Віра ГУСЬКОВА, доцент кафедри штучного інтелекту, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», доктор філософії з комп'ютерних наук.

Владислав ШКОЛЬНІКОВ, завідувач кафедри кримінології та інформаційних технологій, Національна академія внутрішніх справ, доктор філософії в галузі права, доцент.

Богдан ЛИСОВ, аспірант 2-го року навчання, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Артем ХАЛИГОВ, аспірант 3-го року навчання, ІТГІП НАН України

ГІБРИДНА АРХІТЕКТУРА ІНТЕЛЕКТУАЛЬНОЇ ФІЛЬТРАЦІЇ КІБЕРПОДІЙ НА ОСНОВІ RULE-BASED GATE ТА МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ

Сучасні системи інформаційної безпеки функціонують в умовах постійного зростання обсягів цифрових даних, мережевого трафіку, системних журналів, подій автентифікації, повідомлень від засобів моніторингу та спрацювань систем виявлення вторгнень. У таких умовах однією з ключових проблем є не лише виявлення потенційних кіберзагроз, а й оперативна фільтрація великої кількості подій, визначення їхньої критичності та формування пріоритетів для подальшого аналізу. Особливої актуальності ця проблема набуває для об'єктів критичної інфраструктури, державних інформаційних систем, правоохоронних органів та організацій, діяльність яких залежить від безперервності й надійності інформаційних процесів.

Традиційні підходи до аналізу кіберподій часто базуються на правилах, сигнатурах або фіксованих порогових значеннях. Такі методи є швидкими, зрозумілими та відносно простими для впровадження, однак вони мають обмежену здатність виявляти складні, нові або контекстно залежні загрози. З іншого боку, методи машинного навчання та великі мовні моделі відкривають можливості для глибшого аналізу подій, виявлення аномалій, інтерпретації складних взаємозв'язків і формування пояснень. Водночас передача всього потоку кіберподій до складних моделей штучного інтелекту може бути нераціональною через високі обчислювальні витрати, затримки обробки, ризик надмірної кількості хибних спрацювань та складність масштабування [1].

У зв'язку з цим доцільним є використання гібридного підходу, який поєднує швидкість і прозорість rule-based механізмів із аналітичними можливостями моделей штучного інтелекту. Метою роботи є обґрунтування гібридної архітектури інтелектуальної фільтрації кіберподій, у якій первинне сортування подій виконується за допомогою rule-based gate, а складні, неоднозначні або потенційно критичні випадки передаються на подальший аналіз до ML- або LLM-модуля [2].

На рис.1 представлено гібридну архітектуру інтелектуальної фільтрації кіберподій, у якій первинне сортування та маршрутизація подій здійснюється за допомогою rule-based gate, а складні, неоднозначні або потенційно критичні випадки передаються на подальший аналіз до ML/LLM-модуля.

Архітектура включає джерела кіберподій, модуль попередньої обробки, блок оцінювання ризику та модуль підтримки прийняття рішень, що забезпечує формування відповідних дій реагування.

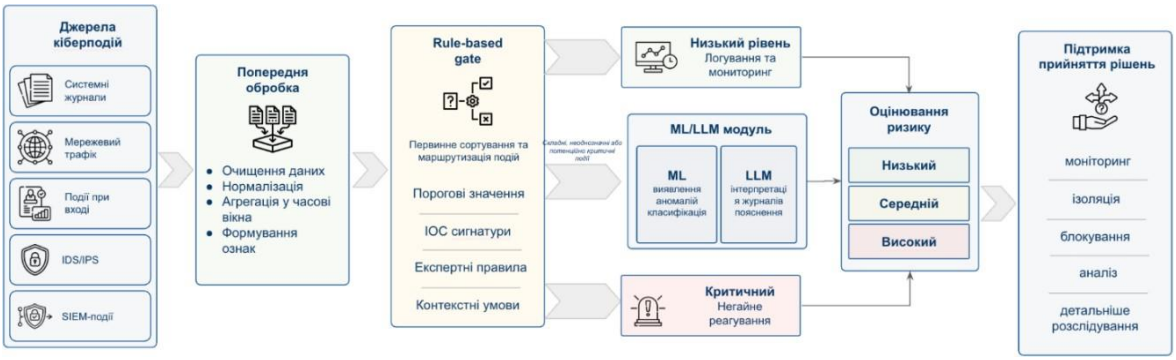

Рисунок 1 – Гібридна архітектура інтелектуальної фільтрації кіберподій на основі rule-based gate та ML/LLM-модуля

Під rule-based gate у межах цієї роботи пропонується розуміти проміжний керуючий шар, який виконує первинну перевірку вхідних кіберподій на основі формалізованих правил, порогових значень, індикаторів компрометації, контекстних обмежень та експертних умов. Такий модуль не замінює штучний інтелект, а виконує функцію попередньої фільтрації, маршрутизації та пріоритизації подій. Його основне завдання полягає у визначенні того, які події можуть бути автоматично віднесені до низькоризикових, які потребують негайної реакції за формальними ознаками, а які мають бути передані до складнішого інтелектуального аналізу [3].

Запропонована архітектура може бути подана у вигляді послідовності таких функціональних компонентів: джерела даних, модуль попередньої обробки, rule-based gate, модуль штучного інтелекту, модуль оцінювання ризику та модуль підтримки прийняття рішень. Джерелами даних можуть бути системні журнали, мережевий трафік, події автентифікації, журнали доступу, повідомлення IDS/IPS, події SIEM-систем, дані про активність користувачів та інші цифрові сліди. На етапі попередньої обробки здійснюється очищення даних, нормалізація, видалення дублікатів, агрегація подій у часові вікна та формування ознак, придатних для подальшого аналізу.

Rule-based gate виконує первинну логічну перевірку подій і забезпечує їх попередню маршрутизацію до відповідного рівня обробки. До прикладу, правила можуть враховувати кількість невдалих спроб входу за певний проміжок часу, доступ до критичних ресурсів у нетиповий час, різке зростання кількості операцій, звернення до системних директорій, появу нетипових процесів, зміну прав доступу, підозрілу географію входу або збіг із відомими індикаторами компрометації [4].

Формалізація таких умов у вигляді правил дозволяє швидко визначати, чи може подія бути оброблена на рівні простого логічного контролю, чи вона потребує додаткового інтелектуального аналізу. Наприклад, якщо кількість невдалих спроб входу перевищує встановлений поріг у межах короткого часового вікна, подія може бути позначена як підозріла та передана до ML-модуля. Інший приклад — одночасне виникнення кількох поведінкових ознак ризику, зокрема використання нового пристрою, нетиповий час доступу та підвищена частота операцій. У такому випадку подія також потребує додаткового оцінювання ризику засобами машинного навчання.

Приклад формалізації правил наведено нижче:

Правило 1: *IF failed_login_attempts > threshold* *AND time_window <= 5 minutes* *THEN mark event as "suspicious"* *AND send to ML module* *ELSE continue standard monitoring*	Правило 2: *IF new_device = true* *AND access_time_is_unusual = true* *AND operation_frequency > normal_frequency* *THEN mark event as "suspicious"* *AND send to ML module for risk assessment* *ELSE continue monitoring*

На основі таких правил події можуть бути розподілені на кілька категорій: очевидно низькоризикові, формально критичні, підозрілі та неоднозначні. Очевидно низькоризикові події можуть залишатися на рівні логування та моніторингу, формально критичні — передаватися на негайне реагування, а підозрілі й неоднозначні — спрямовуватися до ML- або LLM-модуля для глибшого аналізу. Таким чином, rule-based gate виконує роль проміжного фільтра, який зменшує навантаження на ресурсомісткі модулі штучного інтелекту та забезпечує більш раціональну обробку потоку кіберподій [5].

Порівняльну характеристику традиційного rule-based підходу, підходу на основі ML/LLM та запропонованого гібридного рішення наведено в табл. 1.

Таблиця 1 – Порівняльна характеристика підходів до фільтрації кіберподій

Критерій	**Rule-based підхід**	**ML/LLM-підхід**	**Гібридний підхід**
Швидкість обробки	Висока	Середня або низька	Висока для простих подій, поглиблена для складних
Прозорість рішень	Висока	Обмежена	Підвищена за рахунок правил і пояснень
Виявлення нових загроз	Обмежене	Вище	Вище, ніж у rule-based
Обчислювальні витрати	Низькі	Високі	Оптимізовані
Масштабованість	Середня	Залежить від моделі	Висока
Доцільність застосування	Типові та формалізовані події	Складні, неоднозначні випадки	Потоки подій різної складності

Модуль штучного інтелекту у запропонованій архітектурі призначений для поглибленого аналізу тих подій, які не можуть бути однозначно класифіковані на рівні rule-based gate. Залежно від характеру даних і поставленої задачі він може включати різні типи моделей. Для аналізу послідовностей кіберподій доцільним є використання моделей виявлення аномалій, зокрема LSTM Autoencoder, Isolation Forest, кластеризаційних підходів або інших методів машинного навчання. Для класифікації подій можуть застосовуватися Random Forest, Gradient Boosting, XGBoost або нейронні мережі [6].

Окрему роль у такій архітектурі може виконувати LLM-модуль, який доцільно використовувати не для обробки всього потоку подій, а для випадків, що потребують контекстного або семантичного аналізу. Зокрема, він може застосовуватися для інтерпретації текстових журналів, опису інцидентів, узагальнення пов’язаних подій, формування пояснень і підготовки рекомендацій для аналітика. Такий підхід дозволяє уникнути надмірного використання ресурсомістких моделей і водночас забезпечити глибший аналіз складних або неоднозначних кіберподій. Узагальнене призначення окремих інтелектуальних компонентів у межах запропонованої архітектури наведено в табл. 2.

Окремим компонентом архітектури є модуль оцінювання ризику. Його завдання полягає у перетворенні результатів rule-based gate та ML/LLM-аналізу в узагальнений показник ризику. Такий показник може враховувати критичність активу, тип події, частоту повторення, рівень аномальності, наявність відомих індикаторів компрометації, історичний контекст та потенційний вплив на інформаційну систему.

Таблиця 2 – Логіка маршрутизації кіберподій у запропонованій гібридній архітектурі

Тип події	**Ознаки**	**Рішення rule-based gate**	**Подальша дія**
Низькоризикова	Типова активність, відсутність перевищення порогів	Не передавати до ML/LLM	Логування та моніторинг
Формально критична	Збіг з IOC, критичний актив, грубе порушення правила	Позначити як критичну	Негайне реагування
Підозріла	Часткове перевищення порогів, нетипова активність	Передати до ML-модуля	Виявлення аномалій / класифікація
Неоднозначна	Недостатньо контексту, текстові логи, складний опис	Передати до LLM-модуля	Інтерпретація та пояснення
Високоризикова	Комбінація правил і високого ML-score	Підвищити пріоритет	Передати аналітику / ініціювати реагування

У найпростішому вигляді інтегральний показник ризику може бути поданий як зважена сума окремих складових:

$$\text{Risk} = w_1\text{A} + w_2\text{C} + w_3\text{F} + w_4\text{I} + w_5\text{H} ,$$

де: A – рівень аномальності події,
C – критичність активу,
F – частота повторення події,
I – наявність або сила збігу з індикаторами компрометації,
H – історичний контекст поведінки користувача або системи,
w_1–w_5 – вагові коефіцієнти, що визначають внесок кожної ознаки в рівень ризику.

На практиці результат може подаватися у вигляді категорій: низький, середній або високий ризик. Це дозволяє не лише фіксувати факт виявлення підозрілої активності, а й визначати пріоритет реагування.

Завершальним компонентом архітектури є модуль підтримки прийняття рішень, який формує рекомендації відповідно до рівня ризику та типу події. Для низькоризикових подій може бути рекомендовано моніторинг, для середньоризикових – передача аналітику або посилення спостереження, для високоризикових – ініціювання реагування, ізоляція вузла чи блокування облікового запису. Таким чином, запропонована архітектура орієнтована не лише на виявлення кіберподій, а й на підтримку практичних рішень у сфері інформаційної безпеки.

Перевагою підходу є раціональний розподіл навантаження між простими правилами та складними моделями ШІ. Rule-based gate забезпечує швидкість, прозорість і контрольованість первинної обробки, тоді як ML- та LLM-модулі застосовуються для складних або неоднозначних випадків. Водночас ефективність такого підходу залежить від якості й актуальності правил, коректного налаштування порогових значень, контролю вхідних даних і перевірки результатів ML/LLM-аналізу. Тому архітектура має передбачати регулярне оновлення правил, моніторинг якості моделей і участь фахівця з кібербезпеки у перевірці критичних рішень.

Поєднання rule-based механізмів і моделей ШІ дозволяє підвищити стійкість системи до різних типів загроз: від типових спроб несанкціонованого доступу до складних багатоетапних атак. Запропонований підхід забезпечує баланс між швидкістю обробки, якістю аналізу, пояснюваністю результатів і вартістю обчислювальних ресурсів. Його практичне значення полягає у можливості застосування в SOC, SIEM-системах, корпоративному моніторингу та системах захисту об’єктів критичної інфраструктури.

Список використаних джерел

1. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge: MIT Press, 2016. 800 p.
2. Bishop C. M. Pattern Recognition and Machine Learning. New York: Springer, 2006. 738 p.
3. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey. ACM Computing Surveys. 2009. Vol. 41, No. 3. P. 1–58.
4. Buczak A. L., Guven E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys & Tutorials. 2016. Vol. 18, No. 2. P. 1153–1176.
5. Sommer R., Paxson V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. 2010 IEEE Symposium on Security and Privacy. Oakland, 2010. P. 305–316.
6. Lundberg S. M., Lee S.-I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 4765–4774.

Юлія АНТОНОВА, студентка 2-го курсу, спеціальність “Право”, Національна академія Служби безпеки України.

Науковий керівник:
Максим ПАЛЬЧИК, Національна академія Служби безпеки України, кандидат юридичних наук, доцент

МІЖНАРОДНІ ПІДХОДИ ДО ПРАВОВОГО РЕГУЛЮВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ: ЄС, США ТА УКРАЇНИ

Розвиток технологій штучного інтелекту (в подальшому ШІ) на сучасному етапі є одним із ключових факторів трансформації суспільних та правових відносин, який дав поштовх для нових викликів правової системи, особливо у сфері відповідальності, захисту персональних даних, охоплюючи такі важливі сфери як публічне управління, правоохоронну діяльність та судочинство. Водночас розвиток цих технологій впроваджує адаптацію правових норм, створюючи виклики ефективного регулювання [5].

За міжнародними дослідженнями (зокрема документами Європейської комісії та Ради Європи) проаналізовано, що недостатність правової бази призвело до юридичної невизначеності та ризиків порушення прав людини, саме у правозастосування ШІ, тобто міжнародне законодавство потребує адаптації міжнародних стандартів [14] .

Практика Європейського ж суду з прав людини теж свідчить про забезпечення процесуальних гарантій при використанні систем ШІ, зокрема у справах суд неодноразово наголошував на ризиках надмірного використання технологічних засобів та захисту персональних даних. Їх суттю виступають впровадження правових стандартів щодо допустимих меж використання та обробки даних [14] .

В Україні ж як законодавство, так і судова практика не підлаштовані під конкретне регулювання ШІ, воно перебуває на стадії формування [5; 6] . Не існує жодного нормального акта, який визначає правовий статус ШІ чи межі його використання та відповідальності, адже згідно аналізу судових рішень – існує необхідність законодавчих змін для автоматизованих систем у судочинстві, публічному управлінні та захисті персональних даних, яке тягне за собою ускладнення судового контролю.