedings of the 35th Annual Symposium on Foundations of Computer Science, 1994. https://ieeexplore.ieee.org/document/365701
6. Roetteler M., & Beth T. (1998). Polynomial-time solution to the hidden subgroup problem for a class of non-abelian groups. arXiv preprint quant-ph/9812070. https://arxiv.org/abs/quant-ph/9812070
7. Kuperberg G. A subexponential-time quantum algorithm for the dihedral hidden subgroup problem // SIAM Journal on Computing, 2005. No 35(1). P. 170–188. https://epubs.siam.org/doi/10.1137/S0097539703436345
8. Hallgren S., Russell A., & Ta-Shma A. The hidden subgroup problem and quantum computation using group representations // SIAM Journal on Computing, 2003. No 32(4). P. 916–934. https://epubs.siam.org/doi/abs/10.1137/S0097539701391800
9. Bernstein D. J., & Lange T. Post-quantum cryptography // Nature. 2017. No 549(7671). P. 188–194. https://www.nature.com/articles/nature23461
10. Regev O. On lattices, learning with errors, random linear codes, and cryptography // Journal of the ACM. 2005. No 56(6). P. 1–40. https://dl.acm.org/doi/10.1145/1568318.1568324
11. Peikert C. A decade of lattice cryptography // Foundations and Trends in Theoretical Computer Science. 2016. No 10(4). P. 283–424. https://shorturl.at/0CFgn
12. Kotukh Y., Khalimov G. Hard Problems for Non-abelian Group Cryptography // Fifth International Scientific and Technical Conference "Computer and Information systems and technologies". https://doi.org/10.30837/csitic52021232176
13. 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
14. 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
15. Kotukh Y. Quantum cryptoanalysis of prospective asymmetric cryptosystems // Proceedings of conference “Cybersecurity in energy sector”. https://shorturl.at/1pbcK

Надійшла до редколегії 23.09.2024

Відомості про авторів:

Котух Євген Володимирович – канд. техн. наук, доцент, професор кафедри кібербезпеки; Національний технічний університет «Дніпровська політехніка»; Дніпро, Україна; e-mail: yevgenkotukh@gmail.com; ORCID: https://orcid.org/0000-0003-4997-620X

Халімов Геннадій Зайдулович – д-р техн. наук, професор, завідувач кафедри безпеки інформаційних технологій; Харківський національний університет радіоелектроніки; Харків, Україна; e-mail: hennadii.khalimov@nure.ua; ORCID: https://orcid.org/0000-0002-2054-9186

Джура Ілля Євгенович – студент 4-го курсу, Національний Авіаційний Університет; Київ, Україна; e-mail: illya773823@gmail.com; ORCID: https://orcid.org/0009-0002-5470-4479

MEANS OF TELECOMMUNICATIONS
ЗАСОБИ ТЕЛЕКОМУНІКАЦІЙ

УДК 681.3.06:519.248.681 **DOI:10.30837/rt.2024.3.218.09**

І.Є. АНТІПОВ, д.-р техн. наук, О.М. НІКІТІН

ВИЯВЛЕННЯ ШИРОКОСМУГОВИХ СИГНАЛІВ ЗА ОСОБЛИВОСТЯМИ ЇХ СПЕКТРА

Стаття присвячена покращенню енергетичного виявлення широкосмугових сигналів. Запропоновано новий спосіб виявлення, заснований на симетрії спектра ШСС.

Актуальність та постановка задачі

Широкосмугові сигнали знаходять широке застосування в умовах завад у різних системах зв'язку, включаючи Code Division Multiple Access, радіонавігацію та радіолокацію. Вони також використовуються для прихованої передачі інформації та керуючих сигналів, що робить їх важливими у сфері оборони, безпеки та захисту інформації. Тому виявлення цих сигналів є важливим завданням, важливим є не тільки сам факт виявлення, але й його швидкість.

Для виявлення слабких сигналів і сигналів з розподіленим спектром застосовується енергетичне виявлення. Принцип роботи енергетичного виявлення полягає в накопиченні енергії в окремих частотних смугах в інтеграторі. Цей метод вважається класичним, з ним порівнюють за ефективністю та швидкодією всі нові запропоновані методи. Безперечними перевагами енергетичного виявлення є простота реалізації та універсальність. Але цей метод не дозволяє визначити параметри сигналу, крім того, чим слабкіший сигнал, тим більше часу потрібно для його виявлення. Незважаючи на можливі обмеження по точності, енергетичний метод забезпечує баланс між продуктивністю та достатньою ефективністю в більшості сценаріїв моделювання.

Як альтернатива енергетичному виявленню у роботах [2 – 3] пропонуються інші методи виявлення ШСС. Коротко розглянемо їх.

Так, у роботі [2] розглянуто алгоритм виявлення шумоподібних сигналів на основі інваріантного і нелінійного перетворення порядкових статистик спектральних відліків адитивної суміші сигналу і шуму. В процесі аналізу відбувається дискретизація вхідної суміші, обчислення базового перетворення Фур'є, а потім послідовність нелінійних перетворень, що дозволяють виділити «викиди» у спектрі. Як стверджують автори [2], «введення нелінійного перетворення над спектральними відліками дозволило забезпечити компактність і сепарабельність сигнальних відліків відносно шумових (тобто максимізувати відстань між ними). Мірою розрізнення є різниця середніх значень сигнальних і шумових порядкових статистик». Такий метод дозволяє скоротити час, витрачений на виявлення джерела радіовипромінювання у порівнянні з енергетичним виявленням.

У роботі [3] запропоновано метод цифрового спектрального аналізу ШСС, який заснований на розділенні масиву досліджуваного сигналу на ряд підпослідовностей. Ці підпослідовності піддаються БПФ. Згідно з [3] «аналіз ефективності запропонованого методу підтвердив, що вдається підвищити продуктивність спектрального аналізу ШСС у 2–4,6 рази, при кількості відліків 2^{38}», зі зменшенням кількості відліків збільшується продуктивність, «а саме, дорівнює 2–3,4 рази для кількості відліків 2^{20}».

Кожен з розглянутих методів у певних випадках має переваги перед енергетичним виявленням за швидкодією та ефективністю, але має і обмеження по сферах застосування. Також можна сказати, що виявлення ШСС – важливе наукове завдання.

У роботі пропонується розглянути метод виявлення ШСС, заснований на особливостях їх спектрів, який дозволяє не тільки його виявити, але й точно визначити його центральну частоту.

Суть методу

Розглянемо енергетичний спектр фазоманіпульованого ШСС. Згідно з [1] спектральна густина потужності модулюючої послідовності представляє собою вираз

$$|S_{PM}(\omega)| = \sqrt{\sum_{i=1}^{n}\sum_{k=1}^{n} a_i a_k \cos\left((i-k)\omega t_0\right)}. \quad (1)$$

Спектр одиночного імпульсу, з набору яких складається модулююча послідовність, має вигляд

$$S_1(\omega) = A_0 \frac{\sin\left(0{,}5\omega t_0\right)}{0.5\omega}. \quad (2)$$

При перенесенні сигналу на частоту, ω_0 спектр буде представляти собою досить складну функцію частоти, яку можна записати у вигляді

$$|S_{PM}(\omega)| = A_0 \left|\frac{\sin\left(0{,}5(\omega-\omega_0)t_0\right)}{0.5(\omega-\omega_0)}\right| \sqrt{\sum_{i=1}^{n}\sum_{k=1}^{n} a_i a_k \cos\left[(i-k)(\omega-\omega_0)t_0\right]} \quad (3)$$

Як приклад графік для 11-елементної послідовності показаний на рис. 1.

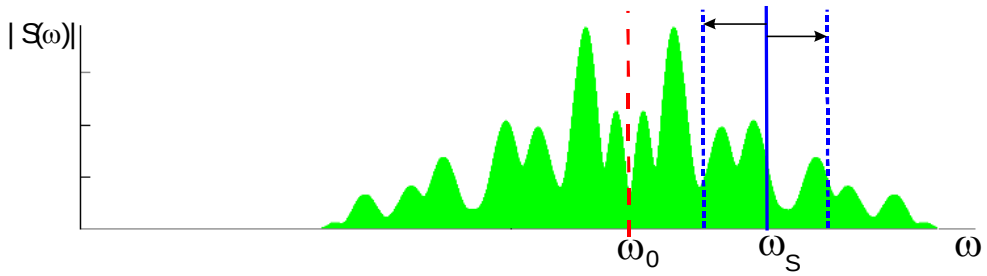


Рис. 1. Представлення енергетичного спектра фазоманіпульованого (ФМ) ШСС

Звернемо увагу на особливості даного спектра. Він являє собою періодичну затухаючу функцію частоти. Чим більше база сигналу, тим більш розподіленим виявляється спектр. Але ми звертаємо увагу на його симетрію відносно середньої частоти. Ця симетрія може бути записана як:

$$S_{PM}(\omega_0 - \Delta\omega) = S_{PM}(\omega_0 + \Delta\omega), \quad (4)$$

де $\Delta\omega = \omega - \omega_0$ – різниця між поточною частотою ω та середньою частотою спектра ω_0. Симетрія відрізняє ФМ ШСС сигнал від шуму та більшості перешкод. Підставивши різницю $\Delta\omega = \omega - \omega_0$ в (4), можна отримати

$$S_{PM}(\omega) = S_{PM}(2\omega_0 - \omega). \quad (5)$$

Але через шум та перешкоди ні центральна частота, ні симетрія, навіть сам ФМ ШСС, можуть бути візуально не помітні на екрані спектроаналізатора. Розглянемо суму добутків лівої та правої частин спектра відносно деякої випадково обраної частоти ω_S:

$$\Pi(\omega_S) = \int_{-\infty}^{\infty} |S(\omega_S + \omega)||S(\omega_S - \omega)| d\omega. \quad (6)$$

Виконавши перетворення, аналогічне (4), (5), отримаємо

$$\Pi(\omega_S) = \int_{-\infty}^{\infty} |S(\omega)||S(2\omega_S - \omega)| d\omega. \quad (7)$$

Тепер врахуємо симетрію. Підставимо (5) в перший добуток підінтегрального виразу (7). Тоді

$$\Pi(\omega_S) = \int_{-\infty}^{\infty} |S(2\omega_0 - \omega)\| S(2\omega_S - \omega)| d\omega . \quad (8)$$

Для обчислення екстремуму виразу (7) потрібно взяти його похідну за ω_S. Навіть у загальному випадку обчислення такого виразу є досить складною задачею, тому ми скористаємося аналогією. Відомо, що вираз для кореляційної функції

$$R(\tau) = \int_{-\infty}^{\infty} f(x) f(x-\tau) dx , \quad (9)$$

досягає максимуму, коли $x = x - \tau$, тобто коли $\tau = 0$. Виходячи з цього, припустимо, що і вираз (7) досягає максимуму коли $\omega_0 = \omega_S$. Тобто, функція (6) досягає максимуму тоді, коли перемноження спектральних відліків відбувається симетрично відносно центральної частоти спектра. Далі можна зробити припущення, що властивість максимуму $\Pi(\omega_S)$ зберігається і при додаванні до шумоподібного ФМ сигналу білого Гаусівського шуму, оскільки він має рівномірну спектральну щільність. Після чого було зроблено друге припущення – дана властивість дозволить виявляти ШСС на фоні шуму, причому не тільки виявляти, але й одразу ж визначати його центральну частоту.

Отже, був запропонований метод виявлення ШСС, що базується на симетрії його спектра, і полягає в тому, що необхідно перебрати всі значення f_S в інтересуючій полосі частот і знайти максимум функції (1). Однак залишається певне обурення, що вплив шуму може "розмити" максимум цієї функції, роблячи його менш помітним.

Алгоритм практичної реалізації

Розглянемо, як має виглядати алгоритм, що реалізує запропонований метод.

1. Виконання БПФ для вхідного сигналу. Формування масиву $S[\omega]$.

2. Вибір початкової та кінцевої частот $\omega_{\min} + d\omega$ та $\omega_{\max} - d\omega$, у межах яких здійснюватиметься пошук, а також крок частоти $d\omega$ (рис. 2).

3. Розділення масиву $S[\omega]$ на праву та ліву частини відносно частоти $\omega_{\min}$ (рис. 3).

4. Знаходження суми добутків лівої та правої частин. Запис отриманого результату в масив $\Pi[\omega]$ (рис. 4).

5. Повторення дій п. 3 та 4 над масивом $S[\omega]$ з кроком $d\omega$ до частоти $\omega_{\max}$.

6. Вибір максимального значення масиву $\Pi[\omega]$ (рис. 5).

7. Визначення відповідної йому частоти ω_0.

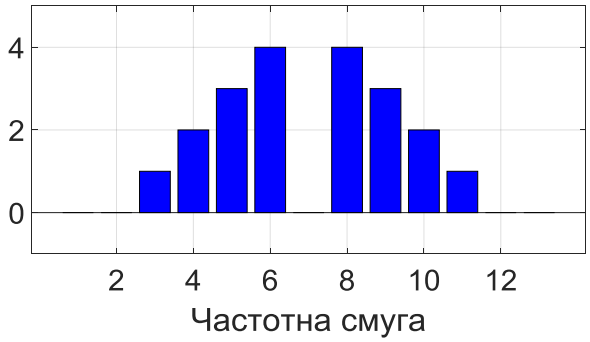


Рис. 2. Представлення масиву $S[\omega]$

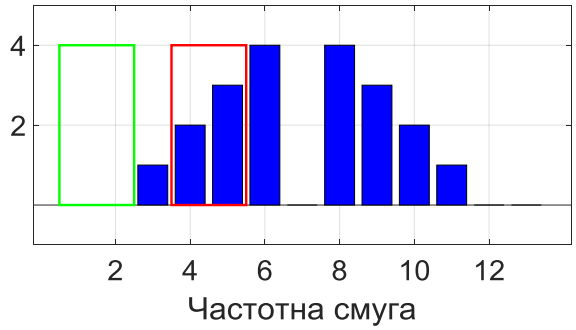


Рис. 3. Сумування масиву $S[\omega]$

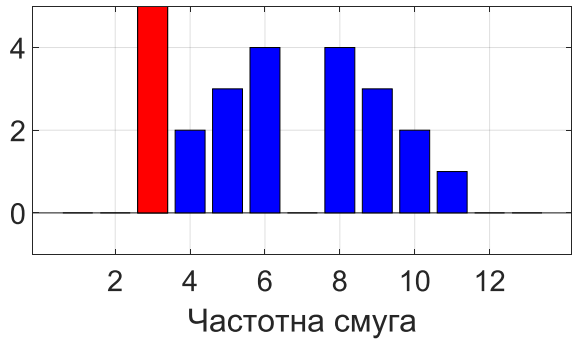


Рис. 4. Занесення нових даних до масиву $\Pi[\omega]$ (червона лінія)

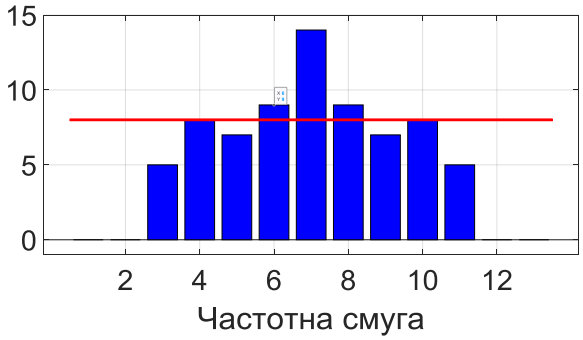


Рис. 5. Заповнення масиву $\Pi[\omega]$, та вибір максимального значення (червона лінія – поріг виявлення)

Моделювання

Для перевірки реалізованості і практичної ефективності запропонованого методу та його алгоритму була розроблена математична модель, що дозволяє формувати ФМ ШСС з різною базою, змішувати його з шумом у заданому співвідношенні та проводити обчислення відповідно до (1). Модель була реалізована у вигляді програми в середовищі MatLab. В моделі було створено дві складові: формування суміші та її подальший аналіз.

1. Формування суміші сигналу та шуму, до складу якої входять: джерело інформації, даний сигнал постійно змінюється після кожного аналізу суміші, для відтворення реальних умов передачі інформації (рис. 6), кодування сигналу та подальша фазова модуляція (рис. 7, де показана лише частина сигналу) і суміш отриманого сигналу з шумом (рис. 8).

2. Аналіз отриманої суміші, яка включала смуговий фільтр (рис. 9), квадратурний детектор (рис. 10), інтегратор (рис.11).

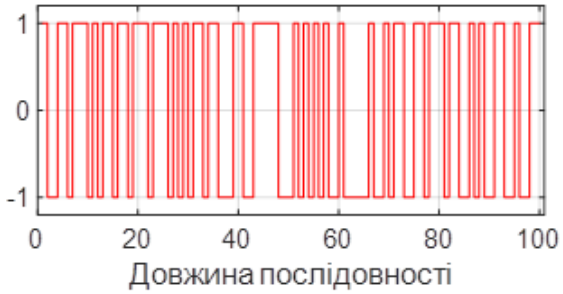


Рис. 6. Закодована послідовність

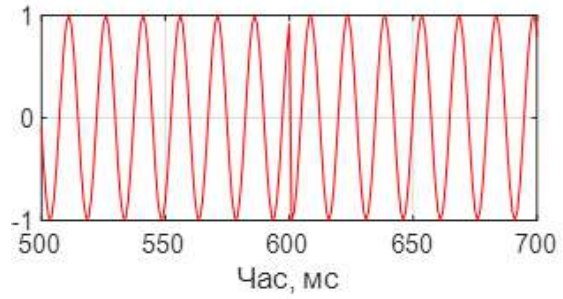


Рис. 7. Фазова модуляція

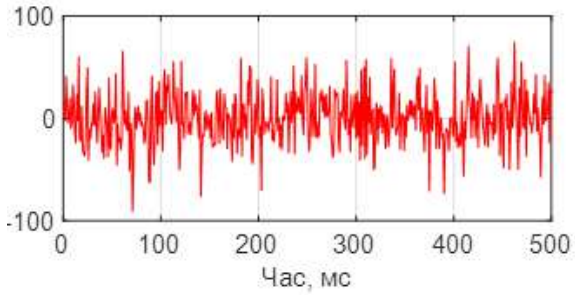


Рис. 8. Суміш корисного сигналу і шуму

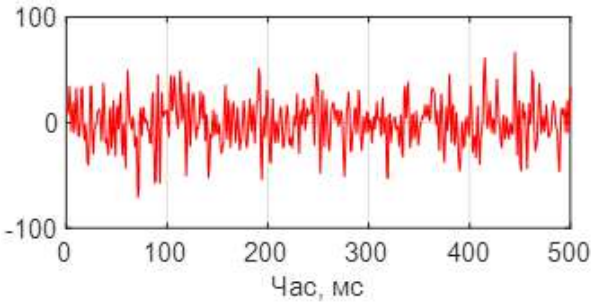


Рис. 9. Реалізація полосового фільтру

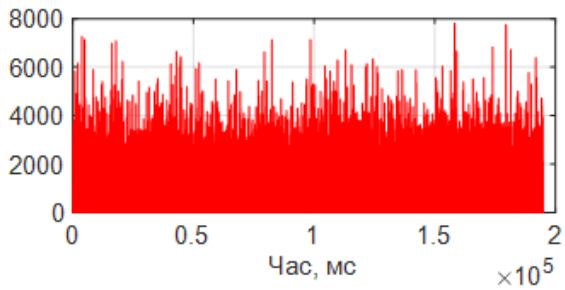


Рис. 10. Реалізація квадратурного детектору

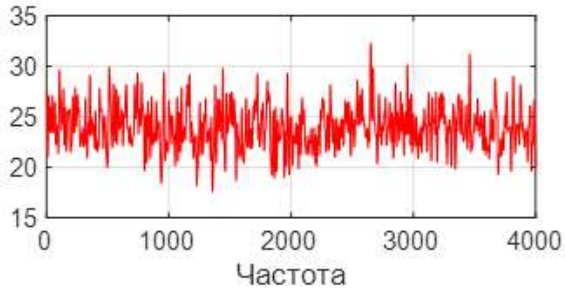


Рис. 11. Реалізація інтегрування за новим методом

Моделювання було виконано для постійної модулюючої послідовності, яка забезпечувала базу 13. При цьому інформаційний сигнал неперервно змінювався, що забезпечувало природню динаміку спектра. Також моделювалося різне співвідношення сигнал/шум. Для оцінки ефективності запропонованого методу моделювання було також виконано для традиційного енергетичного виявлення. Результати моделювання наведено на рис. 12 – 15. Синім кольором показано результати енергетичного накопичення, червоним – запропонований метод.

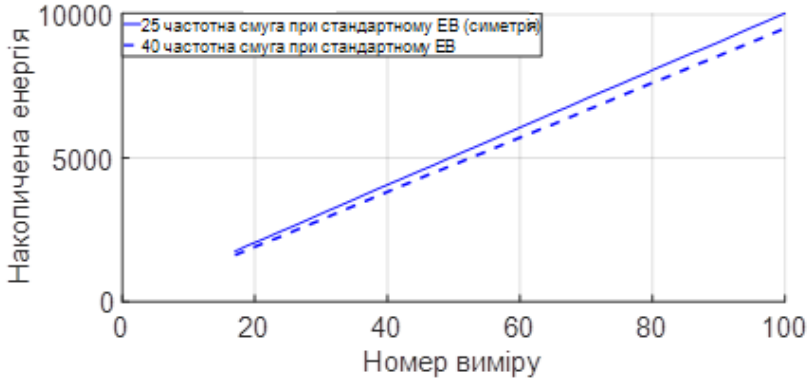


Рис. 12. Результати вимірів за стандартним енергетичному накопиченні при співвідношенням С/Ш 0.4 (суцільна лінія частотна смуга із корисним сигналом, пунктирна із шумом)

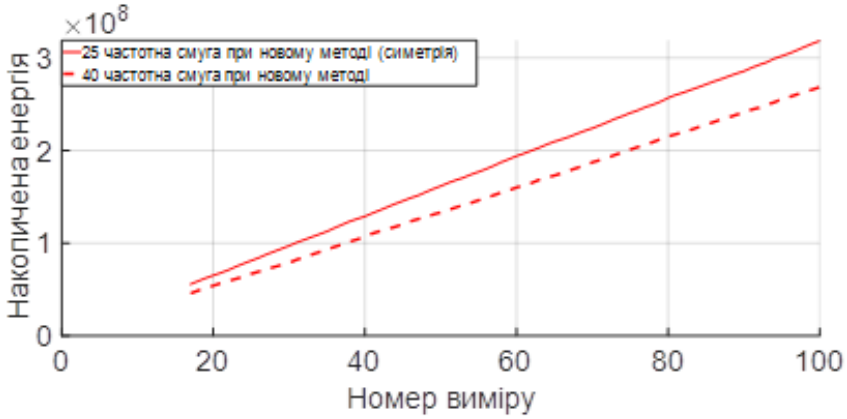


Рис. 13. Результати вимірів при новому методі за співвідношенням С/Ш 0.4 (суцільна лінія частотна смуга із корисним сигналом, пунктирна із шумом)

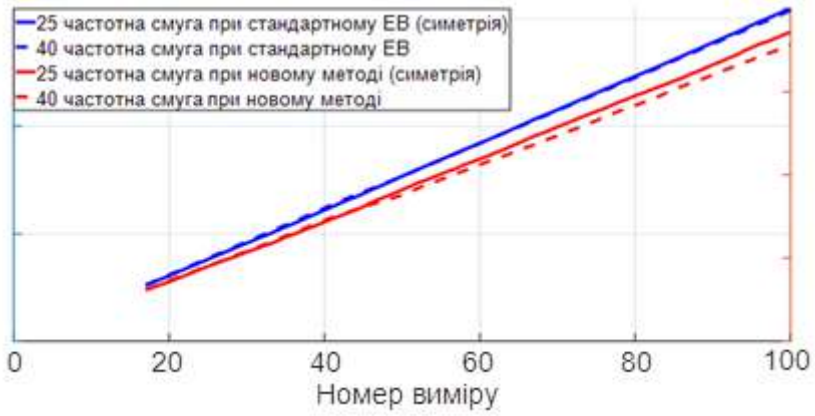


Рис. 14. Порівняння результатів обох методів (синя лінія – стандартне енергетичне накопичення, червона новий метод) при співвідношенням С/Ш 0.2

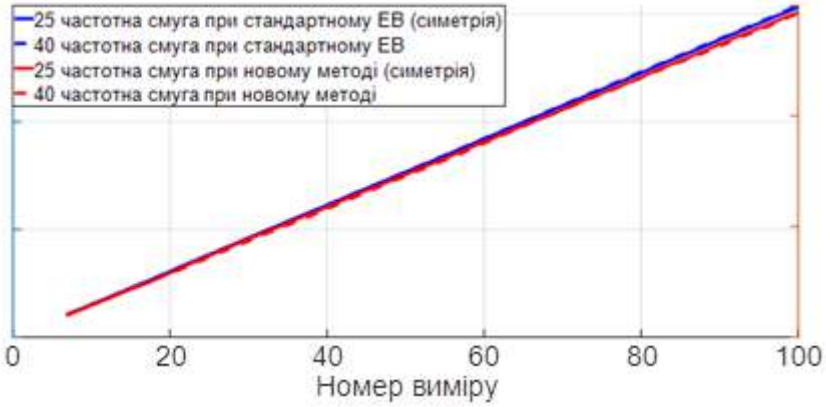


Рис. 15. Порівняння результатів обох методів (синя лінія – стандартне енергетичне накопичення, червона – новий метод) при співвідношенням С/Ш 0.2

Пунктирні лінії відповідають полосам частот, де присутній виключно шум, а суцільна – полосам із ШСС. На рис. 12 представлено результати вимірювання стандартного методу, а на рис. 13 – результати нового методу. З даних рисунків видно, що новий метод починає виділяти корисний сигнал з шумом сильніше порівняно зі стандартним методом. Як видно на рис. 14, метод, заснований на симетрії спектра, раніше починає виділяти ШСС з шумом (47 вимірювань) порівняно з енергетичним виявленням (65 вимірювань). Таким чином, ефективність запропонованого методу на 38,3 % вище, ніж просте енергетичне накопичення. На рис. 15 обидва методи вже не можуть виділити корисний сигнал з шумом після 250 вимірювань (різниці між шумом і корисним сигналом мінімальні), проте варто зауважити, що новий метод більш стійкий до шуму порівняно зі стандартним методом. При стандартному методі корисний сигнал втрачається в шумі після 100 вимірювань, у той час як новий метод все ще намагається виділити корисний сигнал.

Висновки

Отже, у цій роботі:

- розроблено новий метод виявлення ФМ ШСС, заснований на симетрії його спектра, що також дозволяє визначати його центральну частоту;
- запропоновано алгоритм, що реалізує зазначений метод, на основі якого може бути створена комп'ютерна програма;
- розроблено модель прийому та обробки ФМ ШСС в умовах шуму, яка дозволяє порівняти класичний метод енергетичного виявлення з запропонованим методом;

- результати моделювання показали, що запропонований метод швидше виявляє сигнал та може виявляти сигнал при більш тривалих накопиченнях порівняно зі стандартним енергетичним накопиченням;

- під час моделювання було виявлено, що цей метод більш чутливий до перших 100 – 120 вимірювань і може знадобитися до 8 – 15 вимірювань для виявлення сигналу у цей період вимірювань.

Список літератури:

1. Максименко Г. Алгоритм виявлення шумоподібних сигналів, сформованих на основі псевдовипадкових послідовностей // Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні. 2000.С. 107–113.
2. Бортник Г.Г., Васильківський М.В., Коваленко А.В. Цифровий метод спектрального аналізу широкосмугових сигналів // Вісник Хмельницького нац. ун-ту. 2019. Т. 273, №3. С. 92–96.
3. Стеклов В.К., Беркман Л.Н. Теорія електричного зв’язку // Техніка. 2006. С. 552.
4. Amos G., Matlab an introduction with Applications fourth edition // United States, 2010. 430 p.
5. Дробик О.В., Кідалов В.В., Коваль В.В., Костік Б.Я. та інш. Цифрова обробка аудіо- та відеоінформації у мультимедійних системах. Київ : Наук. думка, 2008. 144 с.
6. The MathWorks, Inc. Matlab App Building // United States, release 2023b, 2023. 512 p.
7. The MathWorks, Inc. Matlab Progaming Fundamentals // United States, release 2022b, 2022. 1560 p.
8. Плющ О.Г., Савченко А.С. Дослідження використання широкосмугових сигналів для покращення характеристик адаптивних антених решіток при багатопроменевому розповсюдженні // Наукоємні технології. 2022. Т. 53, №1. С. 31–40.
9. Слободянюк В.В. Метод підвищення роздільної здатності широкосмугових сигналів по частот і доплера на фоні адитивного шуму з невідомою щільністю розподіл у імовірностей // Системи обробки інформації. 2023. Вип. 1, №172. С. 83–91.
10. Слободян М.О., Таранчук А.А., Гавронський В.Є. Генерування широкосмугових хаотичних сигналів для прихованої передачі інформації в телекомунікаційних системах // Вісник Хмельницького нац. ун-ту. 2020. Т. 1, № 4. С. 192–197.
11. Палівода В.С., Хоменко П.В., Березанський Д.О. Перспективи застосування радіостанцій з широкосмуговою адаптивною мережевою формою сигналу в збройних силах україни // Пріоритетні напрямки розвитку телекомунікаційних систем та мереж спеціального призначення, застосування підрозділів, комплексів, засобів зв’язку, автоматизації та кібербезпеки в операції об’єднаних сил. 2020. С. 217–218.
12. Кувшинов О.В., Шишацький А.В., Лютов В.В., Жук О.Г. Аналіз шляхів підвищення скритності широкосмугових систем військового радіозв’язку // Зб. наук. пр. Харк. нац. ун-ту Повітряних Сил. 2017. Т. 50, №1. С. 24–28.
13. Пєвцов Г.В., Яцуценко А.Я., Трофименко Ю.В., Карлов Д.В., Остапова А.М. Енергетичне виявлення радіосигналів в умовах радіоперешкод // Наука і техніка Повітряних Сил Збройних Сил України. 2011. Т. 5, №1. С. 73–77.

Надійшла до редколегії 11.09.2024

Відомості про авторів:

Антіпов Іван Євгенійович – доктор технічних наук, професор, Харківський національний університет радіоедектроніки, завідувач кафедри КРіСТЗІ, Україна; email: ivan.antipov@nure.ua, ORCID: https://orcid.org/0000-0002-9754-4412

Нікітін Олександр Миколайович – Харківський національний університет радіоедектроніки, магістр кафедри КРіСТЗІ, Україна; email: oleksandr.nikitin@nure.ua, ORCID: https://orcid.org/0009-0004-9332-4414

УДК 621.396.96 **DOI:10.30837/rt.2024.3.218.10**

О.В. КАРТАШОВ

МАТЕМАТИЧНА МОДЕЛЬ ЛОКАЦІЙНОГО КАНАЛУ КОНТУРУ АДАПТАЦІЇ СИСТЕМ РАДІОАКУСТИЧНОГО ЗОНДУВАННЯ АТМОСФЕРИ

Вступ

Системи радіоакустичного зондування (РАЗ) атмосфери є ефективним дистанційним засобом отримання інформації про стан процесів, що відбуваються в нижніх шарах атмосфери. Вони дозволяють вимірювати вертикальні профілі температури атмосфери, швидкості вітру, вологості повітря [1–4].

Отримана інформація використовується для забезпечення зльоту та посадки літальних апаратів, прогнозу погоди, прогнозування процесів поширення радіо і акустичних хвиль, вивчення атмосфери [1, 2]. В даний час розвивається теорія радіоакустичного зондування атмосфери, створюються і виробляються станції РАЗ рядом фірм, які пропонуються на ринок [5–7].

Теоретичні і практичні аспекти створення і використання систем РАЗ досліджуються протягом кількох десятиліть, починаючи з 1961 р. [8, 9], проте до цього часу не вдалося подолати низку недоліків, які суттєво обмежують можливості застосування систем РАЗ на практиці [1, 4, 10].

Найбільш важливими серед існуючих обмежень систем РАЗ є вітровий знос плями розсіяного радіосигналу внаслідок переміщення акустичного хвильового пакета під дією вітру та порушення умов Брегга зондувальних сигналів по трасі зондування. Відомі в даний час алгоритми адаптації, які спрямовані на компенсацію впливу порушення умови Брегга на точність та оперативність радіоакустичного зондування, мають недостатню ефективність і не дозволяють суттєво покращити основні характеристики систем РАЗ. Вони не враховують ряд суттєвих особливостей розсіювання радіохвиль на акустичному хвильовому пакеті та створені евристичним шляхом [10–12].

У статті розглядається адекватна математична модель радіоакустичного інформаційного локаційного каналу для використання її в контурі управління частотою зондувального радіосигналу з метою адаптації систем РАЗ до існуючої метеорологічної обстановки.

1. Аналіз публікацій. Відомі алгоритми частотної адаптації систем РАЗ

Вимірювання метеопараметрів в системах РАЗ засновано на отриманні радіосигналу, що розсіюється на імпульсних звукових посиланнях, які випромінюються з поверхні Землі у вертикальному напрямку.

Достатній для наступної обробки рівень розсіяного радіосигналу забезпечується тільки при виконанні умови Брегга [1, 3] між довжинами акустичної та електромагнітної хвиль:

$$\lambda_e = 2\lambda_s \sin\theta,$$

де λ_s – довжина звукової хвилі; λ_e – довжина електромагнітної хвилі; θ – кут розсіювання електромагнітних хвиль.

Довжина хвилі акустичного випромінювання в атмосфері змінюється внаслідок змін температури повітря і швидкості вітру з висотою, що приводить до необхідності підстроювання частот зондувальних сигналів з метою виконання умови Брегга по трасі зондування. В принципі, можливо змінювати адаптивно частоту як акустичного, так і електромагнітного зондувальних сигналів.

В практиці зондування найчастіше використовується метод адаптації до змін метеорологічних умов шляхом підстроювання частоти звукового сигналу зондувального сигналу в ручному режимі. Якщо виконується забезпечення виконання умови Брегга для кожної точки висотного профілю, які іноді називають «майданчиками», то це вимагає значної кількості часу, що може досягати 1-2 годин.

Використовується також методика зондування "точка", що передбачає такий вибір частоти акустичного сигналу, при якій оптимальне співвідношення довжин хвиль акустичного та електромагнітного сигналів забезпечується на деякій середній висоті профілю. При цьому частота звукового сигналу підлаштовується під метеоумови, що спостерігаються на цій висоті, експериментально, по максимуму відбитого радіосигналу [13].

Далі здійснюється випромінювання пакету звукових хвиль та послідовна реєстрація значень доплерівських зсувів частоти через рівні інтервали часу, що відповідають переміщенню звукового пакета на величину його просторової протяжності, у міру поширення пакета звукових хвиль трасою зондування [14]. При цьому похибки вимірювання температури в крайніх точках профілю оказуються досить значними за величиною, порядка одиниць градусів.

Відомі також і алгоритми частотної адаптації систем РАЗ до метеоумов, що змінюються, в яких забезпечення умов Брегга досягається шляхом керування частотою зондувального радіосигналу в міру просування зондувального акустичного хвильового пакета в атмосфері.

У роботі [13] при реалізації алгоритму частотної адаптації систем РАЗ використовується фазове автопідстроювання частоти (ФАПЧ) радіосигналу. В методі здійснюється вимірювання параметра Δ – поточної різниці між значеннями несучої частоти розсіяного радіосигналу $f_{\text{р}}$ і номінальної частоти задаючого генератора радіосигналу у міру поширення звукового імпульсу в напрямку зондування. Далі виконується перетворення значення інформаційного параметру Δ в напругу $U_{\text{упр}}$, яка використовується для управління частотою задаючого генератора радіосигналу.

Інформація про швидкість звуку в атмосфері в даному методі зондування буде полягати в зміні несучої частоти радіосигналу, а значення параметра $f_{\text{р}}$ має залишатися незмінним.

Поточні вимірювання значень частоти генератора електромагнітного випромінювання та зсувів частоти розсіяного сигналу виконувались в системі методом «рахунку нулів», далі отримані значення вводилися в комп'ютер, в якому обчислювалися значення швидкості звуку в атмосфері, а також значення температури середовища по трасі зондування.

У розглянутій схемі зондування за наявності значного початкового розстроювання частот зондувальних сигналів мав місце зрив стеження схеми автопідстроювання за значенням f_s. У разі сильного поривчастого вітру спостерігаються глибокі завмирання амплітуди розсіяного радіосигналу на окремих ділянках траси і стійкість роботи системи РАЗ у такому разі також порушується.

Зазначено [13], що зриви в роботі системи автопідстроювання частоти супроводжуються отриманням значень температури атмосфери, які на кілька градусів перевищують значення температури у сусідніх точках профілю, тобто, формуються аномальні похибки результатів вимірювань.

Таким чином, проаналізовані алгоритми частотної адаптації систем РАЗ не забезпечують необхідних умов до точності вимірювань. Це виникає насамперед внаслідок використання неадекватних уявлень про процес розсіювання радіохвиль на об’єкті, що спостерігається, у вигляді акустичного хвильового пакету, або, іншими словами, внаслідок використання неадекватної математичної моделі каналу систем РАЗ.

Модель інформаційного локаційного каналу в інформаційних радіосистемах значною мірою впливає на функціонування систем [15]. Використання неадекватної моделі каналу призводить до формування додаткових похибок при вимірюванні координат і інших параметрів, для виконання яких існує система, а використання неадекватної моделі каналу в контурі управління об’єктом або в контурі управління параметром при спостереженні за об’єктом, як в системах радіоакустичного зондування атмосфери, не забезпечує ефективного функціону-

вання системи і не дозволяє забезпечити процес відповідного управляння об’єктом або процесом.

2. Постановка задачі дослідження

Значний вплив на ефективність функціонування вимірювальних локаційних радіосистем різного призначення має математична модель інформаційного вимірювального радіоканалу, реалізована в алгоритмі функціонування вимірювальної системи. Математична модель каналу описує особливості механізму поширення хвиль у процесі проходження зондувального сигналу до об'єкта, на якому розсіюються радіохвилі, та назад. Модель визначає також особливості розсіювання радіохвиль на об'єкті, за яким здійснюється спостереження. Таким чином, математична модель каналу описує просторово-часовий сигнал, що надходить на вхід приймальної антени вимірювальної системи. Алгоритм роботи пристрою прийому та обробки сигналу на фоні перешкод синтезується з використанням математичної моделі каналу та особливостей сигналу, що приходить.

Особливості систем радіоакустичного зондування атмосфери полягають перш за все в тому, що в якості радіолокаційної цілі або об'єкта спостереження використовується звукова хвиля, найбільш часто локалізована в просторі, у вигляді акустичного хвильового пакета. Такий об'єкт спостереження має низку істотних особливостей як об'єкт розсіювання радіохвиль. Деякі з цих особливостей виявлено експериментальним шляхом. Однак глибокого вивчення і розуміння такий метод дослідження не може забезпечити.

У зв'язку з цим в статті ставиться задача дослідити канал поширення радіохвиль у радіоакустичних системах зондування атмосфери з використанням теоретичних методів дослідження. Далі отримані результати дослідження будуть використані при побудові алгоритмів та пристроїв обробки сигналів з метою отримання інформації про стан атмосфери, а також у контурі управління параметрами випромінюваного радіосигналу з метою якісного виконання завдання спостереження за акустичним хвильовим пакетом.

3. Математична модель радіоканалу систем РАЗ

Математична модель інформаційного локаційного каналу систем РАЗ [14, 16] дозволяє досліджувати характеристики каналу, визначати вид розсіяного радіосигналу, одержуваного при використанні різних видів зондувальних акустичних та електромагнітних коливань та у різних станах атмосфери.

Як показано в [16], розсіяний на звуковій посилці радіосигнал описується співвідношенням

$$F(r,q) = K\int_0^{\infty} E(2r'-r)S^*(r')e^{jqr'}dr', \qquad (1)$$

де E – комплексна обвідна електричного поля випромінюваної радіохвилі; S – комплексна обвідна випромінюваної акустичної хвилі; $q = 2k_e - k_s$ – параметр розстройки умови Брегга; k_e, k_s – хвильові числа електромагнітної та акустичної хвиль відповідно; r – зміщення взаємодіючих сигналів за поздовжньою просторовою координатою r'; K – амплітудний множник пропорційності. Радіосигнал, що описується у виразі (1) функцією $E(2r'-r)$, стискається вдвічі за просторовою координатою, що описує ефективну взаємодію зондувальних сигналів і формування значення кореляційного інтеграла.

За допомогою теореми Парсеваля отримаємо з (1) співвідношення, де функція $E_1(r)$, що описує розсіяний радіосигнал, визначається через просторові спектри відповідних комплексних обвідних сигналів, які взаємодіють між собою:

$$F(r,q) = \frac{K}{4\pi}\int_{-\infty}^{\infty} S_E\left(\frac{k}{2}\right)S_S^*(k+q)e^{-j\frac{r}{2}k}dk, \qquad (2)$$

тут $\int_{-\infty}^{+\infty} S(r')e^{-jqr'}e^{-jkr'}dr' = S_S(k+q)$, $\int_{-\infty}^{\infty} E(2r'-r)e^{-jkr'}dr' = \frac{1}{2}e^{-j\frac{r}{2}k}S_E\left(\frac{k}{2}\right)$; $k = 2\pi/r'$ – просторова частота.

Далі будемо використовувати більш зручний при виконанні аналізу особливостей розсіяного радіосигналу вираз

$$F(r,q)=\frac{\text{К}}{4\pi}\int_{-\infty}^{\infty}S_E(k)S_S^*(2k+q)e^{-jrk}dk, \tag{3}$$

якій отримано із співвідношення (2) шляхом заміни змінних. У виразі (3) коефіцієнт при аргументі k в спектральної функції електромагнітного сигналу S_E дорівнює одиниці, що є істотним для наступних міркувань.

Опис процесів взаємодії зондувальних сигналів систем РАЗ з використанням частотних уявлень відкриває значні можливості для аналізу властивостей зондувальних коливань, вивчення особливостей розсіяного сигналу, розуміння та правильної інтерпретації накопичених у цій галузі наукових даних, отриманих як теоретичним, так і експериментальним шляхом у ході розвитку даного напряму.

Як випливає з виразу (3), під інтегралом знаходиться добуток енергетичних просторових спектрів процесів $E(r')$ і $S(r'/2)$, що являє собою взаємний енергетичний просторовий спектр цих коливань

$$S_y(k,q)=S_E(k)S_S^*(2k+q). \tag{4}$$

Вирази (2), (3), що визначають розсіяний на звуковій посилці радіосигнал, містять перетворення Фур'є взаємного енергетичного просторового спектру зондувальних коливань.

При $r=0$ в правих частинах виразів (2) і (3) маємо кореляційні інтеграли за хвильовими просторовими частотами спектральних щільностей взаємодіючих сигналів, поєднаних у просторі. При деякому іншому фіксованому значенні параметра зсуву r праві частини (2), (3) також будуть кореляційними інтегралами за частотою k, але вже для сигналів, зміщених на відстань r.

Опис взаємодії зондувальних сигналів систем РАЗ з використанням частотних уявлень – кореляційної функції спектрів сигналів за просторовою частотою та взаємного енергетичного просторового спектру відіграють ключову роль у розумінні особливостей процесів розсіювання та формування відбитої хвилі в системах радіоакустичного зондування атмосфери [2]. Сформований у просторі взаємний енергетичний просторовий спектр взаємодіючих коливань відтворюється на виході спектрального аналізатора системи РАЗ у формі спектру часових частот. Спектроаналізатор призначений для аналізу розсіяного радіосигналу, що приходить на вхід, та для оцінки необхідних інформативних параметрів сигналу – центральної частоти спектру, ширини спектру та ін.

Справді, отриманий в результаті просторово-часової взаємодії хвиль та розсіювання на акустичній хвильовій посилці просторово-часовий електромагнітний сигнал досягає приймальної антени і перетворюється на часове коливання. Спектр хвильових чисел розсіяного радіосигналу перетворюється при прийомі сигналу антеною вимірювального радіоакустичного комплексу в спектр часових частот за правилом $k=2\pi/\lambda \leftrightarrow f=kc/2\pi$. Значенню $k=1$ рад/м ($\lambda=6{,}28$ м), наприклад, відповідає значення частоти $f=47{,}77$ МГц. При цьому форма спектру сигналу не змінюється. Після відповідних перетворень і фільтрації в приймальному пристрої системи сигнал, що несе корисну інформацію про стан атмосфери, піддається переважно перетворенню Фур'є, і далі, по суті, відтворюється взаємний енергетичний просторовий спектр взаємодіючих сигналів у вигляді спектру часових частот [17, 18].

Отже, форма і параметри спектра розсіяного сигналу, сформованого і закладеного в сигнал в процесі розсіювання, містять у собі всю інформацію про стан атмосфери з урахуванням особливостей використовуваних у процесі роботи зондувальних коливань: їх структурних особливостей, параметрів спектрів, значень несучих частот, амплітуд і т.д. Форма і параметри сформованого спектру, що характеризують, наприклад, несиметричність або «скошеність» спектральної функції, кардинально впливають на значення основних інформаційних характеристик систем зондування атмосфери.

Таким чином, розробка як самої системи радіоакустичного зондування, так і алгоритмів, і пристроїв, що реалізують операції вилучення корисної інформації з сигналу, що приходить на вхід, повинні виконуватися з урахуванням особливостей розглянутої взаємодії хвильових процесів в атмосфері, де в розсіяний сигнал закладається корисна інформація.

В системах РАЗ досить поширене використання в якості випромінюваного звукового зондувального сигналу $s(t,r')$ імпульсних акустичних коливань з гармонійним заповненням [19, 20]. Спектр сигналу $S_s(k)$ в цьому випадку є вузькосмуговим у просторі хвильових чисел і отриманий вираз (3) описує основну особливість розсіювання на створюваноій сигналом неоднорідності – насамперед істотну частотну залежність, що спостерігається і яка значно впливає як на енергетичну, так і на інформаційну складові розсіяного сигналу.

Уздовж траси зондування відбувається зміна метеопараметрів, що супроводжується зміною швидкості поширення звуку, яка призводить до деформації (стисненню або розтягуванню) акустичної хвилі вздовж напрямку зондування по координаті r', а отже, к переміщенню спектру сигналу $S_s(k)$ вздовж осі хвильових частот k. Електромагнітна хвиля до впливу змін метеопараметрів уздовж траси зондування практично не схильна і положення спектру радіосигналу $S_e(k)$ на осі частот залишається практично незмінним.

Внаслідок цього максимуми просторових спектрів взаємодіючих сигналів можуть не збігатися (див. рис. 1), а діапазон перекриття спектрів звужуватиметься. Амплітуда розсіяного радіосигналу, що формується, та рівень його спектру при цьому зменшується.

Тут і далі використовуватимемо в якісних міркуваннях для наочності не комплексні обвідні, а сигнали e, s та спектри S_e, S_s, що містять частоти заповнення.

На рис. 1 представлено спектри зондувальних сигналів, що взаємодіють, систем РАЗ. Лінією 1 показаний спектр акустичного зондувального сигналу, лінією 2 – спектр випромінюваного електромагнітного сигналу. Частоти випромінюваних сигналів вибираються таким чином, щоб просторова довжина хвилі звукових коливань була вдвічі меншою за довжину хвилі радіосигналу. Тоді, як видно з рисунку, максимум спектру звукового сигналу буде розміщуватися на частоті, що вдвічі перевищує частоту екстремуму спектру радіосигналу. Протяжність спектру звукової зондувальної посилки зазвичай значно ширше спектру радіосигналу, оскільки радіосигнал має значно більшу просторову протяжність, ніж протяжність звукової посилки внаслідок значної різниці в швидкостях поширення зондувальних коливань, що використовуються, при порівнянній тривалості за часом сигналів. Лінією 3 показаний стислий вдвічі вздовж осі частот k спектр звукової посилки. Стиснення спектру пояснюється присутністю коефіцієнта 2 перед аргументом k спектральної функції $S_S^*(2k+q)$ у виразі (3). Якщо параметр розстройки умови Брегга $q = 2k_e - k_s$ дорівнює нулю, то максимуми спектрів акустичного та радіосигналів збігаються. Якщо значення параметра розстроювання q відмінне від нуля, то максимуми спектрів будуть зміщені по частоті (саме така ситуація – розбіжності максимумів спектрів сигналів, як найбільш загальна, показана на рис. 1).

На рис. 1 форми спектрів 1 і 3 акустичних сигналів – гаусівські, а форма спектру 2 – відповідає прямокутній обвідній зондувального радіоімпульсу.

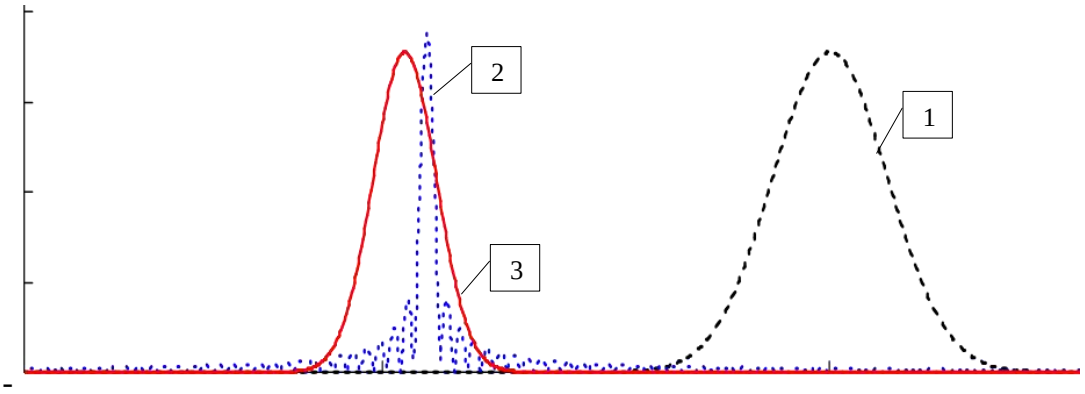


Рис. 1. Ілюстрація взаємодії зондувальних сигналів систем РАЗ у просторі хвильових чисел (1 – спектр акустичного зондувального сигналу, 2 – спектр випромінюваного електромагнітного сигналу, 3 – стиснутий вдвічі вздовж осі частот k спектр звукової посилки)

Механізм взаємодії зондувальних сигналів при розсіюванні хвиль в атмосфері наступний: кожна спектральна складова радіосигналу вибирає з просторового спектру акустичного сигналу складову, що відповідає умові $k \leftrightarrow 2k$. При взаємодії спектр звукового сигналу як би стискається вдвічі під час переходу на частоту k $(k \leftrightarrow 2k)$. Відповідно до цього, якщо в просторі хвильових чисел виконується умова $k_s = 2k_e$ (або в просторі довжин хвиль умова $\lambda_e = 2\lambda_s$), то максимуми спектру радіосигналу і стиснутого в два рази спектру акустичного сигналу збігаються. Однак, зауважимо, що ділянка спектру акустичного сигналу, що вибирається радіосигналом із спектру акустичного коливання в процесі взаємодії, буде вдвічі ширше ($\Delta k_e = 2\Delta k_s$).

Якщо під впливом зміни характеристик середовища спектр звукових коливань $S_s(k)$ зміщується настільки, що спектри сигналів $S_e(k)$ і $S_s(2k)$, що взаємодіють, в просторі хвильових чисел, не будуть перекриватися, то такі зондувальні сигнали в результаті взаємодії не формують розсіяний сигнал. Відповідно до виразу (4) вони стають ортогональними, а акустичний сигнал у такому разі стає повністю прозорим для радіосигналу, що набігає. З іншого боку, в принципі радіосигнал може розсіюватися і на бічних пелюстках спектру акустичного сигналу, або бічні пелюстки спектральної щільності радіосигналу можуть розсіюватися на звуку. Але через досить слабкий рівень розсіяного на звуку радіосигналу в системах РАЗ, практично зареєструвати таке розсіювання не вдається. Тим більше «не видно» результатів розсіяння бічних пелюсток спектру електромагнітної хвилі на бічних пелюстках спектру звукових збурень середовища, хоча таке розсіяння теоретично повинно існувати. У цьому полягає основні енергетичні аспекти взаємодії зондувальних сигналів систем радіоакустичного зондування атмосфери.

Інформаційні аспекти частотних особливостей розглянутого виду розсіяння проявляються в такий спосіб. Якщо максимуми спектрів зондувальних сигналів $|S_e(k)|$ і $|S_s(2k)|$ не збігаються і має місце розстроювання умови Брегга, то максимум просторового спектру розсіяного радіосигналу $\left|S_y(k)\right|$, одержуваного в результаті перемноження спектрів взаємодіючих сигналів, внаслідок несиметричності і нерівномірності спектральної функції звуку $|S_s(2k)|$ в діапазоні перекриття спектрів зміщується додатково на деяку величину Δk.

Це призводить, у свою чергу, до зміщення максимуму часового спектру розсіяного радіосигналу на величину $\Delta\omega = c\Delta k$, яке не є доплерівським зсувом частоти, а підсумовується з ним і призводить до відмінності результуючого зсуву частоти сигналу, отриманого при розсіянні, від суто доплерівського. Аналогічно, можна показати і зміщення центру тяжкості спектральної функції розсіяного сигналу $\left|S_y(k)\right|$ в умовах, що розглядаються. Слід підкреслити, що при переході від просторових частот до часових навіть дуже мале зміщення Δk спектру розсіяного сигналу внаслідок досить великого значення швидкості поширення радіохвиль c трансформується у відчутну різницю $\Delta\omega$ часових частот.

Значення зсуву спектру $\Delta\omega$ при визначенні швидкості звуку по доплерівському зсуву частоти буде проявлятися як систематична похибка у вимірах, наявність якої призводить до помітних похибок при визначенні температури. Формування такої систематичної похибки буде також негативно позначатися на ефективності функціонування алгоритму управління частотою зондувального радіосигналу з метою адаптації систем РАЗ до метеоумов, що змінюються по трасі.

Внаслідок наявності систематичної похибки в результатах вимірювань швидкості звуку в контурі адаптації системи РАЗ до метеообстановки буде знижуватись значення критерія якості системи, якій в детермінованому виді має вигляд,

$$J_M = \{\sum_{i=1}^{M}([\lambda_{ei} - 2\lambda_{si}]^2)\},$$

і в міру зростання миттєвого, термінального значення критерія це приведе к зриву процесу адаптації.

Фінітне значення критерію якості визначає точностні характеристики вимірювання висотного профілю температури атмосфери в заданому висотному діапазоні, а термінальне значення критерію – якості вимірювання температури в конкретної точці профілю.

4. Дослідження інформаційного каналу з використанням тіл розсіювання

Співвідношення (1), (2) визначають результат взаємодії зондувальних коливань систем РАЗ – акустичного і електромагнітного сигналів, а також атмосфери. Вплив середовища при взаємодії сигналів в цих співвідношеннях представлений функціонально лише одним параметром q, якій інтегрує в собі вплив на довжину хвилі акустичного коливання таких характеристик атмосфери як температура середовища, щвидкість вітру, тиск, вологість повітря.

Присутність в виразах (1), (2) тільки одного параметру впливу атмосфери спрощує модель і робить її зручною для проведення досліджень в напрямку теорії систем РАЗ. Функція $F(r,q)$, що описується виразами (1) і (2), характерізує розсіяні радіосигнали у просторовому поданні і має назву функції розсіяння.

Використання просторового подання зондувальних атмосферу сигналів E і S, та, відповідно, двовимірної функції розсіювання, пов'язане з тим, що саме за допомогою такого подання адекватно описується їх взаємодія в середовищі. Наприклад, акустична хвиля з частотою f_s=4 КГц і електромагнітна хвиля, що має частоту f_e=1760 МГц, мають довжини хвиль відповідно λ_e=17 см и λ_s=8,5 см. Вони досить далекі одна від одної в області часових частот, проте близькі в області довжин хвиль та відповідають умові Брегга q=0. Внаслідок цього вони ефективно взаємодіють в атмосфері та формують розсіяний радіосигнал. Отримані за допомогою функції розсіяння просторові характеристики розсіяних сигналів досить просто перетворюються на часові характеристики.

Модуль функції розсіяння $|F(r,q)| = Z(r,q)$ являє собою обвідну, а аргумент – описує фазову структуру розсіяного коливання [2]. Функцію $Z(r,q)$ зручно зображувати у прямокутній системі координат у вигляді 3D поверхні [3], а об'єм, що міститься між поверхнею функції $Z(r,q)$ та площиною (r,q), названий тілом розсіяння (рис. 2). На рис. 2 представлено тіло розсіяння простих за структурою акустичного і електромагнітного сигналів, які мають прямокутні обвідні.

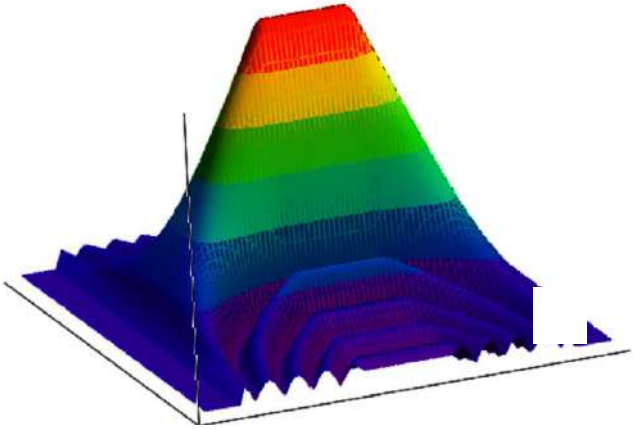


Рис. 2. Тіло розсіяння простих за структурою акустичного та електромагнітного сигналів , які мають прямокутні обвідні

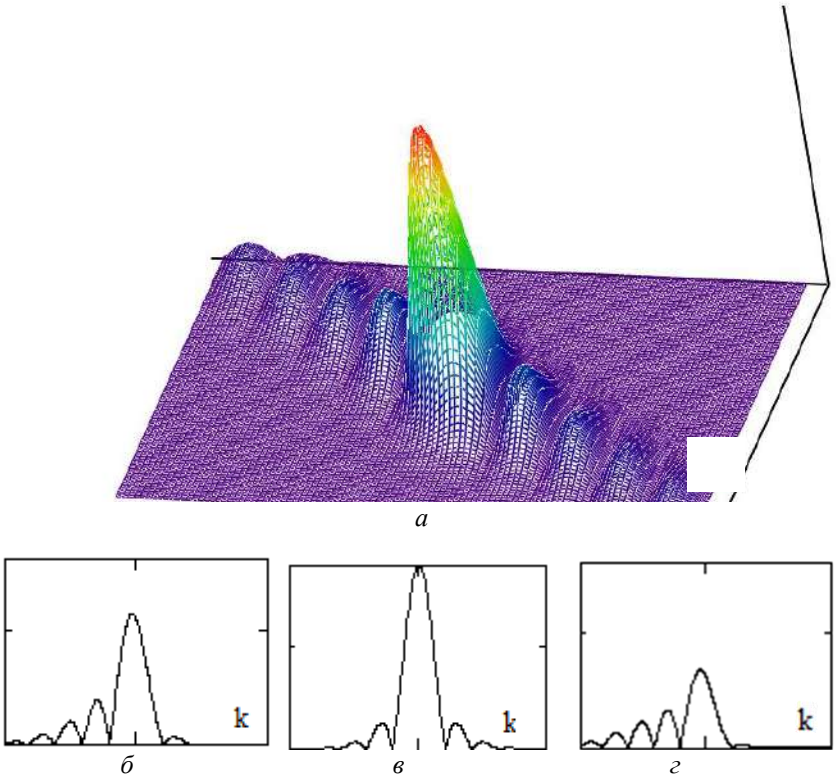
Тіло розсіяння містить в собі інформацію про основні характеристики системи зондування, у якій використовуються обрані зондувальні сигнали. Поряд з 3D поданням, воно може бути охарактеризоване також за допомогою перерізів вертикальними площинами $r = const$, $q = const$. Форма перерізу тіла площиною $q = q_0 = const$ є обвідною розсіяного сигналу для випадку, коли просторові частоти зондувальних акустичного та електромагнітного сигналів розстроєні на задану величину q_0. Аналіз форм відповідних обвідних дозволяє встановити вплив розстройки на характер змін обвідної розсіяного сигналу, на ступінь зменшення амплітуди і, відповідно, на основні характеристики системи зондування, які залежать від амплітуди і особливостей розсіяного сигналу.

Переріз тіла розсіяння площиною $r = 0$ визначає диапазон хвильових чисел Δq параметру розстроювання, в якому відбувається розсіяння. Тому перетин $Z(0,q)$ тіла характеризує зв'язок амплітуди розсіяного сигналу й значень параметра q. Тіло розсіяння дозволяє дослідити, як змінюється амплітуда розсіяного сигналу залежно від значення параметра q.

Рельєф тіла розсіяння сигналів можливо також подавати за допомогою ліній, які формуються при перерізі тіла горизонтальними площинами на певному рівні функції $Z(r,q) = const = Z_c$. Ці замкнуті лінії мають назву діаграм розсіяння. Можуть бути використані перерізи на рівні функції $Z_c = 0{,}5(0{,}7)$, $Z_c = 0{,}1$ або інших рівнях. Кількість градацій, якщо потрібно, можна збільшувати або зменшувати, залежно від завдання дослідження.

Перетин функції вертикальною площиною $Z(r,0)$ формує обвідну в області головної пелюстки, яка має трапеціїдальну форму. При збільшенні значення параметру q обвідна розсіяного сигналу отримає амплітудну модуляцію. Тіло розсіювання має, як бачимо, головну і бокові пелюстки.

В ряді задач дослідження зондувальних сигналів і взагалі систем РАЗ доцільно використати спектральне подання $Z(k,q)$ двовимірної функції розсіювання [2]. Відповідне тіло розсіяння для простих сигналів – акустичного імпульсу з гаусівською обвідною та прямокутного радіоімпульсу зображено на рис. 3, *а*. Перерізи тіла вертикальними площинами $q = const$ являють собою просторові спектри розсіяних сигналів для заданих значень параметру q (рис. 3, *б*, *в*, *г*).



а

б *в* *г*

Рис. 3. Тіло розсіяння прямокутного радіоімпульсу та акустичного імпульсу з гаусівською обвідною в спектральному поданні в координатах Z, k, q (*а*) та його перерізи площинами $q = const$: *б* – $q_1 = 0$; *в* – $q_2 < 0$; *г* – $q_3 < q_2 < 0$

Структура тіла має головну і бокові пелюстки (рис. 3, *а*), які формуються внаслідок розсіювання бокових пелюсток спектру радіосигналу на акустичному імпульсі. Головний і бокові пелюстки розгорнуті на деякий кут, значення якого залежить від протяжності зондувальних акустичного і електромагнітних імпульсів. При $q = 0$ форма спектру розсіяного сигналу є симетричною, а якщо $q \neq 0$, то спектр розсіяного сигналу має несиметричний вигляд, максимум спектру додатково зміщений відносно нуля частот, а значення зміщення залежать від кута повороту тіла розсіяння.

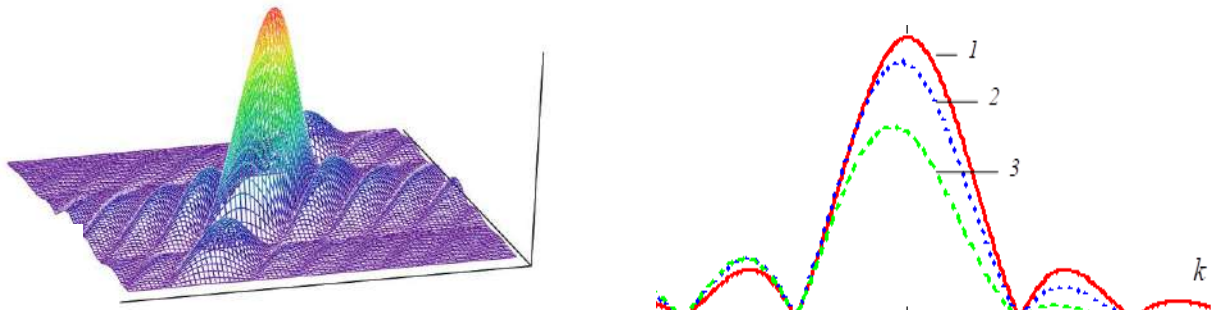
Як показано раніше, спектри розсіяних сигналів систем РАЗ залежать від видів зондувальних сигналів і від стану атмосфери – значень її характеристик. Вони формуються в результаті взаємодії акустичного і електромагнітного сигналів в конкретних атмосферних умовах, відтворюються на виході спектроаналізатору системи у вигляді спектру часових частот і

несуть інформацію про стан атмосфери. Вивчення особливостей спектрів розсіяних сигналів, отриманих для різних видів зондувальних сигналів і при різних станах середовища, дозволяє зрозуміти основні особливості систем РАЗ, їх відмінності від інших інформаційних локаційних систем, а також застосувати отримані знання при проєктуванні станцій РАЗ і алгоритмів їх функціонування.

Для більшості тіл розсіяння перерізи або спектри сигналів при $q \neq 0$ є суттєво несиметричними і характеризуються зміщенням максимуму, а також центру ваги відносно точки $k = 0$. Це зміщення при розсіянні додається до доплерівського зсуву частоти і проявляється при оцінюванні як неінформативна добавка чи похибка (величина похибки визначення температури може досягати одиниць градусів).

Розроблені уявлення дозволили дати зрозумілу фізичну інтерпретацію, пояснити та упорядкувати значну кількість наукових результатів і даних, накопичених за тривалий період розвитку методів. Особливо продуктивними в цьому відношенні виявились частотні зображення.

Тіло розсіяння простих електромагнітного та акустичного імпульсів, які мають прямокутні обвідні, у спектральному поданні $Z(k, q)$ представлено на рис. 4. Спектральне тіло розсіяння аналізованих сигналів має, як видно, декілька рядів бічних піків і розгорнуту на деякий кут основну пелюстку. Один ряд пелюсток формується при розсіюванні головного пелюстка спектру радіосигналу на бокових пелюстках спектру акустичного сигналу, а інші – внаслідок розсіювання бокових пелюсток спектру радіосигналу на основному пелюстку спектру акустичного сигналу.



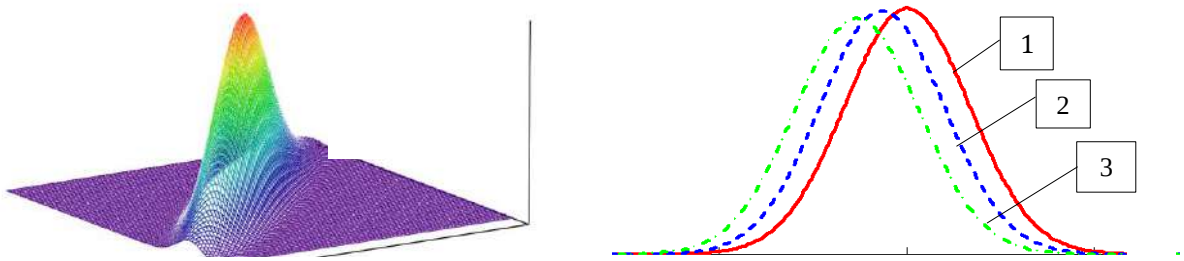
а *б*

Рис. 4. Тіло розсіяння електромагнітного та акустичного імпульсів з прямокутними обвідними у спектральному поданні $Z(k, q)$ (*а*) та перетини тіла розсіяння площинами *q=const* (*б*): 1 – $q_1=0$; 2 – $q_2>0$; 3 – $q_3> q_2>0$

Перетини тіла площинами $q = const$ є просторовими спектрами розсіяних сигналів, що відповідають певним станам атмосфери (рис. 4). Як видно, просторовий спектр розсіяного сигналу при $q = 0$ є симетричним, а при $q \neq 0$ несиметричним і характеризується зсувом максимуму (а також центру ваги) Δk відносно точки $k = 0$. Додатковий неінформативний зсув максимуму спектральної функції при розсіянні радіосигналу на звуковій посилці додається до доплерівського зсуву частоти й проявляється при оцінюванні частоти сигналу, а, отже, оцінюванні швидкості звуку, як похибка при використанні традиційних методів обробки.

Рельєф тіл розсіяння з гаусівськими обвідними (рис. 5), як видно, не має бічних піків через особливости використовуваних коливань (у спектрах зондувальних сигналів гаусівського виду таких бічних пелюсток також немає). Тіла розсіяння більшості сигналів у спектральному поданні мають характерну рису – їхні піки повернені на деякий кут, що характерно і для тіла на рис. 5.

Перетини тіла розсіяння (рис. 5, *а*) площинами *q=const*: 1 – $q_1=0$; 2 – $q_2>0$; 3 – $q_3> q_2>0$ подано на рис. 5, *б*. Просторові спектри розсіяних сигналів, як видно, не мають бокових пелюсток, але для випадків $q \neq 0$ вони також є несиметричними і мають додатковий неінформативний зсув по осі просторових і часових частот.



а *б*

Рис. 5. Тіло розсіяння простих радіо- та акустичного імпульсів, які мають гаусівські обвідні, у спектральному поданні $Z(k,q)$ (*а*) та перетини тіла розсіяння площинами *q=const* (*б*): 1 – $q_1=0$; 2 – $q_2>0$; 3 – $q_3> q_2>0$

Вираз, що описує неінформативне зміщення частоти максимуму спектра розсіяного радіосигналу для тіл, наведених на рис. 5, має вигляд

$$\Delta f = -\frac{2v_e^2}{2\pi(4v_e^2+v_s^2)} \cdot q, \tag{5}$$

де v_e, v_s – параметри, що характеризують просторові тривалості радіо- та акустичного сигналів. Саме значення цих параметрів визначає кут повернення пелюсток тіла розсіяння в спектральному поданні.

Як показують результати розрахунків за допомогою виразу (5), для значень параметрів зондувальних сигналів λ_e =0,68 м (f_e =440 МГц), λ_s =0,34 м (f_s =1 КГц) зміна температури атмосфери з висотою на один градус призводить до додаткового неінформативного зсуву частоти сигналу на Δf ≈0,32 Гц, що відповідає систематичній похибці вимірювання швидкості звуку Δc_s ≈0,11 м/с і похибці вимірювання температури середовища ΔT ≈0,18 С. У реальних умовах зондування значення систематичної похибки результатів вимірювання температури атмосфери може бути значно більше і досягати одиниць градусів.

Висновки

1. Удосконалено математичну модель інформаційного локаційного каналу систем РАЗ, що відрізняється використанням просторових хвильових частотних уявлень зондувальних акустичного та електромагнітного сигналів, які дозволяють фізично змістовно та наочно описати процес їх взаємодії в атмосфері та формування розсіяного радіосигналу при різних видах зондувальних коливань і різних станах атмосфери.

2. Проаналізовано тіла розсіяння в частотному поданні найчастіше використовуваних видів зондувальних акустичних та електромагнітних сигналів – простих за структурою акустичних та електромагнітних імпульсних сигналів з прямокутними та гаусівськими формами обвідних. Досліджено основні властивості та особливості вказаних сигналів. Показано, що при виконанні умови Брегга, коли параметр $q = 2k_e - k_s$ дорівнює нулю, спектр розсіяного сигналу при використанні досліджених сигналів є симетричним, що дозволяє використовувати при обробці і оцінюванні інформативних параметрів відомі класичні методи оцінювання параметрів спектрів сигналів. Коли параметр Брегга $q \neq 0$, спектр сигналу є несиметричним і такий метод обробки сигналів використовувати неможливо.

3. З використанням запропонованої математичної моделі каналу з'ясовано фізичні причини появи різноманітних специфічних похибок в результатах вимірювань швидкості звуку радіоакустичними системами. Зазначені види похибок не можуть бути усунені осередненням одиничних результатів вимірювань, а також не можуть бути усунені методами послідовної оптимальної лінійної фільтрації, які використовуються в контурах керування параметрами систем зондування. Наявність систематичної похибки в контурі адаптації буде негативно позначатися також на ефективності функціонування алгоритму управління частотою зондувального радіосигналу з метою адаптації систем РАЗ до метеоумов, що змінюються по трасі.

4. Сформульовано можливості використання на практиці проаналізованих видів сигналів і вказано, які методи обробки сигналів доцільно використовувати. Визначено рекомендації щодо використання зондувальних сигналів систем РАЗ, режимів роботи та вимірювань у станціях зондування з метою реалізації на практиці методів та алгоритмів частотної адаптації систем до метеорологічної обстановки, що змінюється.

Список літератури:

1. Bradley S. Atmosphere Acoustic Remote Sensing. Principes andApplication. CRC Press. 2007. 267 p.
2. Kartashov V.M., Tikhonov V.A., Oleinikov V.N. Signal processing in radio electronic systems for remote monitoring of the atmosphere. Kharkiv : KNURE, 2014. 312 p.
3. Карташов В.М. Моделі і методи обробки сигналів систем радіоакустичного і акустичного зондування атмосфери. Харків : ХНУРЕ, 2011. 234 с.
4. Lataitis R.J. Theory and Application of a radio-acoustic sounding system (RASS): NOAA Technical Memorandum ERL WPL-230. Nat. Oceanic and Atmos. Admin. Environ, Res. Labs. Boulder, CO, 1993, 207 p.
5. Remtech Radio Acoustic Sounding System (RASS) for remote sensing of temperature. URL: https://remtechinc.com/wp-content/uploads/RASS3.pdf (дата звернення 14.07.2024).
6. Temperature Profiler RASS. URL: https://metek.de/product-group/rass/ (дата звернення 14.07.2024).
7. RASS for Radar Wind Profilers. URL: https://www.scintec.com/catalogs/rass-for-radar-wind-profilers/ (дата звернення 14.07.2024).
8. Smith P. L. Remote measurements of wind velocity by the electromagnetic-acoustic probe. 1. System analysis.1961 // Conf. proc. 5th Annu. convention on military electronics, Wash (D.C.), rep № 419. P. 43–53.
9. Fetter R. V. Remote measurements of wind velocity by the electromagnetic-acoustic probe. II. Experimental system. 1961 // Conf. proc. 5th Annu. convention on military electronics, Wash (D.C.), rep № 419. P. 54–59.
10. Ситнік О.В., Карташов В.М. Радіотехнічні системи : навч. посіб. Харків : Сміт, 2009. 448 с.
11. Chandrasekhar Sarma, T. V., Narayana Rao, D., Furumoto, J., and Tsuda, T. Development of radio acoustic sounding system (RASS) with Gadanki MST radar – first results, Ann. Geophys., 26, 2008, pp. 2531–2542. https://doi.org/10.5194/angeo-26-2531-2008
12. Alexander S. P., Murphy D. J., Klekociuk A. R., High resolution VHF radar measurements of tropopause structure and variability at Davis, Antarctica (69° S, 78° E). Atmos. Chem. Phys., 13, 2013. pp. 3121– 3132. doi:10.5194/acp-13-3121-2013
13. Бабкін С.І., Куценко В.І., Максимова Н.Г. Оцінка похибки двох методик температурного радіоакустичного зондування атмосфери. Експериментальні результати // Радіотехніка. 1988. № 84. С.98–106.
14. Kartashov V.M. Estimation of Signal Parameters Scattered by an Acoustic Wave Packet // Telecommunications and Radio Engineering, 2004. Vol. 61, №2. P. 125–129.
15. Muradyan P., Richard Coulter R. Radar Wind Profiler (RWP) and Radio Acoustic Sounding System (RASS) Instrument Handbook. March, 2020. Environmental Science Division, Argonne National Laboratory. 20 p. URL: https://www.arm.gov/publications/tech_reports/handbooks/rwp_handbook.pdf (дата звернення 14.07.2024).
16. Kartashov V.M. Signal Scattering Functions of Atmospheric Sounding System // Telecommunications and Radio Engineering. 2003, Vol. 59, №7-8-9. P. 88–94.
17. Kartashov V., Babkin S., Kartashov A., Pershyn Y. Development of the Atmosphere Radio-Acoustic Sounding Method in Ukraine and in the World in the Period of 1961-2000 // 2023 IEEE 6th International Conference on Information and Telecommunication Technologies and Radio Electronics, UkrMiCo 2023, 13–15 November 2023, Kyiv, Ukraine. P. 372–376. DOI: 10.1109/UkrMiCo61577.2023.10380339
18. Kartashov V., Oleynikov V., Koryttsev I., Sheiko S., Zubkov O., Babkin S. Processing of Wide Band Acoustic Signals During Detection of Unmanned Aerial Vehicles // 2020 IEEE Ukrainian Microwave Week (UkrMW). Kharkiv, Ukraine, September 21 – 25, 2020. Vol. 1 on 2020 IEEE 12th International Conference on Antenna Theory and Techniques (ICATT). P. 35–39.
19. Developing and Applying Optoelectronics in Machine Vision / O. Sergiyenko, J.C. Rodriguez-Quiñonez. IGI Global, 2016. 341 p.
20. Kartashov V.M., Tikhonov V.A., Voronin V.V. and Tymoshenko L.P. Complex model of random signal in problems of acoustic sounding of atmosphere // Telecommunications and Radio Engineering. 2016. Vol. 75, Iss. 20. P.1885–1892.

Надійшла до редколегії 22.08.2024

Відомості про автора:

Карташов Олександр Володимирович – Харківський національний університет радіоелектроніки, аспірант кафедри медіаінженерії та інформаційних радіоелектронних систем; Україна; email: oleksandr.kartashov@nure.ua; ORCID: https://orcid.org/0000-0002-4618-4787

УДК 615.844 **DOI:10.30837/rt.2024.3.218.11**

О.М. ЗІНЧЕНКО, В.П. ОЛІЙНИК, канд. техн. наук, П.М. ПОДПРУЖНИКОВ

СТАН ТА ПЕРСПЕКТИВИ ЗАСТОСУВАННЯ ЗАСОБІВ ДІАГНОСТИКИ НА ОСНОВІ МЕТОДУ ГАЗОРОЗРЯДНОЇ ВІЗУАЛІЗАЦІЇ

Вступ

Під терміном «діагностика» розуміють процес визначення стану об’єкта дослідження. Результатом діагностики є комплекс якісних і кількісних показників, що дозволяють класифікувати стан об’єкта. У більшості практичних застосувань діагностику пов’язують з об’єктом, стан якого визначається. До поширених видів діагностики належить медична діагностика – комплекс заходів та досліджень, спрямованих на встановлення діагнозу, тобто точної причини захворювання людини, а також змін внутрішнього середовища організму та супутніх захворювань, та призначення ефективного лікування захворювання. Також у сучасній інженерії використовують технічну діагностику – комплекс робіт контролю технічного стану об’єкта. Виконується, в основному, з використанням відповідних методів та засобів вимірювальної техніки. Технічний огляд проводиться у терміни та у випадках, визначених нормативно-правовими актами з охорони праці, організаційно-методичними та експлуатаційними документами [1].

Відповідно до приведених визначень, як для об’єктів живої, так і неживої природи діагностика базується на методі дослідження і засобах його реалізації. Минуло майже три четверті століття як було зареєстровано авторство на метод дослідження, який у науковій термінології отримав назву «Ефект Кірліан» [2]. Він полягає в отриманні зображень випромінювання газового розряду довкола об’єкта дослідження, що знаходиться в змінному електричному полі великої напруженості. У разі, коли об’єктом є живі організми або біологічні структури, використовують назви цього методу – «Кірліан-діагностика» та «Біоелектрографія». Основна діагностична ідея методу – це зв’язок світіння електричного розряду в оточуючому атмосферному повітрі з електрофізичними властивостями внутрішніх та поверхневих структур об’єкта, які у свою чергу інтерпретуються медико-біологічними показниками для живих організмів, або технічними характеристиками для неживих структур [3]. Важливою особливістю цього методу є аналіз не власного зображення об’єкту, а переважно випромінювання оптичного діапазону електромагнітних хвиль, що виникає внаслідок розрядних процесів. Тому можна вважати, що термін «газорозрядна візуалізація» (ГРВ) є коректним для аналізу всіх практичних реалізацій цього методу.

Засоби газорозрядної візуалізації були створені для наукових та прикладних досліджень біологічних, небіологічних об'єктів, стану навколишнього середовища. За допомогою цих засобів можна проводити діагностику організму людини, вивчати рідини та тверді речовини органічного та неорганічного походження (кров, вода, рослинні культури, мінерали тощо) [4]. До недоліків методу відносять залежність інформативних показників зображень не тільки від властивостей об’єкта, а і сталості параметрів оточуючого середовища. Відсутні єдині метрологічні вимоги до технічних засобів ГРВ, що стримує їх практичне застосування.

Метод ГРВ завжди привертав увагу науковців завдяки своїм діагностичним можливостям. Тому об’єктом дослідження в цій роботі є метод газорозрядної візуалізації, предметом – технічні засоби його реалізації.

Мета дослідження – пошук технічних рішень побудови засобів газорозрядної візуалізації для досягнення потенціальних можливостей їх практичного застосування. Для досягнення мети поставлені такі задачі: провести аналіз фізичних та технічних положень газорозрядної

візуалізації; розглянути структуру існуючих засобів ГРВ; визначити перспективні напрями використання ГРВ-засобів та запропонувати шляхи їх модернізації та удосконалення.

Фізичні та технічні основи методу газорозрядної візуалізації

Опубліковано велику кількість робіт, що розглядають фізичні процеси формування ГРВ зображень [5–7]. На етапі розробки нових та модифікації вже наявних засобів газорозрядної візуалізації дуже важливим стає однакова наукова термінологія при плануванні, проведенні та описі результатів діагностики, а також сама назва методу має відображати сутність фізичних процесів.

На сьогодні у світі розроблено сотні практичних модифікацій засобів ГРВ залежно від геометричної форми, параметрів та фізичних властивостей досліджуваних об'єктів живої та неживої природи. При всьому різноманітті конкретних технічних рішень сутність процесу візуалізації у всіх цих модифікаціях однакова і може бути зведена до загальної послідовності процесів.

Газовий розряд виникає в системі, що складається з об'єкта дослідження, носія зображення та електродів, що формують електромагнітне поле (ЕМП). Первинним процесом є взаємодія ЕМП з об'єктом дослідження, ініціювання початкових фаз газового розряду за певної напруженості ЕМП, виникненні емісії заряджених частинок з поверхні об'єкта, що беруть участь у підтримці розряду. Для пояснення деяких кількісних характеристик розрядного процесу використаємо узагальнену будову засобів ГРВ, показану на рис. 1.

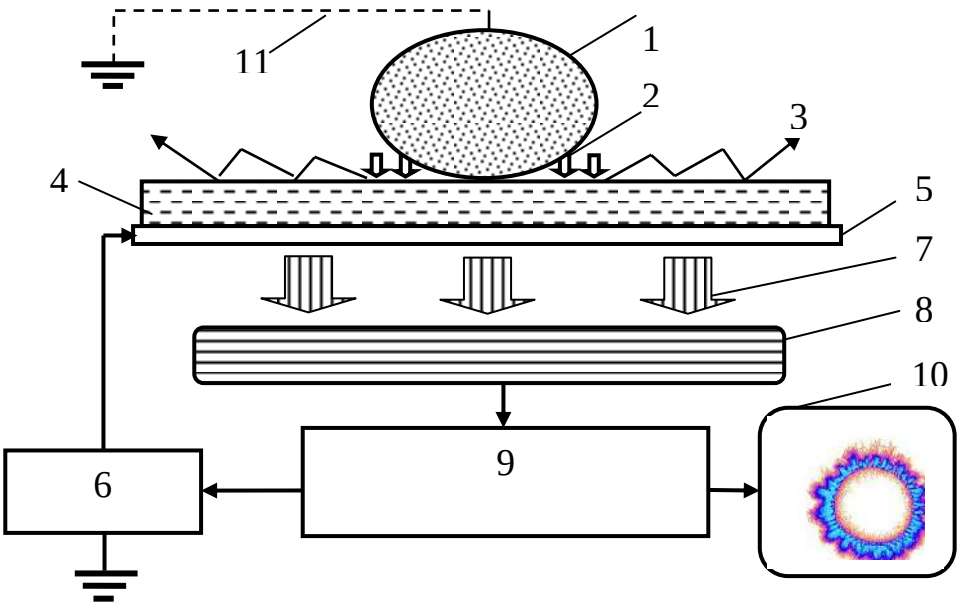


Рис. 1. Структурні елементи засобів газорозрядної візуалізації:
1 – об’єкт діагностики; 2 – лавинний пробій; 3 – ковзний розряд;
4 – прозора діелектрична основа; 5 – прозора провідникова плівка; 6 – джерело високовольтної імпульсної напруги; 7 – світіння в оптичному діапазоні випромінювання; 8 – відеоперетворювач;
9 – комп’ютерний пристрій керування та оброблення інформації; 10 – ГРВ- зображення на моніторі;
11 – коло розрядних струмів

Основне джерело формування зображення – це газовий розряд поблизу поверхні об'єкта, що досліджується. Експериментальні дослідження показали, що можна виділити два основні типи розряду, пов'язані з формуванням газорозрядних зображень: лавинний, що розвивається в обмеженому діелектриком вузькому зазорі, та ковзний на поверхні діелектрика [5–7]. У ході розвитку лавини безупинно збільшується число електронів і позитивних іонів. Із збільшенням числа електронів у голівці лавини зростає напруженість на фронті лавини. Відбувається перехід лавинного розряду в стримерний. Стример являє собою слабоіонізо-

ванний тонкий канал, що світиться, який утворюється в результаті злиття електронних лавин (лавинно-стримерний перехід). Фотони, що виникають у процесі розряду, формують двовимірну картину на носії зображення. Газовий розряд також може впливати на стан об'єкта, викликаючи вторинні емісійні, деструктивні та теплові процеси.

У більшості сучасних реалізацій засобів ГРВ об’єкт діагностики розташовується на сенсорі, що складається з прозорої діелектричної основи (скло, кварцове скло), на зворотну сторону якого напилена прозора провідникова плівка (оксид олова). Джерело високовольтної імпульсної напруги створює напруженість електричного поля достатню для лавинного пробою та ковзного розряду. Коло розрядних струмів має в основному ємнісний характер, але для деяких об’єктів використовують додатковий контакт з «нульовим» потенціалом для збільшення напруженості електричного поля. Випромінювання розряду реєструє відеоперетворювач. У випадку, коли реєстрація світіння відбувається на фотоматеріалі (фотоплівка, фотопапір) з наступною хімічною обробкою, поширено використання назви «кірліанограма». На сучасному етапі побудови засобів ГРВ діагностики функцію відеоперетворювача виконують матричні цифрові камери, доповнені оптичною системою. З відеокамери цифрове зображення світіння розряду подається на пристрій керування процесом візуалізації і після оброблення виводиться на монітор користувача.

Розглянуті технічні рішення реалізовані у програмно-апаратних ГРВ комплексах. Використання RISC мікроконтролера дозволяє проводити управління приладом та обирати робочі параметри від ПК, синхронізувати роботу всіх блоків, а також налаштовувати параметри приладу при проведенні контролю обладнання. Програмована затримка запису відеосигналу по відношенню до поданого імпульсу напруги дозволяє відстежувати динаміку реакції біологічного об'єкта на імпульс збудження.

Більшість ГРВ комплексів мають наступні технічні параметри: амплітуда біполярних імпульсів від 3 до 20 кВ з плавним або ступінчастим регулюванням; тривалість імпульсів 10 мкс; частота проходження імпульсів до 1000 Гц; встановлення часу експозиції у діапазоні 0,1…32 с; здійснення двостороннього зв'язку з комп'ютером за допомогою USB порту, що дозволяє як передавати інформацію (команди) у прилад, так і здійснювати діагностику режимів роботи приладу; кварцова стабілізація всіх параметрів з точністю не гірше 1 %.

У ГРВ засобах для створення електричного поля розряду використовують серію біполярних імпульсів напруги. При кожній полярності імпульсу виникає відповідна фаза розряду, і остаточна картина являє собою суперпозицію зображень від позитивного та негативного розрядів (з урахуванням спотворення електричного поля позитивним поверхневим зарядом, що залишилися після попередніх розрядів). Для ГРВ використовується лише лавинна фаза цього розряду на низькому струмі, при якій інтегральна величина струму в імпульсі не перевищує 50 мА [4, 5].

Для виявлення інформаційної ваги різних компонентів оптичного випромінювання було проведено велику серію робіт з експериментального дослідження спектру світіння різних об'єктів у процесі ГРВ [4]. Інтерес до цього питання стимулювався численними роботами з «ефекту Кірліан», у яких було зазначено, що на кольорових фотографіях світіння спостерігається спектр кольорів, що закономірно залежать від стану досліджуваного об'єкт. У зв'язку з коротким часом розвитку розряду, дослідження цього спектра є складною технічною задачею, при вирішенні якої були використані оптичні фільтри, спектрографи та імпульсні спектрометри. Було встановлено, що спектр випромінювання ГРВ розряду в повітрі переважно займає область від 150 до 800 нм, найбільш активна частина спектру складається в основному з молекулярних смуг другої позитивної системи азоту, а також містить лінії CO, CO_2 та O_2, що зазвичай спостерігається у розряді з невеликим струмом в повітряному середовищі. Основна область спектру випромінювання знаходиться в діапазоні 280…800 нм. У електропозитивних газах (повітря, азот, водень та ін.) вид газорозрядних фігур якісно ідентичний, тоді як введення електронегативних добавок (наприклад, CCl_4) викликає кардинальну зміну всієї фігури: різке зменшення розміру та придушення «тонкої структури» зображення.

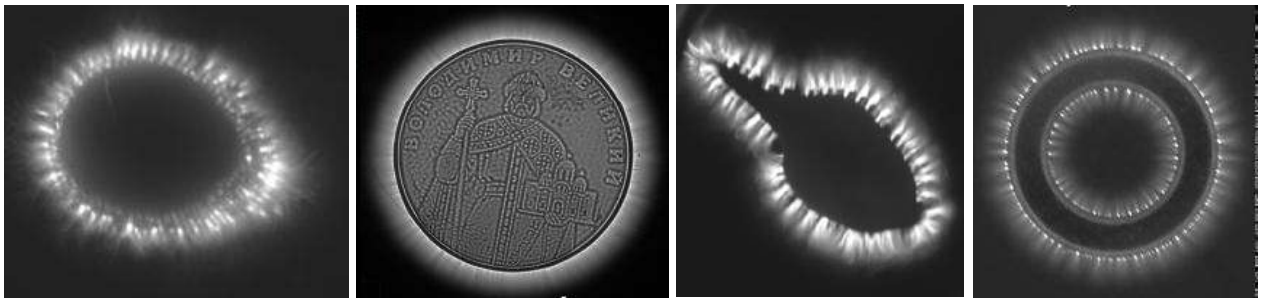
Це пов'язано з трьома основними процесами: поглинанням повільних електронів, що перешкоджає розвитку електронних лавин; поглинанням вторинних електронів, народжених у лавині; спотворенням електричного поля за рахунок негативних іонів.

Як показують експерименти, практично в основі всіх випромінювань тканин організму у видимій та ультрафіолетовій областях спектру лежить той чи інший різновид люмінесценції. У процесі ГРВ може виникати люмінесценція, індукована різними фізичними факторами: ультрафіолетовим та видимим випромінюванням – фотолюмінесценція; іонізуючим випромінюванням – радіолюмінесценція; електричним струмом – електролюмінесценція; хімічними реакціями – хемілюмінесценція.

Як випливає з досліджень, надслабке світіння у видимій та ультрафіолетовій області за певних умов може робити внесок у процеси ГРВ за рахунок фотоіонізації та ініціації електронних лавин [5, 6].

Основні інформативні показники, що визначаються за газорозрядними зображеннями

Цільовим інформаційним джерелом методу ГРВ є ГРВ-грама. ГРВ-грама – це одиничне зображення газового розряду, зафіксованого в будь-який момент часу експозиції електромагнітного поля в області об’єкта [6]. Іноді, ГРВ-граму називають «аурою» об’єкта. На рис. 2 показано ГРВ-грами різних об’єктів, отримані в роботі [8].



а *б* *в* *г*

Рис. 2. ГРВ-грами: *а* – вказівного пальця людини, *б* – металевої монети, *в* – краплі води, *г* – феритового кільця

Поверхневі ГРВ-грами досить складні, тому для виявлення інформації про досліджуваний об’єкт необхідно виявити параметри розряду, які необхідно оцінити [5]. До них відносять:

– параметри, що характеризують розрядні стримери: довжина, кількість, ступінь розгалуження;

– параметри, що характеризують форму газорозрядної фігури: радіус, загальна площа, симетрія, фрактальна розмірність та інші;

– інтегральний струм розряду в різних частотних діапазонах;

– інтегральна інтенсивність світіння розряду;

– спектральний склад випромінювання [9, 10].

При цьому стримером вважається суперпозиція в часі окремих коронних стримерів на півперіодах змінної напруги, тобто інтегральна картина світіння окремих коронних стримерів, що чергуються [11].

Недоліком таких досліджень є неможливість встановлення ідентичності отриманих зображень у різних дослідах, оскільки це пов’язано зі зміною параметрів напруги, частоти, стану атмосфери, часу експозиції фотокамери, енергії виходу з об’єкта, наявності на поверхні забруднень, нерівностей країв об’єктів та його шорсткості, з дією розряду на сам об’єкт тощо.

У ГРВ-грамі інформація про об'єкт проявляється в параметрах зображення за рахунок впливу об'єкта на такі характеристики розряду: інтенсивність світіння, тривалість, частоту проходження і просторовий розподіл окремих лавинних актів, а також спектральний склад випромінювання. Як основні джерела інформативних ознак об'єктів можна виділити такі:

− фактори, що викликають розподіл електричного поля в розрядному зазорі (наприклад, неоднорідність структури поверхні або об’єму), оскільки при однаковій концентрації ініціювальних частинок лавинні розряди виникають переважно в областях з максимальною напруженістю електричного поля і розвиваються більш інтенсивно у порівнянні з сусідніми областями;

− просторова або часова неоднорідність емісійних властивостей поверхні об'єкта, оскільки від цього залежить як величина струму розряду, так і частота проходження розрядів;

− просторова або часова неоднорідність власного газовиділення (або випаровування) поверхні об'єкта, тому що вона впливає на склад газового середовища в зазорі, а отже, на інтенсивність розряду і спектральний склад випромінювання;

− неоднорідність поверхні об'єкта щодо електропровідності або її зміни в часі, оскільки від цього залежить інтенсивність окремих актів розряду та частота їх проходження;

− загальний імпеданс електричного кола, що залежить, при сталості інших параметрів, від електричних властивостей досліджуваного об'єкта, його поверхні та об'єму.

Аналітичні моделі в сукупності з експериментальними даними робіт [9, 10, 12] дозволили зробити висновки, щодо забезпечення інформативності при практичній реалізації методу ГРВ:

− характер фізичних процесів в рамках одиничного акту розряду практично не залежить від форми напруги, що подається (синусоїдального або імпульсного), а визначається перерозподілом електричного поля в розрядному зазорі завдяки накопиченню поверхневого заряду на обмежувальних електродах і досліджуваному об'єкті;

− оптимальним є використання послідовності коротких імпульсів напруги. Зіставлення ГРВ-грам, отриманих при різній тривалості імпульсів T_U, показало, що найкраща якість виявлення металевих включень у товщі діелектрика спостерігається при 5 мкс < T_U < 15 мкс;

− математичний аналіз теплових процесів на поверхні біооб'єкта в умовах отримання ГРВ біоелектрограм показує, що при характерному часі впливу ~ 10^{-5} с потужність в розряді виявляється недостатньою для помітного температурного впливу на об'єкт. Це свідчить про неруйнівний характер діагностики.

У роботі [13] запропоновано безрозмірні показники для оцінювання ГРВ зображення: відношення внутрішньої площі аури до площі досліджуваного об’єкта; зовнішньої площі аури до площі об’єкта; відношення різниці площ внутрішньої і зовнішньої аур до загальної площі аури; відношення різниці площ внутрішньої і зовнішньої аур до площі об’єкта; відношення кількості елементів, що входять до внутрішньої і зовнішньої аур відповідно.

Більшість основних інформативних ознак зображень ГРВ ґрунтуються на аналізі цілого зображення, а не окремих елементів, таких як стримери. Таким чином, інформативні параметри, що можна встановити з аналізу зображень стримерів, недостатньо вивчені та потребують подальших досліджень.

Структура технічних засобів газорозрядної візуалізації

При розгляді різноманіття відомих технічних рішень побудови засобів газорозрядної візуалізації оберемо певні структурні ознаки. Базовою умовою візуалізації є забезпечення напруженості електричного поля достатньої для розряду у повітрі, тому до першої групи ознак віднесемо фізико-технічні властивості які приведені у табл. 1.

Таблиця 1

Фізичні та технічні рішення використані для побудови засобів ГРВ

Формування напруженості електричного поля розряду	**Елементи сенсорного пристрою**	**Спосіб реєстрації зображення**	**Спектральні характеристики випромінювання, що реєструється**	**Динамічні характеристики зображення**
Джерело постійної напруги 3…100 кВ	Двоелектродний: провідник – об’єкт – діелектрик – провідник	Переважно фотоматеріали (фото та рентгенівські плівки, папір, пластини)	Ультрафіолетове, видиме, ближнє інфрачервоне випромінювання, потоки електронів, іонів	Статичні зображення, інтегральна візуалізація визначається тривалістю експозиції
Генератор змінної напруги синусоїдальної форми, амплітуда до 30 кВ, діапазон частот 0,05…200 кГц	Двоелектродний: провідник – об’єкт – прозорий діелектрик – прозорий провідник; або Одноелектродний: об’єкт – прозорий діелектрик – прозорий провідник	Переважно фотоматеріали (фото та рентгенівські плівки, папір, пластини)	Ультрафіолетове, видиме, ближнє інфрачервоне випромінювання, потоки електронів, іонів	Статичні зображення, інтегральна візуалізація визначається тривалістю експозиції
Генератор імпульсної напруги, амплітуда до 25 кВ, частота повторення імпульсів 0,1…500 кГц, тривалість імпульсів більше 1 мкс		Телевізійна відеокамера	Видиме випромінювання	25 кадрів (50 полів) за секунду, стандартний TV-сигнал
		Цифрова відеокамера	Видиме, ближнє інфрачервоне випромінювання	Синхронізація з тривалістю експозиції, максимальна кількість кадрів – 100 за секунду (задається програмно)

При порівнянні двох способів реєстрації – фотоплівка або цифрова камера вибір залежить від того, яке завдання вирішується.

Роздільна здатність фотопаперу становить 300 ліній/мм. Це означає, що при реєстрації зображення світіння на фотопапері шириною 10 см буде розрізнятися 30000 елементів. Сучасні цифрові камери мають до 3000 пікселів у рядку. Значить, якщо за допомогою оптичної системи спрєктувати таке ж за розмірами світіння на приймальну матрицю відеокамери, то маємо програш у роздільності в 10 разів. Тому цифрові камери для досягнення високої роздільної здатності обмежують поле зору до 25 мм.

Принципова відмінність у використанні фотоматеріалів від цифрової камери або телекамери полягає в тому, що відеокамери здійснюють реєстрацію тільки фотонів видимої частини спектру світіння розряду. Фотоматеріали реєструють ширший спектр випромінювання фотонів, потік іонів і потік неелектромагнітного випромінювання. За допомогою фотоматеріалів можна реєструвати процес світіння з непрозорими електродами сенсору, чого неможливо зареєструвати за допомогою відеокамер.

Реєстрація на фотоматеріали ефективніша для реєстрації тонких енергопольових ефектів з високою роздільною здатністю. Недоліки: трудомісткий і тривалий процес хімічного оброблення зображень.

Реєстрація за допомогою відеокамери має велику оперативність і можливість відразу отримувати результат. За рахунок усереднення розрядів від багатьох імпульсів формується зображення що добре відображає загальний стан об’єкта діагностики.

Зважаючи на те, що у науковій термінології для методу газорозрядної візуалізації XX та початку XXI сторіччя затвердився термін «ефект Кірліан» технічні засоби мали назву «Кірліан камера». У більшості цих зразків реєстрація зображення відбувалась на фотоматеріалі. На рис. 3 показана кірліан-камера з реєстрацією на кольоровий фотопапір [14].

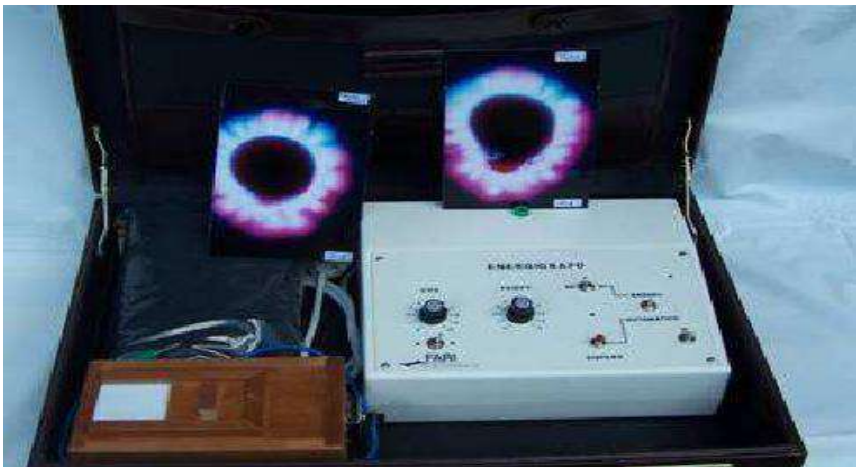


Рис. 3. Кірліан-камера Modelo K200 розробки Рауля Тореса (Аргентина, 2000 – 2005 р.)

Відомо досить багато конструкцій Кірліан-камер, побудованих за цим способом реєстрації в різних країнах світу (Аргентина, Бразилія, Великобританія, Італія, Іспанія, Німеччина, СРСР, США та ін.) [14]. Переважна більшість цих засобів спрямована на дослідження фізіологічного стану людини або окремих патологій за зображеннями світіння газового розряду довкола пальців кінцівок. Це були авторські розробки науковців і дослідників, виготовлені в одиничних або декількох екземплярах.

Суттєві зміни у конструкціях засобів газорозрядної діагностики пов’язані з комп’ютеризацією електроніки і цифровими технологіями оброблення інформації. Поширення персональних комп’ютерів, удосконалення програмного забезпечення та використання цифрових відеокамер забезпечило створення ГРВ апаратно-програмних комплексів. Сучасні розробки мають ГРВ-камеру, що через стандартний USB інтерфейс під’єднується до персонального комп’ютера (ноутбука, планшета). Первинне світіння розряду реєструється цифровою відеокамерою і відображається на моніторі ПК. Відібрані первинні зображення обробляються в ПК за програмним алгоритмом, який визначає необхідні діагностичні показники.

ГРВ обладнання, що набуло поширення в різних напрямах дослідницької діяльності, представлено лінійкою ГРВ приладів [3, 15], їх основні технічні характеристики наведено в табл. 2. Також випускаються допоміжні засоби для дослідження об'єктів навколишнього середовища, що входять в комплектацію набору «ГРВ мінілабораторій». Прилад «ГРВ Експрес» призначений для одночасного зняття ГРВ-грам десяти пальців рук людини.

У 2014 р. розроблено та запущено у виробництво наступне покоління ГРВ приладів «Біо-Велл» з обробкою інформації в Інтернет просторі, що зумовило новий етап у розвитку біоелектрографії [6, 14]. Програмно-апаратний комплекс «Біо-Велл» дозволяє отримані ГРВ-грами зберігати на сервері, де відбувається обробка даних за допомогою аналітичного програмного забезпечення, заснованого на використанні методів обробки зображень, штучного інтелекту та математичного аналізу великих баз даних. Програми, що базуються на більш ніж 30000 вимірюваннях, формують висновки, отримані з порівняння ГРВ-грами випробуваного із усередненими діапазонами величин, притаманними здоровим людям. Автоматична обробка даних дозволяє одночасно обробляти ГРВ-грами кількох випробуваних з можливістю порівняння кількох серій зйомок (наприклад, зроблених у різні дні, до та після різних впливів на організм випробуваного). Ці дані доступні у вигляді графіків чи числових таблиць. Програма також дозволяє розрахувати коефіцієнт, що характеризує рівень стресу.

Таблиця 2

Порівняльні характеристики приладів ГРВ

Функції та характеристики	**ГРВ Міні**	**ГРВ Компакт**	**ГРВ Камера**	**ГРВ Експрес**
				
Діагностика людини	По 1 пальцю	По 1 пальцю	По 1 пальцю	Одночасно 10 пальців
Дослідження речовин	Немає	Тільки вода (за допомогою додаткового пристрою)	Різні речовини (за допомогою ГРВ міні-лабораторії)	Немає
Кількість режимів напруги	1	1	4	1
Амплітуда напруги, кВ	Інформація відсутня	До 5	До 5	Інформація відсутня
Тривалість одиночного імпульсу, мкс	Інформація відсутня	Інформація відсутня	10	Інформація відсутня
Частота послідовності імпульсів, Гц	Інформація відсутня	1024	900–1100	Інформація відсутня
Тривалість автоматичної експозиції, с	Інформація відсутня	Інформація відсутня	0,5/1,0/2,0/32	Інформація відсутня
Вхідний сигнал на пристрій автоматичної обробки	Інформація відсутня	Інформація відсутня	Цифровий відеосигнал в стандарті CYX 4:2:1	Інформація відсутня

Комплекс «Біо-Велл» має гарну відтворюваність основних характеристичних величин світіння – площі (S) та інтенсивності (I) – у серії послідовних вимірів. Для метрологічної перевірки приладу використовується титановий циліндр. Величина розбіжності між послідовними вимірами світіння титанового циліндра для площі не перевищує 5 % ($\Delta S \leq 5$ %), а для інтенсивності – 2 % ($\Delta I \leq 2$ %).

Обробка даних на сервері дозволила вирішити кілька важливих практичних завдань:

– можливість постійної модифікації програмного забезпечення та оперативний апгрейд для користувачів;

– проведення автоматичної метрологічної повірки та налаштування приладів у режим віддаленого доступу;

– захист програмного забезпечення та баз даних від вірусів та атак;

– можливість роботи на різних комп’ютерних платформах;

– зберігання всіх користувачів на сервері;

– можливість обміну даними між різними користувачами;

– можливість проведення мета-аналізу результатів різних користувачів.

Розглянуті вище пристрої потребують доставки об’єкта дослідження або його частини безпосередньо на робочу поверхню сенсора. Система газорозрядної візуалізації «Стример» має модульний принцип побудови, малі габаритгі розміри, наявність виносних електронно-оптичних блоків (рис. 4) [14]. Є блок горизонтального розташування з діаметром робочого

сенсора 60 мм та три блоки типу «пістолет» з діаметрами сенсорів 10, 30 і 60 мм. Особливістю побудови системи «Стример» є розміщення високовольтного трансформатора у виносному блоці, що дозволяє збільшити довжину кабелю з'єднання блока з генератором до 1,5 – 2 м [16].

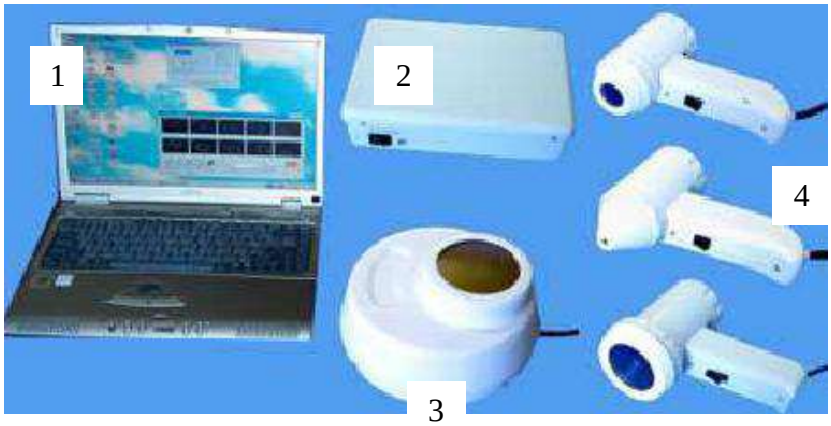


Рис. 4. Система ГРВ «Стример»: 1 – ноутбук із відкритою панеллю управління програми «GDV – Grabber 20», 2 – генератор з програмним керуванням, 3 – блок горизонтального розташування, 4 – три блоки типу «пістолет» із різними діаметрами сенсорів (10, 30, 60 мм)

Горизонтальний блок призначений для традиційної зйомки зображень пальців рук, зразків рідин, мінералів, зерен рослин та інших об'єктів. Блок «пістолет» 60 мм призначається для зйомки зображень пальців рук дорослих пацієнтів із обмеженою рухливістю суглобів. Блок «пістолет» 30 мм призначений для зйомки зображення пальців дітей дошкільного віку. Блок «пістолет» 10 мм призначений для візуалізації акупунктурних точок та впливу на них. Блоки типу «пістолет» можуть бути також використані для візуалізації ефективності терапевтичного впливу на окремі ділянки шкірного покриву людини.

В роботі [16] розглянуто ГРВ пристрій для експрес-оцінювання стану біологічного об'єкта. Технічною особливістю цього засобу є зосередження в корпусі діагностичного щупа основних блоків: сенсора, оптичної системи, цифрової камери, високовольтного генератора, елементів керування. Передача інформації та живлення щупа відбувається через кабель під'єднаний до ПК. Діагностика відбувається переміщенням сенсора щупа по поверхні об'єкта (наприклад – шкірному покриву людини).

Світіння газового розряду в повітрі довкола тестового об'єкта було застосовано для оцінювання загального екологічного стану оточуючого середовища. Структурна схема експериментальної установки приведена на рис. 5 [4].

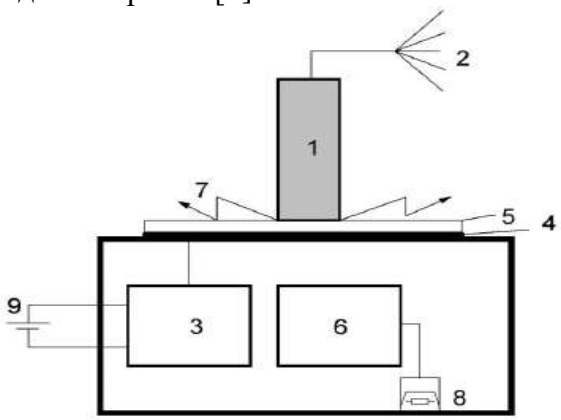


Рис. 5. Схема експериментальної установки: 1 – титановий циліндр (тестовий об'єкт); 2 – антена ГРВ-сенсора; 3 – генератор високовольтних імпульсів; 4 – прозорий струмопровідний шар; 5 – прозорий кварцовий діелектрик; 6 – відеоперетворювач; 7 – газовий розряд; 8 – USB-накопичувач; 9 – автономне джерело живлення

Розряд розвивається за рахунок струмів зміщення між антеною та заземленими або провідними об’єктами в навколишньому просторі. Залежно від наявності полів різної природи в навколишньому середовищі, хімічного складу повітря та стану провідних об’єктів (до яких належать також люди), змінюються умови поширення електромагнітної хвилі в просторі, внаслідок цього струми в системі перерозподіляються, отже, змінюються параметри світіння. Таким чином, дана експериментальна система може реагувати на зміну електричної ємності навколишнього простору та об'єктів, що знаходяться в ній. Для забезпечення розряду використовується послідовність імпульсів напруги амплітудою до – 7 кВ, тривалістю – 10 мкс із частотою – 1 кГц, пачками – 0,5 с, кожні – 5…10 с.

В роботах [9, 17] запропоновано установку для дослідження рідинно-фазних біологічних об’єктів (РФБО). Відмінністю цього технічного рішення є визначення концентрацій хімічних сполук в досліджуваних рідинах по спектральним характеристикам світіння газового розряду. Структурна схема установки показана на рис. 6.

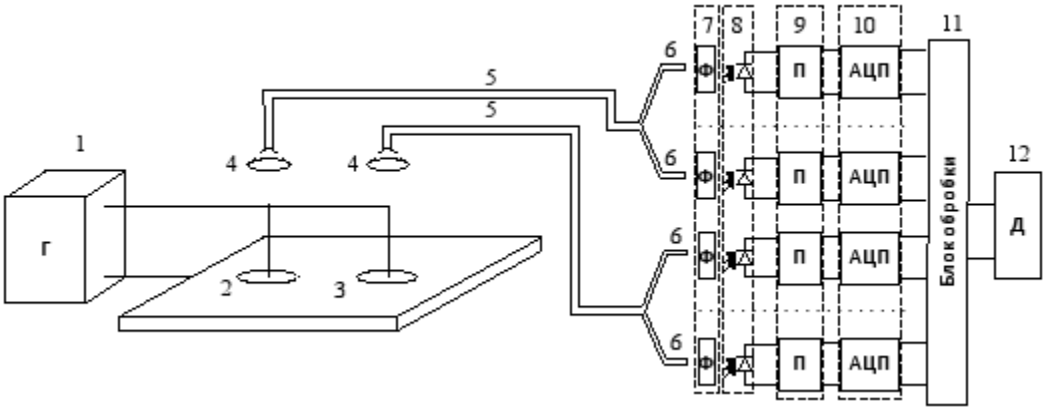


Рис. 6. Структурна схема установки дослідження спектральних складових випромінювання РФБО на основі ГРВ, де 1 – генератор високовольтної високочастотної напруги; 2 – комірка з досліджуваною рідиною; 3 – комірка зі зразковою рідиною; 4 – оптичні лінзи; 5 – багатоканальні світловоди; 6 – виходи каналів багатоканальних світловодів; 7 – світлофільтри; 8 – фотодіоди; 9 – підсилювачі; 10 – АЦП; 11 – блок обробки; 12 – дисплей [17]

Ще одним принциповим рішенням в цій установці є порівняння світіння зразкової рідини і досліджуваної. Це дає можливість частково компенсувати вплив оточуючого середовища і нестабільності вимірювального тракту на результати спектрального аналізу.

Попередньо розглянуті засоби ГРВ були переважно призначені для досліджень об’єктів біологічного походження. Але відомі також застосування методу газорозрядної візуалізації в технічній інженерії. Автором роботи [18] розроблені і впроваджені у практику пристрої (на основі ефекту Кірліан) для неруйнівного контролю матеріалів та конструкцій, що не піддаються контролю традиційними методами. В цих пристроях використовувалась реєстрація зображень на фотоматеріали, що знижувало оперативність технічної діагностики. В роботі [19] пропонується лабораторний стенд для контролю дефектів кремнієвих платин методом газорозрядної візуалізації. При подачі на електроди касети (сенсору) високочастотних імпульсів високої напруги виникає розряд, світіння якого відображає топологічні особливості об’єкту дослідження. Густина та яскравість розрядних стримерів завжди збільшена в місцях дефектів та загострень поверхні об’єкту. Оптико-електронна система у складі об’єктиву, оптичних кілець та телевізійної камери типу Novus NVC-130BH фіксує світіння розряду та перетворює його у телевізійний композитний сигнал (аналогового типу). Телевізійний сигнал надходить до входу адаптера типу Aver TV встановленого у персональний комп’ютер. Монітор ПК забезпечує виведення на екран зображення та результатів його оброблення в числовій та графічній формі.

Таким чином, можна стверджувати, що структура технічних засобів ГРВ змінювалась в залежності від удосконалення пристроїв реєстрації світіння розряду в атмосферному повітрі та потужностей комп'ютерної обробки зображень. Але електрофізичні показники виникнення розряду (амплітуда напруги, тривалість і частота імпульсів) мають експериментально перевірені інтервали.

Напрями застосування засобів газорозрядної візуалізації

Програмно-апаратні ГРВ комплекси та окремі засоби знайшли практичне застосування у наступних основних областях: медицина і біологія, спорт, дослідження рідин і матеріалів, спеціальні напрями [4].

Використання ГРВ-засобів в медицині дозволяє здійснювати: аналіз психологічного та психофізіологічного стану особистості; аналіз вегетативного статусу організму та окремих функціональних систем; моніторинг реакцій організму в процесі терапії; оцінку ймовірності наявності системних порушень роботи органів; оцінку наявності змінених станів; оцінку небезпеки алергенів за параметрами ГРВ свічення зразків крові.

Можна виділити наступні переваги застосування методу ГРВ в медико-біологічній практиці [3]:

– можливість скрінінгу і моніторингу ентропійно-енергетичного гомеостазису всього організму та його окремих систем;

– об'єктивність інформації: незалежність від бажання і досвіду конкретного користувача;

– неінвазивність, безпечність і повна стерильність, зняття інформації тільки з кінцівок пацієнта;

– можливість слідкування за розвитком процесів у часі, співставлення структурних, функціональних та часових процесів в організмі;

– методична простота і зручність: відсутність якихось особливих вимог до приміщення, умов навколишнього середовища;

– використання сучасних методів нелінійної математики для обробки фрактальних зображень і вибору інформації про стан пацієнта;

– наочність та можливість інтерпретації отриманих результатів, зручність їх зберігання та обробки.

Для спортивної діяльності ГРВ-засоби дозволяють: проводити динамічне визначення психофізичного потенціалу спортсмена з метою оперативного контролю рівня функціональних резервів та якості здоров'я в ході навчально-тренувального процесу; забезпечувати експрес діагностику стресостійкості та якості процесів ментальної та психоенергетичної мобілізації з метою прогнозу успішності змагальної діяльності; проводити диференційовану рейтингову оцінку психофізичного потенціалу обстежуваного контингенту спортсменів з метою відбору, підготовки та своєчасної корекції тренувального процесу.

Для дослідження рідиннофозних об'єктів ГРВ-засоби використовують за такими напрямами: діагностика по зразкам біологічних рідин (кров, сеча, слина, мокроти, сльзи, спинномозкова рідина та інші) [6]; виявлення відмінності натуральних та синтетичних масел; оцінка якості косметичних препаратів; дослідження гомеопатичних препаратів.

Неруйнівна дефектоскопія матеріалів і конструкцій засобами ГРВ дозволяє оцінювати структурні неоднорідності діелектричних, напівпровідникових, композитних сполук [21]; властивості зразків дорогоцінного каміння [4]; герметичність зварних механічних конструкцій [18].

Такий широкий спектр застосувань засобів ГРВ обумовлений високою чутливістю до змін емісійних параметрів досліджуваного об'єкта, поміщеного в електромагнітне поле високої напруженості, та застосовуваних методів обробки інформації з урахуванням сучасних теоретичних підходів та методів штучного інтелекту. Прикладами спеціальних напрямів

використання засобів ГРВ є дослідження геоактивних зон та їх впливу на людину, загального екологічного стану довкілля, діагностування стресових станів людини.

Наукові публікації останнього десятиріччя, що пов’язані з методами та засобами газорозрядної візуалізації, були спрямовані на доказовість їх застосування в медичній діагностиці. Автори роботи [20] досліджували методом ГРВ вплив енергетичного стану людини на структуру зразків води. В роботах [21, 22] доведена висока кореляція діагностичних показників ГРВ з показниками ЕКГ, ЕЕГ та акупунктурної діагностики. Дослідження виконані з застосуванням пристрою «GDV Chamber».

У роботі [23] наведені результати систематичного пошуку у дослідницьких базах даних, таких як Google Scholar, PubMed і PsychINFO використання ГРВ за період 2000 – 2022 рр. в англомовних виданнях. Зазначено, що ГРВ має потенціал для ранньої діагностики порушень роботи ендокринної та імунної систем. Рекомендується застосування засобів ГРВ для оцінювання самопочуття здорових людей, спостерігання за ефектами впливу процедур, таких як йога, медитація, акупунктура, цигун, музикотерапія, масаж на енергетику систем людини.

Цікавими є початкові результати досліджень за проєктом «Функціональні зв’язки між параметрами точок акупунктури та нейро-ендокринно-імунною мережею» наведені в роботі [24]. За допомогою ГРВ-засобів зареєстровано зв'язок електропровідності точок акупунктури з особливостями нейро-ендокринно-імунної мережі організму пацієнта. Ці результати є спробою узгодити парадигми західної та східної медицини.

Ще одне перспективне використання ГРВ-засобів полягає в реєстрації біофотонного випромінювання живих організмів, відмічене в роботі [25]. Дійсно, в живих організмах на клітинному рівні відбувається безліч біофізичних процесів в результаті обміну речовин. Ці процеси супроводжуються надслабким широкосмуговим електромагнітним випромінюванням. Фотонна компонента сягає від інфрачервоного до ультрафіолетового та гамма-випромінювань. Фотони цих випромінювань можуть впливати як на виникнення газового розряду довкола об’єкта дослідження, так і на спектральні характеристики його світіння.

Варіанти технічних рішень для засобів ГРВ діагностики

З аналізу відомих вітчизняних та зарубіжних публікацій щодо будови та застосування засобів діагностики на основі методу газорозрядної візуалізації можна визначити перспективи розвитку цього наукового напряму [4].

Розглянемо пріоритетні шляхи модернізації основних компонентів, вузлів та елементів ланцюга газорозрядного процесу, а також етапи подальшої обробки візуальних зображень.

1. Побудова генераторів високої напруги – з використанням L- та /або C-накопичувачів енергії, п’єзоелектричних високовольтних генераторів, трибоелектричних генераторів, електретних генераторів, з перемикачами струму механічного, електричного чи іншого типу.

1.1. Одноімпульсний режим збудження газового розряду.

Використання джерела постійної напруги. Робота на перехідних процесах у момент увімкнення/вимкнення напруги, характеристики визначаються швидкістю перехідних процесів (похідна dU/dt). Параметри, що варіюються: амплітуда піку, величина градієнта наростання/спаду імпульсу, інтегральна потужність імпульсу.

1.2. Режим багатоімпульсного збудження з регульованим числом імпульсів в серії.

Варіюються параметри: амплітуда, шпаруватість імпульсів, їх тривалість, кількість і форма імпульсів.

1.3. Стимуляція розряду періодичним сигналом синусоїдальної чи іншої періодичної форми імпульсу (крім прямокутної) з постійною або змінною в часі амплітудою. Параметри регулювання: амплітуда, частота і форма імпульсів.

1.4. Наявність модуляції, її тип – амплітудна, частотна, амплітудно-імпульсна, частотно-імпульсна, широтно-імпульсна, змішані форми модуляції.

1.5. Режим збудження газового розряду з програмним цифровим формуванням високовольтної напруги, при синхронізації сигналами біологічного зворотного зв'язку.

2. Вимірювальні сенсори.

Прозора для спектру випромінювання газового розряду діелектрична основа сенсору, з товщиною та площою оптимальними для забезпечення необхідної напруженості електричного поля та незначних механічних деформацій при розміщені об'єкта дослідження. Наявність прозорого провідникового електрода (напилення на діелектричну основу). Двоелектродне або псевдоодноелектродне підключення до джерела високовольтної напруги.

3. Відеоперетворювачі та способи знімання інформації.

Твердотільні фотоприймачі та перетворювачі випромінювань: електронно-оптичні та інші перетворювачі, твердотільні детектори, фотоелектронні помножувачі, помножувачі вторинної електронної емісії, ПЗЗ-матриці, світловоди, світлофільтри, дифракційні ґрати, аналізатори спектрального складу випромінювання.

Формування плоских або просторових зображень світіння розряду, контактна або дистанційна зйомка.

4. Способи обробки та інтерпретації експериментального матеріалу. Поєднання з іншими методами діагностики, калібрування, атестація методу та засобів ГРВ. Приведення у відповідність показників отриманого зображення до характеристик об'єкта, що діагностується. Використання методів математичної статистики, оцінювання достовірності, відтворюваності результатів. Встановлення кореляції відгуку з фактором впливу.

5. Комплексування з іншими методами інструментального дослідження. Врахування впливу на процеси ГРВ супутніх фізичних процесів: електромагнітнних випромінюваннь, їх амплітудно-частотних показників; геомагнітних факторів, фонового рівня іонізуючих випромінювань, динаміки змін атмосферного тиску та газових складових повітря.

6. Поєднання з методами та приладами корекції функціонального та фізіологічного стану людини для створення програмно-апаратних комплексів для професійного та домашнього застосування.

7. Створення програмних додатків для цільової ГРВ-діагностики адаптованих для поширених програмних середовищ персональних комп'ютерів, планшетів, мобільних телефонів.

Наведений перелік напрямів є перспективною базою розвитку методів газорозрядної візуалізації. Основним є широкий спектр дослідницьких і практичних задач, які можна вирішувати з використанням ГРВ-засобів.

Висновки

1. Аналіз фізичних та технічних положень газорозрядної візуалізації показав, що при відносній сталості фізичних закономірностей виникнення розряду режими вимірювання інформативних показників зображень в ГРВ-засобах суттєво відрізняються (амплітуда, частота, тривалість, форма змінної напруги).

2. Умови виникнення розряду є чутливими до електрофізичних та емісійних властивостей об'єкту дослідження, але одночасно і до змін фізико-хімічних показників атмосферного повітря.

3. Практично кожна конструкція відрізняється видом сенсору, способом та пристроєм реєстрації первинного випромінювання, алгоритмами оброблення та відтворення діагностичного зображення.

4. В медико-біологічній галузі є найбільша кількість практичних застосуваннь ГРВ-засобів різних конструкцій. Для коректного порівняння результатів діагностики, отриманих різними дослідниками, потрібна стандартизація технічних характеристик ГРВ-засобів та алгоритмів отримання інформативних показників.

5. В технічній діагностиці методи та засоби ГРВ не набули поширення. На базі відпрацьованих рішень для біологічних об'єктів є перспективним використання ГРВ-засобів для контролю складу технічних рідин, неруйнівної дефектоскопії матеріалів і конструкцій.

6. Для використання сучасних інформаційних технологій перевага повинна надаватись побудові апаратно-програмних комплексів ГРВ-діагностики на основі якісних зразків цифрових відеокамер.

7. При проведенні ГРВ-діагностики рекомендується додавати інформацію про характеристики оточуючого атмосферного повітря (температура, вологість, тиск).

Запропоновані технічні рішення потребують експериментальної перевірки. Необхідні порівняльні дослідження діагностичних висновків, отриманих на основі ГРВ-засобів з сертифікованими діагностичними засобами.

Список літератури:

1. Медична діагностика. Технічна діагностика. https://uk.wikipedia.org/wiki
2. Потяженко М. М. Інноваційні методики об’єктивного обстеження з комп’ютерним тестуванням в еволюції реєстрації фізичних феноменів лікарем терапевтичного профілю: історія, реальність, перспективи / М. М. Потяженко, Г. В. Невойт // Медична інформатика та інженерія. 2018. № 4. С. 57–65. DOI: https://doi.org/10.11603/mie.1996-1960.2018.4.9894
3. Oliinyk V., Babakov M., Lomonosov Y., Oliinyk V., Zinchenko O. Modernization of gas discharge visualization for application in medical diagnostics // Technology Audit and Production Reserves. 2020. 4 (1 (66)). P. 21–29. doi: http://doi.org/10.15587/2706-5448.2022.263397
4. Korotkov K.G. Advances in Diagnosis and Monitoring of the human quantum informational field with GDV technique // ParaDigm 2001: Consciousness and Paranormal Phenomena, 2008. https://www.researchgate.net/publication/228831552_Advances_in_Diagnosis_and_Monitoring_of_the_human_quantum_informational_field_with_GDV_technique
5. Method for Determining the Condition of a Biological Object and Device for Making Same / United States Patent (10) Patent No.: US 8,321,010 B2 // Korotkov et al. Date of Patent: Nov. 27, 2012
6. Білинський Й. Й. Методи і засоби газорозрядної візуалізації для аналізу рідиннофазних біооб’єктів : монографія / Й. Й. Білинський, О. А. Павлюк. Вінниця : ВНТУ, 2016. 120 с.
7. Application of Electrophoton Capture (EPC) Analysis Based on Gas Discharge Visualization (GDV) Technique in Medicine: A Systematic Review / K.G. Korotkov, P. Matravers, D.V. Orlov [et al.] // J of Alternative and Complementary Medicine. 2010. Vol. 16. № 1. P. 13–25.
8. Кухтин В. В. Апаратна реалізація і діагностичні можливості методу газорозрядної візуалізації / В. В. Кухтин, П. В. Петельський, Ю. В. Чепурний // Вісник Нац. техн. ун-ту "КПІ". Сер. Радіотехніка. Радіоапаратобудування. 2010. № 42. С. 139–144.
9. Білинський Й. Й. Дослідження спектрів випромінювання рідиннофазних об’єктів при газорозрядній візуалізації / Й. Й. Білинський, О. А. Павлюк, С. В. Юкиш // Технологічний аудит та резерви виробництва. 2014. № 3. С. 58–61.
10. Глухова Н. В. Інформаційна технологія для аналізу кольорових зображень газорозрядного випромінювання / Н. В. Глухова, Л. А. Пісоцька // Перспективні технології та прилади. Луцьк, 2018. Вип. №12. С. 48–52.
11. Reizer Ju.P. Gas Discharge Physics. Springer, 2001. 449 p.
12. Glukhova N.V. Application of wavelets transform for analysis of images of gas-discharge radiation of water N.V. Glukhova, L.A. Pesotskaya, N.G. Kuchuk, J.N. Kharlamova // Системи обробки інформації. 2016. № 2(139). С. 179-185.
13. Xanadu C. Analysis of kirlian images: feature extraction and segmentation / Xanadu C. Halkias, Petros Maragos // IEEE (School of Electrical & Computing Engineering, National Technical University of Athens, Zografou 15773 Athens, Greece). 2004. P. 4.
14. Retrieved from: http://xn--annciowww-68a.bio-well.com/assets/files/kirlian_devices.pdf
15. DEVICE FOR DETERMINING THE STATE OF A BIOLOGICAL SUBJECT United States Patent Application Publication (10) Pub. No.: US 2010/0106424 A1 Korotkov et al.(43) Pub. Date:Apr. 29, 2010 Konstantin Georgievich Korotkov, Saint-Petersburg (RU); Svetlana Alexandrovna Korotkina, Saint Petersburg (RU); Ramiz Ragim-Ogly Jusubov, Saint Petersburg (RU).
16. Пат. 70099 Україна, МПК (2012.01) G03B 41/00, G03G 17/00. Пристрій для експрес-оцінки стану біологічного об’єкта / Семенець В. В., Подпружников П. М.., Левенець О. С. № u 2011 13774; Заявл. 23.11.2011; Опубл. 25.05.2012, Бюл. № 10. 4 с.
17. Kulyk Y. A., Knysh B. P., Maslii R. V., Kvyetnyy R. N., Shcherba V. V., & Kulyk A. I. (2021). METHOD AND GAS DISCHARGE VISUALIZATION TOOL FOR ANALYZING LIQUID-PHASE BIOLOGICAL OBJECTS // Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska 2021. №11(3). P. 22–29. https://doi.org/10.35784/iapgos.2709
18. Романий С. Ф. Неразрушающий контроль материалов по методу Кирлиана / С. Ф. Романий, Э. Д. Черный. Днепропетровск : Изд-во ДГУ, 1991. 144 с.

19. Швайко В. В. Лабораторний стенд для збудження і контролю дефектів в коронному розряді / В. В.Швайко В.В., К. М. Божко // Зб. пр. 14-ї Всеукр. наук.-практ. конф. студентів, аспірантів та молодих вчених у приладобудуванні. 4-5 грудня 2018 р., м. Київ. С. 210–213.

20. Babelyuk V. Y., Dobrovolskiy Y. G., Pidkamin L. I., Popovych I. L., Ushenko Yu. A. Usage of a gas-discharge visualization for an investigation of a human internal energy // Proc. SPIE 11369, Fourteenth International Conference on Correlation Optics, 1136929 (6 February 2020); doi10.1117/12.2553951

21. Babelyuk V., Tserkovniuk R., Babelyuk N., Zukow X., Ruzhylo S., Dubkova G., Korolyshyn T., Hubyts'kyi V., Kikhtan V., Gozhenko A., & Popovych I. The parameters of gas discharge visualization (biophotonics) correlated with parameters of acupuncture points, EEG, HRV and hormones // Journal of Education, Health and Sport. 2021. №11(12). P. 359–373. https://apcz.umk.pl/JEHS/article/view/JEHS.2021.11.12.030

22. Babelyuk V. Y., Popovych I. L., Gozhenko A. I., Dubkova G. I., Kozyavkina O. V., Korolyshyn T. A., Babelyuk N. V., Kovbanyuk M. M., Fihura O. A., Dobrovolskyi Y. G., Zukow W., & Yanchij R. I. (2023). Gas Discharge Vizualization (Electrophotonic Imaging, Kirlianography). Theoretical and Applied Aspects. Feniks. https://doi.org/10.5281/zenodo.7535880

23. Bista S., Jasti N., Bhargav H., Sinha S., Gupta S., Ramarao P., Chaturvedi S., & Gangadhar B. (2022). Applications of Gas Discharge Visualization Imaging in Health and Disease: A Systematic Review. Alternative Therapies in Health and Medicine, AT6764. PMID: 35648690.

24. GOZHENKO Anatoliy, ZANTARAIA Toto and ZUKOW Walery. „Falling values": artifacts or source of unique information? Drastically low electrical conductivity of acupuncture points is accompanied by significant deviations of EEG, HRV, immunity, metabolism and GDV parameters // Quality in Sport. 2024. №17. P.51006. eISSN 2450–3118. https://dx.doi.org/10.12775/QS.2024.17.004 https://apcz.umk.pl/QS/article/view/51006

25. Nevoit G., Bumblyte I., Korpan A., Minser O., Potyazhenko M., Iliev M., Vainoras A., & Ignatov I. The Biophoton Emission in Biotechnological and Chemical Research: from Meta-Epistemology and Meaning to Experiment. Part 1 // Ukrainian Journal of Physics, 2024. №69(3). P. 190. https://doi.org/10.15407/ujpe69.3.190

Надійшла до редколегії 14.09.2024

Відомості про авторів:

Зінченко Олександр Миколайович – Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», аспірант, асистент кафедри радіоелектронних та біомедичних комп'ютеризованих засобів і технологій, Україна; e-mail: a.zinchenko@khai.edu; ORCID: https://orcid.org/0000-0001-5651-8931

Олійник Володимир Петрович – канд. техн. наук, доцент, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», професор кафедри радіоелектронних та біомедичних комп'ютеризованих засобів і технологій, Україна; e-mail: v.oliinyk@khai.edu; ORCID: https://orcid.org/0000-0002-7899-1591

Подпружников Петро Михайлович – Харківський національний університет радіоелектроніки, провідний інженер служби доступності до можливостей навчання «ХНУРЕ без обмежень», Україна; e-mail: petro.podpruzhnykov@nure.ua; ORCID: https://orcid.org/0000-0001-6714-81271

УДК 537.874 **DOI:10.30837/rt.2024.3.218.12**

Є.Є. ДЕМИДЕНКО, В.В. НОВИЦЬКИЙ, Є.М. ОДАРЕНКО, д-р фіз.-мат. наук,
О.О. ШМАТЬКО, д-р фіз.-мат. наук

ХАРАКТЕРИСТИКИ КЕРОВАНОГО БРЕГГІВСЬКОГО ХВИЛЕВОДУ З ГІРОТРОПНИМИ ОБОЛОНКАМИ

Вступ

Фотонно-кристалічні структури інтенсивно досліджуються як в експериментальному, так і в теоретичному аспекті завдяки їх унікальним фізичним властивостям, які дозволяють створювати ефективні функціональні пристрої в різних частотних діапазонах [1, 2]. Терміном «фотонний кристал» зараз фактично позначаються періодичні структури різної розмірності, яка визначається кількістю просторових напрямків, уздовж яких спостерігається періодична зміна параметрів цих структур (зазвичай, показника заломлення).

Базовими пристроями, на основі яких будується весь спектр різноманітних систем, є фотонно-кристалічні хвилеводи та резонатори. Незалежно від розмірності, принцип дії цих пристроїв переважно базується на використанні частотних заборонених зон фотонних кристалів. В межах цих зон розповсюдження хвиль в будь-якому напрямку є неможливим, що дає можливість ефективної локалізації електромагнітної енергії в різноманітних дефектах періодичності фотонних кристалів. Зазвичай фотонно-кристалічні хвилеводи будуються на основі лінійних дефектів періодичності, а фотонно-кристалічні резонатори – на основі локальних дефектів.

Керування характеристиками фотонно-кристалічних пристроїв за допомогою зовнішніх електричних та магнітних полів є важливою практичною проблемою, оскільки дозволяє значно розширити перелік цих пристроїв та забезпечити гнучкість в їх застосуванні. Прикладами керованих фотонно-кристалічних пристроїв є модулятори [3], циркулятори [4], фільтри [5 – 7], логічні елементи [8] тощо. В фотонно-кристалічних пристроях, що керуються електричним полем, переважно застосовуються рідкокристалічні елементи [9, 10]. Для керування магнітним полем застосовуються феритові або плазмоподібні елементи [11, 12], які чутливі до зовнішніх магнітних полів. Періодичні структури, що містять такі магніточутливі елементи, називаються магнітофотонними кристалами.

В даній роботі розглядається планарний хвилевід, оболонки якого складаються з одновимірного магнітофотонного кристалу, в якому діелектричні шари чергуються з гіротропними (феритовими). Така система може бути ідентифікована як Бреггівський відбивний хвилевід [13 – 15], оскільки локалізація електромагнітної енергії в ньому реалізується завдяки практично повному відбиттю хвиль від періодичних багатошарових оболонок.

На основі чисельного розв’язання електродинамічної задачі методом скінченних елементів аналізуються закономірності впливу індукції зовнішнього магнітного поля на дисперсійні характеристики магнітофотонного кристалу та спектральні характеристики планарного Бреггівського хвилеводу.

Модель Бреггівського хвилеводу

Розглянемо модель планарного Бреггівського хвилеводу, оболонка якого містить гіротропні шари. Схема структури із системою координат представлена на рис. 1.

Пустотілий хвилеводний канал з шириною d розташований між двома багатошаровими оболонками, які представляють собою Бреггівські відбивачі з періодом $L = a + b$. На періоді оболонки розташовані два шари, один з яких є діелектричним та ізотропним ($\varepsilon_1 = 2.1,\ \mu_1 = 1$) і має товщину b, а інший (з товщиною a) – гіротропним з діелектричною проникністю $\varepsilon_2 = 12.9$ та магнітною проникністю, яка є тензором.

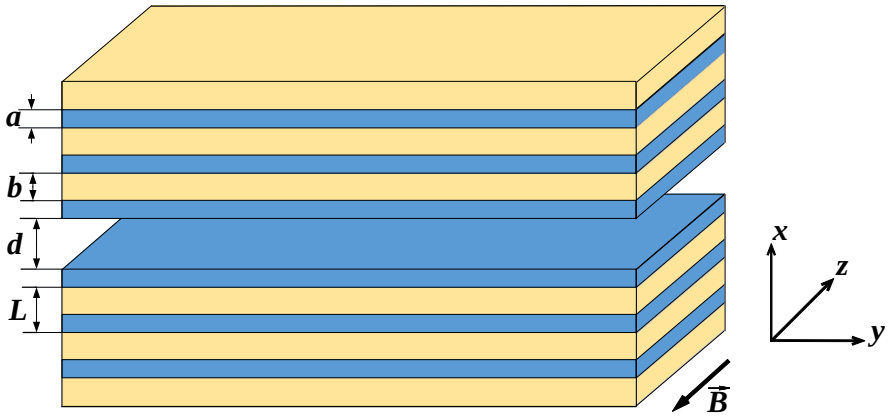


Рис. 1. Схема Бреггівського хвилеводу

Такі матеріальні параметри характерні для двоокису кремнію та фериту відповідно. У випадку, коли зовнішнє статичне магнітне поле спрямоване уздовж від'ємного напрямку координатної осі *Oz* (рис. 1), магнітна проникність фериту може бути представлена в такому вигляді [16]:

$$\overleftrightarrow{\mu}_2 = \begin{Vmatrix} \mu_2 & -i\mu_a & 0 \\ i\mu_a & \mu_2 & 0 \\ 0 & 0 & 1 \end{Vmatrix}, \qquad (1)$$

де

$$\mu_2 = 1 - \frac{\omega_c \omega_m}{\omega^2 - \omega_c^2}, \quad \mu_a = -\frac{\omega \omega_m}{\omega^2 - \omega_c^2},$$

$\omega_c = -eB_0 / m$ – циклотронна або гіромагнітна частота; e та m – від'ємний заряд та маса електрону; $\omega_m = \gamma M_s$; γ – гіромагнітне відношення; $M_s = 2.39 \cdot 10^5$ А/м – намагніченість насичення фериту; B_0 – індукція зовнішнього магнітного поля.

Наявність гіротропних шарів в періодичній багатошаровій структурі дає можливість керування її електродинамічними характеристиками через зміну зовнішнього магнітного поля [17 – 21]. В першу чергу це стосується дисперсійних характеристик, які визначають частотні заборонені зони, в яких неможливе розповсюдження електромагнітних хвиль в періодичних структурах.

Частотні залежності діагональних та недіагональних компонентів тензору магнітної проникності (1), побудовані для різних значень індукції зовнішнього магнітного поля, представлені на рис. 2. Цілком природньо, що зміна значення B_0 супроводжується зсувом феромагнітного резонансу уздовж частотної осі. Таким чином, можна здійснювати керування матеріальними параметрами гіромагнітних шарів і, відповідно, фізичними властивостями періодичної багатошарової структури в цілому. Слід відзначити, що значення елементів тензору магнітної проникності можуть мати як додатні, так і від'ємні значення. Це призводить до реалізації різноманітних власних режимів структури, пов'язаних із поверхневими та об'ємними хвилями [22].

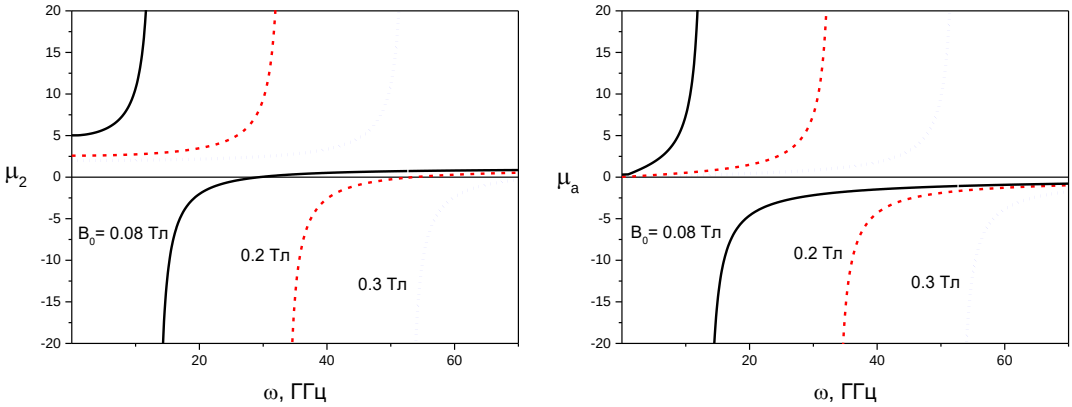


Рис. 2. Частотні залежності компонентів тензору магнітної проникності

Оскільки в даному випадку розглядається двовимірна модель $\left(\partial/\partial z=0\right)$, то при розв'язанні електродинамічної задачі можна виділити дві незалежні поляризації (TM та TE). Виходячи з орієнтації зовнішнього магнітного поля, яка відповідає так званій схемі Войгта [23], та з тензорного характеру магнітної проникності гіротропних шарів періодичної структури, розглядається лише TM поляризація, яка характеризується наявністю лише однієї компоненти електричного поля E_z. Рівняння Гельмгольца в гіротропних шарах для цієї поляризації можна записати таким чином:

$$\frac{\partial^2 \dot{E}_z}{\partial x^2}+\frac{\partial^2 \dot{E}_z}{\partial y^2}+k^2\mu_p\dot{E}_z=0, \tag{2}$$

де $\mu_p=\mu_2\left(1-\mu_a^2/\mu_2^2\right)$; $k=\omega/c$; c – швидкість світла у вакуумі. При розгляді TE поляризації фізичні властивості гіротропного шару відповідають звичайному діелектрику з проникністю $\varepsilon_2=12.9$. У випадку магнітофотонного кристалу з гіроелектричними шарами (наприклад, плазмовими), які характеризуються тензорною діелектричною проникністю, ситуація зворотна – для TM поляризації реалізується діелектричний фотонний кристал. Це цілком закономірний результат, якщо взяти до уваги принцип переставної подвійності однорідних рівнянь Максвелла.

Дисперсійні характеристики одновимірного магнітофотонного кристалу

Для визначення дисперсійних характеристик періодичної багатошарової структури з гіротропними шарами, яка використовується як оболонка пустотілого каналу Бреггівського хвилеводу, використовувалася модель на основі методу скінченних елементів. Ця модель дозволяє враховувати частотну залежність елементів тензору магнітної проникності гіротропних шарів. Для визначення заборонених зон періодичної структури найбільш зручним варіантом дисперсійної характеристики є залежність частоти від хвильового вектору, спрямованого нормально до границь розділу шарів структури. Відповідно до системи координат на рис. 1, це компонента k_x. Результати розрахунків дисперсійної діаграми для різних значень індукції зовнішнього магнітного поля представлені на рис. 3. Уздовж осі абсцис відкладені значення нормованого хвильового числа $k_xL/2\pi$, а уздовж осі ординат – нормована частота $\omega L/(2\pi c)$. Дисперсійна діаграма побудована в межах першої зони Бріллюена.

На рис. 3 представлено перші дві дисперсійні криві, між якими існує частотна заборонена зона.

Збільшення індукції магнітного поля B_0 призводить до суттєвих змін ширини та розташування забороненої зони. Вона зміщується у бік більш високих частот. Крім того, з рис. 3 видно, що збільшення ширини забороненої зони відбувається переважно з високочастотного боку. Оскільки робочий частотний діапазон фотонно-кристалічних пристроїв (хвилеводів, резонаторів тощо) знаходиться в межах частотної забороненої зони базових періодичних структур, то можна припустити, що вплив зміни індукції магнітного поля на спектральні характеристики Брегівського хвилеводу буде здійснюватися за тими ж закономірностями.

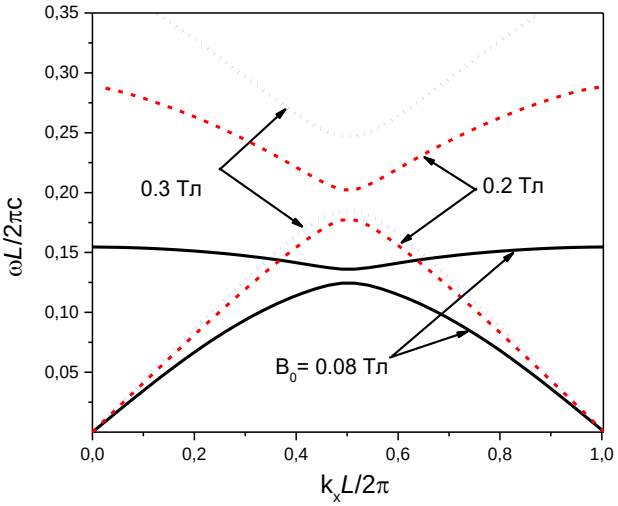


Рис. 3. Дисперсійні діаграми одновимірного магнітофотонного кристалу

Отже, при застосуванні магнітофотонного кристалу для формування оболонки Брегівського хвилеводу за допомогою зміни зовнішнього магнітного поля можна керувати електродинамічними характеристиками цього хвилеводу, зокрема спектральними.

Спектральні характеристики Брегівського хвилеводу

В побудованій моделі Брегівського хвилеводу, схема якого представлена на рис. 1, для чисельних розрахунків використовувалися такі нормовані розміри: a/L = 0.2; d/L = 1.6. В площині y = 0 на краю хвилеводного каналу розташоване джерело випромінювання, яке формується за допомогою активного порту, для якого вказується поляризація випромінювання та вхідна потужність. На протилежному кінці хвилеводу розташований другий порт, який є пасивним.

Спектральна характеристика Брегівського хвилеводу формується як залежність коефіцієнту проходження сигналу T від нормованої частоти. На рис. 4 представлені результати розрахунку цієї характеристики для різних значень індукції зовнішнього магнітного поля. Коефіцієнт проходження представлений у децибелах.

З рисунку видно, що в Брегівському хвилеводі існує смуга пропускання, в межах якої реалізуються досить високі значення коефіцієнту проходження. Збільшення індукції магнітного поля супроводжується зсувом смуги пропускання хвилеводу у бік більш високих частот подібно до того, як це спостерігалося на дисперсійній діаграмі (рис. 3). Крім того, спостерігається збільшення ширини смуги пропускання. Слід відзначити, що частотні характеристики забороненої зони магнітофотонного кристалу на рис. 3 відрізняються від характеристик смуги пропускання Брегівського хвилеводу на рис. 4. Це пояснюється тим, що дисперсійна діаграма магнітофотонного кристалу побудована за умови k_y = 0, тобто за умови відсутності режиму розповсюдження хвиль уздовж шарів структури. Відомо, що в фотонно-кристалічних хвилеводах (наприклад, волоконних) кількість і характеристики заборонених зон змінюються для різних значень поздовжнього хвильового числа (в даному випадку k_y) [24].

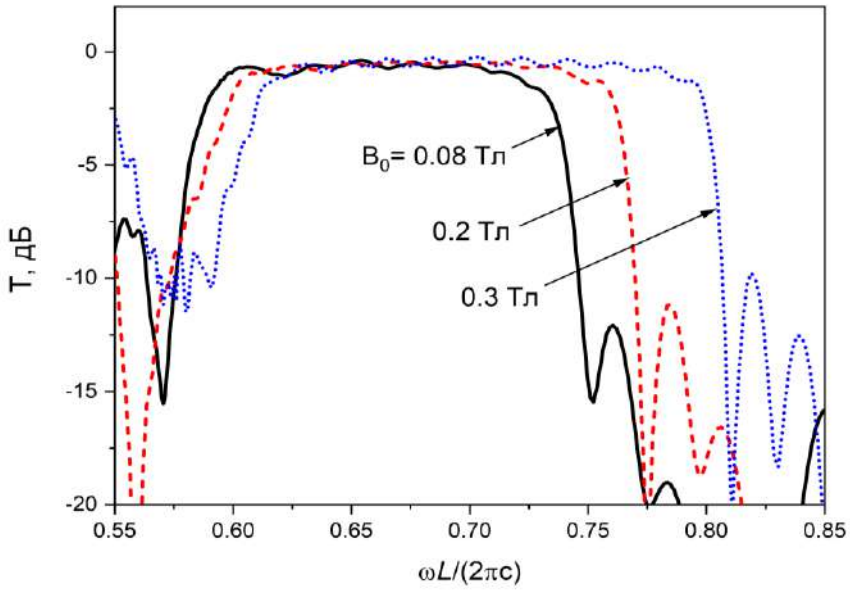


Рис. 4. Спектральні характеристики Бреггівського хвилеводу

В розглянутій моделі хвилеводу не враховуються втрати енергії випромінювання. Такий підхід є цілком обґрунтованим у випадку ефективної локалізації енергії електромагнітного поля в межах пустотілого хвилеводного каналу. Така локалізація енергії підтверджується результатами розрахунків просторового розподілу електричного поля в Бреггівському хвилеводі. Результати цих розрахунків для значень індукції зовнішнього магнітного поля $B_0 = 0.2$ Тл та нормованої частоти $\omega L/(2\pi c) = 0.65$ представлені на рис. 5. З цього рисунку видно, що електричне поле хвилеводної хвилі практично повністю зосереджене в межах пустотілого каналу завдяки дії механізму забороненої зони періодичної шаруватої оболонки.

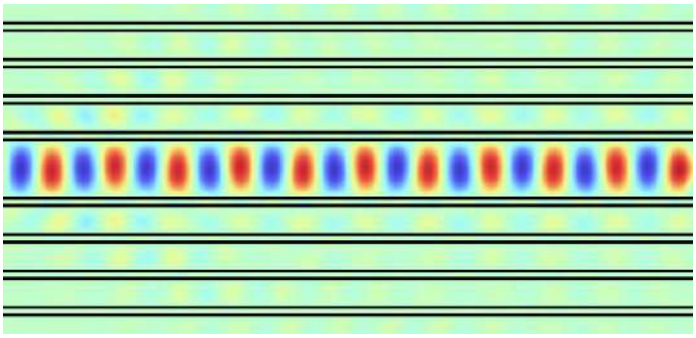


Рис. 5. Просторовий розподіл напруженості електричного поля в Бреггівському хвилеводі

Таким чином, застосування одновимірного магнітофотонного кристалу для формування оболонки Бреггівського хвилеводу дозволяє здійснювати керування його спектральними характеристиками. Отже, на основі таких хвилеводів існує можливість створення фільтрів, ключів, модуляторів та інших функціональних пристроїв (зокрема, формувачів частотних гребінок [25]), керування якими здійснюється зовнішнім магнітним полем.

Висновки

На основі побудованої двовимірної моделі Бреггівського хвилеводу, періодична оболонка якого (магнітофотонний кристал) містить гіротропні шари, чисельно розраховано дисперсійні та спектральні характеристики системи. У якості гіротропного матеріалу використано ферит

з тензорною магнітною проникністю. Результати розрахунків свідчать про можливість керування електродинамічними характеристиками Бреггівського хвилеводу через зміну індукції зовнішнього магнітного поля. Зміна ширини та розташування заборонених зон магнітофотонного кристалу призводить до зсуву смуги пропускання Бреггівського хвилеводу уздовж частотної осі та зміни ширини цієї смуги. Результати розрахунку просторового розподілу електричного поля в хвилеводі свідчать про досить високий ступінь локалізації енергії в пустотілому хвилеводному каналі. Це дозволяє забезпечити достатньо малі втрати в періодичній оболонці хвилеводу з феритовими шарами, які зазвичай суттєво послаблюють потужність електромагнітних сигналів.

Список літератури:

1. Gong Q. and Hu X. (2014) Photonic crystals: Principles and applications. Singapore: Pan Stanford Publishing.
2. Dhanabalan S.S. et al. (2023) Photonic crystal and its applications for Next Generation Systems. Singapore: Springer.
3. Li M. et al. Lithium niobate photonic-crystal electro-optic modulator // Nature Communications. 2020. No 11(1). doi:10.1038/s41467-020-17950-7.
4. Xu B. et al. A terahertz circulator based on Magneto Photonic Crystal Slab // Photonics. 2023. No10(4). P. 360. doi:10.3390/photonics10040360.
5. Tang G. et al. Controllable one-way add-drop filter based on magneto-optical photonic crystal with ring resonator and microcavities // Optics Express. 2022. No 30(16). P. 28762. doi:10.1364/oe.460271.
6. Liu J.-X. et al. (2016) A research of magnetic control ferrite photonic crystal filter // Plasmonics. 2016. No 12(4). P. 971–976. doi:10.1007/s11468-016-0348-5.
7. Wang Y. et al. An electrically controlled tunable photonic crystal filter based on thin-film lithium niobate // Optoelectronics Letters. 2024. No 20(4). P. 200–204. doi:10.1007/s11801-024-3156-8.
8. Pedraza Caballero L.E. and Vilela Neto O.P. A review on Photonic Crystal Logic Gates // Journal of Integrated Circuits and Systems. 2021. No 16(1). P. 1–13. doi:10.29292/jics.v16i1.478.
9. Sung G.-F. et al. Electrically tunable defect-mode wavelengths in a liquid-crystal-in-cavity hybrid structure in the near-infrared range // Materials. 2023. No 16(8). P. 3229. doi:10.3390/ma16083229.
10. Wu C.-Y. et al. Tunable bi-functional photonic device based on one-dimensional photonic crystal infiltrated with a bistable liquid-crystal layer // Optics Express. 2011. No 19(8). P. 7349. doi:10.1364/oe.19.007349.
11. Wu C., Fan J. and Wen G. Magnetically controlled thz three-routing switch based on magnetic photonic crystals // ICEICT 2020 – IEEE 3rd International Conference on Electronic Information and Communication Technology, doi:10.1109/iceict51264.2020.9334253.
12. Dakhlaoui H. et al. Harnessing a dielectric/plasma photonic crystal as an optical microwave filter: Role of defect layers and external magnetic fields // Materials. 2024. No 17(3). P. 559. doi:10.3390/ma17030559.
13. Yeh P. and Yariv A. Bragg Reflection Waveguides // Optics Communications. 1976. No 19(3). P. 427–430. doi:10.1016/0030-4018(76)90115-2.
14. West B.R. and Helmy A.S. Properties of the quarter-wave Bragg reflection waveguide: Theory // Journal of the Optical Society of America B. 2006. No 23(6). P. 1207. doi:10.1364/josab.23.001207.
15. Sashkova Y.V., Odarenko E.N., Shmat'ko A.A. and Afanasieva O.V. Modified Bragg reflection waveguides with binary and ternary claddings // UkrMW 2022 – IEEE 2nd Ukrainian Microwave Week, 2022. doi:10.1109/ukrmw58013.2022.10037133.
16. Pozar D.M. Microwave engineering. New York, 1998. N.Y. : Wiley.
17. Zhang T., Wang G. and Deng D. Switching characteristics of periodically multilayered gyromagnetic metamaterials in waveguide structure // Results in Physics. 2020. No 19. P. 103625. doi: 10.1016/j.rinp.2020.103625.
18. Fu J.-X., Liu R.-J. and Li Z.-Y. Experimental demonstration of tunable gyromagnetic photonic crystals controlled by DC magnetic fields // EPL (Europhysics Letters). 2010. No 89(6). P. 64003. doi:10.1209/0295-5075/89/64003.
19. Kee C.-S. et al. Two-dimensional tunable magnetic photonic crystals // Physical Review B. 2000. No 61(23). P. 15523–15525. doi:10.1103/physrevb.61.15523.
20. Abirami N. and Joseph Wilson K.S. Investigation on photonic band gap of a magneto photonic crystal // Optik. 2020. No 208, 164092. doi: 10.1016/j.ijleo.2019.164092.
21. Ataei E., Sharifian M. and Bidoki N.Z. Magnetized plasma photonic crystals band gap // Journal of Plasma Physics. 2014. No 80(4). P. 581–592. doi:10.1017/s0022377814000105.
22. Shmat'ko A.A., Mizernik V. N. and Odarenko E.N. Surface and bulk modes of magnetophotonic crystals // TCSET 2018 – Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering. 2018. P. 436-440. doi:10.1109/TCSET.2018.8336235.
23. Zhang HF., Liu SB., Kong XK. et al. Photonic band gap of three dimensional magnetized photonic crystal with Voigt configuration // Eur. Phys. J. D. 2013. No 67. P. 169. doi:10.1140/epjd/e2013-40193-3.

24. Lee K.K. et al. (2008) A tale of two limits: Fundamental properties of photonic-crystal fibers // SPIE Proceedings [Preprint]. doi:10.1117/12.778570.

25. Odarenko E., Shmat'ko A., Sashkova Y., Demydenko Y., Novytskyi V. and Shevchenko N. Formation and Tuning of Frequency Comb-Like Signal in Photonic Crystal Coupled-Cavities Waveguides // AICT 2023 – 5th IEEE International Conference on Advanced Information and Communication Technologies, 2023, p. 44–47. doi: 10.1109/AICT61584.2023.10452685.

Надійшла до редколегії 07.09.2024

Відомості про авторів:

Демиденко Євген Євгенович –Харківський національний університет радіоелектроніки, аспірант кафедри фізичних основ електронної техніки; Україна; e-mail: yevhen.demydenko@nure.ua; ORCID: https://orcid.org/0009-0003-5100-3770

Новицький Владислав Віталійович – Харківський національний університет радіоелектроніки, аспірант кафедри фізичних основ електронної техніки; Україна; e-mail: vladyslav.novytskyi@nure.ua; ORCID: https://orcid.org/0009-0004-0346-2259

Одаренко Євген Миколайович – д-р фіз.-мат. наук, професор, Харківський національний університет радіоелектроніки, професор кафедри фізичних основ електронної техніки; Україна; e-mail: yevhen.odarenko@nure.ua; ORCID: http://orcid.org/0000-0001-7656-0440

Шматько Олександр Олександрович – д-р фіз.-мат. наук, професор, Харківський національний університет імені В.Н. Каразіна, професор кафедри фізики надвисоких частот; Україна; e-mail: sh47@ukr.net; ORCID: http://orcid.org/0000-0002-3714-1638

УДК 551.501.7 **DOI:10.30837/rt.2024.3.218.13**

В.В. СЕМЕНЕЦЬ, д-р техн. наук, Т.Е. СТИЦЕНКО, канд. техн. наук,
О.В. ГРИГОР'ЄВ, канд. техн. наук

РОЗРОБКА МОДЕЛІ БІОМЕДИЧНОЇ СИСТЕМИ ЖИТТЄДІЯЛЬНОСТІ ПРИ ВПЛИВІ ЕЛЕКТРОМАГНІТНОГО ВИПРОМІНЮВАННЯ

Вступ

У багатьох системах для оцінки ефективності захисту від електромагнітного випромінювання використовується доза шкідливого впливу як критерій безпеки біомедичної системи. Будь-який шкідливий вплив у технічній системі негативно впливає на життєдіяльність людини, тобто є неминучим злом чи «платою за технічний прогрес» [1–3].

Складні системи біологічного захисту (СБЗ) можна розглядати як системи контролю та управління електромагнітним випромінюванням (ЕМВ). Ці системи можна поділити на автоматичні та ергодичні. В останніх, що отримали в даний час широке поширення, певну роль у контролі та управлінні відіграє інженер-оператор з випробування медичних приладів та систем. Для ергодичних систем необхідна оцінка кількості інформації, яка може бути передана системою інженеру-оператору за допомогою того чи іншого сигналу за певні часові інтервали

У статтях [4–7] описується спосіб оцінки ефективності роботи системи біомедичного захисту з урахуванням оцінки дози шкідливого впливу. При розробці складних систем можливо здійснити математичне моделювання ідеальних та реальних процесів контролю та управління на безперервних і дискретних моделях.

У класичній постановці завдання синтезу системи захисту взагалі не може бути вирішена через відсутність статистичних даних. Тому у статті вона вирішується покроковим способом шляхом дослідження ЕМВ від різних джерел.

Як критерій рівня забезпечення безпеки біомедичної системи життєдіяльності використовується запропонований у роботі [5] критерій – захисна функція, яка дозволяє порівнювати існуючі технічні системи за рівнем їхньої безпеки. Зважаючи на складність та неоднозначність вибору глобального критерію, розглянемо спочатку СБЗ без урахування її зв'язків з іншими елементами системи «біологічний об'єкт-прилади-середовище».

Модель біомедичної системи життєдіяльності

Нехай $D \geq 0$ – доза шкідливого впливу. Індексом «з» далі скрізь позначатиметься значення параметра системи або величини шкідливого фактора, отримане в результаті (і після) дії системи біологічного захисту. Таким чином, $D_з$ – значення D після дії захисту.

Введемо функцію Ф, яка зв'язує ці величини:

$$\Phi(D, V(t), t) = D_з. \tag{1}$$

Назвемо Ф(D, t) захисною функцією системи біологічного захисту.

Очевидно, що $0 \leq \Phi(D, \vec{V}, t) \leq D$ и $\Phi(0, \vec{V}, t) = 0$.

У разі лінійності функції Ф(D, t) по змінній D має місце співвідношення

$$D_з = K_з(\vec{V}, t)D.$$

Функцію

$$K_з(\vec{V}, t) = \hat{O}(1, \vec{V}, t) \tag{2}$$

назвемо коеффіцієнтом захисту.

Ця технічна система захисту (ТСЗ) є структурною, тобто її параметри не залежать від часу. І тут розробка системи зводиться до вибору значень параметрів, які дають мінімум Кз.

Оптимізація ТСБЗ природно розбивається на два ієрархічні рівні: верхній (перший) рівень – структурна оптимізація: вибір типів захисних пристроїв, їх розташування та зв'язків ТСЗ, і нижній (другий) рівень – параметрична оптимізація самих пристроїв за відомої структури ТСБЗ. Нині перший рівень слабо формалізований і виконується переважно вручну, виходячи з особистого досвіду розробника, чи з допомогою експертних оцінок.

Нехай на нижньому (параметричному) рівні оптимізації відомий набір $\{Z_i\}$ із m ($i = 1,\ldots m$) захисних пристроїв, кожний із яких описаний залежністю коефіцієнта захисту $K_з$ від параметрів $H_к$ цього пристрою:

$$\hat{E}_{ç} = Z^{n_i}(H_1,...,H_{n_i}), \tag{3}$$

де n_i – верхній індекс, що означає размірність векторів параметрів i-го пристрою;

$$H^i = \overset{n_i}{\underset{1}{\otimes}}\{H_k^i\}.$$

На множині H^i для кожного індексу і, тобто для i-го пристрою, визначена функція вартості

$$C_i^{n_i}(H_1,...,H_n) > 0.$$

Тут і далі, якщо йдеться про конкретний пристрій, індекс будемо опускати. Символ $\overset{n}{\underset{1}{\otimes}}$ означає декартовий добуток з n множин. Конкретизація параметрів {Hk} залежить від типу пристрою захисту.

Якщо рівень шкідливих чинників трохи коливається біля свого середнього значення, можна орієнтуватися на існуючі норми гранично допустимих шкідливостей. У цьому випадку вихідна вимога до Кз матиме вигляд

$$К_з \le \frac{х_{норм}}{х}.$$

Якщо шкідливі фактори зовнішнього середовища робочого місця змінюються в часі і в операторі «вплив – ефект» спостерігається кумулятивна складова, то доцільний дозовий підхід до вибору Кз.

При дозовому підході в загальному випадку допустимим є наступне наближення для біодії у

$$y = \varphi\left(\int_0^T g(x(\tau))d\tau\right),$$

де φ, g – монотонні незменшуючі функції.

Визначення нормативу дози $D_{норм}$ за зміну знаходиться з рівняння

$$у_{норм} = \varphi(D_{норм}). \tag{4}$$

Так як φ – монотонна функція, то рівняння (4) має єдине рішення

$$D_{норм} = \varphi^{-1}(у_{норм}).$$

Після цього $К_з$ визначається із співвідношення

$$\int_0^T g(К_з х(\tau))d\tau \le D_{норм}. \tag{5}$$

У першому наближенні тут можна обмежитися нагодою $x(\tau) = const$.

Тоді мають місце співвідношення:

$$g(K_3x)T \le D_{норм} \Leftrightarrow g(K_3x) \le \frac{D_{ng}}{T} \Leftrightarrow K_3x \le$$

$$\le g^{-1}\left(\frac{D_{норм}}{T}\right) \Leftrightarrow K_з \le \frac{g^{-1}\left(\frac{D_{норм}}{T}\right)}{x}.$$

З іншого боку, оскільки $\int_0^T g(x)dt = Tg(x) = D$ – величина дози за зміну без застосування захисту, то $x = g^{-1}\left(\frac{D}{T}\right)$.

Таким чином, при дозовому підході в першому наближенні можна прийняти умову

$$K_3 \le \frac{g^{-1}\left(\frac{D_{нор}}{T}\right)}{g^{-1}\left(\frac{D}{T}\right)}. \qquad (6)$$

Зазначимо, що в окремому випадку $g(x) = x$ нерівність (5) має вигляд

$$\int_0^T K_3x(\tau)d\tau \le D_{норм},$$

або

$$K_3D \le D_{норм} \Leftrightarrow K_3 \le \frac{D_{норм}}{D}.$$

Величина D може бути визначена експериментально, а для проєктованих систем – виходячи з узагальненого структурного методу.

При нездатності дозового підходу вибір K_3 повинен спиратися безпосередньо на модель взаємозв'язку "вплив – ефект". Для лінійної динамічної моделі маємо:

$$y(t) = \int_0^t w(\tau)x(t-\tau)d\tau. \qquad (7)$$

Момент t_0 вибирається із умови максимальної біодії $\max_{t\in[0;T]} y(t) = y(t_0)$.

Тоді вимога до $K_з$ визначається із співвідношення

$$K_3 \le \frac{y_{норм}}{\int_0^{t_0} w(\tau)x(t_0 - t)dt} = \frac{y_{норм}}{y_{max}}. \qquad (8)$$

Для проведення цієї оцінки потрібна інформація про характер зміни $x(t)$ протягом робочої зміни T. Для діючих виробництв ці характеристики можна отримати експериментально. Для проєктованих технологічних процесів раціональним є чисельне стохастичне моделювання функції $x(t)$ з урахуванням апріорної інформації.

На етапі технічного проєкту будь-якого із засобів або пристроїв захисту необхідно вирішувати завдання:

$$opt K_3 \big| C \le C^* , \qquad (9)$$

де роль C^* грає величина вартості С на етапі техніко-економічного обгрунтування. У разі оптимізації системи із *m* пристроїв зазначимо, що $K_3 \neq \sum_{i=1}^{m} k_{i3}$.

Щодо критерію вартості С, то він традиційно вважається адитивним:

$$C = \sum_{i=1}^{m} C_i \ . \tag{10}$$

Розглянемо тепер завдання оптимізації верхнього рівня – структурне. Для першого наближення будемо вважати, що розміщення всіх пристроїв оптимально, зазначивши, що в області оптимального геометричного розміщення відомі значні результати, які можуть бути використані при необхідності. Таким чином, наше завдання звелося до оптимального вибору ν пристроїв із сукупності N.

Усього можливостей вибору $\sum_{\nu=1}^{N} C_N^{\nu} = 2^N - 1$, оскільки порядок вибору ролі не грає, а кількість обраних пристроїв (визначається ефективністю та вартістю) може бути будь-яким $1 \le \nu \le N$. Це завдання може бути назване «завданням оптимізації на поєднаннях» (за аналогією з терміном «оптимізація на перестановках») і записується подібно до завдань (9) і (10):

$$1)\ \underset{\Omega_N}{opt}\, K_3 \big| C \le C^*; \qquad 2)\ \min_{\Omega_N} C \big| K_3 \le K^*, \tag{11}$$

де Ω_N – множина всіх поєднань із N по 1≤ν≤N.

Зазначимо, що тут під «пристроєм» можна розуміти і макропристрій. У завданнях (11) передбачається, що оптимізація на нижньому рівні вже здійснено. В іншому випадку змінні K_3 і С будуть нечіткими.

За методикою, викладеною в цій статті, були виміряні ε' і ε'' матеріалу з бікарбонолової нитки при λ = 3,2. З використанням розрахункових співвідношень було розраховано коефіцієнти ослаблення падаючого ЕМВ для плоского екрана від його товщини та побудовано номограми (рис. 1). При цьому, очевидно, експериментальні результати добре узгоджуються з теоретичними.

Так само можна побудувати залежності (щодо базового циліндричного екрану d = 0,8 см) потужності від товщини циліндричного екрана.

З рис. 1 видно, що якщо використовувати матеріал, що екранує, з бікарболонової нитки, ослаблення ЕМВ на 17 дБ забезпечується матеріалом товщиною d = 1,2 см.

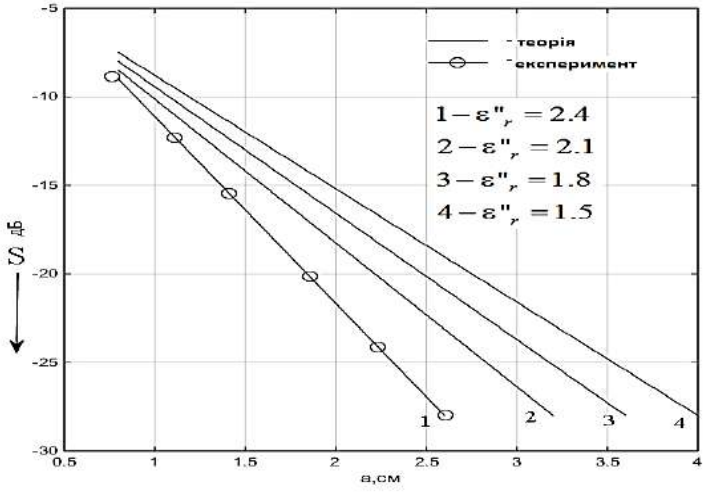


Рис. 1. Залежність коефіцієнта екранування плоского екрана від його товщини

Висновки

Розроблено оцінку ефективності роботи системи біомедичного захисту на основі оцінки дози шкідливого впливу електромагнітного випромінювання, що дозволяє порівняти існуючі технічні системи за рівнем їхньої безпеки.

Розроблено функціональну модель біомедичної системи життєдіяльності з різними джерелами електромагнітного випромінювання НВЧ діапазону на основі детерміністських критеріїв, які дозволяють з певною ймовірністю оцінити виконання завдання функціонування системи життєдіяльності.

Статистичний критерій побудований з урахуванням вартості системи, що дозволяє при побудові систем біологічного захисту на підставі загальних моделей отримати розрахункові формули, що пов'язують нововизначений критерій якості системи з параметрами, які можна порівняно легко виміряти та нормувати з біологічних та санітарних міркувань, що і визначає практичну цінність роботи.

Список літератури:

1. Дзюндзюк Б.В. Основи безпеки эрготичних систем. Київ, 1990. 56 с.

2. Sean E. Maximally Reliable Markov Chains Under Energy Constraints / E. Sean, S. Escolay, M. Eisele, K. Miller, L. Paninski // Neural Computation. 2009. Vol.21, Is.7. P.1863–1912.

3. Watanabe O. Analysis of a randomized local search algorithm for LDPCC decoding problem / O. Watanabe, T. Sawai, H. Takahash i// Lecture Notes in Comp. 2003. Vol. 2827. P.50–60.

4. Grendar M. Maximum Probability and Maximum Entropy methods: Bayesian interpretation // Bayesian inference and Maximum Entropy methods in Science and Engineering. 2005. Vol.2. P. 90–494.

5. Freire N. Entity Recognition and Resolution in Semi-structured Data / N. Freire // JCDL. 2011. P.12–17.

6. Gorban A.N. Entropy: The Markov Ordering Approac / A.N. Gorban, P.A. Gorban, G. Judge // Entropy. 2010. Vol.12. P. 1145–1193.

7. Семенець В.В., Леонідов В.І. Використання мікроконтролера STM32f407vg для дослідження амплітудно-частотних характеристик біологічних тканин // Радіотехніка. 2023. Вип. 214. С. 94–101.

Надійшла до редколегії 11.09.2024

Відомості про авторів:

Семенець Валерій Васильович – д-р техн. наук, професор; Харківський національний університет радіоелектроніки, професор кафедри біомедичної інженерії, Україна; e-mail: valery.semenets@nure.ua; ORCID: https://orcid.org/0000-0001-8969-2143

Стиценко Тетяна Євгенівна – канд. техн. наук, доц., Харківський національний університет радіоелектроніки, завідувач кафедри безпекової інженерії, Україна; e-mail: tatiana.stytsenko@nure.ua; ORCID: https://orcid.org/0000-0003-4530-0253

Григор'єв Олександр Вікторович – канд. техн. наук, доц., Харківський національний університет радіоелектроніки; Україна; e-mail: oleksandr.hryhoryev@nure.ua; ORCID: https://orcid.org/0000-0001-6467-7983

УДК 621.396 + 523.53 **DOI:10.30837/rt.2024.3.218.14**

O.V. HOLOVAN, PhD in Physics and Mathematics,
V.M. KHARCHENKO, PhD in Technical Sciences

ESTIMATING MODEL FOR THE LINEAR ELECTRON DENSITY OF THE TRAIL CREATED BY A METEOROID

Introduction

The linear electron density of the trail (LEDTr) created by a meteoroid when passing the Earth's atmosphere at a given observation altitude is used to estimate the power of the received signal and predict the trail reflectivity in the process of its further transformation.

Unlike many existing formulas for calculating LEDTr and semi-empirical models obtained by fitting to the results of radar observations, the presented technique allows calculations to be carried out for explicitly specified parameters of the meteoroid body and atmosphere. The main physical processes occurring during the interaction of the meteoroid with atoms and molecules of the atmosphere were taken into account.

The obtained relation of LEDTr to altitude corresponds well to the luminosity curves of the meteor trail, and the value of the electron density at the ionization maximum corresponds to the known observational results.

Estimating LEDTr for a given (measured) reflection point provides the ability to calculate the amplitude-time characteristics (ATCH) of radio signals, which allows you to create a model image of the scattered signal.

Comparison of the ATCH received signal with the calculated model images allows make a reasonable assumption about the meteoroid characteristics that generated the ionized trail. Predictive estimates of the mass, density velocity and radiant of the meteoroid that formed the corresponding trail can be made, as well as assumptions about its chemical composition and the genetic relationship of the meteoroid to its parent body.

The purpose of the study is to create a model for estimating the linear electron density of the trail created by a meteoroid, taking into account the physicochemical properties of the meteoroid and atmosphere at the observation altitude, as well as constructing the relation of LEDTr to altitude.

The result of the study is a software-implemented model for estimating LEDTr at the observation altitude for explicitly specified physicochemical properties of the meteoric body and atmosphere.

Calculation methodology for the linear electron density of the trail

Free electrons are formed during the collision of ablated particles of a meteoric body (atoms and molecules) with atmospheric molecules, provided that the particles kinetic energy is sufficient to remove the electron, which is the least tightly bound to the nucleus. The number of collisions is proportional to the concentration of atmospheric particles and midsection of the head part of the meteoric body at the calculated altitude, and the number of electrons generated depends on the chemical composition of the meteoroid and the atmosphere at the observation altitude and is determined by the ionization coefficient.

Calculation of the linear electron density of the trail (LEDTr) involves the estimation of:

- midsection of the head part of the meteoric body at the observation altitude ($S_{MH}(hM)$);
- concentration of particles (numerical density) of the atmosphere at the observation altitude ($N_a(h_M)$);
- values of the integral ionization coefficient at the observation altitude ($\beta_{total}(hM)$).

These parameters depend on the physicochemical properties of the meteoroid (*PCPMet*) and the atmosphere at the altitude of observation of the ionized trail, and LEDTr can be calculated using the expression

$$q(v, m_m, h_\text{M}, PCPMet) = \beta_{total}(v, PCPMet) \times S_{MH}(v, m_m, h_\text{M}, PCPMet) \times N_\text{a}(h_\text{M}) \times l^{-1}, \quad (1)$$

where $\beta_{total}(v, PCPMet)$ is the integral estimate of the ionization coefficient, $N_a(h_M)$ is the numerical density of the atmosphere at the observation altitude h_M, and $l = \Delta h_M / \cos \gamma_m$ is the length of the path along which the electron density was calculated.

The most difficult task is to estimate the $S_{MH}\ (hM)$ at the observation altitude. The value of the meteoroid's midsection at the entrance to the meteor zone without taking into account fragmentation and differential ablation is determined by the expression

$$S_{m0} = A_0 \left(\frac{m_{m0}}{\rho_m}\right)^{\frac{2}{3}}, \quad (2)$$

where A_0 is the shape coefficient, mm0 and ρm are the initial mass and density of the meteoroid, respectively. It is easy to show that for a ball $A_0 \approx 1.2$, and for a cube $A_0 = 1$.

The midsection S_{mM} at the observation point is determined by its residual mass m_{mM} in and expression (2) is converted to the form

$$S_{m\text{M}} = A_0 \left(\frac{m_{m\text{M}}}{\rho_m}\right)^{\frac{2}{3}}. \quad (3)$$

Particles that have evaporated within a given interval $\Delta l = \Delta h / \cos \gamma$ of the motion trajectory form a vapor "cushion" in front of the meteoroid, which leads to a change in the midsection of the meteoroid. In this case, the midsection of the head part at a altitude h_M can be represented by the expression

$$S_{HM}(h_M) = A_0 \left(\frac{m_{m\text{M}}}{\rho_m}\right)^{\frac{2}{3}} + N_{evep}(h_M) \cdot \sigma_{mav}, \quad (4)$$

where $\sigma_{m\ av}$ is the average cross–section for the collision of evaporated particles and $N_{ever}\ (h_M)$ is the number of particles evaporated at altitude h_M. The number of evaporated particles can be represented by the expression

$$N_{ever}(h_M) = \frac{\Delta m_m(h)}{\mu_{mav}} = \frac{[m_m(h_M+\Delta h) - m_m(h_M)] \cdot N_A}{M_{mmol}} = \frac{\Delta m_m(h_M) \cdot N_A}{M_{mmol}} \quad (5)$$

where $\Delta_{mm}\ (h_M)$ is the mass loss at altitude h_M when it changes by Δh , $\mu_{m\ ev}$ is the average absolute mass of evaporated particles, which is defined as the ratio of the molar mass $M_{m\ mol}$ to the Avogadro number N_A = 6.02214082*10^{23} mole^{-1}. The molar mass of any substance, expressed in grams per mole, is numerically equal to the mass of the molecule of this substance, expressed in atomic mass unit, and can be found from the periodic table. Neglecting the mass defect, the relative atomic mass of a molecule is equal to the sum of the relative atomic masses of its constituent elements. It should be noted that if $\Delta_{mm}\ (h_M)$ was calculated with the unit of kg (SI system), the molar mass should be converted to kg/mole – i.e. divide by 1000.

Based on expressions (4) and (5), the midsection $S_{MH}\ (h_M)$ at the observation point can be represented by the expression

$$S_{HM}(h_M) = A_0 \left(\frac{m_{m\text{M}}}{\rho_m}\right)^{\frac{2}{3}} + A_1 \frac{\Delta m_m(h) \cdot N_A}{M_{mmol}} \cdot \sigma_{mav}, \quad (6)$$

where A_1 is the coefficient correcting the increase in the midsection S_{MH} due to evaporation, and $\sigma_{m\ av}$ is the average value of the cross–section of the molecule (atom) of the meteoroid, which depends on its chemical composition.

The effective cross section of a molecule (atom) can be found by sequentially calculating the volume of a mole of a substance *X*, the volume of one molecule of a substance V(X1), and then the

minimum distance d at which the centers of two molecules approach each other during a collision (the effective diameter of the molecule).

The mole volume of a substance *X* can be estimated by calculating its molar mass and dividing it by the density of this substance

$$V(X_{mole}) = \frac{M(X_{mole})}{\rho(X)}. \quad (7)$$

The molar mass of a substance $M(X_{mole})$ in kg/mol is numerically equal to the relative molecular weight of the given substance $\mu_m(X)$ divided by 1000, and its density $\rho(X)$ is selected from the appropriate reference books. An approximate estimate of the volume of one molecule of a substance $V(X_1)$ can be performed, assuming that the molecules are located as close as possible to each other and dividing $V(X_{mole})$ by the Avogadro number N_A

$$V(X_{mole}) = \frac{M(X_{mole})}{\rho(X)}. \quad (8)$$

In this case, the diameter of the molecule d_X is equal to

$$d_{X1} = \sqrt[3]{\frac{6V(X_1)}{\pi}}, \quad (9)$$

and the effective cross–section of a molecule (atom) of a meteoroid σ_{X1} is defined by the expression

$$\sigma_{X1} = \pi d_{X1}^2 = \frac{\pi}{100} \cdot \left[\frac{6\mu_m(X)}{\pi\rho(X)\cdot N_A}\right]^{2/3}, \quad (10)$$

and the average value of the cross section of evaporated particles $\sigma_{X1\,av}$ is equal to

$$\sigma_{X1ev} = \sum_{j=1}^{L} p_j\, \sigma_{X1j} = \sum_{j=1}^{L} p_j \frac{\pi}{100} \cdot \left[\frac{6\mu_m(X_j)}{\pi\rho(X_j)\cdot N_A}\right]^{2/3}, \quad (11)$$

where p_j is the fraction of the corresponding evaporated substance.

Based on expressions (5) and (10), the midsection $S_{MH}(h_M)$ at the observation point can be represented by the expression

$$S_{MH}(h_M) = A_0 \left(\frac{m_{m\mathrm{M}}(h_M)}{\rho_m}\right)^{\frac{2}{3}} + A_1 \frac{\Delta m_m(h)\cdot N_A}{M_{mmolav}} \cdot \sigma_{mev}, \quad (12)$$

where $M_{mmolav} = \sum_{j=1}^{L} p_j\, M_{mj}$.

To implement midsection calculations based on expression (11), it is necessary to obtain the relation of the meteoroid mass to the observation point altitude h_M, which will allow one to estimate the amount of evaporated matter at a given interval Δh. To do this, we will use the basic equations of the physical theory of meteors [1].

The first basic equation of the physical theory of meteors is based on the assumption that the loss of momentum by a meteoroid $m_m\, dv$ is proportional to the momentum of the oncoming air flow:

$$m_m \frac{dv}{dt} = -\Gamma S \rho_a v^2, \quad (13)$$

where m_m is the mass of the meteoroid, v is the velocity, S is the area of the frontal section of the meteoroid (midsection), ρ_a is the density of the atmosphere, Γ is the resistance coefficient (the fraction of the momentum of the incident atoms and atmospheric molecules, which is converted into the deceleration of the body).

The second basic equation of the physical theory of meteors describes the loss of mass of a meteoroid body, provided that all the energy is spent on ablation (evaporation or melting and blowing off the molten film):

$$\frac{dm_m}{dt} = -\Lambda S \rho_a \frac{v^3}{2Q}, \quad (14)$$

where Q is the specific heat of evaporation of the meteoroid substance (J/kg), and Λ is the fraction of the kinetic energy of the oncoming flow of molecules spent on ablation during the time dt.

We relate the change in time and the change in altitude using the relation [2]

$$dt = -\frac{dz}{v\cos\gamma}. \tag{15}$$

Then expressions (13) and (14) are transformed to the form

$$m_m \frac{dv}{dz} = -\frac{\Gamma}{Q_m \cos\gamma} S_m \rho_a v^2, \tag{16}$$

$$\frac{dm_m}{dz} = -\frac{\Lambda}{\cos\gamma} S_m \rho_a v^2. \tag{17}$$

Dividing (17) by (16) we obtain a relation connecting the mass m_m and velocity v of the meteoroid

$$\frac{1}{m_m} dm_m = \frac{\Lambda}{2Q_m\Gamma} v dv. \tag{18}$$

Integrating the left side of (18) in the range from the value of the initial mass of the meteoroid m_{m0} to its mass m_{mM} at the observation point, and the right side in the range from the value of the initial velocity of the meteoroid from v_{m0} to its value v_{mM} at the observation point, we obtain the equation

$$\int_{m_{m0}}^{m_{mM}} \frac{1}{m_m} dm_m = \frac{\Lambda}{2Q_m\Gamma} \int_{v_{m0}}^{v_{mM}} v\, dv. \tag{19}$$

As a result of integration, we obtain the value of the mass of the meteoroid at the observation point M depending on the velocity at this point v_{mM} for a given initial velocity v_{m0}

$$m_{mM} = m_{m0}\, exp\left[-\frac{\Lambda}{2Q\Gamma}\frac{\left(v_{m0}{}^2 - v_{mM}{}^2\right)}{2}\right] \tag{20}$$

To calculate the meteoroid mass at an observation point, it is necessary to estimate its velocity at this point, which depends on the midsection of the meteoroid and the density of the atmosphere. In the first approximation, which is usually sufficient for many purposes of meteor astronomy, from the equation of state and the equation of hydrostatic equilibrium of the atmosphere, we obtain the exponential distribution of the atmospheric density $\rho_a\,(h)$ with respect to altitude h:

$$\rho_a(h) = \rho_b\, exp\left[-\frac{g_b m_{ab}(h-h_b)}{RT_b}\right], \tag{21}$$

where ρ_b, g_b и T_b – atmospheric density, gravitational acceleration and absolute temperature at the h_b selected (base) altitude, m_{ab} – average molecular mass (kg/kmole) of the atmosphere at base altitude, and R = 8314.46 (J/kmole·˚K) is the universal gas constant.

If we choose h_b = 95000 m as the base altitude and use data from [3], formula (21) is converted to the form

$$\rho_a(h) = 1{,}4051 \cdot 10^{-6}\, exp[-0{,}17768 \cdot 10^{-3}(h - 95000)]. \tag{22}$$

As a model for estimating the distribution of atmospheric density, the relation can be used

$$\rho_a(h) = 1{,}405 \cdot 10^{-6}\, exp\left[-\frac{(h-95)}{H_a}\right], \tag{23}$$

where H_a= 46,8273 – 0,95h + 0,0055 h^2 – reduced altitude of the atmosphere (here h and H_a are measured in kilometers), as well as other models presented in [4].

In this study [5], when calculating the meteoroid velocity, it was assumed that its mass m_m and midsection S_m did not change when moving with an initial velocity v_{m0} in a medium with a density ρ_m, and the force of gravity was negligibly small compared to the force of resistance to motion. This made it possible to obtain an approximate estimate of the meteoroid velocity at the observation point represented by the expression

$$v_{mM}(h) = v_{m0}\, exp(-\rho*/2), \tag{24}$$

where $\rho^* = \rho_a(h)/\rho_\beta(h)$, $\rho_\beta(h) = \beta \cos\gamma /h$, $\beta = m_m/S_m$ – ballistic coefficient, S_m – meteoroid midsection.

Taking into account the previously adopted notation, the expression (24) is transformed to the form

$$v_{mM}(h_{aM}, m_{mM}, \gamma_m) = v_{m0}\, exp\left[-\frac{\rho(h_{aM})h_{aM}S_{mM}C_D}{2m_{mM}\cos\gamma_m}\right], \tag{25}$$

where γ_m is the zenith angle of the meteoroid radiant; C_D is a dimensionless aerodynamic resistance coefficient, which, when a meteoroid moves in the range of meteor altitudes, can take values from 2 to 0.5 (depending on the form of the body and the Reynolds number for it). If at the altitude of observation the length free path of molecules is greater than the characteristic size of the body, the coefficient C_D can be taken equal to 2.

Expressing the midsection of the meteoroid S_{mM} terms of its mass m_{mM} and density ρ_m using formula (2) from expression (25) we obtain

$$v_{mM}(h_{aM}, m_{mM}, \gamma_m) = v_{m0}\, exp\left[-\frac{\rho_a(h_{aM})h_{aM}C_D A_0}{2(m_{mM}\rho_m^2)^{\frac{1}{3}}\cos\gamma_m}\right]. \tag{26}$$

The algorithm for calculating the values of m_{mM} and v_{mM} can be implemented based on calculation formulas (20) and (26) using the recurrent method, in which each next member of the sequence is calculated using the result of the calculations of velocity and mass at the previous step. This approach makes it possible to partially remove the restrictions regarding the requirement of constancy of the mass and midsection of the meteoroid during its movement.

To implement the recurrent method of calculating m_{mM} and v_{mM}, you must perform the following steps:

1. Calculate, using expression (26), the value of the velocity v_{mM} at a point close to the upper boundary of the meteor zone at a altitude $h_{max} - \Delta h$. At the same time, the meteoroid velocity at the entrance to the meteor zone v_{m0}, its initial mass m_{m0}, density ρ_m and radiant zenith angle γ_m are assumed to be known, and the altitude of the upper boundary of the meteor zone h_{max}, the coefficients C_D and A_0 are also considered given. The value of atmospheric density at altitude $h_{max} - \Delta h$ can be calculated using expression (22).

2. Calculate the meteoroid mass value at altitude $h_{max} - \Delta h$ using expression (20).

3. Calculate, using expression (26), the value of the velocity v_{mM} at the altitude $h_{max} - i\Delta h$, where $i = 2$. In this case, use the value of the meteoroid mass obtained at the previous step (at the altitude $h_{max} - \Delta h$).

4. Calculate the value of the meteoroid mass at the altitude $h_{max} - 2\Delta h$ using expression (20).

5. Calculate the value of the velocity v_{mM} at the altitude $h_{max} - i\Delta h$, where $i = 3$. In this case, use the value of the meteoroid mass obtained at the previous step (at the altitude $h_{max} - 2\Delta h$).

6. Repeat the calculations of v_{mM} and m_{mM} for $i = 4 \ldots N$, where N is determined from the condition $(h_{max} - N{\cdot}\Delta h) \geq h_{min}$.

To calculate LEDTr using formula (1), it is necessary to determine the number density of the atmosphere Na(hM) (particle concentration) at a given altitude. This can be done on the basis of the reference data [3] or by using the approximate relationship

$$N_a(h_M) = \frac{\rho_a(h_M)N_A}{\mu_{aav}}, \tag{27}$$

where $\mu_{a\,ev}$ – average relative atomic mass of atmospheric particles (kg/kmol), which in the interval of 80 ... 130 km can be approximated by the expression

$$\mu_{aav}(h_M) = \begin{cases} 28{,}964 \\ 28{,}964 - 0{,}14056(95 - h_M) \end{cases} \text{when} \begin{matrix} 95km \geq h_M \geq 80km \\ h_M > 95km \end{matrix} \tag{28}$$

For a complex chemical composition meteoroid, the integral estimate $\beta_{total}(v)$ is a linear combination of ionization coefficients, where each coefficient is taken with a weight p_j proportional to the fraction of atoms present in the composition of the meteoroid

$$\beta_{total}(v) = \sum_j p_j \beta_j\,(v). \tag{29}$$

The value of $\beta_j(v)$ depends on the velocity and ionization potential of the atoms that make up the meteoric body, which can be found in the corresponding reference books.

The works [1, 6–8] present an analysis of various methods for calculating the ionization coefficient, which assume the possibility of obtaining estimates of $\beta(v)$ under various assumptions regarding the physicochemical composition of meteoroids and the atmosphere. Average data for $\beta(v)$ of these elements are presented in Table 1 [1].

Table 1

Element	**Fraction, %**	**Meteoroid velocity *v*, km/s**		
		20	**40**	**70**
O	56,0	0,00016	0,043	0,554
Fe	11,4	0,068	0,595	3,41
Mg	15,4	0,020	0,151	0,81
Ca	0,9	0,082	0,619	1,98
Si	14,4	0,023	0,300	1,70
***β* (*v*)**		**0,0154**	**0,170**	**1,12**

Based on these data in the study [1], an approximating calculation formula for β (v) was obtained

$$\beta(v) = 5{,}4889 \times 10^{-7} \times v^{3{,}42}. \tag{30}$$

This formula can be used to perform approximate calculations of the ionization coefficient $\beta(v)$ for meteoroids of various origins.

For iron meteoroids in the study [7], it was proposed to calculate $\beta(v)$ using the formula

$$\beta_{Fe}(v) = 5{,}96 \times 10^{-6} \times v^{3{,}42}, \tag{31}$$

which is applicable for the velocity range of 20 km/s $\le v \le$ 45 km/s).

Based on expression (1), using relations (10), (11), (22), (20), (26), (28), (30) (or (31)) we obtain a formula that allows us to calculate LEDTr

$$q(v, m_m, h, \gamma) = \beta_{total}(v) \times \left[A_0 \left(\frac{m_{m\text{M}}}{\rho_m}\right)^{\frac{2}{3}} + A_1 \frac{\Delta m_m(h,v,\gamma)\cdot N_A}{\mu_{mev}} \sigma_{X1cp}\right] \times \left[\frac{\rho_a(h_\text{M}) N_A}{\mu_{aev}(h_\text{M})}\right] \times \frac{\cos\gamma}{\Delta h_\text{M}} \tag{32}$$

The average value of the mass of atoms $\mu_{\text{m av}}$ in a molecule of a meteoric substance can be determined based on its chemical formula. For example, a meteoroid like the Chelyabinsk meteorite mainly consists of silicates: olivines ($(Mg,Fe)_2[SiO_4]$ and pyroxenes ($Mg,Fe)_2[Si2O_6]$. An olivine molecule contains 2 Fe atoms = 55.847 amu, 2 Mg atoms = 24.305 amu, 1 Si atom = 28.085 amu, and 4 atoms O = 15.9994 amu (we take the atomic mass from the periodic table). Determine its molar mass $M_{oliv.}$ = 252.3866 amu and the proportion of each atom in the composition of the olivine molecule. After this, it is easy to show that the average mass of atoms in an olivine molecule is $\mu_{m\,av}$ = 36.5785 amu. Density ρ_m for olivine is usually close to 3300 kg/m^3.

Simulation results

Table 2 represents the initial data for calculating LEDTr created by iron meteoroids and meteoroids with a chondritic structure, and Table 3 shows the calculation algorithm using formula (32) and formulas used for calculation.

Table 2

No.	Parameters	Designation, dimension	Value
1	The meteoroid velocity upon entering the meteor zone	v_{m0}, *m/s*	40000
2	Zenith angle of the meteoroid	γ, *degree*	60
3	Initial meteoroid mass	m_{m0}, *kg*	0,001
4	Upper limit of the meteor zone	$h_{M\,max}$, *m*	130000
5	Lower boundary of the meteor zone	$h_{M\,min}$, *m*	80000
6	Mass density of iron meteoroid	$p_{m\,Fe}$, *kg/m*^{3}	7874
7	Iron atom mass	$\mu_{m\,Fe,}$ *amu*	55,8430
8	Mass density of olivine $(Mg,Fe)_2[SiO_4]$	$p_{\text{m олив}}$, *kg/m*^{3}	3300
9	Average atomic mass in an olivine molecule	$\mu_{m\,cp,}$ *amu*	36,5785
10	Aerodynamic resistance coefficient	C_D	2
11	Shape factor	A_0	1,21
12	Correction factor	A_1	1
13	Fraction of the kinetic energy of the oncoming flow of molecules spent on ablation	Λ	1
14	Specific heat of vaporization of meteoroid matter	Q, *J/kg* (m^2/s^2)	$6{,}3\cdot10^6$
15	Atmospheric resistance coefficient, characterizing the fraction of momentum transmitted to the meteoroid by atmospheric particles	Γ	1

Table 3

No.	Actions Performed	Formulas for calculations
1	Set the values of observation altitudes h_M in the meteor zone with a step Δh_M. In this case, $h_{Mi} = h_{M\,max} - i\,\Delta h_M$, where the value of i varies from 0 to $N = (h_{M\,max} - h_{M\,min})/\,\Delta h_M$.	
2	Calculate the average value of the cross section of evaporated particles of the meteoric body $\sigma_{X1\,av}$	(10), (11)
3	Calculate the atmospheric density $\rho_a(h)$ at observation altitudes h_{Mi}.	(22)
4	Calculate the velocity v_{mM} and mass m_{mM} of the meteoroid for the selected altitudes using the proposed iteration method. Take the initial mass and velocity of the meteoroid upon entering the meteor zone from Table 1.	(26), (20)
5	Calculate the meteoroid mass loss Δm_{mM} over the interval Δh in the area of the selected altitude h_{Mi}, subtracting the subsequent value from the previous mass value.	
6	Calculate the average relative atomic mass of atmospheric particles $\mu_{a\,ev}$	(28)
7	Calculate the value of the ionization coefficient $\beta_{total}(v)$for a meteoroid with a complex chemical composition.	(30) or (31) for iron
8	Calculate LEDTr using the proposed formula	(32)

The results of the calculation of LEDTr, performed using Microsoft Excel based on the proposed algorithm for iron meteoroids and meteoroids with a chondritic structure (close in composition to olivine) are presented in Figure 1. The initial data for calculations are taken from Table 2.

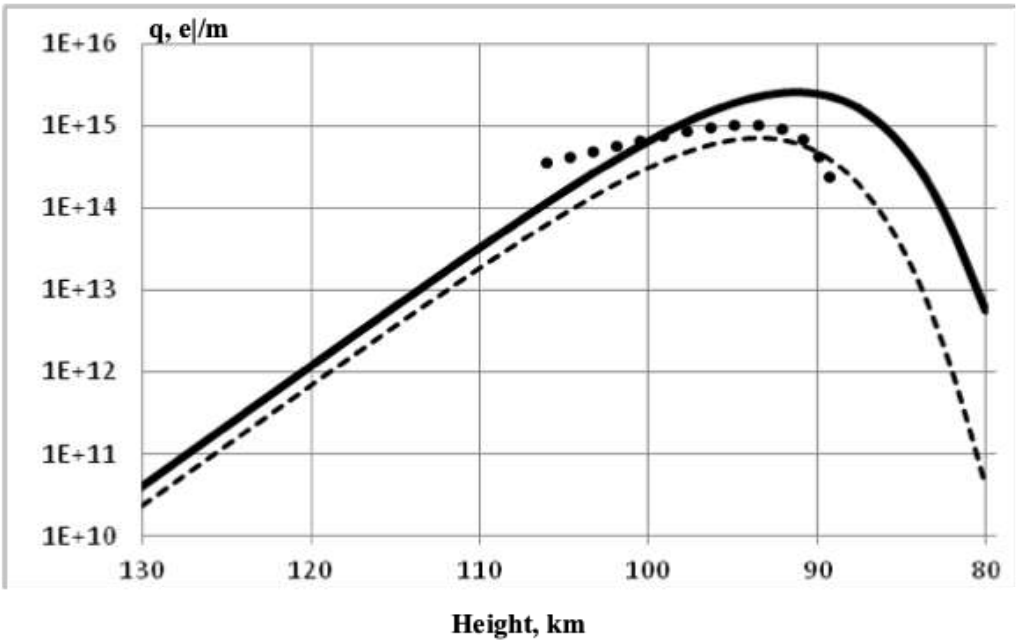


Fig. 1. Relation of LEDTr of an iron meteor body to the altitude of the observation point – solid line. The dotted line shows LEDTr for a meteoroid close in composition to the Chelyabinsk meteorite, and the dots show the result of the calculation using the semi-empirical formula (33)

In the study [9], based on previously obtained results [10 – 13], it was proposed to calculate the linear electron density (in electrons/m) using a semi-empirical formula, valid only in the region of maximum ionization and obtained as a result of fitting to the results of radar measurements

$$q(h_M) = 4{,}03 \times 10^{14} \cdot \frac{m(v_{m0}-8{,}15)^3}{H_M} \cdot \cos\gamma \cdot Z(t), \tag{33}$$

where h_M – altitude above the Earth's surface at the considered point on the trail;
v_{m0} – velocity (in km/s) of the meteoroid upon entering the meteor region;
m – mass (in kg) of the meteoroid upon entering the meteoroid region;
γ – zenith angle of the meteoroid;
$Z(t)$ – function defined by the expression

$$Z(t) = \begin{cases} \frac{9}{4}e^{-t}\left(1-\frac{1}{3}e^{-t}\right)^2 & \text{when } -\ln 3 \le t \le 1.7 \\ 0 & t \le -\ln 3\,, t \ge 1.7 \end{cases} \tag{34}$$

where the relative altitude t is defined as

$$t = \frac{h_M - h_{max}}{H_M}; \tag{35}$$

H_M – reduced altitude of the atmosphere (in km), which is calculated by the formula

$$H_M(h_M) = 6{,}4 + 0{,}09(h_M - 95); \tag{36}$$

h_{max} is the altitude of the maximum ionization (in km), determined by the empirical relation

$$h_{max} = 47{,}4 + 12{,}76 * \ln v_{m0}. \tag{37}$$

The results of calculations using formulas (33) – (37) are represented in Figure 1 by a dotted line. They show a fairly good agreement with the corresponding calculations of LEDTr calculated using the proposed method.

Discussion

When developing the represented model for estimating the linear electron density of the trail (LEDTr), the main physical processes occurring during the interaction of a meteoroid with atoms and molecules of the atmosphere were taken into account. The calculation of LEDTr includes an assessment of midsection of the head part of the meteoric body, the concentration of particles (numerical density) of the atmosphere and the value of the integral ionization coefficient at the observation altitude.

When developing this model, well-known models for estimating the density of the atmosphere at a given altitude and the ionization coefficient, which depends on the chemical composition of the meteoroid, were used. The corresponding calculation formulas were obtained, and a software-implemented calculation algorithm was presented. It is important to note that when estimating the mass loss of the meteoroid, a recurrent method was used to calculate the mass and velocity as its altitude changed. This method made it possible to take into account changes in the mass and midsection of the meteoroid during its movement.

Unlike many existing formulas for calculating LEDTr and semi-empirical models obtained by fitting to the results of radar observations, our technique allows us to carry out calculations for explicitly specified parameters of the meteoric body and atmosphere. The obtained relation of LEDTr to altitude corresponds well to the luminosity curves of the meteor trail, and the value of the electron density at the ionization maximum corresponds to the known results.

Calculations of LEDTr at a given (measured) altitude, performed using the proposed methodology, make it possible to estimate the power of the received signal and predict the reflectivity of the trail in the process of its transformation. This, in turn, makes it possible to calculate amplitude-time characteristics and create a model image of a signal scattered at a given altitude.

The presented model assumes that the meteoroid is a single body, and the main mechanism of ablation is evaporation without crushing or changing the meteoroid shape. However, the model can be adapted for the case of fragmentation of a meteoroid at a given altitude. To do this, it is necessary to calculate LEDTr for each fragment according to the described method, and then use the principle of superposition, according to which the resulting effect of several independent events can be represented by the sum of the effects caused by each event separately. With quasi-continuous crushing of meteoroids, adjustment to the results of radar measurements is necessary. The proposed calculation formula provides a correction factor that allows you to change the value of the meteor body's midsection. This coefficient may differ for meteoroids of different origins.

It should be noted that the presented model for estimating LEDTr can be modified, since it did not take into account the decrease in electron density due to the recombination of positive ions with electrons, the attachment of free electrons to oxygen molecules, as well as ion reactions with ozone. The known estimates of atmospheric density and ionization coefficient used in developing the model can be specified. In this case, the calculation algorithm remains unchanged.

Conclusions

1. The model proposed for estimating the linear electron density of the trail (LEDTr) created by a meteoroid at a given (measured) altitude allows calculations to be carried out for explicitly specified parameters of the meteoroid and atmosphere.

2. The presented methodology and a software-implemented calculation algorithm make it possible to obtain relation of LEDTr to the observation altitude, which qualitatively coincides with the results of radar observations and the luminosity curve of the meteor trail.

3. LEDTr calculations using the proposed method make it possible to estimate the power of the received signal, calculate its amplitude-time characteristics and create a model image of the signal scattered at a given altitude.

4. The known estimates of atmospheric density and ionization coefficient used in developing the model can be refined. In this case, the calculation algorithm remains unchanged.

5. The presented model for estimating LEDTr can be modified, since it does not take into account the decrease in electron density due to the recombination of positive ions with electrons, the attachment of free electrons to oxygen molecules, as well as ion reactions with ozone.

References:

1. .Bronshten V. A. Fizika meteorovykh yavleniy [Physics of Meteor Phenomena]. Moscow : Nauka, 1981. P. 416. (in Russian).
2. Chernogor L. F. Fizicheskie effekty Rumynskogo meteoroida [Physical effects of the Romanian meteoroid] // Kosmichna nauka i tekhnologiya. 2018. No 24(1). P. 49–70. (in Russian).
3. State Standard 4401–73. Standartnaya atmosfera. Parametryi [Standard atmosphere. Parameters]. Moscow : Izdatelstvo standartov [Standards Publishing House], 1977. P. 117. (in Russian).
4. Gorelov D. Yu., & Voloshchuk Yu. I. Issledovanie fizicheskogo faktora obnaruzhimosti [Study of the physical factor of detectability]. Radiotechnika. 2005. No143. P. 215–222. (in Russian).
5. Tirskiy G. A. Vzaimodeistvie kosmicheskikh tel s atmosferami Zemli i planet [Interaction of cosmic bodies with the atmospheres of the Earth and planets] // Sorosovskiy obscheobrazovatelnyiy zhurnal [Soros Educational Journal]. 2000. No 6(5). P. 76–82. (in Russian).
6. Moorhead A. V., Brown P. G., Campbell-Brown M. D., Heynen D., & Cooke W. J. Fully correcting the meteor velocity distribution for radar observing biases // Planetary and Space Science. 2017. No 143. P. 209–217. doi: 10.1016/j.pss.2017.02.002
7. Jones W. (1997). Theoretical and observational determinations of the ionization coefficient of meteors // Monthly Notices of the Royal Astronomical Society. 1997. No 288(4). P. 995–1003. doi: 10.1093/mnras/288.4.995
8. Thomas E., Horányi M., Janches D., Munsat T., Simolka J., & Sternovsky Z. Measurements of the ionization coefficient of simulated iron micrometeoroids // Geophysical Research Letters. 2016. No 43. P.3645–3652. doi:10.1002/2016GL068854
9. Belkovich O. I. Meteornoe rasprostranenie radiovoln [Meteor propagation of radio waves] Zelenodolsk : Izdatelstvo Kazanskogo gosudarstvennogo universiteta [Publishing house of Kazan State University], 2008. P. 48. (in Russian).
10. Eshleman V. R. The theoretical length distribution of ionized meteor trails // J. Atmospheric and Terrestrial Physics. 1957. No 10(2). P. 57-72.
11. Belkovich O. I. Statisticheskaya teoriya radiolokatsii meteorov [Statistical theory of meteor radio location]. Izdatelstvo Kazanskogo gosudarstvennogo universiteta [Publishing house of Kazan State University], 1971. P. 104. (in Russian).
12. Tokhtasev V. S. Obrazovanie i raspod meteornykh sledov [Formation and decay of meteor trails]. Dushanbe, Tajikistan : DONISH [publishing house "Donish"], 1975. (in Russian).

Received 20.09.2024

Information about the authors:

Holovan Olena Viktorivna – PhD in Physics and Mathematics, O.Ya. Usikov Institute of radiophysics and electronics NASU, researcher, Ukraine; e-mail: holovan.helen@gmail.com; ORCID: https://orcid.org/0009-0008-4455-4562

Kharchenko Viktor Mykolayovych – PhD in Technical Sciences, Scientific Research Institute of Radioelectronic techniques, head of the research department, Kharkiv; Ukraine; e-mail: vn6669165@gmail.com

ABSTRACTS РЕФЕРАТИ

SYSTEMS AND METHODS OF INFORMATION PROTECTION
СИСТЕМИ І МЕТОДИ ЗАХИСТУ ІНФОРМАЦІЇ

UDC 004.05

Zero trust architecture: challenges and recommendations for successful implementation / *V.I. Yesin, V.V. Vilihura, D.Y. Uzlov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 7 – 34.

To protect the modern digital enterprise, a new approach is needed today to ensure secure access to your own corporate resources anytime, anywhere, and their efficient operation regardless of where they are located. The traditional perimeter-based network protection model is unable to adapt to the development of modern technologies. Therefore, enterprises have begun to rethink the traditional network security perimeter, leaning toward a new concept and architecture of protection. Such a concept is currently the security paradigm called "zero trust". The zero trust concept attracts special attention of researchers and practitioners, as it is able to meet new requirements for information security and cybersecurity. One of the factors contributing to the demand for zero trust architecture is the increased complexity and heterogeneity of modern IT systems. However, despite the popularization of this concept and the obvious security benefits from its use, there are certain difficulties in its implementation in enterprises. Deploying a zero trust architecture is quite complex from both a technical and organizational point of view. At the same time, interested representatives of enterprises are not yet fully aware of the advantages and disadvantages of the zero trust concept, which significantly hinders its application, which is still in the process of development. The main serious factors hindering the implementation of the zero trust concept are the lack of information for choosing a zero trust solution and the insufficient number of qualified specialists in this area. That is, today there is a problem associated with a certain lack of awareness about the concept and zero trust architecture (about their theoretical and practical significance) for choosing the right solution when building a security system for a corporate information system in modern conditions. This paper aims to solve this problem by summarizing existing research and the experience of various international companies that are implementing this approach in practice. The purpose of this paper is to assist IT enterprise information security professionals in the selection and application of enterprise-relevant, forward-looking zero trust architectures to increase the cybersecurity of the enterprise information system. This paper briefly discusses the conceptual zero trust architecture, its main logical components, deployment models, threats associated with zero trust architecture, and some recommendations for successful implementation of zero trust architecture in the IT enterprise.

Key words: zero trust; zero trust architecture; zero trust architecture deployment models; information security; cybersecurity.

2 tabl. 9 fig. Ref: 41 items.

УДК 004.05

Архітектура нульової довіри: проблеми та рекомендації щодо успішного впровадження / *В.І. Єсін, В.В. Вілігура, Д.Ю. Узлов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 7– 34.

Щоб захистити сучасне цифрове підприємство, сьогодні необхідний новий підхід, який дозволить забезпечити безпечний доступ у будь-який час і в будь-якому місці до власних корпоративних ресурсів, а також їхнє ефективне функціонування незалежно від того, де вони розташовані. Традиційна модель захисту мережі на основі периметра не спроможна адаптуватися до розвитку сучасних технологій. Тому підприємства стали переосмислювати традиційний периметр безпеки мережі, схиляючись до нової концепції та архітектури захисту. Такою концепцією в даний час стала парадигма безпеки, що отримала назву «нульова довіра». Концепція нульової довіри привертає особливу увагу дослідників і практиків, оскільки вона здатна задовольнити нові вимоги до інформаційної безпеки, кібербезпеки. Одним із факторів, що сприяють затребуваності архітектури нульової довіри, є підвищена складність та гетерогенність сучасних ІТ-систем. Однак, незважаючи на популяризацію цієї концепції та очевидні переваги у сфері безпеки від її застосування, на підприємствах виникають певні складнощі щодо її реалізації. Розгортання архітектури нульової довіри є досить складним як з технічної, так і з організаційної точки зору. Водночас зацікавлені представники підприємств ще не повною мірою усвідомлюють переваги та недоліки концепції нульової довіри, що значно стримує її застосування, яке й так ще перебуває в процесі розвитку. Основними серйозними факторами, що перешкоджають впровадженню концепції нульової довіри, є брак інформації для вибору рішення з нульовою довірою та недостатня кількість кваліфікованих фахівців у цій сфері. Тобто сьогодні існує проблема, пов'язана з певним дефіцитом поінформованості про концепцію та архітектуру нульової довіри (про їх теоретичну та практичну значущість) для вибору правильного рішення при побудові системи безпеки корпоративної інформаційної системи в сучасних умовах. Ця стаття націлена на вирішення цієї проблеми шляхом узагальнення наявних досліджень та досвіду різних міжнародних компаній, які впроваджують даний підхід на практиці. Мета роботи – допомогти спеціалістам з інформаційної безпеки ІТ-підприємств у виборі та застосуванні актуальних для підприємства перспективних архітектур нульової довіри, які дозволять підвищити кібербезпеку корпоративної інформаційної системи. У стислому викладі розглядаються концептуальна архітектура нульової довіри, її основні логічні компоненти, моделі розгортання, загрози, пов’язані з архітектурою нульової довіри, а також деякі рекомендації щодо успішного впровадження архітектури нульової довіри на ІТ-підприємстві.

Ключові слова: нульова довіра; архітектура нульової довіри; моделі розгортання архітектури нульової довіри; інформаційна безпека; кібербезпека.

Табл. 2. Іл. 9. Бібліогр.: 41 назв.

UDC 004.056.5

Using zero watermarks for image authorship and multi-factor authentication / *V.O. Poddubnyi, R.Y. Gvozdov, O.V. Sievierinov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 35 – 43.

The relevance of this work lies in the revealed possibilities of using zero watermarks to confirm the authorship of digital images and the use of zero watermarks for images as a method of multifactor authentication.

The paper proposes a modernized zero watermark algorithm for testing the capabilities of zero watermarks. The proposed multifactor authentication algorithm using zero watermarks.

Tests have shown that null watermarks are not suitable for confirming authorship, as they cannot guarantee the distinction between similar images or between images of the same author. However, there is a prospect of using this technology in multi-factor authentication schemes, in combination with other authentication methods (such as password, biometrics, tokens).

Key words: watermark; zero watermark; authentication; authorship; DWT; SVD; CA; NCC; PSTR.

1 tabl. 6 fig. Ref: 14 items.

УДК 004.056.5

Використання нульових водяних знаків для підтвердження авторства зображень та багатофакторної автентифікації / *В.О. Поддубний, Р.Ю. Гвоздьов, О.В. Сєвєрінов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 35 – 43.

Актуальність роботи полягає у виявленні можливостей використання нульових водяних знаків для підтвердження авторства цифрових зображень та використання нульових водяних знаків для зображень в якості методу багатофакторної автентифікації. Запропоновано модернізований алгоритм нульового водяного знаку для тестування можливостей нульових водяних знаків. Запропоновано алгоритм багатофакторної автентифікації з використанням нульових водяних знаків. Проведені тестування виявили, що нульові водяні знаки не підходять для підтвердження авторства, оскільки не можуть гарантувати відмінність між подібними зображеннями або між зображеннями одного автора. Проте існує перспектива використання даної технології в схемах багатофакторної автентифікації, в поєднанні з іншими методами автентифікації (такими як пароль, біометрія, токени).

Ключові слова: водяний знак; нульовий водяний знак; автентифікація; авторство; DWT; SVD; CA; NCC; PSTR.

Табл. 1. Іл. 6. Бібліогр.: 14 назв.

UDC 004.056.5

Methods of information protection based on quantum image steganography / *O.I. Fediushyn, Y.V. Holovko, A.O. Smirnov, V.M. Sukhoteplyi, O.V. Chechui* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 44 – 55.

The purpose of the article is to study information protection methods based on quantum image steganography. Traditional image protection techniques have limitations in terms of resilience and reliability. The use of quantum computing, particularly the NEQR model, offers significant advantages in enhancing the security and efficiency of watermarks and steganography. The article analyzes existing approaches to image representation in quantum states and demonstrates their superiority in digital image protection. A simulation of steganography using quantum algorithms is presented, showing improved information security.

The article will be useful for information security professionals involved in data protection using modern quantum technologies.

Key words: steganography; watermarks; images; qubits; hiding.

1 tabl. 14 fig. Ref: 10 items.

УДК 004.056.5

Методи захисту інформації на основі квантової стеганографії зображень / *О.І. Федюшин, Є.В. Головко, А.О. Смірнов, В.М. Сухотеплий, О.В. Чечуй* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 44 – 55.

Метою статті є дослідження методів захисту інформації на основі квантової стеганографії зображень. Традиційні методи захисту зображень мають обмеження щодо стійкості та надійності. Використання квантових обчислень, зокрема моделі NEQR, забезпечує значні переваги для підвищення безпеки та ефективності водяних знаків і стеганографії. Стаття аналізує існуючі підходи до представлення зображень у квантових станах та демонструє їхню перевагу в захисті цифрових зображень. Представлено симуляцію стеганографії за допомогою квантових алгоритмів, яка підтверджує покращення безпеки інформації.

Стаття буде корисною фахівцям у галузі інформаційної безпеки, які займаються питаннями захисту даних за допомогою сучасних квантових технологій.

Ключові слова: стеганографія; водяні знаки; зображення; кубіти; приховування.

Табл. 1. Іл. 14. Бібліогр.: 10 назв.

UDC 621.3.095.221

Technical channel of information leakage by side electromagnetic re-radiation of auxiliary technical means and systems / *V.I. Zabolotnyi, A.M. Oleynikov, D.M. Zabolotnyi, A.K. Kustov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 56 – 63.

The design and implementation of a complex of technical information protection requires the identification and analysis of technical channels of information leakage. The work is devoted to the analysis of the channel of side electromagnetic re-radiation of auxiliary technical means and systems provided by the main technical means. The article

presents approaches to the methodological support of the process of creating and applying separate quantitative models of TKVI, which are formed by the processes of the formation of side electromagnetic radiation, their impact on the random antenna of auxiliary technical means and systems, re-radiation of the field outside the controlled zone, their further composition with the radiation field of the main technical tool. Research is based on the theory of the electromagnetic field, current organizational documents in the field of technical information protection practice.

Expedient areas of further research arising from related problems, namely the development of quantitative models, have been identified: leakage of information due to the phenomenon of electric induction of the electric component of the field on a random electric antenna; leakage of information due to the reradiation of the electric field component of the electric field antenna created by the OTZ magnetic antenna and vice versa.

Key words: spurious electromagnetic radiation; technical protection of information; technical channel of information leakage.

3 fig. Ref: 8 items.

УДК 621.3.095.221

Технічний канал витоку інформації побічними електромагнітними перевипромінюваннями допоміжних технічних засобів і систем / *В.І. Заболотний, А.М. Олейніков, Д.М. Заболотний, А.К. Кустов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 56 – 63.

Проєктування та впровадження комплексу технічного захисту інформації потребує виявлення і аналізу технічних каналів витоку інформації. Робота присвячена аналізу каналу побічних електромагнітних перевипромінювань допоміжних технічних засобів і систем, наведених основними технічними засобами.

Наведено підходи до методичного забезпечення процесу створення і застосування окремих кількісних моделей ТКВІ, що утворюються процесами формування побічних електромагнітних випромінювань, їх впливом на випадкову антену допоміжних технічних засобів і систем, перевипромінювань поля за межі контрольованої зони, їх подальшого складення з полем випромінення основного технічного засобу. Дослідження основані на теорії електромагнітного поля, чинних організаційних документів в області практики технічного захисту інформації. Визначено доцільні напрями подальших досліджень, що витікають з пов’язаних проблем, а саме розробки кількісних моделей: витоку інформації за рахунок явища електричної індукції електричної складової поля на випадкову електричну антену; витоку інформації за рахунок перевипромінювання електричною антеною ДТЗС, створеною магнітною антеною ОТЗ складової електричного поля і навпаки.

Ключові слова: побічне електромагнітне випромінення; технічний захист інформації; технічний канал витоку інформації.

Іл. 3. Бібліогр.: 8 назв.

UDC 621.372(075)

Detection of web attacks via HTTP requests using NLP techniques / *M.S. Kavetskiy, V.I. Ruzhentsev* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 64 – 75.

The work focuses on improving web attack detection methods through the analysis of HTTP traffic using natural language processing (NLP) techniques and transformer-based models, particularly BERT. The relevance of research in the field of web attack detection is underscored by the significant achievements of the model trained on an extended dataset containing 195,000 records. The developed model based on BERT demonstrates high efficiency in detecting web attacks due to its deep contextual understanding and modern Word Piece tokenizer, which better handles rare words. Unlike methods such as Doc2Vec, LSTM-CNN, or Isolation Forest, our model accounts for global word relationships enhancing its accuracy.

Some previous studies have limitations; notably, some do not utilize state-of-the-art architectures, which limits their ability to achieve high model accuracy. Additionally, while using modern architectures, other studies operate with small datasets, limiting their capability to effectively detect various attack types and ensure high detection quality. In the context of these challenges, the created model was trained on an extended dataset, resulting in significantly better performance compared to leading analogs in the field of web attack detection. The high balanced accuracy of the model at 0.9998 confirms its effectiveness and reliability, making it a potentially important tool for cyber security applications.

Key words: HTTP request; NLP; BERT; transformer architecture; web attack detection; machine learning.

3 fig. Ref: 9 items.

УДК 621.372(075)

Виявлення веб-атак по HTTP запитам з використання технік NLP / *М.С. Кавецький, В.І. Руженцев* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 64 – 75.

Робота зосереджена на вдосконаленні методів виявлення веб-атак через аналіз HTTP-трафіку з використанням технік обробки природної мови (NLP) та моделей на базі трансформерів, зокрема BERT. Актуальність дослідження в галузі виявлення веб-атак підкреслюється значними досягненнями моделі, натренованої на розширеному датасеті, який містив 195 тисяч записів. Розроблена модель на основі BERT демонструє високу ефективність у виявленні веб-атак завдяки глибокому контекстному розумінню та сучасному токенізатору WordPiece, який краще обробляє рідкісні слова. На відміну від методів Doc2Vec, LSTM-CNN або Isolation Forest, наша модель враховує глобальні взаємозв'язки між словами, що підвищує її точність.

Деякі попередні дослідження мають свої недоліки, зокрема, деякі з них не використовують новітні архітектури, що обмежує їх здатність досягти високої точності моделі. Крім того, інші дослідження, хоч і використовують сучасні архітектури, працюють з малими датасетами, що обмежує їхню здатність ефективно виявляти різноманітні типи атак та забезпечувати високу якість їх виявлення. Створена модель тренована на розширено-

му датасеті, що дозволило досягти значно кращих результатів у порівнянні з провідними аналогами в галузі виявлення веб-атак. Висока збалансована точність моделі на рівні 0,9998 підтверджує її ефективність та надійність, роблячи її потенційно важливим інструментом для кібербезпекових застосувань.

Ключові слова: HTTP запит; NLP; BERT; архітектура transformer; виявлення веб-атак; машинне навчання.
Іл. 3. Бібліогр.: 9 назв.

UDC 004.056.55

Refining security assessments of quantum-resistant asymmetric encryption standards taking into account the structure of q-ary lattices / *S.O. Kandii, I.D. Gorbenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 76 – 92.

The article provides a comprehensive analysis of modern quantum-resistant cryptographic standards security based on lattice theory, such as DSTU 8961:2019 ("Skelya"), CRYSTALS-Kyber, and CRYSTALS-Dilithium. These standards are becoming increasingly popular for solving various practical tasks due to their resistance to attacks that can be implemented on quantum computers. As quantum computing gradually transitions from the theoretical to the practical realm, there is an urgent need for the development and improvement of security models capable of addressing these new challenges. This article focuses on applying the cryptanalysis model developed in previous works to specific cryptographic standards based on lattices. Special attention is given to refining security estimates by considering the algebraic structure of q-ary lattices, which form the foundation of the cryptographic problems underlying these standards. It was found that when considering the algebraic structure of q-ary lattices, security estimates differ significantly from those obtained using the GSA model. In particular, for the key encapsulation mechanisms of DSTU 8961:2019 and CRYSTALS-Kyber, the difference between the estimates in these two models can range from 20 to 30 bits of security, with the refined estimates indicating that existing attacks are less effective than previously thought. It was also revealed that for NIST Level 1 security, decoding attacks show better performance compared to embedding attacks, whereas for NIST Level 5 security, the effectiveness of decoding attacks decreases significantly, falling behind embedding attacks. Thus, the results highlight the importance of accounting for the algebraic structure of lattices in obtaining more accurate security assessments. This allows for a better understanding of potential threats and the optimization of existing lattice-based cryptographic transformations.

Key words: NTRU; SIS; LWE; cryptanalysis; lattice cryptography; Crystals-Dilithium.
16 tabl. 8 fig. Ref: 12 items.

УДК 004.056.55

Уточнення оцінок безпеки квантово-стійких стандартів асиметричного шифрування з врахуванням структури q-арних решіток / *С.О. Кандій, І.Д. Горбенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 76 – 92.

Проведено всебічний аналіз безпеки сучасних квантово-стійких криптографічних стандартів, заснованих на теорії решіток, таких як ДСТУ 8961:2019 («Скеля»), CRYSTALS-Kyber та CRYSTALS-Dilithium. Ці стандарти набувають все більшої популярності для вирішення різних практичних завдань завдяки їхній стійкості до атак, що можуть бути реалізовані на квантових комп'ютерах. Оскільки квантові обчислення поступово переходять із теоретичної площини в практичну, виникає нагальна необхідність у розробці та вдосконаленні моделей безпеки, здатних протистояти цим новим викликам. Статтю присвячено застосуванню розробленої у попередніх роботах моделі криптоаналізу до конкретних криптографічних стандартів, заснованих на решітках. Особливу увагу приділено уточненню оцінок безпеки з урахуванням алгебраїчної структури q-арних решіток, на яких базуються криптографічні проблеми, що лежать в основі зазначених стандартів. Встановлено, що при врахуванні алгебраїчної структури q-арних решіток оцінки безпеки значно відрізняються від тих, що були отримані за допомогою моделі GSA. Зокрема, для механізмів інкапсуляції ключів ДСТУ 8961:2019 та CRYSTALS-Kyber різниця між оцінками у цих двох моделях може становити від 20 до 30 біт безпеки, причому уточнені оцінки свідчать про те, що існуючі атаки є менш ефективними, ніж вважалося раніше. Було виявлено, що для 1 рівня безпеки NIST атака декодування показує кращі результати порівняно з атакою вкладення, тоді як для 5 рівня безпеки ефективність атак декодування значно знижується, поступаючись атакам вкладення. Таким чином, отримані результати свідчать про важливість урахування алгебраїчної структури решіток для отримання точніших оцінок безпеки. Це дозволяє краще зрозуміти природу потенційних загроз та оптимізувати існуючі криптографічні перетворення на основі решіток.

Ключові слова: NTRU; SIS; LWE; криптоаналіз; криптографія на решітках; Crystals-Dilithium.
Табл. 16. Іл. 8. Бібліогр.: 12 назв.

UDC 004.056:519.2

Probabilistic properties of solutions to the equation system of keystream generators with irregular motion / *A.M. Alekseychuk, I.V. Samoylov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 93 – 102.

The traditional basis for constructing modern stream ciphers is keystream generators, which are based on linear shift registers and nonlinear complexity elements. One of the well-known methods for increasing the resistance of such generators, in particular, against algebraic and correlation attacks, is using irregularity into register's motion process. The most popular keystream generators with irregular motion, used in stream ciphers are A5/1, Alpha1, LILI-128 and others, were thoroughly studied in the 1990s and 2000s. However, specialists' interest remains relevant, it's evidenced by recent publications dedicated to new attacks on the A5/1 cipher and some other stream ciphers, constructed based on keystream generators with irregular motion. On the one hand, this known methods for covering wide range of different

keystream generators, but on the other hand, it reduces the accuracy and adequacy of conclusions about the resistance of specific ones.

The main result of the article is probabilistic properties of solutions to the system of keystream generation equitation of keystream generators with irregular motion. In particular, a matrix representation for the average number of solutions to these systems of equitation has been obtained and conditions have been established under which a combinational generator with external control is irreversible by Huffman. Additionally, sufficient conditions for the exponential growth of the average number of solutions to keystream generation of this generator, depending on the length of output sequence, have been obtained, along with analytical expressions and estimates of probability distributions for sums of random vectors produced by motion control blocks of a certain class of combinational keystream generators. The obtained results can be applied to solving the problems of evaluation the resistance of keystream generators with external controlled motion and justifying the requirements for the cryptographic parameters of complexity nodes in such generators in such generators, which determine their resistance to correlation attacks.

Key words: cryptographic information protection; stream cipher; keystream generator with irregular motion; finite automaton; Huffman irreversibility; keystream generation equation system; resistance justification.

Ref: 18 items.

УДК 004.056:519.2

Ймовірнісні властивості розв'язків систем рівнянь гамоутворення генераторів гами з нерівномірним рухом / *А.М. Олексійчук, І.В. Самойлов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 93 – 102.

Традиційною основою для побудови сучасних потокових шифрів є генератори гами, які базуються на лінійних регістрах зсуву та нелінійних елементах ускладнення. Одним з відомих методів підвищення стійкості таких генераторів, зокрема, до алгебраїчних та кореляційних атак, є введення нерівномірності в процес руху регістрів. Найвідоміші на сьогодні генератори гами з нерівномірним рухом, які застосовуються у потокових шифрах A5/1, Alpha1, LILI-128 та деяких інших, ретельно досліджено ще у дев'яності та нульові роки. Проте інтерес фахівців до таких генераторів зберігається і сьогодні, про що свідчать, зокрема, нещодавні публікації, присвячені новим атакам на шифр A5/1 та деякі інші потокові шифри, побудовані на базі генераторів гами з нерівномірним рухом. При цьому відомі методи оцінювання стійкості таких генераторів, розвинуті переважно в дев'яності роки, базуються на спрощеному описі їх функціонування, що, з одного боку, це надає змогу охопити широкий клас різноманітних генераторів гами, але, з іншого – знижує точність та адекватність висновків про стійкість окремих з них.

Основними результатами статті є ймовірнісні властивості розв'язків систем рівнянь гамоутворення генераторів гами з нерівномірним рухом. Зокрема, отримано матричне представлення для середнього числа розв'язків зазначених систем рівнянь та встановлено умови, за яких комбінувальний генератор із зовнішнім управлінням є необоротним за Гаффманом. Отримано також достатні умови експоненційного росту середнього числа розв'язків системи гамоутворення цього генератора від довжини його вихідної послідовності, аналітичні вирази та оцінки розподілів ймовірностей сум випадкових векторів, які виробляються блоками управління рухом певного класу комбінувальних генераторів гами. Отримані результати можуть бути застосовані при розв'язанні задач оцінювання стійкості генераторів гами із зовнішнім управлінням рухом та обґрунтування вимог до криптографічних параметрів вузлів ускладнення таких генераторів гами, що визначають їхню стійкість відносно кореляційних атак

Ключові слова: криптографічний захист інформації; потоковий шифр; генератор гами з нерівномірним рухом; скінченний автомат; оборотність за Гаффманом; система рівнянь гамоутворення; обґрунтування стійкості.

Бібліогр.: 18 назв.

UDC 621.391:519.2

The problem of finding periodicity in quantum cryptanalysis of group cryptography algorithms / *Y. Kotukh, G. Khalimov, I. Dzhura* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 103 – 109.

The article explores the fundamental role of quantum period determination algorithms, particularly in the context of cryptography. The work focuses on quantum algorithms that exploit periodicity, such as Shor's algorithm, which is central to efficient integer factorization. It highlights the challenges quantum algorithms face when applied to non-Abelian groups such as the Suzuki, Hermitage, and Rie groups, which exhibit complex periodic structures that are difficult to solve with existing quantum methods. The study delves into the structure and properties of these groups, explaining the complexity of their representations and the challenges posed by the quantum Fourier transform (QFT) in these cases. This contrasts the relative ease with which Abelian groups can be handled by quantum algorithms with the exponential complexity encountered with non-Abelian groups. The study provides a comparative analysis of the computational complexity between classical and quantum approaches to find the period in different types of groups, highlighting that while quantum algorithms offer exponential speedup for Abelian cases, non-Abelian structures remain a frontier for further research. The conclusion calls for continued research in quantum representation theory and cryptanalysis, especially for non-Abelian groups where current quantum methods have not yet provided efficient solutions. The problem of finding the period has been identified as critical to the advancement of both quantum computing and cryptographic applications.

Key words: quantum period search problem; post-quantum protection; hidden subgroup problem.

Table 1. Ref.: 15 items.

УДК 621.391:519.2

Проблема знаходження періодичності в квантовому криптоаналізі алгоритмів групової криптографії / *Є.В. Котух, Г.З. Халімов, І.Є. Джура* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 103 – 109.

Стаття досліджує основоположну роль квантових алгоритмів визначення періоду, зокрема в контексті криптографії. Робота зосереджена на квантових алгоритмах, які використовують періодичність, таких як алгоритм Шора, який є центральним для ефективної цілочисельної факторизації. Він підкреслює проблеми, з якими стикаються квантові алгоритми при застосуванні до неабелевих груп, таких як групи Сузукі, Ерміта та Рі, які демонструють складні періодичні структури, які важко вирішити за допомогою існуючих квантових методів. Дослідження заглиблюється в структуру та властивості цих груп, пояснюючи складність їх уявлень і проблеми, які представляє в цих випадках квантове перетворення Фур'є (QFT). Це протиставляє відносну легкість, з якою абелеві групи можуть бути розглянуті за допомогою квантових алгоритмів, з експоненціальною складністю, яка зустрічається з неабелевими групами. Дослідження забезпечує порівняльний аналіз обчислювальної складності між класичним і квантовим підходами для пошуку періоду в різних типах груп, підкреслюючи, що хоча квантові алгоритми пропонують експоненціальне прискорення для абелевих випадків, неабелеві структури залишаються межею для подальших досліджень. Висновок вимагає продовження досліджень квантової теорії репрезентації та криптоаналізу, особливо для неабелевих груп, де поточні квантові методи ще не забезпечили ефективних рішень. Проблема пошуку періоду визначена як критична для просування як квантових обчислень, так і криптографічних програм.

Ключові слова: проблема пошуку квантового періоду; постквантовий захист; проблема прихованої підгрупи.

Табл. 1. Бібліогр.: 15 назв.

MEANS OF TELECOMMUNICATIONS
ЗАСОБИ ТЕЛЕКОМУНІКАЦІЙ

UDC 681.3.06:519.248.681

Detection of broadband signals by their spectral features / *I.E. Antipov, O.M. Nikitin* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 110 – 117.

The paper is about a method for detecting wideband signals on a noise background, which uses the symmetry property of its spectrum. The authors of the article proposed the method. An analysis of literary sources has shown that the detection of WBS is relevant for electronic warfare and information security tasks, and the speed of detection is very important. Also, based on a literature review, it was concluded that the energy detection method is not effective enough, including due to the need for long-term accumulation of signal energy. As the SNR deteriorates, the required accumulation time increases. In paper proposes to use the fact that the energy spectra of most WBS, in particular phase-shifted one, are symmetrical with respect to their central frequency.

The proposed method involves applying a Fast Fourier Transform to the analyzed mixture of signal and noise. Next, a function is equal to the sum of the products of spectral samples taken symmetrically with respect to the expected central frequency calculated. It is analytically shown that such processing makes it possible to detect the PM WBS on a white Gaussian noise background of based on the characteristic maximum of this function. If this function is calculated only for noise (without a signal), then the maximum does not arise, since white noise has a uniform spectral density.

If the value of the average frequency of the signal is not known, then, according to the proposed algorithm, it is necessary to perform several cycles of calculations for different estimated average frequencies. Thus, both the presence of a signal in the mixture and the value of its central frequency are detected. Computer modeling, the results of which are also presented in the paper, shows that the use of the proposed method can reduce the accumulation time and increase resistance to noise exposure, compared to the energy detection method.

Key words: noise-like signal; detection against noise background; energy detection; mathematical modeling; spectral characteristics.

15 fig. Ref: 13 items.

УДК 681.3.06:519.248.681

Виявлення широкосмугових сигналів за особливостями їх спектра / *І.Є. Антіпов, О.М. Нікітін* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 110 – 117.

Стаття присвячена методу виявлення широкосмугових сигналів і натомість шуму, у якому використовується властивість симетрії його спектра. Метод запропонований авторами статті.

Аналіз літературних джерел показав, що виявлення ШПС є актуальним для задач радіоелектронної боротьби та захисту інформації, причому дуже важливою є швидкість виявлення. Також на підставі огляду літератури зроблено висновок, що метод енергетичного виявлення недостатньо ефективний, у тому числі через необхідність тривалого накопичення енергії сигналу. При погіршенні відношення сигнал/шум потрібний час накопичення зростає. У роботі запропоновано використати той факт, що енергетичні спектри більшості ШПС, зокрема фазоманіпульованих ШПС, симетричні відносно своєї центральної частоти.

Запропонований метод передбачає застосування швидкого перетворення Фур'є до аналізованої суміші сигналу та шуму. Далі обчислюється функція, що дорівнює сумі результатів перемноження спектральних відліків, взятих симетрично відносно центральної частоти, що очікується. Аналітично показано, що така обробка дозволяє виявити ФМ ШПС на тлі білого Гаусового шуму за характерним максимумом цієї функції. Якщо ж цю

функцію обчислювати для шуму без сигналу, то максимум не виникає, оскільки білий шум має рівномірну спектральну щільність. Якщо значення середньої частоти сигналу невідомо, то, згідно з запропонованим алгоритмом, необхідно зробити кілька циклів обчислень для різних передбачуваних середніх частот. Таким чином, виявляється факт наявності сигналу в суміші і значення його центральної частоти. Комп'ютерне моделювання, результати якого представлені у статті, показує, що застосування запропонованого методу дозволяє зменшити час накопичення і збільшити стійкість до шумового впливу, порівняно з методом енергетичного виявлення.

Ключові слова: шумоподібний сигнал; виявлення на тлі шуму; енергетичне виявлення; математичне моделювання; спектральні характеристики.

Іл. 15. Бібліогр.: 13 назв.

RADIO ELECTRONIC SYSTEMS
РАДІОЕЛЕКТРОННІ СИСТЕМИ

UDC 621.396.96

Mathematical model of the location channel of the contour of the adaptation of radioacoustic atmospheric sounding systems / *A.V. Kartashov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 118 – 128.

The system of radio acoustic sounding (RAS) of the atmosphere is used in the process of solving actual scientific and applied tasks to ensure flights of aircraft, weather forecasting, studying the atmosphere, etc. However, the effectiveness of the existing radio-acoustic means is insufficient and there are practical needs for developing more promising structures and algorithms that will be implemented during the construction of specific types of sounding stations. Among the existing limitations of radio acoustic sounding systems, the main one is the violation of the Bragg condition, which determines the ratio between the lengths of acoustic and electromagnetic waves.

The article considers the method of adapting RAS systems by changing the frequency of the sounding radio signal in order to fulfill the Bragg condition as the emitted acoustic pulse signal moves along the sounding path. It is shown that to increase the efficiency of the frequency adaptation method, it is necessary to use an adequate mathematical model of the information location channel, which describes the main features of the scattering of radio waves on the acoustic wave parcel, in the control circuit of the frequency of the sounding radio signal.

The mathematical model of the location channel is considered. The study of the main types of probing acoustic and electromagnetic signals was performed using scattering bodies in the spectral representation. It has been revealed that in the presence of the Bragg condition detuning, the spectrum of the radio signal scattered on the sound has an asymmetric shape, and this is the main reason for the appearance of systematic errors in the results of sound speed measurements. Eliminating the identified errors will improve the efficiency of the adaptation devices and the RAS system as a whole.

Key words: remote sensing of the atmosphere; method; algorithm; adaptation; estimation of parameters; management; model, location information channel; temperature; sounding signal.

5 fig. Ref: 20 items.

УДК 621.396.96

Математична модель локаційного каналу контуру адаптації систем радіоакустичного зондування атмосфери / *О.В. Карташов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 118 – 128.

Системи радіоакустичного зондування (РАЗ) атмосфери використовується в процесі вирішення актуальних науково-прикладних завдань з забезпечення польотів літальних апаратів, прогнозу погоди, вивчення атмосфери та ін. Проте ефективність існуючих радіоакустичних засобів є недостатньою і існують потреби практики з розробки більш перспективних структур та алгоритмів, що реалізуватимуться при побудові конкретних видів станцій зондування. Основним серед існуючих обмежень систем радіоакустичного зондування є порушення умови Брегга, що визначає співвідношення між довжинами акустичної та електромагнітної хвиль.

В статті розглядається метод адаптації систем РАЗ шляхом зміни частоти зондувального радіосигналу з метою виконання умови Брегга у міру переміщення випромінюваного акустичного імпульсного сигналу трасою зондування. Показано, що для підвищення ефективності методу частотної адаптації необхідно використовувати в контурі управляння частотою зондувального радіосигналу адекватну математичну модель інформаційного локаційного каналу, яка описує основні особливості розсіювання радіохвиль на акустичної хвильової посилці.

Розглянуто математичну модель локаційного каналу, з використанням тіл розсіяння в спектральному поданні виконано дослідження основних видів зондувальних акустичних і електромагнітних сигналів. Виявлено, що при наявності розстройки умови Брегга спектр розсіяного на звуці радіосигналу має несиметричну форму і це є основною причиною появи систематичних похибок в результатах вимірювань швидкості звуку. Усунення виявлених похибок дозволить підвищити ефективність функціонування пристроїв адаптації та системи РАЗ в цілому.

Ключові слова: дистанційне зондування атмосфери; метод; алгоритм; адаптація; оцінка параметрів; управляння; модель, локаційний інформаційний канал; температура; зондувальний сигнал.

Іл. 5. Бібліогр.: 20 назв.

PHYSICS OF DEVICES, ELEMENTS AND SYSTEMS
ФІЗИКА ПРИЛАДІВ, ЕЛЕМЕНТІВ І СИСТЕМ

UDC 615.844

Status and prospects for the use of diagnostic tools based on the method of gas-discharge visualization / *O.M. Zinchenko, V.P. Oliinyk, P.M. Podpruzhnykov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 129 – 143.

Diagnostic tools based on the gas-discharge visualization method were created for scientific and applied research of biological and non-biological objects, the state of the environment. The diagnostic essence of the method is the connection of the glow of an electric discharge around the object in the atmospheric air with the electrophysical properties of the internal and surface structures of the object, which in turn are interpreted as medical and biological indicators for living organisms, or technical characteristics for non-living ones structures. With the help of these tools, you can diagnose the human body, study liquids and materials of organic and inorganic origin. The disadvantages of the method include the dependence of informative indicators of discharge images not only on the properties of the object, but also on the constancy of the parameters of the surrounding environment. There are no uniform metrological requirements for the technical means of gas discharge visualization, which prevents their practical application. The purpose of the research is to find technical solutions for the construction of gas-discharge visualization tools to achieve the potential possibilities of their practical application. The analysis of the physical and technical conditions of gas-discharge visualization showed that with the relative constancy of the physical regularities of the discharge, the modes of measurement of the informative indicators of the images in these means differ significantly (amplitude, frequency, duration, form of alternating voltage). The known designs differ in the type of sensor, method and device of primary radiation registration, algorithms for reproduction and processing of the diagnostic image. It is proposed to standardize the technical characteristics of these tools and algorithms for obtaining informative indicators to compare correctly the diagnostic results obtained by different researchers (especially in the medical field). It is advisable to expand the use of gas-discharge visualization tools for controlling the composition of technical liquids, non-destructive flaw detection of materials and structures. It is recommended to add information about the characteristics of the surrounding atmospheric air (temperature, humidity, pressure) when conducting diagnostics.

Key words: diagnostics; object; gas; discharge; electric; radiation; spectrum; pulse; high-voltage; frequency.

2 tabl. 6 fig. Ref: 25 items.

УДК 615.844

Стан та перспективи застосування засобів діагностики на основі методу газорозрядної візуалізації / *О.М. Зінченко, В.П. Олійник, П.М. Подпружников* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 129 – 143.

Засоби діагностики на основі методу газорозрядної візуалізації створені для наукових та прикладних досліджень біологічних і небіологічних об’єктів, стану навколишнього середовища. Діагностична сутність методу – зв’язок світіння електричного розряду довкола об’єкта в атмосферному повітрі з електрофізичними властивостями внутрішніх та поверхневих структур об’єкта, які у свою чергу інтерпретуються медико-біологічними показниками для живих організмів, або технічними характеристиками для неживих структур. За допомогою цих засобів можна проводити діагностику організму людини, вивчати рідини і матеріали органічного та неорганічного походження. До недоліків методу відносять залежність інформативних показників зображень розряду не тільки від властивостей об’єкта, а і сталості параметрів оточуючого середовища. Відсутні єдині метрологічні вимоги до технічних засобів газорозрядної візуалізації, що стримує їх практичне застосування. Мета дослідження – пошук технічних рішень побудови засобів газорозрядної візуалізації для досягнення потенціальних можливостей їх практичного застосування. Аналіз фізичних та технічних положень газорозрядної візуалізації показав, що при відносній сталості фізичних закономірностей виникнення розряду режими вимірювання інформативних показників зображень в цих засобах суттєво відрізняються (амплітуда, частота, тривалість, форма змінної напруги). Відомі конструкції відрізняються видом сенсору, способом та пристроєм реєстрації первинного випромінювання, алгоритмами відтворення та оброблення діагностичного зображення. Пропонується для коректного порівняння результатів діагностики отриманих різними дослідниками (особливо в медичній галузі) провести стандартизацію технічних характеристик цих засобів та алгоритмів отримання інформативних показників. Доцільно розширити використання засобів газорозрядної візуалізації для контролю складу технічних рідин, неруйнівної дефектоскопії матеріалів і конструкцій. Рекомендується додавати інформацію про характеристики оточуючого атмосферного повітря (температура, вологість, тиск) при проведенні діагностики.

Ключові слова: діагностика; об’єкт; газ; розряд; електричний; випромінювання; спектр; імпульс; напруга; частота.

Табл. 2. Іл. 6. Бібліогр.: 25 назв.

UDC 537.874

Characteristics of a controlled Bragg reflection waveguide with gyrotropic cladding / *Y.Y. Demydenko, V.V. Novytskyi, Y.M. Odarenko, O.O. Shmat’ko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 144 – 150.

A planar Bragg waveguide with a hollow channel, the cladding of which contains periodically arranged dielectric and gyrotropic ferrite layers, has been theoretically investigated. A two-dimensional model of the waveguide is developed taking into account the frequency dependence of the components of the magnetic permeability tensor of gyrotropic layers. Results of numerical calculations (using the finite element method) show regularities of influence of the external

magnetic field (pointed to the direction transverse to the longitudinal axis of the structure) induction on the dispersion characteristics of the Bragg waveguide cladding and its spectral characteristics. The orientation of the external magnetic field corresponds to the Voigt configuration. An increase in the induction of the magnetic field leads to a significant transformation of the band gap of the waveguide channel cladding. There is an increase in its width and a shift towards higher frequencies. The results of the calculations of the spectral characteristics indicate the corresponding changes that occur in the frequency band of the Bragg waveguide. It is shown that the increase in the width of the transmission zone of the waveguide occurs mainly in the high-frequency region of the considered part of the spectrum. The calculated spatial distribution of the electric field in the studied structure indicates a high degree of localization of electromagnetic energy in the hollow waveguide channel within its transmission zone. This makes it possible to simplify the model of the studied structure due to the exclusion from consideration of losses in gyrotropic layers. On the basis of controlled Bragg waveguides, a variety of functional devices in the microwave and optical ranges can be developed, the operating characteristics of which change under the influence of an external magnetic field.

Key words: Bragg waveguide; magnetophotonic crystal; gyrotropic layers; ferrite; bandwidth.

5 fig. Ref: 25 items.

УДК 537.874

Характеристики керованого Брегґівського хвилеводу з гіротропними оболонками / *Є.Є. Демиденко, В.В. Новицький, Є.М. Одаренко, О.О. Шматько* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 144 – 150.

Теоретично досліджено планарний Брегґівський хвилевід з пустотілим каналом, оболонки якого містять періодично розташовані діелектричні та гіротропні феритові шари. Побудована двовимірна модель хвилеводу, в якій враховується частотна залежність компонентів тензору магнітної проникності гіротропних шарів. Результати чисельних розрахунків (методом скінченних елементів) демонструють закономірності впливу індукції зовнішнього магнітного поля (спрямованого у напрямку поперечному до поздовжньої осі структури) на дисперсійні характеристики оболонки Брегґівського хвилеводу та його спектральні характеристики. Орієнтація зовнішнього магнітного поля відповідає конфігурації Войгта. Збільшення індукції магнітного поля призводить до суттєвої трансформації забороненої зони оболонки хвилеводного каналу. Спостерігається збільшення її ширини та зсув у бік більш високих частот. Результати розрахунків спектральних характеристик свідчать про відповідні зміни, які відбуваються із частотною зоною пропускання Брегґівського хвилеводу. Показано, що збільшення ширини зони пропускання хвилеводу відбувається переважно у високочастотній області розглянутої ділянки спектру. Розрахований просторовий розподіл електричного поля в досліджуваній структурі свідчить про високий ступінь локалізації електромагнітної енергії в пустотілому каналі хвилеводу в межах його зони пропускання. Це дозволяє спростити модель досліджуваної структури через виключення із розгляду втрат в гіротропних шарах. На основі керованих Брегґівських хвилеводів можуть бути розроблені різноманітні функціональні пристрої мікрохвильового та оптичного діапазонів, робочі характеристики яких змінюються під впливом зовнішнього магнітного поля.

Ключові слова: Брегґівський хвилевід; магнітофотонний кристал; гіротропні шари; ферит; смуга пропускання.

Іл. 5. Бібліогр.: 25 назв.

UDC 551.501.7

Development of a model of a biomedical system of vital activity under the influence of electromagnetic radiation / *V.V. Semenets, T.E. Stytsenko, A.B. Grigoriev* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 151 – 155.

An assessment of the effectiveness of the biomedical protection system has been developed based on the assessment of the dose of the harmful effects of electromagnetic radiation, which allows comparing the existing technical systems according to their level of safety. A functional model of the biomedical vital system with various sources of electromagnetic radiation of the microwave range has been developed based on deterministic criteria, which allow evaluating with a certain probability the performance of the task of functioning of the vital activity system. The statistical criterion is built taking into account the cost of the system, which allows us (when building biological protection systems based on general models) to obtain calculation formulas connecting the newly defined system quality criterion with parameters that can be relatively easily measured and normalized from biological and sanitary considerations, which determines the practical value of the work.

Key words: structural optimization; protective device; harmful factor; monotonic function; biomedical protection; criterion.

1 fig. Ref: 7 items.

УДК 551.501.7

Розробка моделі біомедичної системи життєдіяльності при впливі електромагнітного випромінювання / *В.В. Семенець, Т.Е. Стиценко, О.В. Григор'єв* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 151 – 155.

Розроблено оцінку ефективності роботи системи біомедичного захисту на основі оцінки дози шкідливого впливу електромагнітного випромінювання, що дозволяє порівняти існуючі технічні системи за рівнем їхньої безпеки. Розроблено функціональну модель біомедичної системи життєдіяльності з різними джерелами електромагнітного випромінювання НВЧ діапазону на основі детерміністських критеріїв, які дозволяють з певною ймовірністю оцінити виконання завдання функціонування системи життєдіяльності.

Статистичний критерій побудований з урахуванням вартості системи, що дозволяє при побудові систем біологічного захисту на підставі загальних моделей отримати розрахункові формули, що пов'язують ново визначений критерій якості системи з параметрами, які можна порівняно легко виміряти та нормувати з біологічних та санітарних міркувань, що і визначає практичну цінність роботи.

Ключові слова: структурна оптимізація; захисний пристрій; шкідливий фактор; монотонна функція; біомедичний захист; критерій.

Іл. 1. Бібліогр.: 7 назв.

RELATED PROBLEMS OF RADIO ENGINEERING
СУМІЖНІ ПРОБЛЕМИ РАДІОТЕХНІКИ

UDC 621.396 + 523.53

Model for estimating the linear electron density of the trail created by a meteoroid / *O.V. Holovan, V.M. Kharchenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №218. P. 156 – 165.

An assessment of the linear electron density of the trail (LEDTr) created by a meteoroid when passing through the Earth's atmosphere at a given observation altitude is used to calculate the power of the received radio signal and predict the reflectivity of the trail in the process of its further transformation.

In the published works, the formulas for calculating the LEDTr do not show its relation to altitude, and there are no explicitly specified parameters of the meteoroid body and atmosphere.

When developing the presented model for estimating LEDTr, the main physical processes occurring during the interaction of a meteoroid with atoms and molecules of the atmosphere were taken into account. The proposed calculation of the LEDTr includes an assessment of the midsection of the meteoroid's head, the concentration of particles (numerical density) in the atmosphere, and the value of the integral ionization coefficient at the observation altitude. The necessary calculation formulas were obtained, and a software-implemented calculation algorithm was presented.

Unlike to the well-known formulas for calculating LEDTr and semi-empirical models obtained by fitting to the results of radar observations, the presented technique allows calculations to be carried out for explicitly specified parameters of the meteoric body and atmosphere. It is important to note that when estimating the mass loss of the meteoroid, a recurrent method was used to calculate the mass and velocity as its altitude changed. This method made it possible to take into account changes in the mass and midsection of the meteoroid during its movement. The obtained relation of LEDTr to altitude corresponds well to the luminosity curves of the meteor trail, and the value of the electron density at the ionization maximum corresponds to the known observational results.

Estimating the LEDTr for a given (measured) reflection point makes it possible to calculate the amplitude-time characteristics (ATCH) of radio signals, which allows for the creation of a model image of a scattered signal. Comparison of the ATCH of the received signal with the calculated model images allows one to make informed assumptions about the characteristics of the meteoroid that generated the ionized trail.

Key words: meteoroid; linear electron density of the trail; midsection; ionization coefficient; numerical density of the atmosphere.

3 tabl. 1 fig. Ref: 12 items.

УДК 621.396 + 523.53

Модель оцінки лінійної електронної щільності сліду, що створюється метеороїдом / *О.В. Головань, В.М. Харченко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 218. С. 156 – 165.

Оцінку лінійної електронної щільності сліду (ЛЕЩСл), що створюється метеороїдом при проходженні атмосфери Землі на заданій висоті спостереження, використовують для розрахунку потужності радіосигналу, що приймається, і прогнозування відбиваючої здатності сліду в процесі його подальшої трансформації.

В опублікованих роботах у формулах для розрахунку ЛЕЩСл не представлено її залежність від висоти, відсутні явно задані параметри метеорного тіла та атмосфери. При розробці представленої моделі оцінки ЛЕЩСл враховувалися основні фізичні процеси, що відбуваються при взаємодії метеороїду з атомами та молекулами атмосфери. Запропонований розрахунок ЛЕЩСл заснований на оцінці міделя головної частини метеорного тіла, концентрації частинок (числової щільності) атмосфери та значення інтегрального коефіцієнта іонізації на висоті спостереження. Були отримані необхідні розрахункові формули та представлений програмно реалізований алгоритм розрахунку. На відміну від відомих формул для розрахунку ЛЕЩСл і напівемпіричних моделей, отриманих підгонкою до результатів радіолокаційних спостережень, представлена методика дозволяє проводити розрахунки для явно заданих параметрів метеорного тіла і атмосфери. Важливо відзначити, що в оцінці втрати маси метеороїда використали рекурентний метод обчислення маси та швидкості при зміні його висоти. Цей метод дозволив врахувати зміни маси та міделю метеороїду в процесі його руху.

Отримана залежність ЛЕЩСл від висоти добре відповідає кривим світності метеорного сліду, а значення електронної щільності в максимумі іонізації відповідає відомим результатам спостережень.

Оцінка ЛЕЩСл для заданої (виміряної) точки відображення забезпечує можливість розрахунку амплітудно-часових характеристик (АЧХ) радіосигналів, що дозволяє створити модельний образ розсіяного сигналу. Зіставлення АЧХ прийнятого сигналу з розрахованими модельними образами дозволяє робити обґрунтовані припущення про характеристики метеороїду, що породив іонізований слід.

Ключові слова: метеороїд; лінійна електронна щільність сліду; міделі; коефіцієнт іонізації; чисельна щільність атмосфери.

Табл. 3. Іл. 1. Бібліогр.: 12 назв.

COLLECTION OF SCIENTIFIC PAPERS
RADIOTEKHNIKA
Issue 218
In English and Ukrainian

ЗБІРНИК НАУКОВИХ ПРАЦЬ
РАДІОТЕХНІКА
Випуск 218
Англійською та українською мовами

Коректор Л.І. Сащенко

.

Підп. до друку 30.09.2024. Формат 60x90/8. Папір офсет. Гарнітура Таймс. Друк. ризограф.
Ум. друк. арк**.** 10,4. Обл.-вид. арк. 9,9. Тираж 300 прим. Зам. № 123. Ціна договір.

Харківський національний університет радіоелектроніки (ХНУРЕ)
Просп. Науки, 14, Харків, 61166.

Оригінал-макет підготовлено і збірник надруковано у ПФ „Колегіум”,
Свідоцтво про внесення суб’єкта видавничої діяльності до Державного реєстру видавців.
Сер. ДК №1722 від 23.03.2004.