УДК 004.056.53:004.7 DOI 10.18372/2073-4751.87.21681

Дячук О. Ю.[1],
orcid.org/0000-0002-6996-4700,
e-mail: kkik_doyu@ztu.edu.ua**,**
Кухарчук А. А.[1],
orcid.org/0009-0006-9843-1838,
e-mail: kb221_kaa@student.ztu.edu.ua**,**
Колощук М. С.[1],
orcid.org/0009-0001-5825-2054,
e-mail: kkik_kms@ztu.edu.ua**,**
Шелуха О. О.[1], к.т.н.**,**
orcid.org/0000-0002-6088-8262,
e-mail: oleksii.shelukha@ztu.edu.ua**,**
Квасніков В. П.[2], д.т.н.,
orcid.org/0000-0002-6525-9721,
e-mail: kvp@kai.edu.ua

BASELINE-ОРІЄНТОВАНА СИСТЕМА ВИЯВЛЕННЯ АНОМАЛІЙ МЕРЕЖЕВОГО ТРАФІКУ З КОРЕЛЯЦІЄЮ ПОДІЙ У SIEM WAZUH

[1]Державний університет «Житомирська політехніка»
[2]Національний університет «Київський авіаційний інститут»

Вступ

Корпоративні мережі залишаються основним об'єктом атак, які розгортаються поступово: після первинного проникнення зловмисник виконує розвідку внутрішньої інфраструктури, переміщується між сегментами, встановлює канал керування та виводить дані. Кожен із цих етапів залишає слід у мережевому трафіку, проте засоби виявлення, побудовані на сигнатурах, фіксують лише ті події, для яких заздалегідь описано шаблон. Нова комунікаційна пара хостів, SSH-сесія з користувацького сегмента чи періодичні з'єднання з зовнішнім вузлом не відповідають жодній сигнатурі, хоча саме такі відхилення часто передують інциденту [1, 2].

Поведінковий підхід розв'язує цю проблему інакше: замість пошуку відомого шкідливого шаблону система порівнює поточну активність із моделлю нормальної поведінки конкретної мережі – baseline. Відхилення від baseline (нові взаємодії, нетипові сервіси, активність поза робочими годинами, статистичні сплески) стає підставою для формування події безпеки навіть тоді, коли атака раніше не спостерігалася [3, 4]. Нормативні документи закріплюють цей підхід як складову архітектури захисту: NIST SP 800-94 розглядає аномалійно-орієнтоване виявлення як окремий клас IDPS [5], NIST SP 800-137 вимагає безперервного моніторингу стану інформаційних систем [6], а ISO/IEC 27001:2022 відносить моніторинг мережевої активності до обов'язкових механізмів системи управління інформаційною безпекою [7].

Практична реалізація поведінкового аналізу в корпоративному сегменті стикається з двома труднощами. По-перше, потрібен сенсор, здатний перетворювати сирий трафік на структуровані журнали з'єднань, DNS-запитів і сервісів – цю роль виконує мережевий аналізатор Zeek [8], який

пасивно обробляє копію трафіку та формує події через власний Notice Framework [9, 10]. По-друге, потрібна платформа, що централізовано декодує ці події, застосовує правила, поєднує пов'язані відхилення в один сценарій і візуалізує результат – таку функцію виконує SIEM-платформа Wazuh [11, 12]. Обидва інструменти мають відкритий код, що робить рішення відтворюваним для організацій без бюджету на комерційні NDR-продукти. Тим не менш їх інтеграція для baseline-орієнтованого виявлення вимагає додаткового проєктування, оскільки готових детекторів новизни комунікацій, часових політик чи кореляційних сценаріїв ані Zeek, ані Wazuh «з коробки» не надають.

У статті описано архітектуру, детектори та експериментальну перевірку системи виявлення поведінкових аномалій трафіку корпоративної мережі, побудованої на інтеграції Zeek і Wazuh із формалізованою baseline-моделлю, кастомними Zeek-скриптами, Wazuh-декодерами, правилами та кореляційними сценаріями.

Мета

Метою дослідження є розробка та експериментальна перевірка архітектури системи виявлення поведінкових аномалій трафіку корпоративної мережі, яка поєднує пасивний аналіз трафіку засобами Zeek, формалізовану модель нормальної поведінки (baseline), кастомні детектори відхилень і централізовану обробку подій у SIEM-платформі Wazuh з кореляцією пов'язаних аномалій та їх класифікацією за матрицею MITRE ATT&CK.

Для досягнення мети розв'язано такі завдання: побудовано модель baseline нормальної поведінки сегментованої корпоративної мережі; розроблено проєкти та реалізовано набір кастомних Zeek-детекторів для виявлення аномалій новизни, політик доступу та статистичних відхилень; розроблено Wazuh-декодери, спеціалізовані правила та кореляційні сценарії; проведено експериментальну верифікацію повного ланцюга виявлення аномалій мережевого трафіку на симуляційному стенді GNS3/VMware і проаналізовано джерела хибнопозитивних спрацювань та напрями калібрування системи.

Аналіз останніх досліджень і публікацій

Систематизацію методів виявлення аномалій закладено в класичній роботі Chandola, Banerjee і Kumar [3], де визначено типи аномалій (точкові, контекстні, колективні) та базові класи методів їх виявлення. Стосовно мережевих систем виявлення вторгнень Maseer зі співавторами [1] у метааналізі 2024 року показали, що більшість сучасних досліджень зосереджена на ML-моделях, які оцінюються на статичних датасетах, тоді як питання відтворюваного розгортання поведінкового виявлення в реальній інфраструктурі залишається менш опрацьованим. Lyu, Gharakheili та Sivaraman [2] у огляді безпеки корпоративних мереж окремо виділяють моніторинг поведінки активів (asset behavioral monitoring) як напрям, що потребує поєднання пасивної телеметрії з централізованою кореляцією подій.

Розвиток ентропійних та статистичних методів виявлення аномалій трафіку висвітлено в роботі Hsiao і Lee [13], а ідентифікацію аномальних станів для систем виявлення вторгнень систематизовано в монографії Корченко [14]. Maosa, Ouazzane і Ghanem [15] запропонували ієрархічну модель кореляції подій безпеки для виявлення загроз у реальному часі, яка обґрунтовує доцільність поєднання одиничних алертів у складені сценарії – саме цей принцип покладено в основу кореляційних правил пропонованої системи.

Інструментальну базу дослідження описано в документації Zeek – зокрема щодо Notice Framework як механізму формування подій [9] і SumStats Framework як засобу статистичної агрегації в реальному часі [16], – а також у документації Wazuh щодо архітектури платформи [11], синтаксису правил [12] та інтеграції з матрицею MITRE ATT&CK [17]. Контекст суцільного шифрування веб-трафіку, який робить інспекцію вмісту дедалі менш

результативною та підсилює роль аналізу метаданих, кількісно охарактеризовано в попередній праці авторів щодо ефективності протоколів TLS 1.2 і TLS 1.3 [18]. Найближчими до предмета цієї статті є нещодавні роботи, в яких baseline будується поверх Zeek-телеметрії. Nguyen зі співавторами [19] розгорнули у виробничому середовищі модуль динамічного поведінкового baseline на основі глибинного прогнозування часових рядів, сформованих із журналів Zeek, з візуалізацією відхилень на аналітичних панелях; Moomtaheen зі співавторами [20] застосували алгоритм Extended Isolation Forest до Zeek conn.log із розміткою подій за тактиками MITRE ATT&CK. Обидва підходи, однак, спираються на ML-моделі з обмеженою пояснюваністю окремого алерту та не охоплюють правилової обробки й кореляції подій у SIEM: інтеграція Zeek і Wazuh у публікаціях розглядається переважно на рівні технічних настанов із пересилання журналів [21], тоді як формалізація baseline-моделі, склад детекторів новизни та кореляційні сценарії для такої зв'язки системно не описані. Заповнення цієї прогалини: від побудови моделі нормальної поведінки мережі до експериментальної перевірки ланцюга виявлення аномалій з правилами та високою пояснюваністю і є предметом цієї статті. Узагальнене позиціювання запропонованої системи відносно наведених підходів подано в табл. 1. Підсумовуючи, на відміну від наявних робіт запропонована система: формально описує нормальну поведінку мережі кортежем $B = (P, S, D, W, A, \Theta)$; реалізує детектори як детерміновані предикати над компонентами B; інтегрує Zeek і Wazuh на рівні власних декодерів і спеціалізованих правил, а не готових пакетів детекції; та поєднує одиничні відхилення кореляційними сценаріями із зіставленням тактикам MITRE ATT&CK.

Таблиця 1. Позиціювання запропонованої системи відносно відомих підходів до поведінкового аналізу Zeek-телеметрії

Підхід	**Модель нормальної поведінки**	**Метод виявлення**	**Пояснюваність алерту**	**Кореляція подій**	**MITRE ATT&CK**
Nguyen et al. [19]	Динамічний baseline (DL-прогноз часових рядів)	Машинне навчання	Низька (відхилення прогнозу)	Немає	Немає
Moomtaheen et al. [20]	Неявна (навчальна вибірка)	Extended Isolation Forest	Низька (Anomaly score)	Немає	Розмітка датасету
Інтеграція Zeek–Wazuh [21]	Немає (готові пакети детекції)	Правила	Висока	Обмежена	Частково
Запропонована система	Формалізований кортеж B (статична + динамічна частини)	Детерміновані предикати + SumStats	Висока (Alert зводиться до предиката над B)	Правила 110200–110202	Так

Основна частина

Архітектура системи та середовище розгортання. Система розгорнута на симуляційному стенді GNS3 із віртуальними машинами VMware, що моделює сегментовану корпоративну

мережу. Внутрішня інфраструктура поділена на чотири VLAN: серверний сегмент VLAN 10 (192.168.10.0/28, вузол 192.168.10.10 із сервісами NGINX, dnsmasq та навчальним веб-застосунком DVWA), користувацький сегмент VLAN 20 (192.168.20.0/24), моніторинговий сегмент VLAN 30 (192.168.30.0/29, у якому розміщено Wazuh Manager і Zeek-сенсор) та адміністративний сегмент VLAN 40 (192.168.40.0/30; префікс /30 обрано свідомо – сегмент розрахований на єдину робочу станцію адміністратора). Зовнішню взаємодію та вузол атаки моделює сегмент 10.255.255.0/24, підключений через NAT. Міжсегментний трафік контролюється списками доступу ACL, які водночас залишають відкритими напрями, потрібні для моделювання розвідки та звернень до внутрішніх сервісів.

Сегментація тут виконує не лише захисну, а й аналітичну функцію: нормальність події залежить від того, до якого сегмента належить її джерело. SSH-підключення з адміністративного VLAN є штатною операцією, тоді як таке саме підключення з користувацького сегмента класифікується як порушення політики доступу. Саме тому адресний план і ролі сегментів входять до складу baseline-моделі, а не залишаються зовнішнім контекстом.

Zeek-сенсор отримує копію трафіку через SPAN-порт комутатора [22] і працює цілком пасивно: жодного впливу на продуктивність чи доступність мережі він не створює. Кастомні детектори формують події через Notice Framework у файл notice.log у форматі JSON. Wazuh Agent, розгорнутий на вузлі сенсора, зчитує цей журнал і передає записи на Wazuh Manager, де кастомні декодери витягують поля події, а правила визначають тип аномалії, рівень критичності та відповідність технікам MITRE ATT&CK. Оброблені алерти через Filebeat індексуються у Wazuh Indexer і відображаються в Wazuh Dashboard (рис. 1).

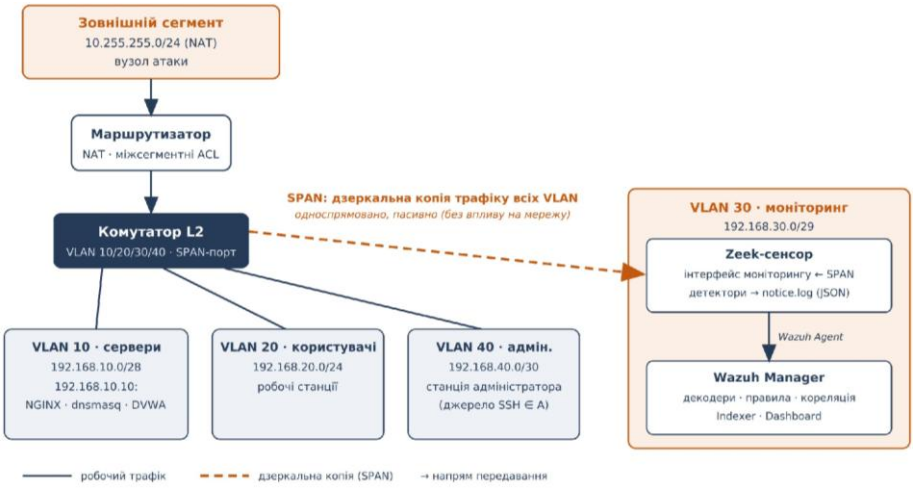

Рис. 1. Узагальнена топологія мережі з підсистемою моніторингу Zeek–Wazuh (підключення сенсора через SPAN)

Модель нормальної поведінки. Baseline формалізовано як сукупність семи логічних груп параметрів, які в математичній моделі нижче згруповано в шість компонент кортежу, кожна з яких має власне джерело даних у журналах Zeek: відомі комунікаційні пари «джерело – призначення» та дозволені сервіси для цих пар (conn.log), перелік DNS-доменів нормальної роботи (dns.log), часові вікна активності хостів, дозволені джерела адміністративного SSH-доступу, статистичні пороги повторюваності подій та допустимі обсяги переданого трафіку. Первинний baseline побудовано обробленням журналів conn.log і dns.log

Python-скриптом за період штатної роботи стенда, упродовж якого генерувалася легітимна активність усіх передбачених типів: звернення користувацького вузла до внутрішнього веб-сервера, DNS-запити через внутрішній сервер dnsmasq і адміністративний SSH-доступ з відповідного сегмента. Оскільки стенд є керованим середовищем, репрезентативність baseline забезпечувалася не календарною тривалістю спостереження, а сценарним покриттям усіх перелічених типів легітимної активності. Критерієм достатності періоду збору в такій постановці слугує стабілізація статичної частини моделі – припинення появи нових елементів множин P, S і D при продовженні спостереження; для розгортання у виробничій мережі це означає, що період побудови baseline має охоплювати щонайменше повний робочий цикл організації з її періодичними процесами, після чого модель фіксується й надалі оновлюється контрольовано. Отже, статична частина baseline змінюється лише в межах таких контрольованих оновлень, тоді як повторна побудова B з нуля виконується один раз – при перенесенні системи в іншу мережу.

Принциповим у моделі є поділ параметрів на дві частини. Статична частина (пари, сервіси, домени, часові вікна, джерела адміністративного доступу) описує допустимий простір взаємодій і дозволяє виявляти аномалії новизни – появу того, чого в мережі раніше не було. Динамічна частина (пороги кількості унікальних портів, повторюваних DNS-запитів, відповідей NXDOMAIN, обсягу передавання) описує допустиму інтенсивність подій і дозволяє виявляти накопичувальні аномалії, для яких підозрілою є не сама поява події, а її інтенсивність – частота повторень або сумарний обсяг. Baseline при цьому не є універсальним: він прив'язаний до конкретної топології, адресації та ролей вузлів, і одна й та сама дія в різних сегментах отримує різну оцінку.

Формально baseline подамо кортежем $B = (P, S, D, W, A, \Theta)$, де $P \subseteq H \times H$ – множина відомих комунікаційних пар на множині хостів H; $S \subseteq P \times \Pi$ – дозволені сервіси (порти з множини Π) для відомих пар; D – множина легітимних DNS-доменів; $W: H \to 2^T$ – відображення, що зіставляє хосту його вікна нормальної активності на часовій осі T; $A \subseteq H$ – множина дозволених джерел адміністративного SSH-доступу; $\Theta = (\theta_ports, \theta_dns, \theta_nx, \theta_bytes)$ – вектор статистичних порогів. Перші п'ять із семи описаних вище груп параметрів відповідають компонентам P, S, D, W, A, а дві останні – пороги повторюваності подій і допустимі обсяги передавання – об'єднані у векторі Θ. Кожен детектор новизни чи політики є предикатом над подією e та статичною частиною B: детектор нової комунікаційної пари істинний за умови $(src(e), dst(e)) \notin P$, детектор нового сервісу – за умови $(src(e), dst(e)) \in P \wedge (src(e), dst(e), port(e)) \notin S$, детектор порушення SSH-політики – за умови $proto(e) = ssh \wedge src(e) \notin A$, детектор часової аномалії – за умови $t(e) \notin W(src(e))$, детектор нового домену – за умови $dom(e) \notin D$. Статистичні детектори визначено над ковзним вікном тривалості τ: подія формується, коли агрегат A_τ (кількість унікальних портів призначення, повторюваних запитів до одного домену, відповідей NXDOMAIN або сумарний обсяг переданих байтів) перевищує відповідну компоненту Θ. Кореляційне правило є кон'юнкцією предикатів над множиною вже сформованих алертів у вікні τ_c: правило 110200 істинне, якщо для одного джерела в межах вікна зафіксовано алерти Network_Beaconing і Large_Transfer; правила 110201 і 110202 побудовані аналогічно над парами «нова пара + сканування» та «позанормовий час + SSH-порушення».

Таке подання дає дві методологічні переваги. По-перше, воно відокремлює

модель нормальної поведінки від механізму виявлення: адаптація системи до іншої мережі зводиться до повторної побудови кортежу B за журналами штатного періоду без зміни предикатів детекторів. По-друге, воно забезпечує високу пояснюваність: для кожного алерту можна вказати, який саме предикат і на якому елементі B порушено, – на відміну від ML-підходів [19, 20], де аномальний скор не розкладається на перевірювані умови. Ціна такої пояснюваності – залежність якості виявлення від повноти та актуальності B, що й визначає головне джерело хибнопозитивних спрацювань, розглянуте далі. Структуру моделі та побудований на ній трирівневий ланцюг виявлення узагальнює рис. 2.

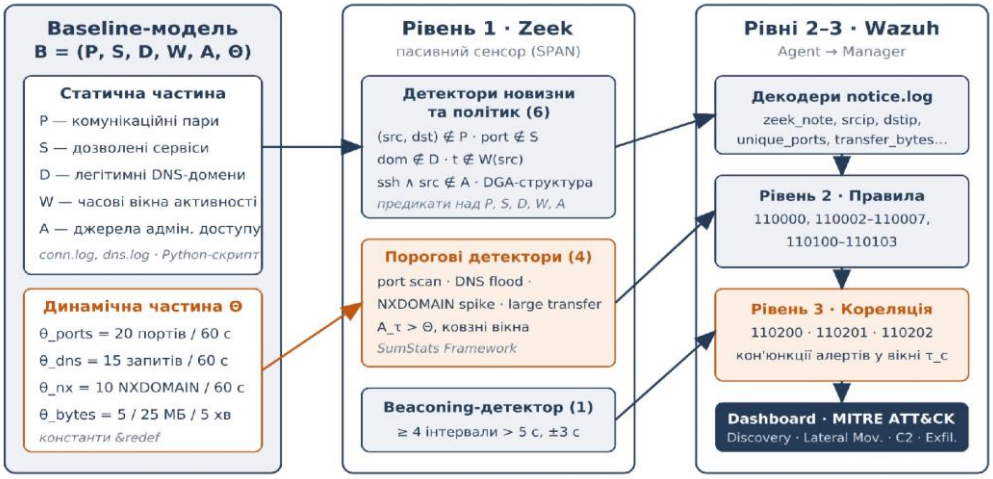

Рис. 2. Формалізована baseline-модель і трирівневий ланцюг виявлення

Детектори, декодери та правила. Перший рівень виявлення реалізують одинадцять кастомних Zeek-детекторів (табл. 2). Детектори новизни та політик порівнюють події conn.log і dns.log зі статичною частиною baseline; три з чотирьох порогових детекторів – сканування портів, повторюваних DNS-запитів і великих передавань – побудовано на SumStats Framework [16], який агрегує події за ковзними часовими вікнами; детектор сплеску NXDOMAIN, пороговий за характером, реалізований підрахунком відповідей безпосередньо в обробнику подій dns.log, що й відображено в табл. 2.

У реалізованій конфігурації компоненти вектора Θ мали такі значення: θ_ports = 20 унікальних портів призначення у вікні 60 с; θ_dns = 15 повторюваних запитів до одного домену у вікні 60 с; θ_nx = 10 відповідей NXDOMAIN у вікні 60 с; θ_bytes = 5 МБ для вихідного передавання з користувацького вузла та 25 МБ для серверного у вікні 5 хв. Евристика DGA-подібності доменного імені є детермінованою: за кожну з чотирьох ознак (довжина понад 25 символів, наявність цифр, суцільна алфавітно-цифрова послідовність від 12 символів, чергування «літера–цифра–літера») нараховується один бал, і подія формується для домену поза множиною D із сумарним балом не менше трьох. Детектор beaconing фіксував щонайменше чотири послідовні міжз'єднувальні інтервали тривалістю понад 5 с зі стабільністю в межах ±3 с. Наведені значення параметрів підібрано емпірично під час налагодження стенда як компроміс між чутливістю детекторів і кількістю хибнопозитивних спрацювань на легітимній активності; вони відповідають умовам стенда і не претендують на

універсальність — їх перегляд під профіль конкретної мережі є штатною частиною калібрування. Усі пороги оголошено константами з можливістю перевизначення (&redef), тож калібрування під іншу мережу не потребує зміни коду детекторів.

Таблиця 2. Кастомні Zeek-детектори системи

Детектор	Джерело даних	Умова виявлення
New_Communication_Pair	conn.log	Пара «джерело – призначення» відсутня в baseline
New_Service_For_Known_Pair	conn.log	Для відомої пари хостів з'явився новий порт/сервіс
SSH_From_Non_Admin_VLAN	conn.log	SSH-підключення не з адміністративного сегмента
Activity_Outside_Normal_Hours	conn.log	Активність хоста поза нормальними часовими вікнами
New_DNS_Domain	dns.log	DNS-запит до домену, відсутнього в baseline
Suspicious_DNS_Query	dns.log	Домен має підозрілу (DGA-подібну) структуру
DNS_NXDOMAIN_Spike	dns.log	Сплеск відповідей NXDOMAIN від одного джерела
Network_Beaconing	conn.log	Повторювані з'єднання зі стабільним інтервалом
SumStats_Port_Scan	SumStats	Перевищення порогу унікальних портів за інтервал
SumStats_DNS_Query_Flood	SumStats	Багаторазові повторювані запити до одного домену
SumStats_Large_Transfer	SumStats	Перевищення порогу обсягу переданих даних

Другий рівень реалізовано у Wazuh. Кастомні декодери розбирають JSON-запис notice-події та витягують поля: zeek_note (тип події, за яким обирається правило), srcip, dstip, dstport, dnsquery, dns_score, dns_count, nxdomain_count, unique_ports, transfer_bytes і zeek_sub. Для кожного типу аномалії спроєктовано спеціалізоване правило: діапазон 110002–110007 обробляє події новизни та політик (нова комунікаційна пара, SSH не з адміністративного сегмента, підозрілий DNS-запит, новий DNS-домен, активність поза нормальними годинами, новий сервіс для відомої пари), правило 110000 – network beaconing, а діапазон 110100–110103 – статистичні події (port scan, NXDOMAIN spike, DNS query flood, large transfer). Рівні критичності диференційовано: події зміни baseline отримують середній рівень, тоді як сканування портів, порушення SSH-політики, beaconing і великі передавання – підвищений, оскільки такі події характерні для активної фази розвідки, спроб несанкціонованого доступу чи виведення даних.

Третій рівень – кореляційні правила, які поєднують кілька одиничних відхилень в один критичніший сценарій: правило 110200 реагує на поєднання network beaconing із великим обсягом передавання (типовий ланцюг «канал C2 + ексфільтрація»), правило 110201 – на нову комунікаційну пару разом зі скануванням портів (мережева розвідка), правило 110202 – на SSH-доступ з неадміністративного сегмента в поєднанні з активністю поза робочими годинами (несанкціонований адміністративний доступ у нетиповий час). Частину правил зіставлено з тактиками MITRE ATT&CK

[17, 23]: port scan і нові комунікаційні напрями віднесено до Discovery, SSH з неадміністративного сегмента в межах прийнятої на стенді сегментної політики інтерпретовано як Lateral Movement (техніка Remote Services: SSH); в іншому контексті, зокрема для з'єднання із зовнішнього вузла, та сама подія може відповідати тактикам Initial Access чи Valid Accounts, DGA-подібні DNS-запити, сплески NXDOMAIN і beaconing – до Command and Control, тоді як поява нового DNS-домену сама по собі трактується як подія новизни, а не як ознака каналу керування, large transfer – до Exfiltration.

Методика експерименту. Верифікацію виконано за чотирнадцятьма сценаріями: одинадцять перевіряли одиничні детектори (нова комунікаційна пара; SSH з неадміністративного сегмента; активність поза нормальними годинами; новий DNS-домен; підозрілий DNS-запит; DNS query flood; NXDOMAIN spike; новий сервіс для відомої пари; сканування портів; передавання великого обсягу; network beaconing), три – кореляційні правила 110200–110202. Для кожного сценарію генерувалася тестова активність (зокрема з вузла атаки в зовнішньому сегменті) і контролювався повний ланцюг виявлення: поява notice-події в Zeek, передавання запису у Wazuh, коректність декодування полів, спрацювання очікуваного правила та відображення алерту в Dashboard (рис. 3).

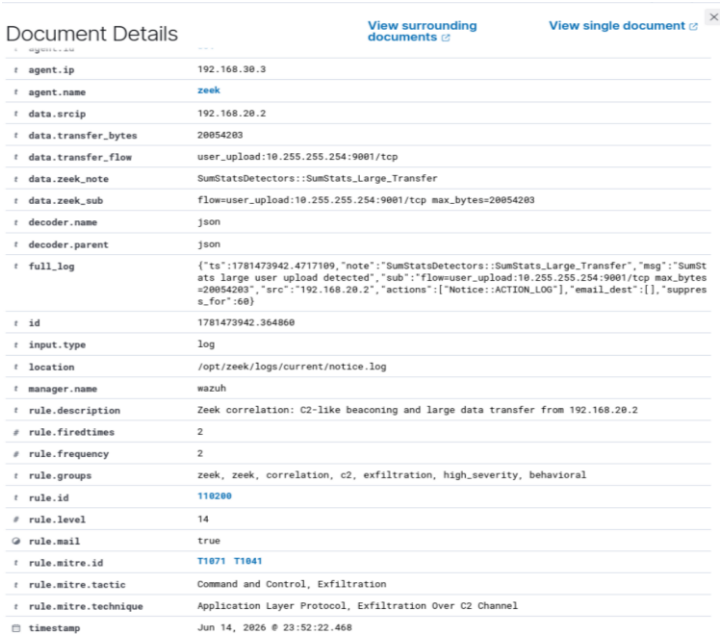

Рис. 3. Спрацювання кореляційного правила 110200 (beaconing + large transfer)

Оцінювання було якісним: для кожного сценарію фіксувався результат спрацювання (true positive) і визначалися умови, за яких те саме правило може дати хибнопозитивний алерт.

Результати

Підсумкову оцінку спрацювань наведено в табл. 3. Для одинадцяти з чотирнадцяти сценаріїв зафіксовано коректне спрацювання без застережень; для трьох детекторів – New_DNS_Domain, Network_Beaconing і New_Service_For_Known_Pair – спрацювання коректні для змодельованих аномалій, але визначено умови можливих false positives у разі легітимних змін нормальної активності. Усі три кореляційні правила сформували очікувані алерти, що підтвердило працездатність ланцюга «тестова

активність → Zeek notice → декодування
→ правило → кореляція → алерт».

Таблиця 3. Оцінка спрацювання алертів за сценаріями

Сценарій	**Правило Wazuh**	**Результат**	**Оцінка**
Нова комунікаційна пара	110002	Алерт сформовано	TP
SSH не з адміністративного сегмента	110003	Алерт сформовано	TP
Активність поза нормальними годинами	110006	Алерт сформовано	TP
Новий DNS-домен	110005	Алерт сформовано; можливі FP для легітимних нових доменів	TP*
Підозрілий DNS-запит	110004	Алерт сформовано	TP
DNS query flood	110102	Алерт після перевищення порогу	TP
DNS NXDOMAIN spike	110101	Алерт сформовано	TP
Новий сервіс для відомої пари	110007	Алерт сформовано; можливі FP при легітимному сервісі	TP*
Сканування портів	110100	Алерт після перевищення порогу унікальних портів	TP
Трафік великого обсягу	110103	Алерт при великому upload	TP
Network beaconing	110000	Алерт сформовано; можливі FP для періодичних сервісів	TP*
SSH + активність поза годинами	110202	Кореляційний алерт сформовано	TP
Нова пара + port scan	110201	Кореляційний алерт сформовано	TP
Beaconing + великий обсяг трафіку	110200	Кореляційний алерт сформовано	TP

* Спрацювання коректне для змодельованої аномалії; водночас визначено умови, за яких те саме правило може дати хибнопозитивний алерт на легітимних змінах активності (аналіз наведено нижче).

За повний цикл виконання одиничних аномалій Wazuh Dashboard чотирнадцяти тестових сценаріїв панель зафіксувала 354 алерти, панель

кореляційних сценаріїв – 37 алертів. Кореляційні алерти формуються додатково до одиничних, тому співвідношення 354:37 не слід трактувати як виміряний коефіцієнт зниження навантаження на аналітика – воно значною мірою визначене складом самих тестових сценаріїв. Ці числа ілюструють лише те, що рівень сценаріїв продукує суттєво компактніший і пріоритетніший потік подій, з якого доцільно починати розбір, зберігаючи можливість перейти від сценарного алерту до базових подій; коректна кількісна оцінка зниження навантаження потребує окремого вимірювання на реальному трафіку. Часова шкала одиничних алертів (рис. 4) має вигляд окремих піків, що відповідає послідовному виконанню сценаріїв; серед вузлів призначення переважає серверний вузол 192.168.10.10, на який були спрямовані сценарії сканування, доступу до веб-сервера та передавання даних. Розподіл кореляційних алертів (рис. 5) за рівнями критичності зміщений у бік підвищених значень, а їх зіставлення з техніками MITRE ATT&CK дозволяє розглядати алерти не як ізольовані технічні події, а як фрагменти можливого сценарію атаки.

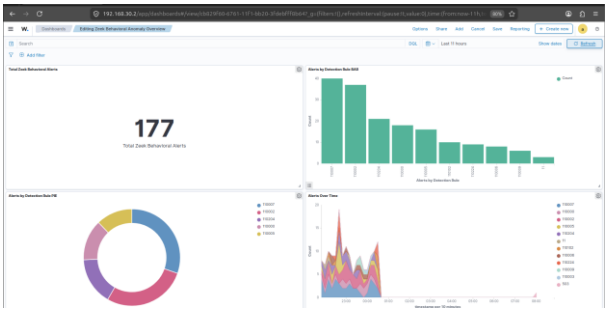

Рис. 4. Wazuh Dashboard одиничних поведінкових аномалій

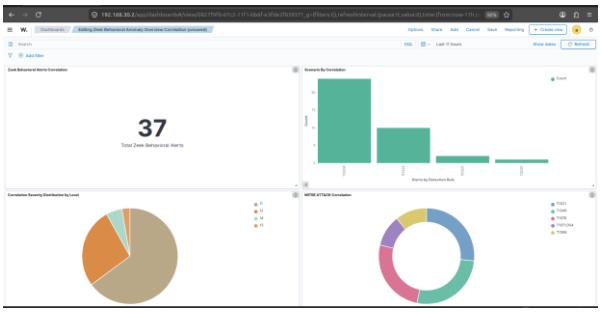

Рис. 5. Wazuh Dashboard кореляційних аномалій

Аналіз false positives і калібрування. Найчутливішими до легітимних змін виявилися детектори, що реагують на новизну відносно baseline. Для New_DNS_Domain хибні спрацювання створюють звернення до легітимних доменів, не зафіксованих під час формування baseline (домени оновлень, хмарних сервісів, CDN); протидія – періодичне оновлення DNS-частини baseline і ведення білого списку підтверджених доменів без автоматичного додавання всіх нових. Для Network_Beaconing джерелом надлишкових спрацювань є регулярна службова активність внутрішніх вузлів (перевірки доступності, фонові з'єднання застосунків), тому тюнінг має враховувати не лише стабільність інтервалу, а й напрям взаємодії, адресу та порт призначення і роль вузла. Для New_Service_For_Known_Pair спрацювання можливі при появі легітимного сервісу після змін конфігурації, що потребує уточнення переліку дозволених сервісів між відомими парами.

Статистичні пороги калібруються під конкретну мережу: занижений поріг унікальних портів перетворює звичайну активність на «сканування», завищений – пропускає повільну розвідку. Для large transfer істотним є напрям передавання: великий download із веб-сервера відповідає нормальній роботі сервісу, тоді як великий upload з користувацького вузла має вищу аналітичну цінність як потенційна ексфільтрація. Кореляційні

правила в цьому контексті виконують і функцію зниження шуму: поєднання двох середньокритичних відхилень в один сценарій зменшує кількість подій, що потребують уваги аналітика, без втрати виявлювальної здатності.

Окремо зазначимо властивість, істотну для мереж із переважно шифрованим трафіком: усі детектори системи працюють на метаданих з'єднань – адресах, портах, часових характеристиках та обсягах – і на DNS-подіях, а не на вмісті пакетів, тому поширення TLS [18] не знижує їхньої чинності, на відміну від сигнатурних підходів, які потребують інспекції корисного навантаження. Обмеженням тут залишається шифрування самого DNS (DoH/DoT), яке виводить DNS-аномалії з поля зору пасивного сенсора та потребує або контролю точок виходу DoH-трафіку, або журналів внутрішнього резолвера.

Висновки та перспективи подальших досліджень

Розроблено та експериментально перевірено систему виявлення поведінкових аномалій трафіку корпоративної мережі на основі інтеграції мережевого аналізатора Zeek і SIEM-платформи Wazuh. Наукова новизна полягає у формалізації baseline-моделі нормальної поведінки сегментованої мережі у вигляді кортежу B = (P, S, D, W, A, Θ) з поділом на статичну частину (комунікаційні пари, сервіси, DNS-домени, часові вікна, джерела адміністративного доступу) та динамічну частину (статистичні пороги), у поданні детекторів як предикатів над подією та компонентами B, а також у побудові на цій моделі трирівневого пояснюваного ланцюга виявлення: одинадцять кастомних Zeek-детекторів, спеціалізовані Wazuh-правила (110000, 110002–110007, 110100–110103) і кореляційні правила 110200–110202 із зіставленням алертів тактикам MITRE ATT&CK. На відміну від ML-орієнтованих підходів до аналізу Zeek-телеметрії, кожен алерт системи зводиться до перевірюваної умови, а адаптація до іншої мережі – до повторної побудови B без зміни предикатів.

Експериментальна перевірка за чотирнадцятьма сценаріями підтвердила коректне спрацювання повного ланцюга виявлення: одинадцять сценаріїв дали true positive без застережень, для трьох детекторів новизни (New_DNS_Domain, Network_Beaconing, New_Service_For_Known_Pair) спрацювання коректні з визначеними умовами можливих false positives. За період експериментів зафіксовано 354 одиничні алерти та 37 кореляційних; кореляційні сценарії підтвердили здатність системи поєднувати розрізнені відхилення в змістовні ланцюги «розвідка», «несанкціонований позаробочий SSH-доступ» і «канал C2 з ексфільтрацією».

Дослідження має обмеження, які визначають межі узагальнення результатів. Перевірку виконано на симуляційному стенді GNS3/VMware зі змодельованими сценаріями, а не на реальному продакшн-трафіку; оцінка спрацювань є якісною (TP та умови FP), кількісні метрики precision/recall на репрезентативному трафіку не обчислювалися, а повторюваність спрацювань за багаторазового виконання однакових сценаріїв окремо не оцінювалася; тривала експлуатація з еволюцією baseline не досліджувалася. Машинне навчання в системі не застосовувалося – виявлення побудоване на детермінованих правилах і статистичній агрегації, що, з іншого боку, забезпечує високу пояснюваність кожного алерту.

Практичне значення результатів визначається переносністю рішення: програмні компоненти системи — Zeek-скрипти детекторів, Wazuh-декодери, правила та кореляційні сценарії — не залежать від конкретної мережі, оскільки вся мережева специфіка зосереджена в кортежі B; тому переносною є саме

програмна реалізація, тоді як модель B за означенням специфічна для мережі: розгортання в іншій інфраструктурі зводиться до збирання журналів штатного періоду та побудови нового baseline. Подальші дослідження логічно спрямувати на розгортання системи на реальному трафіку з кількісною оцінкою частоти хибнопозитивних спрацювань, порівняльне оцінювання на тих самих сценаріях із готовими пакетами детекції Zeek і сигнатурними системами класу Suricata, автоматизацію періодичного оновлення baseline, а також на розширення архітектури: доповнення поведінкового рівня сигнатурним IDS/IPS і міжмережевим екраном у межах багаторівневої архітектури захисту периметра, підхід до побудови якої запропоновано в попередній праці авторів [24], застосування ML-моделей (Isolation Forest, автоенкодери) поверх Zeek-журналів та збагачення алертів даними threat intelligence (MISP) з подальшою SOAR-автоматизацією реагування на критичні кореляційні сценарії.

Література

1. Maseer Z. K., Kadhim Q. K., Al-Bander B., Yusof R., Saif A. Meta-analysis and systematic review for anomaly network intrusion detection systems: detection methods, dataset, validation methodology, and challenges. IET Networks. 2024. Vol. 13, № 5–6. P. 339–376. DOI: https://doi.org/10.1049/ntw2.12128

2. Lyu M., Gharakheili H. H., Sivaraman V. A Survey on Enterprise Network Security: Asset Behavioral Monitoring and Distributed Attack Detection. IEEE Access. 2024. Vol. 12. P. 89363–89383. DOI: https://doi.org/10.1109/ACCESS.2024.3419068

3. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey. ACM Computing Surveys. 2009. Vol. 41, № 3. Article 15. P. 1–58. DOI: https://doi.org/10.1145/1541880.1541882

4. Baselining. BlackFog Cybersecurity 101. URL: https://www.blackfog.com/cybersecurity-101/baselining/ (дата звернення: 06.06.2026).

5. Scarfone K., Mell P. Guide to Intrusion Detection and Prevention Systems (IDPS) : NIST Special Publication 800-94. Gaithersburg : NIST, 2007. DOI: https://doi.org/10.6028/NIST.SP.800-94

6. Dempsey K. et al. Information Security Continuous Monitoring (ISCM) for Federal Information Systems and Organizations : NIST Special Publication 800-137. Gaithersburg : NIST, 2011. DOI: https://doi.org/10.6028/NIST.SP.800-137

7. ISO/IEC 27001:2022. Information security, cybersecurity and privacy protection – Information security management systems – Requirements. International Organization for Standardization, 2022. URL: https://www.iso.org/standard/27001 (дата звернення: 16.06.2026)

8. What is Zeek? Zeek Documentation. URL:https://docs.zeek.org/en/master/about/what.html (дата звернення: 09.06.2026).

9. Notice Framework. Zeek Documentation. URL:https://docs.zeek.org/en/master/frameworks/notice.html (дата звернення: 10.06.2026).

10. weird.log and notice.log. Zeek Documentation. URL: https://docs.zeek.org/en/master/reference/logs/weird-and-notice.html (дата звернення: 10.06.2026).

11. Wazuh Architecture. Wazuh Documentation. URL: https://documentation.wazuh.com/current/getting-started/architecture.html (дата звернення: 10.06.2026).

12. Rules Syntax. Wazuh Documentation. URL: https://documentation.wazuh.com/current/user-manual/ruleset/ruleset-xml-syntax/rules.html (дата звернення: 10.06.2026).

13. Hsiao H.-W., Lee Y.-Z. Entropy-Based Anomaly Detection for Cybersecurity Threats in Network Traffic. Multidisciplinary Social Networks Research. MISNC 2025. Cham : Springer, 2026. CCIS. Vol. 2729. P. 106–117. DOI: https://doi.org/10.1007/978-3-032-09945-7_10

14. Корченко А. О. Методи ідентифікації аномальних станів для систем виявлення вторгнень : монографія. Київ : ЦП «Компринт», 2019. 361 с.

15. Maosa H., Ouazzane K., Ghanem M. C. A Hierarchical Security Event Correlation Model for Real-Time Threat Detection and Response. Network. 2024. Vol. 4, № 1. P. 68–90. DOI: https://doi.org/10.3390/network4010004

16. Summary Statistics. Zeek Documentation. URL: https://docs.zeek.org/en/master/frameworks/sumstats.html (дата звернення: 10.06.2026).

17. MITRE ATT&CK framework. Wazuh Documentation. URL: https://documentation.wazuh.com/current/user-manual/ruleset/mitre.html (дата звернення: 10.06.2026).

18. Дячук О. Ю., Колощук М. С., Рудюк Б. М., Квасніков В. П. Ефективність криптографічних протоколів TLS 1.2 і TLS 1.3 у сучасних веб-додатках: аналітичний та практичний аспекти. Проблеми інформатизації та управління. 2026. № 2(86). С. 52–60. DOI: https://doi.org/10.18372/2073-4751.86.21273

19. Nguyen G., Dlugolinsky S., Tran V., López García Á. Network security AIOps for online stream data monitoring. Neural Computing and Applications. 2024. Vol. 36, № 24. P. 14925–14949. DOI: https://doi.org/10.1007/s00521-024-09863-z

20. Moomtaheen F., Bagui S. S., Bagui S. C., Mink D. Extended Isolation Forest for Intrusion Detection in Zeek Data. Information. 2024. Vol. 15, № 7. Article 404. DOI: https://doi.org/10.3390/info15070404

21. Network security monitoring with Wazuh and Zeek. Wazuh Blog. URL: https://wazuh.com/blog/network-security-monitoring-with-wazuh-and-zeek/ (дата звернення: 16.06.2026).

22. SPAN, RSPAN, ERSPAN. Cisco Learning Network. URL: https://learningnetwork.cisco.com/s/article/span-rspan-erspan (дата звернення: 10.06.2026).

23. MITRE ATT&CK. The MITRE Corporation. URL: https://attack.mitre.org (дата звернення: 16.06.2026).

24. Кухарчук А. А., Дячук О. Ю., Колощук М. С. Дослідження та реалізація системи комплексного захисту периметру корпоративної мережі на основі багаторівневих підходів до кібербезпеки. Інтегровані інтелектуальні робототехнічні комплекси (ІІРТК-2026) : тези міжнар. наук.-практ. конф. (м. Київ, 19–20 травня 2026 р.). Київ : Київський авіаційний інститут, 2026. С. 509–510.

Дячук О. Ю., Кухарчук А. А. Колощук М. С., Шелуха О. О., Квасніков В. П. BASELINE-ОРІЄНТОВАНА СИСТЕМА ВИЯВЛЕННЯ АНОМАЛІЙ МЕРЕЖЕВОГО ТРАФІКУ З КОРЕЛЯЦІЄЮ ПОДІЙ У SIEM WAZUH

У роботі розроблено та експериментально перевірено систему виявлення поведінкових аномалій трафіку корпоративної мережі на основі інтеграції мережевого аналізатора Zeek і SIEM-платформи Wazuh. Запропоновано формалізовану baseline-модель нормальної поведінки сегментованої мережі у вигляді кортежу множин і порогів, що поєднує статичну частину (комунікаційні пари, дозволені сервіси, DNS-домени, часові вікна активності, джерела адміністративного доступу) та динамічну частину (статистичні пороги повторюваності подій і обсягів передавання); детектори подано як предикати над подією та компонентами моделі, що забезпечує високу пояснюваність кожного алерту.

На основі моделі реалізовано трирівневий ланцюг виявлення: одинадцять кастомних Zeek-детекторів (аномалії новизни, порушення політик доступу, статистичні відхилення на базі SumStats Framework), кастомні Wazuh-декодери та спеціалізовані правила (110000, 110002–110007, 110100–110103), а також кореляційні правила 110200–110202, які поєднують окремі відхилення в сценарії мережевої розвідки, несанкціонованого SSH-доступу в нетиповий час та каналу керування з ексфільтрацією. Алерти зіставлено з тактиками MITRE ATT&CK (Discovery, Lateral Movement, Command and Control, Exfiltration).

Експериментальна верифікація на симуляційному стенді GNS3/VMware за чотирнадцятьма сценаріями підтвердила коректне спрацювання повного ланцюга виявлення: одинадцять сценаріїв класифіковано як true positive без застережень, для трьох детекторів новизни визначено умови можливих хибнопозитивних спрацювань і напрями калібрування (оновлення baseline, білі списки доменів, урахування напряму та ролі вузла, налаштування порогів). Зафіксовано 354 одиничні та 37 кореляційних алертів;

кореляційний рівень формує компактний пріоритетний потік подій для першочергового розбору. Окреслено обмеження дослідження та напрями розвитку системи на реальному трафіку.

Ключові слова: *поведінковий аналіз; аномалії мережевого трафіку; baseline; Zeek; Wazuh; кореляція подій; сегментація мережі; MITRE ATT&CK.*

Diachuk O. Yu., Kukharchuk A. A. Koloshchuk M. S., Shelukha O. O., Kvasnikov V. P.

A BASELINE-ORIENTED NETWORK TRAFFIC ANOMALY DETECTION SYSTEM WITH EVENT CORRELATION IN THE WAZUH SIEM

The paper presents the design and experimental verification of a behavioral anomaly detection system for corporate network traffic based on the integration of the Zeek network analyzer and the Wazuh SIEM platform. A formalized baseline model of normal behavior for a segmented network is proposed as a tuple of sets and thresholds, combining a static part (known communication pairs, allowed services, DNS domains, activity time windows, administrative access sources) and a dynamic part (statistical thresholds for event repetition and transfer volumes); detectors are defined as predicates over an event and the model components, which makes every alert fully explainable.

A three-level detection chain is implemented on top of the model: eleven custom Zeek detectors (novelty anomalies, access policy violations, and statistical deviations based on the SumStats Framework), custom Wazuh decoders with specialized rules (110000, 110002–110007, 110100–110103), and correlation rules 110200–110202 that combine individual deviations into scenarios of network reconnaissance, unauthorized off-hours SSH access, and a command-and-control channel with exfiltration. Alerts are mapped to MITRE ATT&CK tactics (Discovery, Lateral Movement, Command and Control, Exfiltration).

Experimental verification on a GNS3/VMware simulation testbed across fourteen scenarios confirmed the correct operation of the full detection chain: eleven scenarios were classified as true positives without reservations, while for three novelty detectors the conditions of possible false positives and tuning directions were identified (baseline updates, domain whitelists, consideration of traffic direction and host role, threshold calibration). The dashboards registered 354 single and 37 correlation alerts; the correlation layer provides a compact, prioritized event stream for first-line triage. The study limitations and directions for deploying the system on real traffic are outlined.

Keywords: *behavioral analysis; network traffic anomalies; baseline; Zeek; Wazuh; event correlation; network segmentation; MITRE ATT&CK.*

Стаття подана до редакції: 13/07/2026
Стаття прийнята до опублікування:15/07/2026
Стаття опублікована: 10/11/2026

Стаття поширюється на умовах ліцензії CC BY 4.0