УДК 004.05 **DOI:10.30837/rt.2025.1.220.02**

В.В. БОРОДАВКА, В.І. ЄСІН, д-р техн. наук

ДОЦІЛЬНІСТЬ ВИКОРИСТАННЯ МОЖЛИВОСТЕЙ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ЗАБЕЗПЕЧЕННЯ КІБЕРБЕЗПЕКИ ПІДПРИЄМСТВА, ЯКА ҐРУНТУЄТЬСЯ НА КОНЦЕПЦІЇ НУЛЬОВОЇ ДОВІРИ

Вступ

Швидкість розвитку та зміни кіберпростору в останні три-п’ять років вражають не тільки недосвідчених користувачів, а й спеціалістів у галузі інформаційних технологій та інформаційної безпеки. Відбувається стрімкий розвиток не лише обсягу оброблюваних даних, кількості пристроїв, підключених до Інтернету або чисельності застосунків та сервісів, а й самих концепцій і технологій. Всеосяжна цифровізація та перехід бізнесу в онлайн, прискорені пандемією та війною, багато в чому сприяли розвитку цієї тенденції.

Широке використання мов програмування, фреймворків і середовищ розробки, розвиток хмарної інфраструктури та технологій віртуалізації й контейнеризації дають змогу збирати нові застосунки у безпрецедентно короткий термін, в тому числі за допомогою штучного інтелекту (ШІ). З іншого боку, з такою ж швидкістю множаться і кіберзагрози, оскільки зловмисники використовують аналогічні високоефективні інструменти розробки, але у своїх протиправних цілях. Це виводить рівень кіберпротидії на новий рівень: якщо раніше протистояння зі зловмисниками можна було описати як боротьбу розумів і налаштованих засобів захисту інформації, то тепер це вже можна назвати своєрідним поєдинком технологій, у якому задіяний ШІ. Звіт компанії (лідера у сегменті автоматизованої перевірки безпеки) Pentera [1] про стан тестування на проникнення за 2024 р. проливає світло на нагальні проблеми та зміну парадигм кібербезпеки в глобальних організаціях.

Розмови про практичне використання ШІ у сфері кібербезпеки ведуться вже тривалий час, але лише нещодавно ці інструменти стали доступними на ринку. Сьогоднішній рівень зрілості таких рішень дозволив забезпечити впровадження їх у корпоративні середовища для прийняття важливих рішень, причому цілком з виправданими витратами на їх реалізацію, в тому числі враховуючи можливості для зловмисників, що знову відкрилися. А саме, ефективна та своєчасна протидія атакам стає можливою лише із застосуванням технологій ШІ.

Згідно зі звітом «The macroeconomic impact of artificial intelligence» від міжнародної аудит-консалтингової корпорації PwC [2] прогнозується, що до 2030 р. завдяки прискореному розвитку систем ШІ глобальний ВВП може зрости на 14 % (це близько 15,7 трлн доларів США). У сфері комп’ютерних наук ШІ означає здатність машин виконувати завдання, для яких, зазвичай, вимагається наявність людського інтелекту. При цьому ШІ включає багато суміжних областей та технологій, таких як машинне навчання, глибоке навчання, нейронні мережі, обробка природних мов та інші.

На тлі швидкого розвитку інноваційних технологій ШІ набуває широкого застосування для виявлення кіберзагроз і забезпечення посиленого захисту від кібератак, а також сприяє прийняттю обґрунтованих та скоординованих управлінських рішень. Розширюючи інструментарій для виявлення й запобігання кіберзагрозам, автоматизуючи рутинні завдання та скорочуючи час реагування на кіберінциденти, технологія ШІ може значно зміцнити захист від кібератак, що дозволить мінімізувати загрозливі тенденції у цій сфері. За даними компанії IBM [3], збитки від порушення даних у всьому світі можуть бути знижені, якщо організації застосовуватимуть автоматизовані рішення безпеки. Організації, які не застосовували автоматизацію в безпеці, зазнали витрат у зв’язку з порушеннями, які були на 95 % вищими, ніж порушення в організаціях з повністю розгорнутою автоматизацією.

Поряд з цим, щоб захистити сучасне цифрове підприємство, необхідна комплексна стратегія для безпечного доступу у будь-який час і в будь-якому місці до корпоративних ресурсів підприємства (застосунків, застарілих / успадкованих систем, даних, пристроїв тощо) неза-

лежно від того, де вони розташовані [4, 5]. Дійсно, розвиток хмарних обчислень, Інтернету речей, бізнес-партнерів та зростаючої кількості віддалених співробітників підвищує складність захисту цифрових активів підприємства, оскільки точок входу, виходу та доступу до даних істотно збільшилося, а існуючі рішення не завжди здатні реагувати на динамічні зміни, через те, що часто ґрунтуються на статичних наборах правил, брандмауерах, віртуальних приватних мережах (Virtual Private Network – VPN). Тому підприємства стали переосмислювати традиційний підхід до захисту, що базується на забезпеченні безпеки периметра мережі, схиляючись до нової концепції та архітектури захисту. Такою концепцією зараз стала парадигма безпеки, відома як «нульова довіра» (Zero Trust – ZT). Її особливістю можна вважати жорсткіший принцип «ніколи не довіряти, завжди перевіряти». Тобто згідно з її основною ідеєю – не існує областей, які заслуговують на довіру. Нульова довіра – це не єдина архітектура, а набір керівних принципів для робочого процесу, проектування системи та операцій, які можна використовувати для покращення стану безпеки будь-якої класифікації або рівня чутливості [6].

Виходячи з сказаного, спираючись на принципи цієї концепції і задіявши можливості ШІ (технології ШІ можуть ефективно використовуватися для своєчасного виявлення вразливостей у різних інформаційних системах та мережах), можна спробувати побудувати відповідну сучасним вимогам адаптивну і стійку систему безпеки, яка постійно буде перевіряти кожну взаємодію та швидко реагувати на загрози, що знову з’являються.

У рамках вирішення зазначеної задачі розглянемо деякий підхід, що спирається на принципи концепції нульової довіри та потенціал ШІ, а саме – можливість вирішення актуального завдання, яке полягає у подальшому удосконаленні моделей і методів захисту від кібератак та зловживань шляхом інтегрування концептуальних принципів нульової довіри та ШІ. Такий підхід може дозволити забезпечити більш динамічний та адаптивний захист за рахунок постійної перевірки автентичності користувачів і процесів, незалежно від їх попереднього статусу довіри, та більш ефективно ідентифікувати потенційні загрози в режимі реального часу й реагувати на них раніше, ніж вони можуть завдати шкоди.

1. Огляд концепції нульової довіри

Як відомо [7], традиційна мережева безпека ґрунтується на концепції периметра безпеки, згідно з якою мережа поділяється на дві частини: внутрішню довірену мережу та зовнішню недовірену мережу. Відповідно до цього підходу добре структурована архітектура безпеки розглядає безпеку мережі як багаторівневу систему, де кожен периметр захищає область, яку він покриває. За даними дослідників із компанії Akamai [8], периметр – укріплена ділянка мережі, яка може включати граничні маршрутизатори, міжмережеві екрани (брандмауери, файрволи), системи виявлення вторгнень (Intrusion Detection System – IDS), системи запобігання вторгненням (Intrusion Prevention System – IPS), VPN, програмну архітектуру, демілітаризовану зону (Demilitarized Zone – DMZ) та віртуальні локальні мережі (Virtual Local Area Network – VLAN).

Міжмережеві екрани або фаєрволи захищають активи, ізолюючи приватні мережі від публічних шляхом фільтрації трафіку та блокування доступу до недовірених джерел або IP-адрес. Однак, якщо зловмиснику вдається прорвати захисний периметр, фаєрвол не може зупинити його в подальших діях у внутрішній мережі. VPN часто використовується для доступу до віддалених мереж, створюючи захищене з’єднання між локальною та віддаленою мережею за допомогою шифрування даних. Хоча цей метод ефективний для захисту комунікацій, він становить загрозу для корпоративних активів, оскільки VPN використовує статичну автентифікацію та не може безперервно перевіряти особу користувача та надійність кінцевого пристрою під час сесії. Також VPN не здатен визначати та обмежувати рівень доступу користувачів, що дозволяє їм необмежено користуватися внутрішніми ресурсами після підключення. Мережа DMZ додає ще один рівень безпеки для внутрішньої мережі. Однак, як і у випадку з іншими методами, надмірна залежність від фаєрволів може призвести до виник-

нення певних проблем. Якщо зловмисник зможе обійти фаєрвол, наприклад, через фішингові електронні листи, що дозволить отримати доступ до внутрішньої мережі, то DMZ не зможе протидіяти такій загрозі. Крім того, DMZ не здатен виявляти атаки довірених пристроїв на інші довірені пристрої.

Архітектура безпеки периметра забезпечувала ефективний захист від типових загроз, таких як шкідливе програмне забезпечення, фішингові атаки, атаки на відмову в обслуговуванні та атаки нульового дня. Однак зі зростанням обсягу даних, що переміщуються у хмарні сервіси, та з розширенням кількості користувачів, включно з пристроями Інтернету речей (Internet of Things – IoT), традиційне розуміння мережевого периметра змінилося, що зробило зовнішні атаки більш складними та динамічними. Поверхня атак збільшилася, багато загроз походять зсередини, і захист на основі периметра більше не може ефективно протидіяти внутрішнім загрозам, оскільки являє собою одно-направлений захист і безсилий проти атак зсередини мережі.

Концепція нульової довіри, «ніколи не довіряй, завжди перевіряй», була вперше запропонована у 2010 р. [9] для розв'язання проблем, спричинених внутрішніми загрозами для організацій. Нульова довіра відноситься до двох основних областей: автентифікації та авторизації [6]. В її основі лежить ідея обмеження неявної довіри як визнання обмеженості використання одиничних статичних засобів захисту у великій мережі.

Порівнюючи традиційну модель безпеки та модель нульової довіри можна побачити (табл. 1 [5, 10]), що традиційний підхід на основі захисту згідно з периметром ефективно захищає від зовнішніх атак, але ігнорує внутрішні атаки.

Таблиця 1

Порівняння традиційної моделі безпеки та моделі нульової довіри

Характеристика	**Традиційна модель безпеки**	**Модель нульової довіри**
Підхід	Довіряй але перевіряй.	Нікому не довіряй і все перевіряй.
Межа довіри	Зовнішня (немає довіри). Внутрішня (довірена).	Мікросегментація (мережі поділяються на дрібніші сегменти або безпечні зони, щоб обмежити бічне переміщення загроз; кожен сегмент має свої засоби керування доступом користувачів та політики безпеки).
Мережева архітектура	Модель «замок та рів» з підвищеним акцентом на захист периметра.	Децентралізована та мікросегментована, з детальним контролем доступу.
Контроль доступу	Контроль доступу на основі IP.	Керування доступом, орієнтоване на дані (з урахуванням ідентифікаційних та контекстно-залежних даних).
Автентифікація	Однократна після перевірки при початковому доступі та статична.	Перед доступом та постійна (мультимодальна та динамічна) перевірка.
Керування безпекою	Індивідуальний моніторинг та видимість.	Видимість, автоматизація та оркестрування поведінки, пристроїв, сервісів та безпеки.
Політика безпеки	Заздалегідь встановлені правила та загальна політика.	Деталізовані правила та адаптивні політики (оцінка рівня безпеки).
Шифрування зв'язку	Зовнішня мережа (шифрування). Внутрішня (без шифрування).	Повне шифрування трафіку.
Реагування на порушення	Після того, як периметр порушено, зловмисники отримують можливість діяти вільно.	Навіть якщо порушення сталося, переміщення зловмисників ретельно відстежуються.

Традиційна архітектура захисту не спроможна ефективно протистояти сучасним кібератакам, привілейовані шляхи доступу стають більш ризикованими й ускладнюють захист згідно з периметром від несанкціонованих атак з боку легітимних внутрішніх користувачів. У той час як використання, наприклад, принципу найменших привілеїв та мікросегментація архітектури нульової довіри (Zero Trust Architecture – ZTA) ефективно обмежує привілеї внутрішніх користувачів і дозволяє уникати ризиків необмеженого бічного переміщення (lateral movement) користувачів у мережі.

Принципи нульової довіри не є новими, але їх унікальність полягає в комплексному застосуванні для захисту корпоративних ресурсів. На відміну від традиційних систем, де права доступу визначаються статично відповідно до посадових обов'язків, підхід концепції нульової довіри передбачає динамічне прийняття рішень через центр політики на основі внутрішніх правил і зовнішніх даних. Хоча автоматизація є ключовим елементом таких систем, вони також забезпечують можливість ручного втручання на окремих етапах перед запуском автоматичних процесів реагування [5].

У спеціальній публікації NIST 800-207 [6] нульова довіра описується як парадигма кібербезпеки, яка зміщує акцент захисту. Нульова довіра зосереджена на захисті ресурсів (активів, служб, робочих процесів, мережевих облікових записів тощо), а не на сегментах мережі, оскільки мережеве розташування більше не вважається основним компонентом безпеки ресурсу. Згідно з концепцією нульової довіри активам або обліковим записам користувачів не надається ніякої неявної довіри на основі їхнього фізичного або мережевого розташування. Натомість автентифікація та авторизація передбачають певні кроки перед тим, як отримати доступ до корпоративного ресурсу [6]. На рис. 1 представлено основні логічні компоненти архітектури нульової довіри, а також взаємозв'язок між ними.

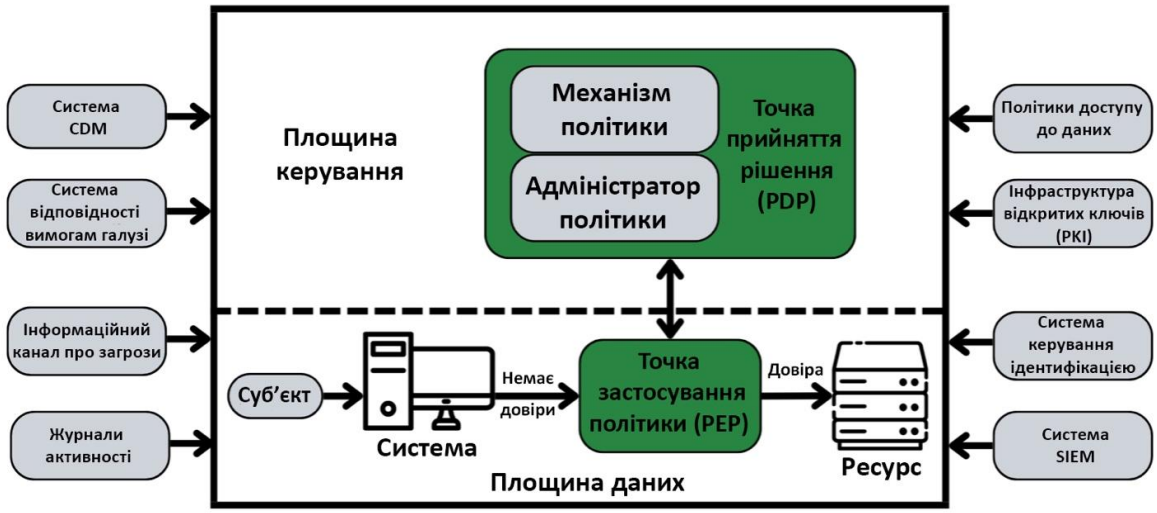

Рис. 1. Основні логічні компоненти ZTA

В даній схемі існує поняття суб'єкта, яке визначається NIST [6] як користувач, сервіс, застосунок або пристрій, що працює в комп'ютерній системі (або разом з нею) і має доступ до корпоративного ресурсу (Enterprise Resource). Цей ресурс може бути застосунком, даними, документом або робочим навантаженням, які знаходяться під контролем організації та захищені системою нульової довіри. Вважається, що суб'єкт працює в ненадійному середовищі в ненадійній мережі, і йому дозволено отримати доступ до ресурсу лише через точку застосування політики (Policy Enforcement Point – PEP). PEP контролює доступ суб'єкта до ресурсу через те, що NIST [6] називає неявною зоною довіри (implicit trust zone). PEP не зберігає і не визначає політику – цю роботу виконує точка прийняття рішення (Policy Decision Point – PDP). Варто звернути увагу, що суб'єкт взаємодіє з корпоративним ресурсом через так звану площину даних, яка відрізняється від площини керування – як зазначає NIST [7], PDP і PEP «взаємодіють у мережі, яка є логічно відокремленою і не є безпосередньо доступною для активів і ресурсів організації. Площина даних використовується для трафіку даних застосунків».

Крім цих основних компонентів на підприємстві, що реалізує ZTA, є ряд зовнішніх компонентів, які сприяють реалізації безпеки з нульовою довірою. А саме – існує кілька локальних та зовнішніх джерел даних, що надають вхідні дані та правила політик, які використовуються механізмом політик при прийнятті рішень про доступ. Серед них: система керування ідентифікацією (ID management system), інфраструктура відкритих ключів (Public Key

Infrastructure – PKI), система аналізу загроз (інформаційні канали про загрози), підсистема реєстрації подій і моніторингу, підсистема безперервної діагностики та усунення наслідків, система управління інформацією та подіями безпеки (Security Information and Event Management – SIEM), система безперервної діагностики та пом'якшення наслідків – Continuous Diagnostics and Mitigation (CDM). Ці елементи є важливими вхідними даними (контекстом) для системи нульової довіри і, безумовно, впливають на її практичні рішення.

У свою чергу, компанія Forrester запропонувала [11] розширену архітектуру нульової довіри (Zero Trust Extended – ZTX), що охоплює ширше коло потоків даних, зокрема тих, що проходять через локальні мережі, хмарні сервіси, зовнішні застосунки, сайти та різні види кінцевих пристроїв, зокрема датчики IoT тощо. Компанія Gartner, ґрунтуючись на принципі безперервного адаптивного оцінювання ризику та довіри, також висунула ідею [12] про ZTX. Покращена модель нульової довіри розширює архітектуру, запропоновану NIST, забезпечуючи більш повну обізнаність про ситуацію і враховує практичні особливості введення в експлуатацію. Згідно з такою моделлю у процесі вибору рішень беруть до уваги і суб'єкт, і кінцеву точку, включно зі ступенем їхньої відповідності вимогам безпеки. Приблизно в той самий час Google розпочала розробку своєї реалізації концепції безпеки з нульовою довірою BeyondCorp [13], яка перенесла контроль доступу з периметру мережі на окремі пристрої та користувачів. BeyondCorp, базуючись на політиках автентифікації, авторизації та контролю доступу, гарантує доступ до корпоративних ресурсів лише авторизованим користувачам і пристроям. Додатковий імпульс розвитку концепції надали публікації NIST [6] та NCCoE (National Cybersecurity Center of Excellence) [15], які акцентували увагу на ZTA для захисту корпоративних даних [5].

Деякі спеціалісти з безпеки компанії Oracle пов'язують, наприклад, дії щодо забезпечення безпеки на основі принципів концепції нульової довіри з використанням чотирьох ключових напрямків (рис. 2), зокрема [14]: запобігання атакам (Attack Prevention), виявлення інцидентів (Incident Detection), усунення інцидентів (Incident Response) та аналіз подій (Event Analysis). Такий підхід підкреслює безперервний цикл захисту даних, інтегруючи системи керування ідентифікацією та доступом (Identity and Access Management – IAM), автоматизоване виявлення загроз, безпеку застосунків та моніторинг мережі.

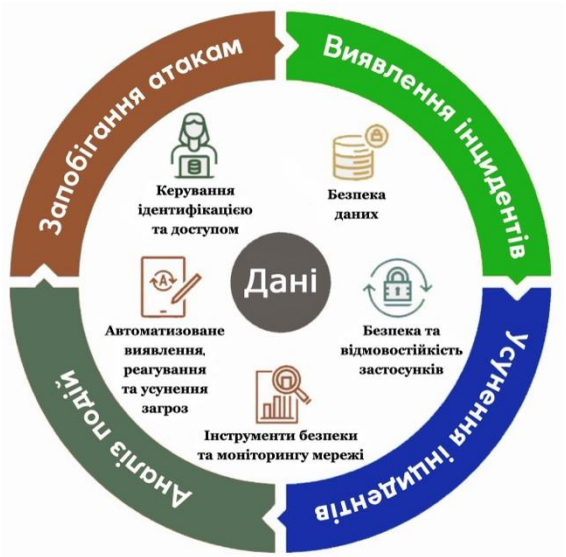

Рис. 2. Ключові напрями підходу до захисту

У 2023 р. компанія Fortinet провела дослідження серед компаній-лідерів у сфері безпеки з 31 країни, що охоплювало різні галузі, включно з державним сектором. Результати показали, що лише 28 % організацій впровадили повноцінне рішення нульової довіри для мінімізації кіберризиків [16]. Це свідчить про те, що значна частина компаній ще не реалізувала комплексні заходи ZTA, що залишає їх вразливими до сучасних кіберзагроз. Проте впроваджен-

ня подібних рішень не лише дозволить підвищити рівень безпеки, а й забезпечить значні операційні та економічні переваги для підприємства, що робить ZTA ключовим елементом ефективного управління кіберризиками. Однією з головних переваг концепції нульової довіри є оптимізація витрат на забезпечення кібербезпеки, скорочення обсягу робіт із дотримання нормативної відповідності та ефективне використання ресурсів. Крім аспектів безпеки, ZTA відіграє ключову роль у підтримці цифрової трансформації бізнесу, а організації отримують додаткові конкурентні переваги, такі як підвищення операційної гнучкості, ефективне управління ресурсами та здатність швидко адаптуватися до змін у динамічному цифровому середовищі [17].

В цілому ж можна констатувати, що реалізація концепції нульової довіри передбачає впровадження комплексних заходів безпеки, спрямованих на мінімізацію ризиків і забезпечення захисту ресурсів організації в умовах сучасного цифрового середовища. І одним із ключових компонентів у цьому аспекті є безперервний моніторинг та аудит, які дозволяють виявляти підозрілу активність та аномалії шляхом аналізу даних із мережевого трафіку, журналів активності користувачів та систем захисту кінцевих точок. Це забезпечує реальний час для оцінки загроз, реагування на інциденти (Incident Response – IR), проведення цифрової криміналістики (digital forensics) та аналізу кіберзагроз. Поряд із цим важливим є IAM, яка визначає правила доступу до ресурсів і передбачає перевірку кожного запиту на основі багатофакторної автентифікації та динамічної оцінки ризиків. Ця система забезпечує високу точність контролю доступу за рахунок інтеграції параметрів, таких як ідентифікатор та пароль користувача, пристрій, час поточного місцезнаходження, права доступу тощо. Іншим критично важливим елементом ZTA є аналіз поведінки користувачів та сутностей (пристроїв, застосунків та інших об’єктів в інформаційній системі; User and Entity Behavior Analysis – UEBA), який дозволяє виявляти аномальні дії, наприклад спроби несанкціонованого доступу або передачі даних. Це допомагає ідентифікувати загрози, включаючи розподілені атаки на відмову в обслуговуванні, атаки грубої сили, витоки даних та інсайдерські загрози, забезпечуючи пріоритетність реагування на найбільш серйозні ризики. Водночас важливим є і реагування на інциденти (IR), що передбачає швидке виявлення атак, визначення їх масштабів, нейтралізацію наслідків та запобігання повторенню. Організації можуть мінімізувати вплив кіберінцидентів, скоротити час простою і захистити свою репутацію, впроваджуючи відповідний план IR. З іншого боку, система аналізу загроз сприяє збору та аналізу інформації про дії зловмисників і використовувані ними вразливості. Сюди входить інформація про те, що роблять зловмисники, як вони це роблять і які вразливості вони використовують. Ця інформація допомагає організаціям проактивно захищати свої активи, підвищуючи стійкість до сучасних кіберзагроз. На думку експертів з інформаційної безпеки [18], вище вказані компоненти безпеки зазвичай вважаються ключовими та важливими і мають бути дотримані при будь-якій реалізації концепції нульової довіри. У свою чергу, компанія ManageEngine для впровадження ZTA та підвищення ефективності безпеки в організаціях запропонувала комплексний підхід [19]. Одним із ключових компонентів комплексного підходу ManageEngine вважає багатофакторну автентифікацію (Multi-Factor Authentication – MFA), що надає додатковий рівень перевірки ідентифікації. Це значно ускладнює несанкціонований доступ навіть у разі компрометації облікових даних, наприклад, через фішингові атаки. MFA має застосовуватися для різних операційних систем (Windows, macOS, Linux), VPN-сервісів (Fortinet, Cisco AnyConnect, Pulse, OpenVPN) та кінцевих точок із підтримкою RADIUS (Citrix Gateway, VMware Horizon, Microsoft RDP). Як додаткові фактори при автентифікації можна використовувати відбитки пальців, розпізнавання обличчя, а також застосунки на кшталт Microsoft Authenticator, Google Authenticator, Duo Security або YubiKey Authenticator. Однак, з огляду на зростаючі вимоги до безпеки та зручності, сучасні підходи до автентифікації спрямовані на мінімізацію залежності від традиційних паролів. У цьому контексті безпарольна автентифікація, що реалізована через відкриті стандарти (OAuth, OpenID Connect, SAML), дозволяє значно знизити ризики, пов’язані з крадіжкою паролів, а механізм єдиного входу

(Single Sign-On – SSO) забезпечує централізований доступ до множини ресурсів без необхідності повторного введення облікових даних. Іншим важливим елементом ZTA в рамках комплексного рішення від ManageEngine є реалізація широко відомого принципу найменших привілеїв, який полягає у наданні користувачам доступу лише до тих ресурсів, які необхідні для виконання їхніх завдань. Це зменшує ризик витоку інформації у разі компрометації облікового запису. Реалізація цього принципу передбачає автоматизоване керування дозволами, створення шаблонів для нових співробітників на основі їхніх посадових обов'язків і тимчасове надання доступу до певних ресурсів. У рішеннях UEBA, запропонованих тією ж компанією ManageEngine, аномальна поведінка користувача виявляється шляхом аналізу шаблонів та відхилень від нормальної поведінки (наприклад, спроби несанкціонованого доступу чи передачі даних). При цьому інтеграція UEBA з уніфікованим рішенням SIEM (Security Information and Event Management – система керування інформацією та подіями безпеки) дозволяє автоматизувати аналіз загроз, систематизувати події за ступенем ризику та забезпечувати реагування в режимі реального часу шляхом надсилання сповіщень командам безпеки.

Таким чином, концепція нульової довіри є сучасним підходом до кібербезпеки, який забезпечує комплексний захист даних через постійний моніторинг, керування доступом та швидке реагування на загрози. При цьому слід зазначити, що дана концепція продовжує еволюціонувати, оскільки постачальники та організації зі стандартизації постійно оновлюють і вдосконалюють її специфікації та реалізації, визнаючи її кардинальною зміною в підході до кібербезпеки [5, 18]. Тому, з огляду на зростаючу складність атак, природно, як окремі традиційні методи кібербезпеки, так і комплексний захист, заснований на концепції нульової довіри, повинні доповнюватися інноваційними технологіями, зокрема ШІ, які здатні допомогти організаціям забезпечити постійний моніторинг систем і швидке реагування на нові загрози. Таким чином, концепція нульової довіри в поєднанні з можливостями ШІ, може стати не лише технологічним рішенням, але й стратегічним підходом, що дозволить організаціям ефективно протистояти сучасним кіберзагрозам, забезпечуючи безпеку даних на всіх етапах їх обробки, передачі та зберігання.

2. Напрями застосування штучного інтелекту в рамках концепції нульової довіри

З розвитком ШІ еволюціонують і ризики. Широка доступність таких інструментів, як ChatGPT, Google Gemini та інших, надала зловмисникам можливість швидко підвищувати складність кібератак. Тому фахівці з безпеки повинні активно та оперативно впроваджувати сучасні підходи та стратегії для захисту ресурсів підприємств від подібних загроз. Для боротьби зі зловмисниками та програмами генеративного ШІ дуже важливо, щоб розробка, вдосконалення та впровадження надійних засобів захисту здійснювалися постійно.

Найпоширенішою сферою застосування технологій ШІ з метою захисту інформації є автоматизовані та інформаційні системи, включно з комп'ютерними мережами різної архітектури. До основних завдань захисту інформації з використанням ШІ можна віднести:

- виявлення та запобігання витокам конфіденційних даних;
- виявлення кібератак та шкідливих програм;
- виявлення модифікованих даних або повідомлень;
- підвищення надійності та кіберстійкості систем, сервісів і мереж;
- оцінка ризиків кібербезпеки.

Недоліки традиційних систем безпеки багато в чому пов'язані з тим, що вони засновані на статичних правилах. Тобто, використовуються заздалегідь визначені методи виявлення загроз і реагування на них. Це тягне за собою обмеження, зокрема нездатність реагувати на нові загрози, оскільки з появою нових загроз правила повинні оновлюватися вручну. Іншим обмеженням є обсяг даних. Наявні системи безпеки можуть генерувати великі обсяги даних, які складно аналізувати в реальному часі. Крім того, системи, засновані на статичних правилах, можуть виявитися неефективними під час виявлення складніших атак, наприклад тих, що використовують ШІ для імітації звичайної поведінки користувача.

Можливість швидкої обробки великих масивів даних для завчасного реагування – це одна з ключових переваг ШІ. За допомогою ШІ можна в режимі реального часу аналізувати великі обсяги інформації, насамперед мережевого трафіку, виявляти аномалії та незвичні активності. ШІ може аналізувати дані з декількох джерел, включно з мережевим трафіком та системними журналами, для виявлення подій, що виходять за рамки норми. При виявленні кібератак або шкідливих програм основним сценарієм застосування технології ШІ може бути визначення аномалій у поведінкових моделях користувачів інформаційних систем. Наприклад, ШІ може виявити незвичні моделі поведінки або аномалії, які можуть вказувати на кібератаку, зокрема, передачу великих обсягів даних у зовнішню мережу, нестандартні спроби входу в систему, дії пов'язані з внутрішніми загрозами (доступ співробітників до даних у неробочий час або отримання доступ до даних, до яких зазвичай немає дозволу).

Застосування ШІ в рамках ZTA є перспективним напрямком у галузі кібербезпеки, що дозволить підвищити рівень захисту корпоративних систем від сучасних загроз. Модель нульової довіри виходить з того, що жоден користувач або пристрій не повинні отримувати автоматичну довіру – навіть якщо вони знаходяться всередині корпоративної мережі. В такому середовищі ШІ може відіграти ключову роль, допомагаючи автоматизувати процеси моніторингу, виявлення загроз і своєчасного реагування. Наприклад, якщо система ШІ виявляє спробу кібератаки, вона може автоматично заблокувати доступ до скомпрометованої системи, запобігаючи подальшим ризикам, а також надсилати сповіщення адміністраторам безпеки, надаючи їм інформацію про інцидент.

На рис. 3 показано основні логічні компоненти ZTA, які можуть загалом використовувати алгоритми ШІ, зокрема: механізм політики, точка застосування політики (PEP) та деякі зовнішні компоненти.

Рис. 3. Логічні компоненти ZTA, які можуть використовувати алгоритми ШІ

2.1. Механізм політики

Механізм політики є ключовим компонентом ZTA, оскільки відповідає за остаточне ухвалення рішень щодо надання або обмеження доступу до ресурсів. Він використовує політики підприємства, а також інформацію із зовнішніх джерел (наприклад, системи CDM, служби аналізу загроз тощо) як вхідні дані для алгоритму довіри для надання, заборони або скасовування доступу до ресурсу. Для прийняття рішень механізм політики використовує алгоритм довіри, який враховує різні джерела даних, такі як роль користувача, інформацію про поведінку, дані про ресурс (наприклад, тип операційної системи, рівень виправлень/оновлень), а також контекстні атрибути (час, місцезнаходження тощо). На основі цих даних алгоритм обчислює рівень довіри до суб'єкта та визначає, чи можна надати доступ до ресурсу, обмежити його або повністю заборонити [17].

Алгоритм довіри може бути реалізований у різний спосіб. Наприклад, у підході, заснованому на критеріях, доступ надається лише у випадку, якщо всі встановлені умови виконані. Ці умови можуть включати вимоги до автентифікації, параметри безпеки ресурсу тощо. Інший підхід, заснований на оцінках, передбачає розрахунок рівня довіри шляхом аналізу

значень різних атрибутів з урахуванням їх вагових коефіцієнтів. Якщо отримана оцінка перевищує порогове значення, доступ надається; у протилежному випадку запит відхиляється або рівень доступу обмежується. Крім того, алгоритми довіри можуть бути сингулярними або контекстними. Сингулярний підхід не враховує історичну інформацію про користувача, що дозволяє пришвидшити процес прийняття рішень, але може знизити ефективність виявлення аномалій або зловмисних дій. Навпаки, контекстний підхід аналізує історичні моделі поведінки користувача, що дозволяє виявляти нетипові дії та потенційні загрози. Наприклад, якщо користувач раптово намагається отримати доступ до ресурсу з нового місцезнаходження або в незвичний час, система може ініціювати додаткові перевірки або повністю заблокувати доступ.

Таким чином, механізм політики та алгоритм довіри в ZTA забезпечує гнучкий та адаптивний підхід до контролю доступу, що дозволяє ефективно реагувати на зміни в поведінці користувачів та наявні загрози. Автоматизація цього процесу є складним завданням, проте її реалізація, насамперед із використанням можливостей ШІ, значно підвищує ефективність політик безпеки та забезпечує надійний захист критичних ресурсів. ШІ здатен здійснювати постійний аналіз та оцінку рівня довіри до користувачів і пристроїв у корпоративному середовищі. В свою чергу, використання передових методів машинного навчання, таких як алгоритми кластеризації, дозволяє більш ефективно оцінювати рівень довіри та адаптивно регулювати доступ до ресурсів. Дослідники [20] проаналізували умови, за яких алгоритми оцінки довіри на основі машинного навчання можуть підвищити надійність пристроїв у розподіленій системі. У цій роботі вони виконали комплексний аналіз сучасних підходів до оцінки довіри, виділили ключові вимоги, яким повинні відповідати такі методи, та розробили критерії оцінки їх ефективності. Вони також класифікували існуючі підходи за сценаріями застосування, розглянувши широкий спектр методів машинного навчання, що використовуються для оцінки довіри у різних галузях, включно з багатокомпонентними системами, сервісами та мережами. Як результат, автоматизація процесів та інтеграція механізмів ШІ та машинного навчання в алгоритми оцінки довіри, незважаючи на свою складність, стає ключовим фактором підвищення ефективності механізму політики в рамках ZTA та відкриває нові можливості для створення більш надійних та адаптивних систем контролю доступу.

2.2. Точка застосування політики

PEP є ключовим компонентом ZTA, відповідальним за підключення, моніторинг та завершення з'єднань між суб'єктом та корпоративним ресурсом [17]. Усі комунікації, що проходять через PEP, контролюються відповідно до встановлених політик безпеки, що забезпечує динамічне управління доступом. Одним із ключових аспектів роботи PEP є контроль доступу, що є одним з основних механізмів захисту в рамках ZTA. Традиційні механізми контролю доступу обмежують дії користувача або застосунку всередині мережі, однак з появою, наприклад, технології 5G та IoT, кількість інтелектуальних пристроїв, що під'єднуються до мережі, зростає, що призводить до розширення поняття контролю доступу. Зараз контроль доступу більше не обмежується лише регулюванням доступу до даних для користувачів і застосунків. В архітектурі ZTA контроль доступу переосмислюється таким чином, що лише автентифіковані та авторизовані суб'єкти можуть отримати доступ до системи (можливе динамічне надання або відкликання доступу). Суб'єктами виступають як користувачі, так і застосунки (або сервіси) або їх комбінації з пристроїв, серверів, сервісів, застосунків тощо. Тоді як система може являти собою пристрій, такий як ноутбук, мобільний телефон, віртуальна машина, контейнер тощо [17].

Основними моделями контролю доступу, які використовуються в ZTA, вважаються моделі контролю доступу на основі ролей (Role-based Access Control – RBAC) та атрибутів (Attribute-based Access Control – ABAC) [18]. Їх використання дає змогу автоматично призначати ролі та дозволи користувачам на основі їх статичних і динамічних атрибутів для більш детального та гнучкого контролю доступу. Статичні атрибути можуть містити ідентифікато-

ри користувача або пристрою, тоді як динамічні атрибути охоплюють такі фактори, як час та місце запиту на доступ.

Однак, слід зауважити, що концепція нульової довіри не визначає RBAC як кращу методологію керування доступом [9]. Моделі контролю доступу на основі ролей вимагають постійного ручного втручання при створенні, видаленні та керуванні ролями, що створює труднощі в масштабованості та адаптації системи до змін в організаційній структурі. Статичні ролі, в свою чергу, можуть не відповідати поточним потребам користувачів, що призводить до надмірного або недостатнього доступу. Крім того, традиційні RBAC-рішення часто не враховують контекстні атрибути, що знижує ефективність управління доступом в динамічних середовищах. Тому необхідно застосовувати підходи, що забезпечують автоматизоване коригування ролей та дозволів відповідно до актуальних умов використання ресурсів. Одним із таких підходів є інтелектуальне призначення ролей на базі ШІ [21], що передбачає динамічний розподіл ролей користувачам на основі контекстних факторів, що покращує управління доступом і підвищує рівень безпеки. ШІ здатний аналізувати поведінку та історичну інформацію користувачів, щоб пропонувати відповідні призначення ролей, а також визначати ролі з високим рівнем ризику та небезпечні комбінації ролей. Подібна адаптивна система на основі ШІ може виявляти закономірності між користувачами зі схожими посадовими обов'язками, що сприяє більш точному та обґрунтованому призначенню ролей, також автоматично оцінювати сценарії використання ролей, виявляти надмірні або конфліктні права доступу та пропонувати оптимальні зміни для їх коригування. Це дозволяє організаціям зменшити ризики, пов'язані з надмірними дозволами, що можуть спричинити витік даних або несанкціонований доступ, а також забезпечити не лише підвищення рівня безпеки, але й зменшити адміністративне навантаження на систему контролю доступом.

З іншого боку, формально, роль можна розглядати як особливий вид атрибута, і тому ABAC можна трактувати як підмножину RBAC. Насправді, як стверджують фахівці [18], архітектури нульової довіри – це найбільш ефективний спосіб керування доступом на основі атрибутів.

Застосування ШІ також є перспективним напрямом і в системі керування привілейованим доступом (Privileged Access Management – PAM). Алгоритми ШІ здатні аналізувати та навчатися на основі шаблонів входу привілейованих користувачів [22], встановлюючи базовий рівень стандартної поведінки та виявляючи аномалії, що може свідчити про потенційні загрози безпеці. Однією з ключових переваг використання ШІ в PAM є можливість прогнозування, тобто аналізуючи історичні дані та виявляючи закономірності, він може передбачати ймовірні загрози ще до їх реалізації, що дозволяє організаціям завчасно вживати заходи для усунення потенційних загроз. У межах системи PAM, ШІ може удосконалювати механізми надання та відкликання привілейованого доступу, забезпечуючи обґрунтованість, відповідність політикам безпеки та мінімізуючи ризики несанкціонованого використання. Додатково ефективні рішення системи PAM повинні містити оцінку ризиків на основі поведінкових факторів та історичної інформації кожного користувача, а інтеграція ШІ з механізмами моніторингу користувачів може своєчасно виявляти аномальні шаблони поведінки та потенційні загрози безпеці, підвищуючи ефективність управління привілейованим доступом.

Підходи, що базуються на алгоритмах ШІ, можуть значно скоротити час і підвищити точність призначення прав доступу порівняно з традиційними ручними методами, що є важливим кроком у забезпеченні більш ефективної та адаптивної безпеки в рамках ZTA.

2.3. Зовнішні компоненти

Окрім основних компонентів на підприємстві, важливу роль у реалізації ZTA відіграють зовнішні компоненти, які забезпечують організацію додатковою інформацією для ухвалення рішень щодо безпеки. Використання можливостей ШІ у цих системах дозволяє автоматизувати аналіз великих обсягів даних, виявляти аномалії, прогнозувати потенційні загрози та

адаптивно коригувати політики доступу, що підвищує ефективність контролю та забезпечує проактивний підхід до кібербезпеки.

2.3.1. Керування ідентифікацією

Керування ідентифікацією є ключовою компонентою ZTA, забезпечуючи точну автентифікацію та авторизацію користувачів і пристроїв для доступу до ресурсів організації. Тому, враховуючи важливість точної автентифікації ZTA, доцільним є впровадження MFA, через те, що вона додає додаткові рівні перевірки. Окрім базового методу автентифікації (наприклад, пароля), MFA передбачає використання одноразового пароля (One Time Password – OTP), що генерується на основі часу в мобільному пристрої користувача або надсилається на електронну пошту чи в SMS повідомленні [23]. Крім того, MFA включає різні додаткові принципи автентифікації, які застосовуються до процесу входу в систему, дозволяючи перевірити користувача через кілька незалежних каналів і підтвердити його особу за допомогою додаткових атрибутів. У зв'язку з цим, у межах ZTA реалізується багаторівневий підхід до керування ідентифікацією, що включає як біометричну автентифікацію, так і автентифікацію на фізичному рівні.

Біометричні характеристики людини, такі як голос, райдужна оболонка ока чи відбитки пальців, є унікальними для кожної особи [24]. Застосування же технологій ШІ при автентифікації за допомогою голосу дозволяє підвищити її точність, зокрема шляхом навчання системи розрізняти голоси окремих осіб і мінімізувати ризики при спробах його підробки. При цьому машинне навчання забезпечує обробку великих обсягів даних та здатність адаптуватися до змін навколишнього середовища, що включає виявлення специфічних мовних особливостей, таких як акцент або емоційні варіації в голосі [25]. Аналогічно, при автентифікації за допомогою розпізнавання обличчя або райдужної оболонки ока технології ШІ порівнюють зібрані біометричні дані з базами даних для визначення відповідності. Відмінності в структурі обличчя чи райдужної оболонки є унікальними для кожної особи, що робить ці методи високоточними для ідентифікації. Завдяки можливостям ШІ, подібні системи можуть точно визначати відмінності в цих характеристиках навіть за умов поганої освітленості чи змін у зовнішньому вигляді користувача. Це дозволяє ефективно використовувати такі методи в системах безпеки, де вони можуть замінювати традиційні способи доступу, такі як пропуски чи паролі, знижуючи ймовірність несанкціонованого доступу та підвищуючи рівень захисту [26].

Однак традиційні методи біометричної автентифікації, що базуються на поверхневих ознаках, поступово втрачають ефективність. Це пояснюється тим, що такі характеристики, як відбитки пальців, можуть бути легко викрадені або підроблені. У зв'язку з цим увага дослідників усе більше зосереджується на використанні біометричних даних, що відображають внутрішні характеристики організму, які важче підробити або скопіювати. Такі дані можуть бути зібрані за допомогою сучасних носимих смарт-пристроїв, однак ключовим викликом залишається ефективна класифікація та ідентифікація цих даних. Методи ШІ є перспективним рішенням цієї проблеми. Вони дозволяють автоматично обробляти дані, отримані від носимих пристроїв, і використовувати їх для ідентифікації особи. Зокрема, у роботі [27] проаналізовано застосування алгоритмів машинного навчання для забезпечення постійної (безперервної) багатофакторної автентифікації, що включає кілька біометричних ознак. Для автоматизованої ідентифікації користувачів зібрані біометричні характеристики спочатку перетворюються у числовий формат і класифікуються за допомогою алгоритмів машинного навчання під наглядом, таких як алгоритм KNN або дерева рішень. Ці методи надалі використовуються для побудови моделей автентифікації. Разом з тим, глибокі нейронні мережі, такі як згорткові нейронні мережі (Convolutional Neural Networks – CNN), дозволяють автоматично виділяти та навчатися на біометричних характеристиках користувача, які отримуються від біометричних пристроїв моніторингу, що значно підвищує точність та надійність ідентифікації. Крім того, для вирішення проблем, пов'язаних із недостатнім обсягом біометричних даних, застосовуються підходи глибокого передавального навчання (Deep Transfer Learning –

DTL). Такі методи дозволяють використовувати попередньо навчені моделі для адаптації до нових даних, забезпечуючи високу продуктивність навіть у разі обмежених ресурсів для навчання.

Інший підхід, пов'язаний з безперервною автентифікацією, спрямований на забезпечення постійного підтвердження ідентичності кінцевих точок протягом усього сеансу зв'язку. Для його реалізації перспективним вважається метод використання автентифікації на рівні пристроїв на основі фізичного рівня (Physical Layer Authentication – PLA), де алгоритми ШІ дозволяють ефективно виділяти характеристики пристроїв із комунікаційних каналів і безперервно перевіряти ідентичність як користувачів, так і пристроїв [28]. Дослідження [29] містить детальний огляд сучасних технологій на основі PLA. Автори підкреслюють, що традиційні криптографічні методи автентифікації мають низку суттєвих обмежень, таких як низька сумісність, ненадійність і висока складність впровадження. Тому у цьому контексті вони вважають, що PLA на основі пристроїв виступає перспективним доповненням, оскільки такий метод дозволяє використовувати унікальні фізичні властивості середовища та пристроїв.

Однак порівняно з властивостями пристроїв, характеристики каналу зв'язку є значно складнішими для копіювання або імітації, що забезпечує вищий рівень безпеки при автентифікації пристроїв. Виділення характеристик каналу здійснюється на основі комунікаційного потоку між пристроями, і їх ручна класифікація є практично неможливою через значний обсяг і складність даних. Для вирішення цієї проблеми у межах ZTA перспективним рішенням було б застосування технологій ШІ та машинного навчання, які можуть автоматично аналізувати характеристики каналів зв'язку, ідентифікувати пристрої та значно підвищувати ефективність і точність їх автентифікації.

2.3.2. Журнали активності

ZTA потребує постійного моніторингу всіх дій користувачів, пристроїв та сервісів. Навіть після успішної автентифікації користувачів або пристроїв необхідно відстежувати їхню поведінку для виявлення аномалій та потенційних загроз. Це дозволяє ідентифікувати підозрілі дії легітимних користувачів або скомпрометованих пристроїв. У цьому контексті журнали активностей відіграють ключову роль у забезпеченні безпеки, оскільки дозволяють аналізувати поведінку (маючи зворотний зв'язок про роботу системних компонентів та користувачів), вчасно ідентифікувати аномальну активність при отриманні доступу до сервісу та оперативно реагувати на загрози. Журнали зазвичай реєструють роботу системи у вигляді часових послідовностей. І в цьому зв'язку методи аналізу журналів активності, засновані на контрольованому навчанні (supervised learning) можуть дуже допомогти в автоматизації при визначенні аномальних ознак в подібних часових рядах. Наприклад, модель Deeplog [30] фокусується на аналізі виявлених аномалій в файлах журналів. Проте Wang Y. M. та Ji Z. X. [31] зазначили, що продуктивність Deeplog є незадовільною, тому вони провели її оптимізацію та, використовуючи алгоритм виявлення відхилень, запропонували напівконтрольовану модель для виявлення аномалій. Однак методи машинного навчання, що покладаються на мітки, не задовольняють вимогам систем виявлення аномалій у реальному часі, оскільки процес маркування даних є трудомістким. Для вирішення цієї проблеми доцільно застосовувати неконтрольовані методи аналізу журналів [32], які можуть ефективно виявляти аномалії без необхідності визначати явні ознаки, скорочувати час навчання та підвищувати ефективність обробки даних. На рис. 4 представлено схему моделі неконтрольованого аналізу журналів.

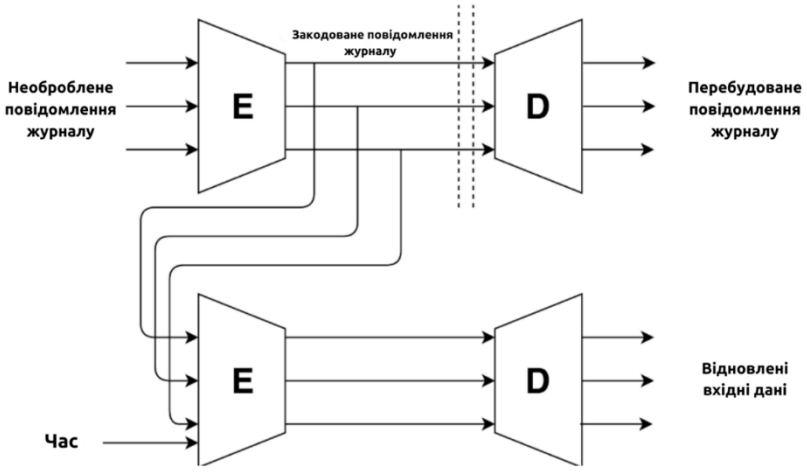

Рис. 4. Схема моделі неконтрольованого аналізу журналів

Дана модель визначає початковий автокодувальник **E**, що вставляє повідомлення журналу, тренується на тексті журналу (без позначки часу), щоб навчитися додавати повідомлення журналу фіксованої розмірності. Після навчання дешифратор **D** відкидається, після чого автокодувальник **E** для виявлення аномалій навчається на основі вкладених повідомлень та числової часової мітки повідомлення. Далі обчислюється міра відстані між входами та виходами, і вхід вважається аномальним, якщо його міра відстані перевищує відповідним чином вибраний поріг. Головна інновація цієї роботи полягає в тому, що модель не накладає вимог на структуру повідомлень журналу і не вимагає попередньої обробки повідомлень журналу, що забезпечує гнучкість до будь-якого типу журналів.

В цілому ж, застосування неконтрольованих методів при аналізі журналів активності дозволяє ефективно виявляти аномалії без потреби в маркуванні даних, що є суттєвою перевагою для систем, які працюють в режимі реального часу. Проте ці методи не завжди здатні забезпечити повну оцінку потенційних загроз, що виникають у результаті аномальної активності. З огляду на це, для підвищення точності оцінки загроз доцільно поєднувати методи аналізу журналів активності з механізмом постійного моніторингу, який дозволяє відстежувати доступ до ресурсів, виявляти потенційні загрози з боку вже авторизованих суб'єктів і тим самим запобігати їхнім протиправним діям. У цьому сенсі ШІ може сприяти підвищенню ефективності цього процесу. А саме, ШІ забезпечує можливість безперервного моніторингу, що дозволяє адаптивно контролювати доступ на основі поведінки користувачів та пристроїв. Це сприяє створенню більш динамічного та безпечного середовища контролю доступом. Завдяки аналізу активності, системи на базі ШІ спроможні виявляти аномальну або потенційно небезпечну поведінку, що може свідчити про несанкціонований доступ чи компрометацію облікових записів. При цьому слід мати на увазі, що поведінка користувачів, яка, в тому числі, може бути об'єднана в деякі групи, може змінюватись в залежності від конкретних умов, що створює нові виклики у виявленні аномалій. Тому для вирішення цієї проблеми у свій час був запропонований раніше згаданий метод кластеризації [33], заснований на аналізі траєкторій поведінки користувачів у програмному середовищі, в якому дані про доступ та операції користувачів у матриці траєкторій підлягають відповідній трансформації, в результаті якої визначається подібність поведінки. З метою покращення точності виявлення аномалій у поведінці користувачів та пристроїв в системах, що здійснюють моніторинг, можуть ефективно використовуватись алгоритми глибокого навчання (Deep Learning – DL). Наприклад, алгоритми DL, такі як мережі на основі довгої короткочасної пам'яті (Long short-term memory – LSTM) [34], що дозволяють ефективно виокремлювати часові ознаки в поведінці користувачів і пристроїв, запропонований авторами роботи [35] метод виявлення аномалій у мережевій поведінці користувачів, який базується на гібридному алгоритмі машинного навчання, або модель виявлення аномальної поведінки користувачів, яка використовує

LSTM для моделювання шаблонів поведінки в рамках активності користувачів, зокрема під час їх комунікації з сервісами та ресурсами [36]. Також може бути використана вдосконалена модель виявлення загроз, яка базується на використанні двонаправленої мережі LSTM (Bi-LSTM) для більш ефективного вибору ознак, а також метод опорних векторів (Support Vector Machine – SVM) для класифікації поведінки користувачів на звичайну (normal) або зловмисну (malicious) [37].

Однак навіть за умов високої точності в аналізі поведінки користувачів та пристроїв, самостійне функціонування таких систем не гарантує комплексного керування загрозами без належної класифікації та централізованого керування виявленими інцидентами. Більше того, у разі ідентифікації аномалій вони не завжди автоматично корелюються з іншими подіями безпеки, що ускладнює формування загальної картини загроз. Відсутність своєчасного сповіщення адміністраторів безпеки або невчасне застосування контрзаходів може призвести до негативних наслідків для інформаційної інфраструктури організації. Для вирішення цих проблем важливим є інтеграція систем моніторингу користувачів та пристроїв із зовнішніми компонентами архітектури нульової довіри, зокрема з SIEM.

2.3.3. SIEM

Автоматизована SIEM (*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*)) є ефективним рішенням проблеми автоматичного виявлення аномалій або загроз у поведінці не тільки користувачів, а й пристроїв. Дане рішення може забезпечити ефективну класифікацію та управління цими подіями безпеки. Більше того, при виявленні подій безпеки відбувається автоматичне попередження адміністратора з безпеки для вживання контрзаходів.

Однак існуючі системи SIEM мають недоліки через обмеженість в аналізі великих обсягів даних [38]. Тому до таких систем доцільно впроваджувати технології машинного навчання.

Наприклад, автори роботи [39] впровадили технології машинного навчання в SIEM і довели можливість ефективного аналізу великих об'ємів даних у цій системі. В інший роботі [40] автори зосередилися на поєднанні даних з різних джерел для вдосконалення SIEM та застосували моделі виявлення вторгнень на основі нейронних мереж (рис. 5).

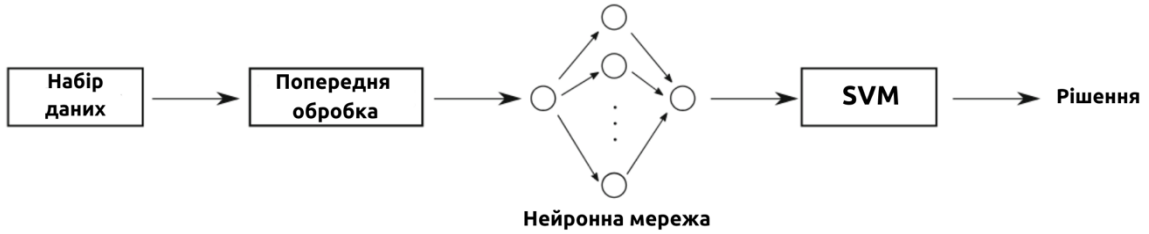

Рис. 5. Система розпізнавання на основі SVM-нейромережі

Запропонована система використовувала набір даних NSL-KDD (Network Security Layer-Knowledge Discovery Database) як вхідний набір даних, де перший шар моделі використовував нейронні мережі для класифікації системних подій на зловмисні та звичайні, а другий шар використовував метод SVM для підвищення продуктивності класифікації. У рамках даного дослідження визначено низку параметрів, які мають суттєвий вплив на продуктивність нейронних мереж у завданнях виявлення вторгнень. До таких параметрів належить кількість обраних ознак або атрибутів, які визначають обсяг вхідних даних і можуть значно впливати на точність класифікації. Важливим фактором є нормалізація даних, яка забезпечує уніфікацію масштабу вхідних параметрів, сприяючи кращій збіжності алгоритму, а також характеристики архітектури нейронної мережі, зокрема кількість вузлів у прихованому шарі, що визначає складність і здатність моделі до узагальнення. Крім того, важливо враховувати вибір функції активації, яка впливає на нелінійність мережі, а також параметри швидкості

навчання та моменту часу, що оптимізують процес збіжності. Ці аспекти визначають ефективність роботи моделі та її здатність до адаптації в умовах варіативності вхідних даних. Для перевірки цих припущень у роботі [40] було реалізовано 37 нейронних мереж прямого поширення (Feed-Forward Neural Networks – FFNN) з використанням різних алгоритмів навчання та функцій вартості, а отримані результати підтвердили критичну важливість цих параметрів для досягнення високої точності. Зокрема, встановлено, що вибір відповідної функції вартості дозволяє підвищити точність алгоритму *traingd* (алгоритм оберненого градієнтного спуску для навчання нейронної мережі) на 17,74 %. Аналіз набору даних NSL-KDD показав, що алгоритм *trainrp* (алгоритм стійкого оберненого поширення для навчання нейронної мережі) демонструє швидшу збіжність та вищу точність класифікації порівняно з іншими методами навчання. Крім того, точність *trainrp* виявилася співставною, а подекуди й вищою, ніж у більш складних моделей на основі нейронних мереж, що підкреслює його потенціал для покращення ефективності та швидкості класифікації вторгнень.

У роботі [41] був розроблений набір інструментів для автоматичної класифікації подій на основі машинного навчання, за допомогою якого проводилися експерименти з різними алгоритмами машинного навчання на декількох наборах даних, з метою – знайти найбільш ефективну модель класифікації даних. Результати проведених тестів показали, що метод опорних векторів (SVM) досяг найкращого показнику ефективності на наборі даних *TippingPoint* (набір IP-адрес та DNS-імен, які представляють потенційні ризики для мережевої безпеки) і становив 95,08 %.

Певні автоматизовані рішення SIEM активно застосовуються у різних практичних сценаріях, зокрема, в критичній інфраструктурі. Наприклад, Hindy H. та інші [42] розробили SIEM для виявлення аномальних подій у системі водопостачання, що контролюється системою SCADA (Supervisory Control And Data Acquisition). Використовуючи машинне навчання, дослідники класифікували дані про атаки на 14 різних сценаріях, які автоматично повідомлялися операторам служби безпеки. Представлені експериментальні сценарії охоплювали широкий спектр подій, від відмов обладнання до саботажу, й містили три дослідження із застосуванням шести методів машинного навчання. У першому експерименті оператору повідомлялася лише наявність аномалії у вигляді двійкового результату, без деталізації її типу. Другий експеримент уточнював пошкоджений компонент, надаючи інформацію про дані одного або кількох датчиків. Третій експеримент, що забезпечив найвищу ефективність, деталізував аномалію до рівня конкретного сценарію, дозволяючи операторам вживати коригувальних заходів. Загальна точність досягала 94 % для двійкової класифікації та 95,64 % для визначення сценаріїв, причому алгоритми класифікації KNN (K-Nearest Neighbors Algorithm), Decision Trees і Random Forests перевершили Gaussian Naive Bayes і SVM. Алгоритм KNN показав найвищу точність виявлення аномалій та ідентифікації конкретних сценаріїв атак у всіх експериментах. Результати підтвердили важливість використання рівня довіри для підвищення інформативності повідомлень, а також актуальність збільшення обсягу даних і побудови гібридних моделей для оптимізації класифікації подій і сценаріїв. Хоча запропонована модель прискорювала реагування на мережеві атаки, автори зазначили її обмеження у виявленні нових сценаріїв загроз.

Попри високу точність класифікації подій в системах критичної інфраструктури, важливим викликом залишається ефективна обробка сповіщень у масштабних системах безпеки. Дослідження Feng C. та інших [43] акцентує увагу на проблемі високої частоти хибних сповіщень у наявних SIEM, що значно перевищує можливості обробки операційними центрами безпеки (Security Operation Center – SOC). Для її розв’язання запропоновано систему, яка завдяки алгоритмам машинного навчання суттєво знижувала частоту хибних спрацьовувань, підвищуючи ефективність реагування на загрози. Сильною стороною таких підходів є здатність навчатися як на попередніх, так і на поточних даних для прогнозування майбутніх інцидентів безпеки, що може допомогти аналітикам швидше ідентифікувати та реагувати на загрози.

Досягнення досліджень в області аналітики та машинного навчання для SIEM знайшли своє продовження в сучасних системах безпеки для оркестрації, автоматизації та реагування (Security Orchestration, Automation, and Response – SOAR). SOAR-системи у поєднанні зі ШІ та машинним навчанням, демонструють значний потенціал у виявленні, пом'якшенні та запобіганні кіберзагрозам. Постачальники даних інструментів інтегрують алгоритми ШІ та машинного навчання для підвищення ефективності роботи аналітиків SOC [44]. Водночас виникають питання ефективності алгоритмів класифікації та визначення пріоритетності інцидентів, а також необхідність подальших досліджень щодо впровадження глибокого посиленого навчання (Deep Reinforced Learning – DRL) у SOAR-системи. Для вирішення даної проблеми використовують нові моделі інтерпретовного ШІ (Explainable Artificial Intelligence – XAI), які забезпечують зрозумілість та прозорість у поясненні результатів прогнозування. Зокрема, XAI дозволяє аналітикам безпеки виявляти помилкові прогнози шляхом оцінки надійності результатів. Це значно зменшує кількість хибних спрацьовувань і підвищує ефективність роботи SOC, оскільки аналітики отримують змогу швидше та точніше реагувати на реальні загрози [45].

В цілому ж, можна помітити, що сучасні виклики, пов'язані зі складністю виявлення та нейтралізації кіберзагроз, вимагають комплексного підходу. Зокрема, для ефективного функціонування SOC необхідна інтеграція алгоритмів машинного навчання на всіх етапах роботи – від збору та обробки даних до аналізу загроз і управління інформаційною безпекою. Наприклад, для цього у роботі [46] пропонується багаторівнева модель SOC, яка включає збір даних, їх обробку, застосування алгоритмів машинного навчання для аналізу загроз та використання інформаційних панелей для прийняття рішень. Серед алгоритмів машинного навчання дослідники пропонують використовувати як контрольовані алгоритми (SVM, дерева рішень), так і неконтрольовані (кластеризація KNN). В іншої роботі [47] автори запропонували впровадити інструменти на основі ШІ для диференціації загроз і зменшення втоми від численних сигналів зі сповіщеннями безпеки для аналітиків SOC. Також, вони акцентували увагу на необхідності використання алгоритмів SVM для аналізу попередніх даних та оцінки рівня загроз.

Таким чином, інтеграція ШІ та машинного навчання, включно з підходами DRL, дозволяє автоматизувати рутинні процеси, одночасно підвищуючи ефективність безпеки. А всі згадані вище дослідження, що демонструють, як застосування алгоритмів ШІ та машинного навчання може сприяти підвищенню точності виявлення загроз та зниженню кількості помилкових спрацьовувань, зайвий раз підтверджують це та доводять доцільність використання можливостей ШІ в рамках ZTA.

2.3.4. Аналіз загроз

Інформаційний(і) канал(и) про загрози / система аналізу загроз (Threat Intelligence – TI) є одним із ключових елементів в рамках ZTA, що сприяє захищеності підприємства від зовнішніх і внутрішніх атак. Це інформація з внутрішніх або зовнішніх джерел, яка допомагає механізму політики приймати рішення про доступ. З огляду на це, основні техніки системи TI можна поділити на два основні напрямки: використання інформаційних каналів для отримання даних з різних джерел, і безпосередньо аналіз великих обсягів даних для виявлення аномалій. Перший напрямок включає використання служб, які збирають дані про нещодавно виявлені атаки, вразливості, шкідливе програмне забезпечення чи інші загрози з різних джерел. Другий напрямок зосереджений на аналізі даних у внутрішніх системах для виявлення аномальних дій, які можуть вказувати на потенційні атаки, що дозволяє своєчасно реагувати на підозрілі події та мінімізувати шкоду від кібератак.

Інформація, отримана завдяки TI, дозволяє не лише виявляти атаки та вразливості, а й прогнозувати їх розвиток та завчасно оцінювати потенційні ризики. Однак оскільки джерел такої інформації може існувати багато, виникає природна необхідність впровадження автоматизованих систем, у тому числі або насамперед, заснованих на використанні ШІ, здатних збирати дані TI та оцінювати їх достовірність. У зв'язку з цим, наприклад, у роботі [48] авто-

ри розглянули сучасні технології у сфері аналізу загроз та запропонували нову модель прогнозування загроз на основі функції оцінки ризику та вдосконаленого алгоритму апостеріорної ймовірності Баєса (Enhanced Naive Bayes Posterior Probability – ENBPP) з використанням машинного навчання. Запропонований ними алгоритм об'єднує модифіковану функцію ENBPP з адаптованою функцією оцінки ризику, що забезпечує підвищення точності прогнозування загроз та скорочення часу обробки. Для аналізу ефективності рішення вони використали п'ять різних наборів даних, які містили 328 814 зразків загроз. Отримані результати засвідчили про перевагу запропонованого підходу, оскільки точність прогнозування зросла до 92–96 %, а середній час обробки знизився з 0,043 до 0,028 секунди порівняно з альтернативними методами. Крім того, алгоритм продемонстрував здатність ефективно долати проблеми, пов'язані із залежністю від заздалегідь визначених шаблонів дій зловмисників і порогових значень у множинних сценаріях прогнозування. В цілому ж новий підхід здатний забезпечити більш надійний механізм аналізу та прогнозування загроз у різних сценаріях.

Хакерські форуми є також одним із ключових джерел даних для TI, про що свідчить, зокрема, дослідження, проведене в роботі [49], спрямоване на автоматизоване отримання відповідної інформації про кіберзагрози з хакерських форумів за допомогою гібридного процесу машинного навчання. Авторами [49] було розроблено дворівневий алгоритм, що поєднує метод SVM для фільтрації нерелевантних повідомлень та моделювання латентних/прихованих розподілів Дирихле (Latent Dirichlet Allocation – LDA) для кластеризації релевантних постів за відповідною тематикою. У межах експерименту проаналізовано мільйон постів із реального хакерського форуму, що дозволило ідентифікувати такі ключові загрози, як витік облікових даних, шкідливі проксі-сервери та програмне забезпечення. Результати продемонстрували ефективність методу, оскільки використання SVM для відсіювання нерелевантних повідомлень зменшило обсяг даних для аналізу більш ніж на 90 %, що дозволило значно скоротити час обробки. Наприклад, середній час для моделювання тем за допомогою LDA знизився з 238 до 16 хвилин за умови обмеження словникового запасу до 50 000 слів. У сукупності тематичне моделювання дозволило виділити ключові теми, такі як обговорення скомпрометованих облікових записів, IP-адрес шкідливих проксі та програм, що робить цю методику цінним інструментом для виявлення оперативної інформації про загрози. А модель продемонструвала високу гнучкість у питанні налаштування параметрів, що дозволяє в майбутньому адаптувати її до різних типів даних і завдань. Крім того, варто зазначити, що в рамках даної системи важливим є участь людини-експерта для остаточної інтерпретації результатів даного тематичного моделювання, оскільки навіть найбільш досконалі алгоритми можуть не враховувати специфіку рідкісних або складних тем, наприклад, вразливості нульового дня. Але загалом, дана гібридна система є ефективним рішенням для швидкого вилучення важливої інформації, яка може бути доповнена до традиційних засобів захисту для підвищення їх ефективності.

Застосування сучасних методів для аналізу загроз на основі вдосконалених алгоритмів оцінки ризиків та автоматизованих систем обробки великих обсягів даних, сприяє покращенню виявлення та прогнозування потенційних атак. Однак ефективність таких підходів значною мірою залежить від якості вхідних даних, адаптивності моделей до нових типів загроз і можливості інтеграції результатів у комплексні системи кіберзахисту. Отже, можна зробити висновок, що потрібні додаткові дослідження у напрямку вдосконалення наявних підходів в аналізі загроз (і в першу чергу тих, що спираються на можливості ШІ), що дозволяють більш детально оцінити ризики для системи, визначити можливі вектори атак, встановити пріоритетність загроз та розробити відповідні заходи захисту або стратегії запобігання.

Таким чином, розглянуті дослідження та підходи до впровадження компонентів ZTA демонструють ключову роль машинного навчання та ШІ у забезпеченні концепції нульової довіри. Використання ШІ в основних і зовнішніх компонентах ZTA, таких як алгоритм довіри, контроль доступу, SIEM, керування ідентифікацією, журнали активності (в тому числі, моніторинг користувачів та пристроїв) та аналіз загроз, дозволяє автоматизувати аналіз ве-

ликих обсягів даних, виявляти аномалії та адаптивно реагувати на потенційні загрози. Алгоритми ШІ забезпечують постійний моніторинг, оптимізацію політик безпеки та скорочення часу реагування на інциденти, що значно підвищує ефективність захисту. Завдяки цьому модель ZTA у поєднанні з технологіями ШІ створює гнучку й надійну систему кібербезпеки, здатну протистояти сучасним загрозам у динамічному цифровому середовищі.

3. Виклики та подальший напрямок досліджень

З розвитком IoT, хмарних та інших інноваційних технологій, багато пристроїв можуть бути інтегровані до єдиної централізованої системи управління. Однак застаріла інфраструктура, програми, сервіси та інші елементи можуть не відповідати принципам нульової довіри, оскільки в них відсутні концепції мінімальних привілеїв, захисту від бічного переміщення, а також немає динамічної автентифікації. Через це застарілі системи залишаються вразливими до широкого спектра кіберзагроз. Можливим рішенням є додавання модуля для автентифікації до центральної системи управління з подальшим визначенням привілеїв. Хоча це рішення частково вирішує проблему застарілих систем, воно вимагає від суб'єкта прямого проходження через всю інфраструктуру, що порушує принцип мікросегментації, який є також основоположним у концепції ZTA. Такий підхід окреслює значні виклики для інтеграції застарілих систем у сучасні моделі безпеки, засновані на концепції нульової довіри, та вимагає подальших досліджень у напрямку адаптації існуючої інфраструктури до нових стандартів кіберзахисту.

З іншого боку, впровадження ZTA потребує поступового переходу, оскільки різкий перехід може викликати значні труднощі, зокрема збільшення кількості помилкових спрацювань, зниження продуктивності системи та опір з боку персоналу. Тобто впровадження ZTA потребує ретельного планування та має бути поетапним, з урахуванням особливостей підприємства, його бізнес-процесів, критичних активів та рівня кіберзагроз (при цьому з мінімальним негативним впливом на роботу підприємства). Особливо актуальним є питання адаптації персоналу до нових принципів роботи, зокрема шляхом розробки ефективних стратегій навчання та автоматизованих засобів підтримки прийняття рішень у процесі переходу. Крім того, перспективним напрямом є дослідження методів оцінки ефективності впровадження ZTA та розробка відповідних метричних показників, які дозволять об'єктивно аналізувати та оцінювати рівень захищеності підприємств.

Ще одним важливим викликом для ZTA є стандартизація даних. Інформація про загрози може надходити з різних джерел, проте на сьогодні в ZTA немає єдиного стандарту, який може бути використаний алгоритмами оцінки довіри. Це впливає на достовірність даних та фінальну оцінку ефективності, оскільки вхідні дані надаються різними джерелами, а відсутність уніфікованого формату для зібраних даних ускладнює використання відповідних алгоритмів. Важливим аспектом цієї проблеми, також є моніторинг мережевого трафіку та поведінки користувачів у системі, який значною мірою залежить від журналів активності, що надаються різними інструментами безпеки. Це вимагає від систем використовувати різні алгоритми оцінки довіри для адаптації до різних форматів даних. Такий підхід не лише ускладнює процес аналізу та оцінки, але й призводить до зниження продуктивності моделі оцінки довіри. Крім того, кожне джерело даних в ZTA має власні операційні політики та стандарти, що створює додаткові труднощі при автоматизації процесів. Відсутність єдиної політики, яка б регулювала взаємодію компонентів ZTA, зокрема, політики шифрування, специфікації коду тощо, ускладнює впровадження ефективних механізмів кібербезпеки.

Дані виклики підкреслюють необхідність в уніфікації підходів до обробки та стандартизації даних для забезпечення ефективної роботи в рамках моделі нульової довіри. Подальші дослідження доцільно було б направити, за можливістю, на створення єдиних стандартів, які б забезпечили узгодженість даних та процедур безпеки, що дозволило б підвищити ефективність, точність моделей оцінки та узгодженість процесів автоматизації безпеки в рамках ZTA.

ZTA в поєднанні з елементами ШІ має значні переваги в автоматизації процесів, однак покладатися виключно на рішення прийняті ШІ може бути ризиковано, оскільки можливі помилкові або упереджені висновки, що в результаті може призводити до хибних спрацьовувань або невиправданих відмов. У зв'язку з цим, важливим є інтеграція людського досвіду в процес ухвалення рішень. Включення людини до циклу (human-in-the-loop) допоможе знизити або усунути ризики, пов'язані з помилками ШІ. Наприклад, якщо система ZTA на базі ШІ неправомірно відмовляє в доступі легітимному користувачеві, експерт може переоцінити рішення та надати зворотний зв'язок, що дозволить підвищити точність спрацювань системи. В результаті, система зі ШІ буде навчатися на своїх помилках, а, отже, покращувати свою ефективність. Таким чином, використання підходу human-in-the-loop є перспективним напрямом для покращення точності і ефективності рішень ШІ в ZTA. Водночас, ефективність цього підходу значною мірою залежатиме від якості та доступності даних для навчання моделей ШІ, які, у свою чергу, значною мірою залежатимуть від великих наборів даних для навчання. При цьому, слід звернути увагу і на те, що скомпрометовані дані теж можуть серйозно вплинути на ефективність систем на базі ШІ.

Однією з потенційних загроз, як відомо, є атаки типу отруєння даних (data poisoning), коли шкідливі дані вводяться до навчальної вибірки для маніпулювання рішеннями системи, що призводить до некоректних результатів. Для мінімізації ризиків подібних атак необхідно впроваджувати надійні методи очищення даних та перевірки їх цілісності та якості, використовувати набори даних із різних джерел, а також застосовувати методи випадкової вибірки для зниження впливу потенційних загроз. Враховуючи, що якість рішень ШІ залежить від достовірності вхідних даних, важливим аспектом є також забезпечення коректної взаємодії між компонентами безпеки, що здійснюють збір та аналіз даних. У цьому контексті інтеграція систем SIEM та SOAR у ZTA постає як окремий виклик, де ключовою задачею є забезпечення узгодженості між автоматизованими процесами даних систем та принципами ZTA.

SIEM, як система збору й аналізу даних про події безпеки, часто стикається з проблемами уніфікації форматів даних, що ускладнює автоматичну кореляцію подій між різними джерелами. Це призводить до підвищення кількості хибних спрацювань та перевантаження аналітиків, особливо в умовах обробки великих обсягів різнорідної інформації. У свою чергу, SOAR забезпечує автоматизацію IR, але залежить від якості вхідних даних і сценаріїв, створених для обробки загроз. Недоліки у стандартизації та адаптації алгоритмів, які використовуються в цих системах, можуть знижувати точність та ефективність обробки інцидентів у реальному часі. Тому для подолання даних викликів доцільно було б певним чином стандартизувати дані, що надходять до SIEM, а також удосконалити механізми обробки даних, щоб забезпечити відповідність принципам нульової довіри. Наприклад, алгоритми аналізу подій мають бути адаптовані до динамічних політик доступу, характерних для ZTA, а процеси автоматизації SOAR повинні враховувати необхідність постійної перевірки довіри суб'єктів і пристроїв. Подальші дослідження можуть бути спрямовані на інтеграцію алгоритмів машинного навчання для оптимізації виявлення аномалій і мінімізації хибних спрацювань, а також на розробку єдиних протоколів обміну даними між системами SIEM, SOAR та компонентами ZTA. У зв'язку з цим, інтеграція SIEM та SOAR у рамках ZTA вимагає не лише технологічної адаптації, але й розробки уніфікованих стандартів для забезпечення узгодженості й ефективності.

Таким чином, використання можливостей ШІ в рамках ZTA для забезпечення кібербезпеки підприємств є актуальним питанням, але водночас вимагає подальших досліджень. Подальші дослідження мають бути спрямовані на створення надійної методології для впровадження ZTA та алгоритмів ШІ в її компоненти, розробку ефективних механізмів захисту від атак на моделі машинного навчання та стандартизацію обміну даними для підвищення узгодженості та ефективності кібербезпеки. Успішна реалізація даних підходів сприятиме розвитку надійних систем кібербезпеки, здатних реагувати на сучасні загрози в умовах постійної мінливості та динаміки загроз.

Висновки

1. Для забезпечення надійного захисту сучасного цифрового підприємства необхідна комплексна стратегія, концепція, яка забезпечує безпечний доступ до корпоративних ресурсів у будь-який час і в будь-якому місці. Враховуючи сучасні виклики, такою концепцією є парадигма безпеки «нульової довіри». Вона є ефективним підходом для захисту інформаційних систем від новітніх загроз, де кожен доступ має контролюватися та проходити верифікацію.

2. ШІ відіграє важливу роль у розвитку автоматизованих рішень для ZTA, зокрема в таких сферах, як виявлення атак, керування доступом, моніторинг, аналіз загроз тощо. Проте, незважаючи на значний прогрес у цій сфері, залишаються невирішеними питання щодо повної інтеграції ШІ в процес автоматизації для всіх компонентів ZTA. Як показують результати досліджень у цьому напрямі у світі лише в окремих логічних компонентах ZTA використовуються деякі методи автоматизації на основі можливостей ШІ. Насправді ж, механізми, що використовують ШІ для автоматизації процесів виявлення загроз та управління доступом у рамках ZTA, можуть істотно допомогти розробникам та спеціалістам з безпеки досягти більшої ефективності при впровадженні ZTA на підприємствах, підвищити ефективність захисту їх інформаційних систем загалом. Це дозволить створити більш надійні та масштабовані механізми захисту, що відповідають вимогам сучасного цифрового середовища.

3. Представлені результати аналізу сучасного стану використання ШІ в рамках концепції нульової довіри показують, що існуючі виклики та напрямки потребують подальшого дослідження можливостей застосування ШІ для автоматизації прийняття рішень. Встановлено, що подальші дослідження доцільно зосередити на розробці нових методів для покращення взаємодії між ZTA та ШІ, стандартизації механізмів обміну даними та методології поступового впровадження ZTA. При цьому слід не забувати про важливість так званого підходу human-in-the-loop для підвищення точності рішень ШІ, що в цілому сприятиме створенню більш адаптивних, надійних та ефективних систем кібербезпеки.

Застосування технологій ШІ у рамках концепції нульової довіри відкриває значні перспективи для розвитку більш надійних та масштабованих систем захисту інформаційних ресурсів, що є надзвичайно актуальним у сучасному цифровому кіберпросторі.

Список літератури:

1. Pentera: The State Of Pentesting 2024 Survey Report. (2024). URL: https://pentera.io/resources/reports/the-state-of-pentesting-2024-survey-report/.
2. PwC: The macroeconomic impact of artificial intelligence (2018). URL: https://www.pwc.co.uk/economic-services/assets/macroeconomic-impact-of-ai-technical-report-feb-18.pdf.
3. Cost of a data breach (2024). URL: https://www.ibm.com/reports/data-breach.
4. National Cybersecurity Center of Excellence (NCCoE). Implementing a Zero Trust Architecture. URL: https://www.nccoe.nist.gov/projects/implementing-zero-trust-architecture.
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. https://doi.org/10.30837/rt.2024.2.217.03.
6. Rose S., Borchert O., Mitchell S., & Connelly S. Zero Trust Architecture. NIST Special Publication 800-207 // National Institute of Standards and Technology. 2020. 59 p. https://doi.org/10.6028/NIST.SP.800-207.
7. Ahmed I., Nahar T., Urmi S. S., Taher K. A. Protection of Sensitive Data in Zero Trust Model // Proceedings of the International Conference on Computing Advancements. 2020. Vol. 63. P. 1–6. https://doi.org/10.1145/3377049.3377114.
8. Zero Trust Security (2024). URL: https://www.akamai.com/solutions/security/zero-trust-security.
9. Kindervag J. No More Chewy Centers: Introducing The Zero Trust Model Of Information Security. Forrester Research, For Security & Risk Professionals. URL: https://media.paloaltonetworks.com/documents/Forrester-NoMore-Chewy-Centers.pdf.
10. Sarkar S., Choudhary G., Shandilya S. K., Hussain A., Kim H. Security of Zero Trust Networks in Cloud Computing: A Comparative Review // Sustainability. 2022. 14(18). 11213. https://doi.org/10.3390/su141811213.
11. Cunningham C., Balaouras S., Barringham B., Dostie P. The Zero Trust eXtended (ZTX) Ecosystem. Extending Zero Trust Security Across Your Digital Business. Forrester Research, Inc. Cambridge, MA. 2018. URL: https://www.cisco.com/c/dam/m/en_sg/solutions/security/pdfs/forrester-ztx.pdf.

12. Fisher B. Forrester's Zero Trust or Gartner's Lean Trust? 2019.
URL: https://blogs.cisco.com/security/forresters-zero-trust-or-gartners-lean-trust.
13. Ward R., Beyer B. Beyondcorp // A new approach to enterprise security. 2014. 39(6). P. 6–11.
14. Oracle. Zero-trust security model. 2024. URL: https://www.oracle.com/nl/security/what-is-zero-trust/.
15. National Cybersecurity Center of Excellence (NCCoE). Implementing a Zero Trust Architecture. URL: https://www.nccoe.nist.gov/projects/implementing-zero-trust-architecture.
16. Fortinet. The State of Zero Trust. Report. 2023.
URL: https://www.fortinet.com/content/dam/fortinet/assets/reports/report-state-of-zero-trust.pdf.
17. Єсін В. І., Вілігура В. В., Узлов Д. Ю. Архітектура нульової довіри: проблеми та рекомендації щодо успішного впровадження. Радіотехніка. 2024. Вип. 218. С. 7–34. https://doi.org/10.30837/rt.2024.3.218.01.
18. Garbis J., Chapman J. W. Zero Trust Security: An Enterprise Guide. Berkeley, CA : Apress, 2021. 300 p.
19. ManageEngine: Adopting Zero Trust to safeguard against generative AI cyberthreats (2024).
URL: https://www.manageengine.com/active-directory-360/ebooks/zero-trust-approach-to-combating-gen-ai-cyberattacks.html.
20. Wang J. W., Jing X. Y., Yan Z., Fu Y. L., Pedrycz W., Yang L. T. A survey on trust evaluation based on machine learning // ACM Computing Surveys. 2020. 53(5). P. 1–36. https://doi.org/10.1145/3408292.
21. Ajish D. The significance of artificial intelligence in zero trust technologies: a comprehensive review // Journal of Electrical Systems and Inf Technol. 2024. 11(30). P. 1–23. https://doi.org/10.1186/s43067-024-00155-z.
22. Rangaraju S. Secure by intelligence: enhancing products with AI-driven security measures // EPH – International Journal of Science and Engineering. 2023. 9(3). P. 36–41. https://doi.org/10.53555/ephijse.v9i3.212.
23. Suleski T., Ahmed M., Yang W., Wang E. A review of multi-factor authentication in the Internet of Healthcare Things // Digital Health. 2023. Vol. 9. P. 1–20. https://doi.org/10.1177/20552076231177144.
24. Borodavka V., Tsuranov M. Biometrics: analysis and multi-criterion selection // The 9th IEEE International Conference on Dependable Systems, Services and Technologies, DESSERT, Kyiv, Ukraine. 2018. P. 334–339. https://doi.org/10.1109/DESSERT.2018.8409152.
25. Bodepudi A., Reddy M., Gutlapalli S. S., & Mandapuram M. Voice Recognition Systems in the Cloud Networks: Has It Reached Its Full Potential? // Asian Journal of Applied Science and Engineering. 2019. 8(1). P. 51–60. https://doi.org/10.18034/ajase.v8i1.12.
26. Stouffer C. What is facial recognition and how does it work? 2023. URL: https://us.norton.com/blog/iot/how-facial-recognition-software-works.
27. Ryu R., Yeom S., Kim S. H., Herbert D. Continuous multimodal biometric authentication schemes: A systematic review // IEEE Access. 2021. Vol. 9. P. 34541–34557. https://doi.org/10.1109/ACCESS.2021.3061589.
28. Germain K. S., Kragh F. Mobile physical-layer authentication using channel state information and conditional recurrent neural networks. In Proceedings of the 93rd IEEE Vehicular Technology Conference, Helsinki, Finland. 2021. P. 1–6. https://doi.org/10.1109/VTC2021-Spring51267.2021.9448652.
29. Meng R., Xu B., Xu X., Sun M., Wang B., Han S., Lv S., Zhang P. A survey of machine learning-based physical-layer authentication in wireless communications // Journal of Network and Computer Applications. 2024. 111 p. https://doi.org/10.48550/arXiv.2411.09906.
30. Du M., Li F. F., Zheng G. N., Srikumar V. DeepLog: Anomaly detection and diagnosis from system logs through deep learning // Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, Dallas, USA. 2017. P. 1285–1298. https://doi.org/10.1145/3133956. 3134015.
31. Wang Y. M., Ji Z. X. Design and implementation of a semi-supervised anomaly log detection model DDA // Proceedings of International Conference on Computer Communication and Artificial Intelligence, Guangzhou, China. 2021. P. 86–90. https://doi.org/10.1109/CCAI50917.2021. 9447533.
32. Bursic S., Cuculo V., D'Amelio A. Anomaly detection from log files using unsupervised deep learning // Proceedings of International Symposium on Formal Methods, Porto, Portugal. 2019. P. 200–207. https://doi.org/10.1007/978-3-030-54994-7_15.
33. Tang Y. P., Ma B. X., Wu Z. Research on user clustering algorithm based on software system user behavior trajectory // Proceedings of the 2nd International Conference on Big Data Technologies, Jinan, China. 2019. P. 11–14. https://doi.org/10.1145/3358528.3358572.
34. Zhao Z., Chen W., Wu X., Chen P.C.Y., Liu J. LSTM network: A deep learning approach for Short-term traffic forecast // IET Intelligent Transport Systems. 2017. 11(2). P. 68–75. https://doi.org/10.1049/iet-its.2016.0208.
35. Singh M., Mehtre B. M., Sangeetha S. User behavior profiling using ensemble approach for insider threat detection // Proceedings of the 5th IEEE International Conference on Identity, Security, and Behavior Analysis, Hyderabad, India. 2019. P. 1–8. https://doi.org/10.1109/ISBA.2019. 8778466.
36. Sharma B., Pokharel P., Joshi B. User behavior analytics for anomaly detection using LSTM autoencoder-insider threat detection // Proceedings of the 11th International Conference on Advances in Information Technology, Bangkok, Thailand. 2020. Vol. 5. P 1–9. https://doi.org/10.1145/3406601.3406610.
37. Singh M., Mehtre B. M., Sangeetha S. User behaviour based insider threat detection in critical infrastructures // Proceedings of the 2nd International Conference on Secure Cyber Computing and Communications, Jalandhar, India. 2021. P. 489–494. https://doi.org/10.1109/ICSCCC51823.2021.9478137.

38. Marchal S., Jiang X. Y., State R., Engel T. A big data architecture for large scale security monitoring // Proceedings of IEEE International Congress on Big Data, Anchorage, USA. 2014. P. 56–63. https://doi.org/10.1109/BigData. Congress.2014.18

39. Li T. M., Yan L. M. SIEM based on big data analysis // Proceedings of the 3rd International Conference on Cloud Computing and Security, Nanjing, China. 2017. P. 167–175. https://doi.org/10.1007/978-3-319-68505-2_15.

40. El Hajji S., Moukafih N., Orhanou G. Analysis of neural network training and cost functions impact on the accuracy of IDS and SIEM systems // Proceedings of the 3rd International Conference on Codes, Cryptology, and Information Security, Rabat, Morocco. 2019. P. 433–451. https://doi.org/10.1007/978-3-030-16458-4_25.

41. Hossain S. M. M., Couturier R., Rusk J., Kent K. B. Automatic event categorizer for SIEM // Proceedings of the 31st Annual International Conference on Computer Science and Software Engineering, Toronto, Canada. 2021. P. 104–112. https://dl.acm.org/doi/10.5555/3507788.3507803.

42. Hindy H., Brosset D., Bayne E., Seeam A., Bellekens X. Improving SIEM for critical SCADA water infrastructures using machine learning // Proceedings of International Workshop on Security and Privacy Requirements Engineering, Barcelona, Spain. 2019. P. 3–19. https://doi.org/10.1007/978-3-030-12786-2_1.

43. Feng C., Wu S. N., Liu N. W. A user-centric machine learning framework for cyber security operations center. In Proceedings of IEEE International Conference on Intelligence and Security Informatics, Beijing, China. 2017. P. 173-175. https://doi.org/10.1109/ISI.2017.8004902.

44. Kinyua J., Awuah L. AI/ML in security orchestration, automation and response // Future research directions. Intelligent Automation & Soft Computing. 2021. 28(2). P. 527–545. https://doi.org/10.32604/iasc.2021.016240.

45. Aslam N., Khan I.U., Mirza S., AlOwayed A., Anis F.M., Aljuaid R.M., Baageel R. Interpretable Machine Learning Models for Malicious Domains Detection Using Explainable Artificial Intelligence (XAI) // Sustainability. 2022. 14(12), 7375. P. 1–22. https://doi.org/10.3390/su14127375.

46. Yeshwanth M.V., Kalluri R., Rao M.S., Kumar R.K.S., Bindhumadhava B.S. Adoption and Assessment of Machine Learning Algorithms in Security Operations Centre for Critical Infrastructure // Pillai R.K., Ghatikar G., Sonavane V.L., Singh B.P. (eds) ISUW 2020. Lecture Notes in Electrical Engineering, Springer, Singapore. 2022. № 847. P. 395–407. https://doi.org/10.1007/978-981-16-9008-2_38

47. Ban T., Ndichu S., Takahashi T., Inoue D. Combat security alert fatigue with AI-assisted techniques // CSET –21: Cyber Security Experimentation and Test Workshop. 2021. P. 9–16. https://doi.org/10.1145/3474718.3474723.

48. Sentuna A., Alsadoon A., Prasad P. W. C., Saadeh M., Alsadoon O. H. A novel enhanced naive Bayes posterior probability (ENBPP) using machine learning: Cyber threat analysis // Neural Processing Letters. 2021. № 53(1). P. 177–209. https://doi.org/10.1007/s11063-020-10381-x.

49. Deliu I., Leichter C., Franke K. Collecting cyber threat intelligence from hacker forums via a two-stage, hybrid process using support vector machines and latent dirichlet allocation // Proceedings of IEEE International Conference on Big Data, Seattle, USA. 2018. P. 5008–5013. https://doi.org/10.1109/BigData.2018.8622469.

Надійшла до редколегії 10.01.2025

Відомості про авторів:

Бородавка Владислав Вячеславович – Харківський національний університет імені В. Н. Каразіна, аспірант кафедри кібербезпеки інформаційних систем, мереж і технологій навчально-наукового інституту комп'ютерних наук та штучного інтелекту; Україна; e-mail: vladyslav.borodavka@karazin.ua; ORCID: https://orcid.org/0009-0002-3885-1364

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