DOI: 10.35681/1560-9189.2025.27.2.345591

УДК 004.942

Л. Ю. Гальчинський, В. В. Личик
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Проспект Берестейський, 37, 03056 Київ, Україна
e-mail: hleonid@gmail.com, lychyk.vlad@gmail.com

Використання моделі мереж Баєса для раннього оцінювання загроз кібератак об'єкта електроенергетики

Представлено результати дослідження розробки інструмента для полегшення виявлення кібератак на ІТ-мережу енергооб'єкта. Встановлено, що модель баєсових мереж має значні перспективи для раннього оцінювання загроз кібератак об'єкта електроенергетики. Як методологію використано модель кібератаки у вигляді орієнтованого ациклічного графа, а для оцінювання рівня загроз запропоновано баєсову мережу. Для апріорного оцінювання ймовірностей використано систему метрик CVSS. Проведено оцінку чутливості моделі до змін апріорних параметрів умовних розподілів.

Ключові слова: *ІТ-мережа енергооб'єкта, граф атаки, баєсова мережа, чутливість, кіберстійкість.*

Вступ

Сучасні енергетичні мережі є критично важливою інфраструктурою, і будь-які загрози для них можуть суттєво вплинути на національну безпеку та економічну стабільність. Доконаним фактом є те, що цифрові технології мережевого зв'язку в енергетичних системах суттєво покращують функції моніторингу та контролю. Проте, з іншого боку значно зросла ймовірність кіберзагроз, які можуть привести до значних збоїв і можуть мати катастрофічні наслідки.

Характерною особливістю кібератак на енергооб'єкти є їхня багатоетапність і значна невизначеність у намірах зловмисника та в можливостях систем захисту протягом атаки. При цьому, чим далі зловмисникам вдалося проникнути всередину мережі енергооб'єкта, тим більший рівень загроз. А відтак це питання потребує пошуку інструментів для зниження рівня невизначеності для раннього виявлення кіберзагрози.

Аналіз досліджень і публікацій

Приклади кібератак на енергосистеми та їхні наслідки добре відомі. Це зокрема, кібератаки BlackEnergy на енергосистему України у 2015 та 2016 роках, які спричинили масове відключення електроенергії в Україні, що торкнулося сотень тисяч людей. У квітні 2022 року угрупування Sandworm, яке міжнародні дослідники класифікують як підрозділ ГРУ РФ, спробувало відключити електроенергію в Україні, використовуючи шкідливе програмне забезпечення Industroyer2 — як удосконалену версію застосованого групою ПЗ для відключення електроенергії в Україні в 2016 році. Все це мотивує на пошуки ефективних засобів захисту та протидії такого роду загрозам, а це, у свою чергу, потребує ретельного вивчення особливостей механізму проходження кібератак у середовищі інфраструктури енергосистеми. Нове покоління енергетичних систем передбачає глибоку інтеграцію власне енергомереж із сучасними технологіями мережевого зв'язку і управління, та вважається невід'ємною частиною енергетичної інфраструктури [1, 2].

Основним джерелом вразливостей об'єктів критичної інфраструктури є стандартизовані протоколи зв'язку та програмні застосунки, які використовують механізми обміну між хостами IT-мереж. Об'єкти електроенергетики, як один із найбільш важливих представників критичної інфраструктури, включають у себе це джерело, що дозволяє зловмисникам отримувати несанкціонований доступ і контроль над активами, що, в свою чергу, суттєво впливає на енергосистеми у випадку успішної атаки [3]. Ці нові енергосистеми включають як традиційну енергомережу інфраструктури, так і комунікаційні мережі для моніторингу і управління. Тому зловмисники можуть використовувати вразливості в центрах управління, або ж підстанціях, щоб отримати права адміністратора через інтерфейси людина-машина. Далі, отримавши такий доступ, вони можуть маніпулювати командами керування обладнанням. А за допомогою протоколів Ethernet може бути захоплений контроль, наприклад над системою автоматичних вимикачів у фізичній енергосистемі. У цьому випадку може бути реалізований катастрофічний сценарій дисбалансу між виробництвом електроенергії і навантаженням, що спричиняє нестабільність мережі.

Усі електроенергетичні підприємства мають місію постачання електроенергії користувачам через національний енергоринок. Для постачання потрібно ефективно виконувати три головні функції: генерацію, розподіл і передачу електроенергії. Усі вони важливі, однак практика останніх десятиліть показує, що найбільший показник кіберзагроз припадає на функцію розподілу.

Використання баєсових мереж для кібербезпеки почалося відносно недавно. В першу чергу це стосується робіт Фріго, в яких вперше були отримані результати застосування моделей баєсових мереж для оцінки безпеки комп'ютерних мереж [4]. У цих роботах показано, що інтеграція сучасних мережевих комунікаційних технологій в енергетичні системи покращує функції моніторингу та керування. Однак, таке впровадження збільшує ймовірність кіберзагроз, підвищуючи ризик катастрофічних збоїв. Якщо вразливості системи будуть атаковані, це загрожуватиме безпечній і стабільній роботі енергетичних систем і спричинить серйозні наслідки [5]. З урахуванням технологічних трендів енергосистеми нового покоління, які переплітають традиційні енергосистеми із су-

часними мережевими комунікаційними технологіями, розглядаються як високо взаємопов'язана та взаємозалежна інфраструктура [6]. При цьому інші автори зазначають, що будь-який збій компонента може суттєво вплинути на всю енергосистему. При цьому значний вклад у розуміння природи невизначеностей при атаках на ІТ-мережі критичної інфраструктури викладено у роботі [7]. Однак, питання про раннє виявлення кібератаки на ІТ-мережу енергооб'єкта все ще досліджено недостатньо.

Формулювання цілей статті

Метою роботи є розробка математичної моделі раннього виявлення кіберзагроз на енергооб'єкт на основі Баєсової мережі, та встановлення її придатності для практичного використання. Така модель може бути використана як інструмент для зменшення рівня невизначеності при кібератаці на енергооб'єкт і прийняття подальших рішень щодо релевантних засобів для ліквідації, або пом'якшення рівня загроз для енергооб'єкта.

Виклад основного матеріалу

Сучасні енергомережі є зразком кіберфізичних систем (КФС), в яких складно об'єднані власне об'єкти електроенергетичної інфраструктури та цифрові комунікаційні технології. При цьому ієрархічна структура енергетичної КФС ретельно окреслена: від центрів керування до фізичних пристроїв.

На найвищому рівні розташовані центри управління, що охоплюють головні розподільчі станції, сервери SCADA та сервери інформаційної системи управління.

Під цим знаходиться рівень мережі енергоспоживання, що складається з різних комунікаційних пристроїв і протоколів, таких як: маршрутизатори, інтелектуальні лічильники та виконавчі механізми. Цей рівень сприяє передачі даних, включаючи вимірювальну інформацію та команди керування, між центрами управління та розподіленими підстанціями, забезпечуючи безперебійний зв'язок у системі.

Архітектура складається з компонентів інформаційних технологій (ІТ) та операційних технологій (ОТ). Проте неможливо однозначно розділити енергетичну архітектуру на домени ІТ та ОТ, оскільки в реальному світі ІТ-пристрої повинні взаємодіяти з компонентами ОТ. Так само, аналіз і впровадження кібербезпеки повинні враховувати як аспекти ІТ, так і аспекти ОТ у комплексній оцінці. Тому аналіз процесів кібератак має розглядати корпоративну область як точку входу, а на наступних етапах атаки проходять через різні мережі і можуть досягати середовищ керування та процесів.

Еталонною моделлю, яка зазвичай використовується для представлення основних рівнів ІКС, є модель Пердью (PERA). Компоненти кожного рівня взаємодіють за допомогою локальних мереж, а зв'язок між рівнями використовує, можливо, гетерогенні технології, що реалізують як локальні, так і глобальні мережі.

На найвищому рівні розташовані центри управління, що охоплюють головні розподільчі станції, сервери SCADA та сервери інформаційної системи управління, яка відома як ERP-система. Ці центри утворюють керуючу мережу енергетичної КФС, яка відповідає за обробку і управління даними по всій мережі, моніторинг робочого стану та забезпечення стабільності та безпеки електропостачання.

Наступний рівень — це мережі енергоспоживання, з різних комунікаційних пристроїв і протоколів. Основна функція цього рівня полягає в передачі даних, зокрема вимірювальних, і передачі команд керування між центрами управління та розподіленими підстанціями.

Далі розташований рівень розподіленої підстанції, оснащений вторинними пристроями, такими як віддалені термінальні блоки (RTU) та фідерні термінальні блоки (FTU). Ці блоки діють як інтерфейс між фізичним та інформаційним рівнями, збираючи дані в режимі реального часу з мережі та виконання керуючих команд з інформаційного рівня, відіграючи вирішальну роль у регулюванні роботи енергосистеми.

Найнижчий рівень складається з фізичних об'єктів, включаючи первинне електричне обладнання, таке як генератори, трансформатори, лінії електропередачі та навантаження. Ці компоненти електрично з'єднані між собою в межах фізичного рівня, утворюючи основу енергосистеми, яка безпосередньо відповідає за виробництво, передачу та розподіл електричної енергії.

Метою всіх кібератак на енергосистему, зокрема на підсистему розподілу електроенергії, є захоплення контролю для подальшого руйнівного впливу на баланс споживання в електромережі. Мережі енергосистеми (МЕС) підтримують обмін даними між глобальними мережами Інтернет і користувачами у певних регіональних межах до 10 км, зі швидкістю передачі даних від 100 кбіт/с до 10 Мбіт/с.

МЕС сприяють двонаправленій передачі даних між пристроями користувачів і концентраторами даних або підстанціями, забезпечуючи такі послуги як дистанційний облік, контроль використання та моніторинг польового обладнання через мережу обліку. Типовими цілями атак на МЕС є електричні підстанції і розподільчі центри.

Характерною особливістю кібератак на енергосистему є багатоетапність і відносно довгий час їхньої тривалості. При цьому традиційні для кібербезпеки оцінки ризику є необхідними, але недостатніми, бо глобальною метою таких атак є руйнування об'єкта електроенергетики, тобто мова йде про виживання системи. В цьому контексті на передній план виходить оцінка рівня кіберстійкості [8] та пошук засобів протидії зловмисному програмному забезпеченню. У свою чергу це потребує виявлення і оцінки виду та рівня кіберзагроз у реальному часі, та розробки порядку реагування на ці кіберінциденти [9]. Аналіз безпеки в режимі реального часу є набагато складнішим і невизначеним процесом, ніж детерміністичні міркування, або класичне статистичне оцінювання. Захисник наперед не знає поведінку зловмисника, зокрема вибір напряму розвитку атаки. Відтак має місце невизначеність, яка пов'язана з невідомою поведінкою. Кібератаки не завжди гарантовано будуть успішними, тому існує невизначеність, що пов'язана з недосконалою природою зловмисного програмного забезпечення. Крім того, можливості спостереження системи захисту щодо актуальної активності атаки обмежені, і в результаті ми маємо невизначеність, що пов'язана з хибнопозитивними та хибнонегативними результатами датчиків системи виявлення вторгнень (IDS).

Проте, якщо уявити кібератаку на енергосистему як послідовність кроків, кожен з яких націлений на певний актив, то найбільш релевантною моделлю буде ациклічний спрямований граф, в якому закодована логічна причинність зловмисника. Такий підхід має перспективу для створення практичних інструментів захис-

ту. Однак модель атаки у вигляді спрямованого графа за всієї наглядності недостатня — як модель для оцінки загроз на систему розподілу електроенергії, оскільки є якісним засобом. Далі розглянемо можливості кількісної оцінки загроз з урахуванням топології комунікацій у середовищі ІТ-системи розподілу, наявності вразливостей у хостах мережі та ймовірнісної природи логіки кроків кібератаки.

Граф атак на основі баєсової мережі

Використання баєсових мереж з метою отримання оцінок і прийняття рішень для складних систем в умовах невизначеності набуло широкої популярності в різних сферах [10]. Застосування цієї методології для кіберзахисту критичної інфраструктури з'явилося порівняно недавно і ще не набуло широкого поширення у практиці кіберзахисту критичної інфраструктури. Баєсова мережа (БМ) — це ймовірнісна графічна модель, де орієнтований ациклічний граф використовується для з'єднання дискретних випадкових величин, які є залежними [11, 12]. Спільну ймовірність моделі можна отримати за допомогою набору умовних розподілів, які локально задані в кожному вузлі змінної відносно набору батьківських вузлів (тобто, спільний розподіл факторизується за допомогою добутку таких локальних розподілів). У контексті можливих сценаріїв атаки мають бути враховані потенційні вразливості вузла атаки [13, 14]. Щоб обчислити локальний умовний розподіл ймовірностей (CPT) вузла, адміністратору необхідно оцінити ймовірність успіху, коли зловмисник використовує відому вразливість. Для цього треба мати оцінку вразливості даного вузла. Таблиця умовних імовірностей (CPT) BAG обчислюється як комбінований ефект вразливості в мережі. Локальний CPT вузлів з логічними умовами «І» та «АБО» в BAG обчислюється згідно рівняння, яке викладене нижче. Логічне «І» означає всі дозволи, які достатні для компрометації вузла X_i. Логічне «АБО означає, що будь-якого одного дозволу достатньо для компрометації вузла. Безпосередніми батьками вузла X_i є pa_i, а ймовірність використання вразливості v_j дорівнює pv_j вузла.

$$p(X_i|pa_i) = \begin{cases} 0, & \exists X_j \in pa_i | X_j = 0, \\ \Pi_{j:X_j} p_{v_j}, & otherwise; \end{cases} \quad (1)$$

$$p(X_i|pa_i) = \begin{cases} 0, & \forall X_j \in pa_i | X_j = 0, \\ 1 - \Pi_{j:X_j}(1 - p_{v_j}), & otherwise. \end{cases}$$

Оскільки адекватність моделі критично залежить від апріорної оцінки ймовірності вразливості, у вузлах мережі необхідно використати наявні дані про ці вразливості. У подальшому, для оцінки ймовірності успішності атаки на даний вузол, ми будемо використовувати метрики, які визначені в Системі загальної оцінки вразливостей NIST (CVSS) [15],

Оцінка CVSS — це десяткове число за шкалою від 0 до 10. Вона складається з трьох груп — базової, часової та екологічної. Базові метрики кількісно визначають внутрішні характеристики вразливості за допомогою двох компонент:

1) компонента експлуатованості, що складається з вектора доступу (B/AV), складності доступу (B/AC) та екземплярів автентифікації (B/AU);

2) компонента впливу, що виражає потенційну шкоду конфіденційності (B/C), цілісності (B/I) та доступності (B/A).

Часові метрики кількісно визначають динамічні аспекти вразливості на середовище організації, такі як рівень експлуатації і доступність виправлень. Ці метрики враховують доступність інструментів і методів, що підлягають експлуатації (TE), рівень виправлення (TRL) і достовірність звіту (TRC). Метрики середовища кількісно визначають два аспекти впливу, які залежать від середовища, що оточує організацію. У цьому дослідженні нас цікавить оцінка ймовірності, тому оцінка груп впливу та метрики середовища у даному дослідженні не враховуються. У подальших дослідженнях часові та екологічні метрики можуть бути враховані. Більш докладно застосування метрик NIST (CVSS) буде показано на реальному прикладі.

З урахуванням зазначеної можливості апріорної початкової оцінки CPT вводиться поняття байєсівської графової моделі, базове визначення якого дано в роботі [16]. Баєсова мережа визначається орієнтованим ациклічним графом (DAG) над вузлами, що представляють випадкові величини, та дугами, які позначають умовні залежності між парами вузлів. Нехай множина $X = \{X_1, \dots, X_n\}$ дискретних змінних, де кожна змінна X_i може приймати значення з кінцевої області. Баєсова мережа — це пара (S, P), де S — мережева структура, яка кодує набір тверджень про умовну незалежність змінних у X, а P — набір локальних розподілів імовірностей, які пов'язані з кожною змінною. Тобто,

$$P = \{P_i\}, \text{ де } P_i = (X_i | Pa_i), \; Pa_i \text{ позначає батьків вузла } X_i \text{ в } S.$$

Потреба закодувати внесок різних умов безпеки, різні властивості мережі та здійснення різних способів компрометації системи зловмисником, під час компрометації системи, привела дослідників [17] до необхідності розширення поняття баєсівських мереж і до визначення баєсівського графа атак (BAG). У свою чергу, визначення баєсівського графа атак уже ґрунтується на визначенні шаблону атрибута, який дозволяє нам класифікувати ці властивості мережі для подальшого аналізу.

Визначення: шаблон атрибута — це загальна властивість мережі, яка включає такі компоненти (але, водночас, не обмежується ними):

1) системні вразливості;

2) небезпечні системні властивості, такі як небезпечна політика безпеки, пошкоджений дозвіл на доступ до файлів/пам'яті, або ж доступ на читання-запис у файловій структурі;

3) небезпечні мережеві властивості, такі як небезпечний стан мережі, небезпечні властивості брандмауера, небезпечний дозвіл на доступ до пристроїв чи периферійних засобів;

4) привілеї доступу, такі як: обліковий запис користувача, гостьовий обліковий запис або обліковий запис root.

Шаблон-атрибут допомагає нам класифікувати властивості мережі, які можуть бути корисними для зловмисника. Наприклад, «вразливість переповнення буфера SSH на FTP-сервері» можна розглядати як екземпляр шаблону системних вразливостей.

Атрибут — це випадкова величина за розподілом Бернуллі, яка представляє стан екземпляра шаблону атрибута. Таким чином, атрибут S пов'язаний зі станом Істина $\left(S = \frac{1}{T}\right)$ або Хибність $\left(S = \frac{0}{F}\right)$ та ймовірністю $Pr(S)$. Стан означає значення істинності твердження, що лежить в основі екземпляра шаблону атрибута. Наприклад, екземпляр S: «доступ користувача на локальній машині» є атрибутом, коли він

пов'язаний зі значенням істинності, яке вказує на те, чи має зловмисник доступ користувача до локальної машини. Відношення між атрибутами визначаються елементарними атаками, тобто елементарна атака дозволяє зловмиснику скомпрометувати атрибут S_{post} з S_{pre} з ненульовою ймовірністю успіху. S_{post} та S_{pre} є елементами множини атрибутів S. Тоді баєсівський граф атаки визначається як кортеж $BAG = (S, \tau, \varepsilon, P)$, де S — набір атрибутів, а A — набір елементарних атак, визначених на S, де:

1) $S = N_{internal} \cup N_{external} \cup N_{external}$ позначає набір атрибутів S_i, для яких $Ba \in A|Si = post(a)$. $N_{internal}$ позначає набір атрибутів S_j, для яких $\exists a_1, a_2 \in \in A|[S_j = pre(a_1)]$ та $S_j = post(a_2)$. $N_{terminal}$ позначає набір атрибутів S_j, для яких $Ba \in A|S_k = pre(a)$;

2) $\tau \subseteq S \times S$. Впорядкована пара $(S_{pre}, S_{post}) \in \tau$, якщо ${S_{pre}}' \to a$ BAG з класичного графа атаки. Ймовірності фіксуються в локальному умовному розподілі ймовірностей вузла. LCPD — це набір значень імовірності, що визначають імовірність компрометації вузла, враховуючи різні комбінації станів його батьків.

Використовуючи модель баєсівського графа атаки, можна провести декілька видів аналізу. Наприклад, можна обчислити розподіл імовірностей значень будь-якого вузла. Метою обчислення може бути спостереження значення будь-якого набору інших вузлів у BN. У прогнозному аналізі спостережувана змінна представляє причину, а запитувана змінна — наслідок. У діагностичному аналізі спостережувана змінна представляє наслідок, а запитувана змінна — причину. У змішаному діагностичному/прогнозному аналізі як причини, так і наслідки набору запитуваних змінних є частиною доказів. Оскільки ми зосередимося на задачі визначення рівня загроз у процесі атаки, в першу чергу нас цікавить прогнозний аналіз імовірності успішної атаки. Ця досить складна задача потребує вирішення двох основних підзадач:

1) вирахування ймовірності досягнення певного вузла мережі;
2) оцінки рівня загроз при досягненні певного вузла мережі.

Оцінка рівня загроз прямо зв'язана з оцінкою ймовірності досягнення певного вузла мережі. Оцінка рівня загроз при досягненні певного вузла мережі ще потребує визначення додаткової метрики, яку ми розглянемо далі.

Моделювання кібератаки на мережу енергооб'єкта

Кібератаки на енергооб'єкт, як уже зазначалося, мають складний характер, який передбачає декілька етапів, які, у свою чергу, можуть мати багато кроків. Мета зловмисників при атаці на функцію розподілу полягає в тому, щоби крок за кроком, переміщуючись IT-мережею, отримати права, а далі отримати доступ до безпосереднього управління OT. І зрештою, захопивши управління MTU та RTU, провести руйнівні маніпуляції для обладнання енергооб'єкта. Оскільки наміри зловмисників наперед невідомі, дізнатися про ціль атаки можна буде тільки на заключному етапі, коли вже складно щось вдіяти. Тому важливо ідентифікувати загрозу на ранніх етапах. З нашої точки зору, атака на енергосистеми складається з трьох основних етапів:

1) етапу проникнення та закріплення в трастовій зоні;
2) етапу латерального руху з метою підвищення привілеїв;
3) етапу реалізації цілей атаки.

На рис. 1 зображено схему IT-мережі енергооб'єкта, який перебуває під кібератакою. Розглянемо корпоративну мережу, що складається з декількох взаємопов'язаних підмереж, кожна з яких захищена внутрішніми та зовнішніми брандмауерами. Ця мережа призначена для підтримки критично важливих бізнес-функцій, та включає різні типи серверів і робочих станцій, як показано на рис. 1. Топологію та конфігурацію мережі було розроблено на основі аналізу реальних мережевих середовищ енергопідприємств з матеріалів Агентства з кібербезпеки та безпеки інфраструктури [18] та документації реальних інфраструктурних об'єктів з власних українських джерел.

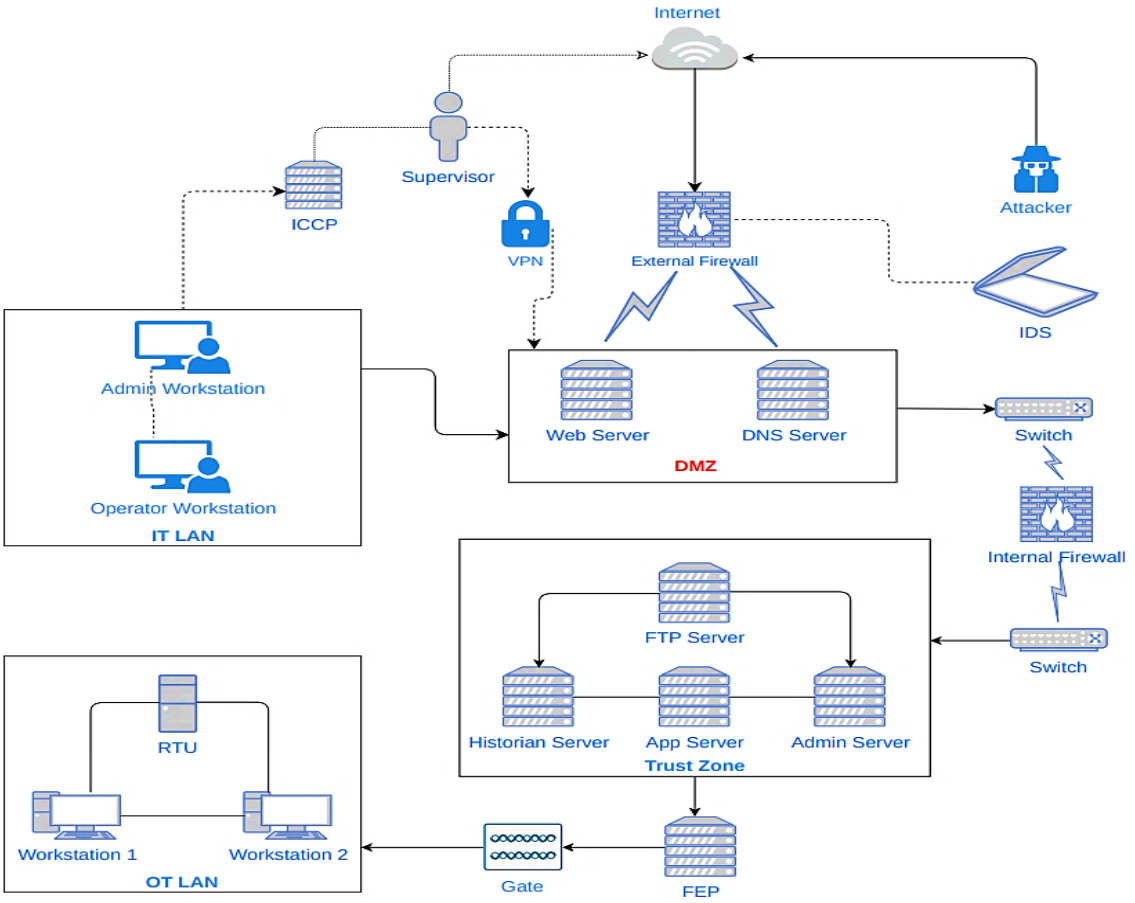

Рис. 1. Схема IT-мережі енергооб'єкта

Ціллю атаки є проникнення зловмисного експлойту в локальну мережу OT для контролю над RTU. Проте, ця подія може статися на третьому етапі, як вже зазначалося раніше. Спочатку шкідлива програма має проникнути в демілітаризовану зону (DMZ) IT. Після цього, користуючись вразливостями хостів, має потрапити у трастову зону. Останнім кроком на цьому етапі має бути створення програми, відомої як троянський кінь, яка утворить процес, мета якого підвищити свій рівень прав шляхом латерального руху. Це і є другий етап атаки. І вже на третьому етапі шкідлива програма за запрограмованим сценарієм оволодіє інформаційними ресурсами корпоративної мережі, зокрема, так званим Jump-сервером, за допомо-

гою якого може проникнути в середовище ОТ для виконання руйнівної місії. Специфіка кібератаки, цілеспрямованої на систему розподілу електроенергії проявляється саме на третьому етапі. Проте, вимоги безпеки говорять, що чим раніше буде виявлено загрозу, тим краще. Тому, спочатку треба зосередитися на першому етапі, тобто забезпечити раннє виявлення кібератаки. Для того, щоби побудувати модель атаки, необхідно провести аналіз вразливостей у хостах ІТ. Досить тривала практика боротьби з кібератаками вказує на те, що корінь вразливості програмного забезпечення в комп'ютерних мережах треба шукати в протоколах передачі пакетів між хостами та в утилітах, які забезпечують мережевий зв'язок між хостами.

Мережа ІТ енергоб'єкта може мати різні конфігурації з набором спеціалізованих серверів, а також робочих станцій. Взаємозв'язок між хостами не є уніфікованим. У табл. 1 (власна розробка) показано можливі протоколи комунікацій між різними серверами.

Водночас, ще одним важливим аспектом для оцінки комунікації між серверами є програмні інтерфейси та засоби. Загалом, більшість серверів з'єднуються через різноманітні специфічні API та RCE. При побудові баєсової моделі кібератаки ми враховуємо, що переважна більшість серверів і робочих станцій ІТ енергосистем базуються на ОС Windows. Якщо зосередитися лише на ранньому виявленні кібератаки на мережу енергооб'єкта, то високу ймовірність має наступний сценарій: на першому кроці буде проникнення в мережу, через проходження повз захист зовнішнього фаєрвола. Далі зловмисник після звернення до DNS-сервера спочатку скомпрометує Web Server, віддалено використовуючи вразливість CVE-2021-31166 для отримання локального доступу на сервері. Оскільки Web Server має доступ до File Server через протокол NFS, він може спробувати змінити дані на файловому сервері. Існує як мінімум два способи досягнення цього. Якщо у сервісі NFS є вразливості, він може спробувати використати їх і отримати локальний доступ на машині; або якщо таблиця експорту NFS налаштована неправильно, він може змінити файли на сервері через протокол NFS, використовуючи програмну оболонку типу NFS Shell. Як тільки він зможе змінити файли на файловому сервері, зловмисник може встановити троянську програму у виконувані бінарні файли на File Server (на одній із робочих станцій). Тепер атакуючий може чекати, доки необережний користувач на Workstation ініціює шкідливе програмне забезпечення, що породить процес, який у ході латерального руху буде підвищувати рівень своїх привілеїв, можливо аж до рівня адміністратора системи, щоби перейти до своїх руйнівних дій. Оскільки в цьому аспекті діє принцип «чим раніше – тим краще», вкрай важливо допомогти системі безпеки виявити загрозу ще до переходу на етап латерального руху. Проблема, яку потрібно вирішити — це якомога досконало визначити ситуацію, яка є суттєво невизначеною. Як справедливо зазначалось у [19], система захисту має справу з трьома видами невизначеності:

1) структури атак містять невід'ємну невизначеність, оскільки більшість атак не мають 100 % гарантії успіху;

2) оскільки спостереження захисника щодо потенційної активності атаки є обмеженими, тому маємо невизначеність, яка залежить від хибнопозитивних і хибнонегативних результатів датчиків системи виявлення вторгнень (IDS);

3) невідомий вибір зловмисника, отже кібератаки не гарантовано мають бути успішними.

Таблиця 1. Протоколи комунікацій між серверами

Сервер	WEB	DNS	ICCP	Application	Admin	FEP	FTP	Historian
WEB	—	UDP, TCP/HTTPS (DoH)	через проміжний застосунок	HTTP, HTTPs	HTTP/HTTPS	via HTTP	NFS	HTTPS, HTTPS, OPC, Proprietary
DNS	UDP, TCP/HTTPS (DoH)	—	пряма комунікація неможлива	UDP, TCP/HTTPS (DoH)	пряма комунікація неможлива	UDP, TCP/HTTPS (DoH)	—	пряма комунікація неможлива
ICCP	через проміжний застосунок	пряма комунікація неможлива	—	Gate OPC UA/TASE.2	TCP/IP connection	TASE.2	—	HTTPS (DoH), DNS over TLS (DoT) та DNSCrypt.
Application	HTTP, HTTPs	UDP, TCP/HTTPS (DoH)	Gate OPC UA/TASE.2	—	HTTP/HTTPS Proprietary, FTP	HTTPs	FTP FTPs	OPC DCOM
Admin	HTTP/HTTPS	пряма комунікація неможлива	TCP/IP connection	HTTP/HTTPS, Proprietary, FTP	—	FTP, SFTP, FTPS	пряма комунікація неможлива	FTP, SFTP, FTPS
FEP	HTTP protocols	UDP, TCP/HTTPS (DoH)	TASE.2	HTTPs	FTP, SFTP, FTPS SSL/TLS	—	TASE.2	FTP
FTP	NFS	UDP, TCP/HTTPS (DoH)	HTTPs	FTP FTPs	пряма комунікація неможлива	TASE.2	—	FTP
Historian	HTTPS, OPC, Proprietary	пряма комунікація неможлива	HTTPS (DoH), DNS over TLS (DoT) та DNSCrypt.	OPC DCOM	FTP, SFTP, FTPS	FTP	FTP	—
Application	через проміжний застосунок	UDP, TCP/HTTPS (DoH)	Gate OPC UA/TASE.2	—	HTTP/HTTPS Proprietary, FTP	via HTTP	NFS	HTTPS, OPC, Proprietary
Admin	HTTP/HTTPS	пряма комунікація неможлива	TCP/IP connection	HTTP/HTTPS, Proprietary, FTP	—	FTP, SFTP, FTPS SSL/TLS	пряма комунікація неможлива	FTP, SFTP, FTPS
FEP	HTTP protocols	UDP, TCP/HTTPS (DoH)	TASE.2	HTTPs	FTP, SFTP, FTPS	—	TASE.2	FTP
FTP	NFS	UDP, TCP/HTTPS (DoH)	HTTPs	FTP FTPs	пряма комунікація неможлива	TASE.2	—	FTP
Historian	HTTPS, OPC, Proprietary	пряма комунікація неможлива	HTTPS (DoH), DNS over TLS (DoT) та DNSCrypt.	OPC DCOM	FTP, SFTP, FTPS	FTP	FTP	—

Модель баєсових мереж дає можливість закодувати усі три типи невизначеності шляхом емпіричного визначення параметрів умовного розподілу, які можна визначити за допомогою значень, вирахуваних з метрик CVSS. При цьому, уточнення цих параметрів апостеріорного розподілу можна провести, використовуючи теорему Баєса, за рахунок отриманих даних від системи захисту.

На рис. 2 показано модель баєсової мережі кібератаки на IT-підсистему енергоб'єкта, реалізована за допомогою пакету GeNIe Academic Vesion. Розроблена баєсова модель дозволяє проаналізувати природу проходження багатоетапної атаки як імовірнісного каскаду, де успіх кожного наступного кроку залежить від результату попереднього.

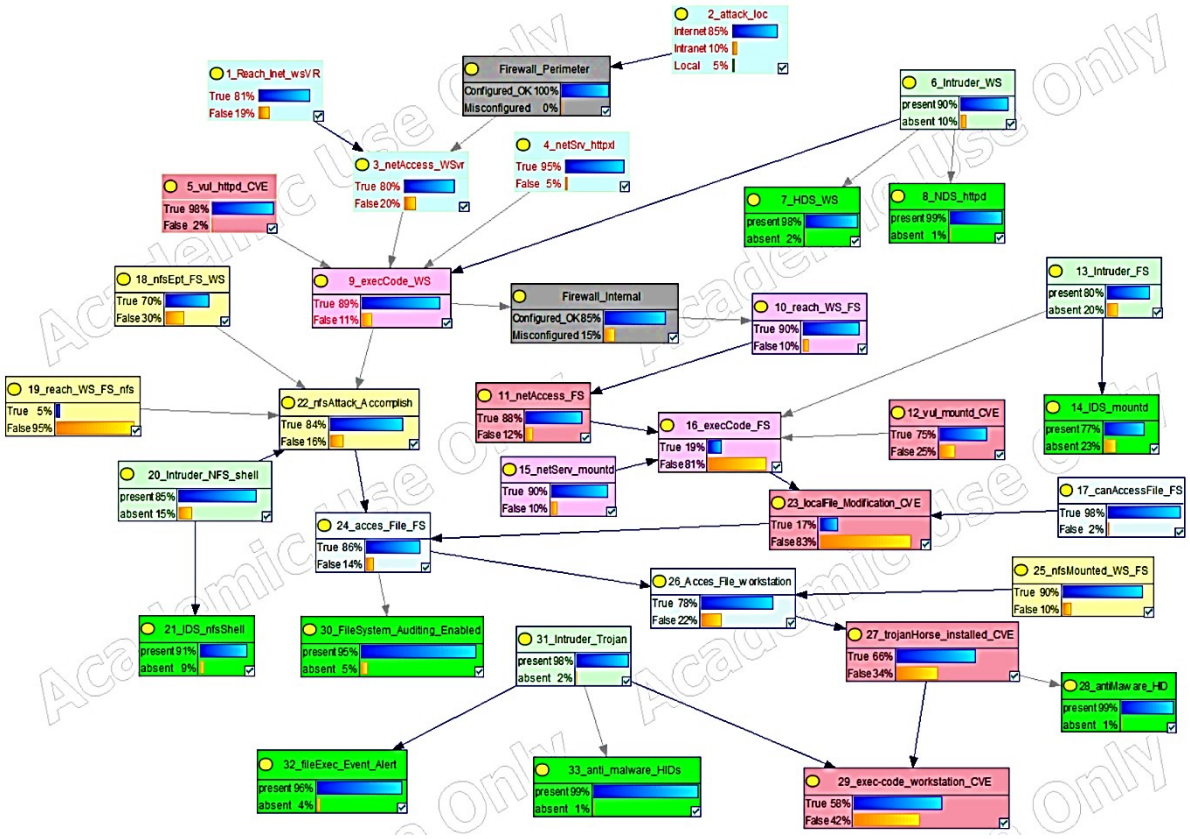

Рис. 2. Баєсова модель атаки на IT-мережу

Розроблена баєсова модель дозволяє проаналізувати природу проходження багатоетапної атаки як імовірнісного каскаду, де успіх кожного наступного кроку залежить від результату попереднього. Так, вузли, що позначають початкові передумови атаки, такі як її походження, виділені блакитним кольором. Послідовні дії зловмисника та проміжні стани компрометації позначені жовтим і рожевими кольорами, а ключові вразливості (CVE), що експлуатуються, — червоним. Мережеві фаєрволи, що виступають як основні бар'єри між сегментами, представлені сірим кольором, тоді як кінцеві системи контролю та моніторингу, такі як антивірусне ПЗ та HIDS, — зеленим. Окрему групу складають спеціалізовані вузли, що функціонують як узагальнюючі індикатори стану компрометації. Вони позначені світло-зеленим кольором і носять префікс «Intruder». Ці вузли функціонують як логічні

перемикачі фактору атаки: стан «present» означає наявність дій зловмисника на даному етапі та «активує» залежні гілки залежних передумов і подальших дій. Тоді як стан «absent» блокує подальше поширення ризику по відповідній гілці, імітуючи відсутність загрози.

При цьому модель враховує три фундаментальні вектори потенційної локації атаки, що задаються у вузлі attack_loc: ззовні (Інтернет), зсередини корпоративної мережі (Інтранет) і безпосередньо з локального хоста (Локальна мережа). Це дозволяє оцінювати як зовнішні загрози, так і ризики, що пов'язані з діями інсайдерів і скомпрометованими внутрішніми системами, а також ризиками, які викликані реальними вразливостями (CVE). Процес атаки розглядається як послідовність етапів: подолання периметра, бічне переміщення всередині мережі та, зрештою, доставка та виконання шкідливого програмного забезпечення на кінцеву точку. Модель комплексно відображає цей шлях, дозволяючи оцінити стійкість системи до такого типу загроз. Для оцінки стійкості системи було відпрацьовано чотири ключові сценарії.

1) ***Штатний режим***

Цей сценарій представляє поточний базовий, тобто «штатний» стан системи. Апріорні ймовірності для вузлів вразливостей встановлені на основі експертних оцінок, а для всіх засобів контролю (брандмауерів, систем моніторингу та цілісності, антивірусного ПЗ) задана висока ймовірність коректної роботи та налаштування. Розрахунок за таких умов демонструє базовий, інтегральний ризик для мережі з урахуванням усіх діючих захисних механізмів при атаці з найбільш імовірного вектора (Інтернет). Даний розрахунок моделі показує, що кінцева ймовірність успішної компрометації робочої станції (execCode_workstation) становить 58 %.

2) ***Відмова систем запобігання та виявлення***

Даний сценарій імітує ситуацію, коли зловмисник використовує передові техніки для обходу активних засобів захисту на хостах. У моделі це реалізується шляхом встановлення ймовірності для вузлів моніторингу у стан absent (відсутній/не спрацював) на рівні, близькому до еквіваленту 1.0. Такий підхід дозволяє оцінити, наскільки інфраструктура мережі залежить від цих засобів контролю, і що станеться, якщо вони будуть нейтралізовані. Оскільки, при імітації сценарію, в якому засоби активного захисту на хостах не спрацьовують, ризик значно зростає. Отже, кінцева ймовірність компрометації у цьому випадку досягає 78 %.

3) ***Відмова мережевого захисту (відключені брандмауери)***

Аналізується критичність мережевої сегментації і захисту периметра. Зокрема, моделюється ситуація повної відмови, або ж критичної помилки в конфігурації периметрового та внутрішнього брандмауерів. Це встановлюється шляхом задання для вузлів Firewall_Perimeter та Firewall_Internal стану Misconfigured з імовірністю 1.0. Такий сценарій демонструє максимальну вразливість системи до мережевих атак. За умови повної відмови мережевого захисту модель демонструє, що ймовірність успіху атаки становить 62 %.

4) ***Валідація моделі у стані спокою (відсутність зловмисної активності)***

Цей додатковий сценарій слугує як підтвердження коректності логіки моделі. У ньому імітується відсутність активності зловмисника та передумов для атаки. Дані умови досягаються шляхом встановлення ймовірностей для ключових вузлів вразливостей на нульовий рівень. Усі засоби контролю при цьому залишаються у

штатному, робочому режимі. У підсумку, отримані результати моделювання підтверджують логіку роботи мережі: кінцева ймовірність компрометації прямує до нуля і становить 1 %, а ймовірність спрацювання систем сповіщень відповідає їхньому базовому рівню хибних тривог. Це демонструє, що модель коректно та послідовно реагує на відсутність початкової загрози.

Аналіз розрахунків щодо кібератаки на основі моделі баєсової мережі показує, що є підмножина вузлів, які можна характеризувати як критичні. Досягнення критичного вузла зловмисником показує зростання рівня загроз, при цьому модель баєсової мережі дозволяє оцінити ймовірність такої події. У множину критичних вузлів входять:

— вузол події: «Захоплення Web Server»;
— вузол події: «Успішна атака на File Server»;
— вузол події: «Виконання шкідливого коду на робочій станції»;
— вузол події: «Інсталяція Trojan Horse».

Якщо отримати оцінку ймовірності для кожної із цих подій, можна оцінити загальний рівень загроз для ІТ-мережі. Однак виникає питання про рівень довіри до оцінки ймовірності, яка отримана на підставі емпіричних даних. Ключовим питанням тут є наскільки чутливою є похибка оцінки ймовірностей від похибок апріорних параметрів.

Оцінювання чутливості моделі

Придатність показаної вище моделі BN для оцінювання рівня загроз може бути піддана під сумнів за умови, якщо невеликі неточності параметрів апріорних розподілів можуть привести до суттєвих змін оцінки ймовірностей у вузлах мережі. Зрозуміло, що експертний характер визначення цих параметрів в основі своїй є наближеним методом. Використання баз даних CVSS очевидним чином ситуацію покращує, але не дає гарантії надання оцінки параметрів розподілу із заданою точністю. Тобто, виникає питання про робастність вихідних імовірностей баєсової мережі при певній невизначеності параметрів розподілу, яку можна дослідити, виконавши аналіз чутливості мережі. Для баєсової мережі виконання аналізу чутливості дає уявлення про співвідношення між параметрами ймовірності цієї мережі та її апостеріорними маргінальними значеннями. Найпростіший випадок аналізу — це однофакторний аналіз, за яким вивчається один параметр. Суттєво складнішим для дослідження є більш загальний *n*-факторний аналіз спільного впливу неточностей у наборі декількох параметрів. Вирахування чутливості від зміни певних залежних параметрів в основі своїй вимагає вирахування значень параметра у двох точках: початковій і зміщеній за рахунок зміни незалежних параметрів. Для виконання аналізу чутливості оцінок імовірностей у баєсовій мережі необхідно цілеспрямовано змінювати значення параметрів розподілу, а далі розрахувати наскільки змінилися відповідні ймовірності. Така процедура має передбачати багаторазовий перерахунок імовірностей і таким чином може вимагати великої кількості оцінок мережі, або повних поширень. Критичним фактором тут є метод обчислень, бо метод грубої сили є надто трудомістким, щоб мати якесь практичне застосування.

Низка авторів доклала немалих зусиль [20, 21], однак їм лише частково вдалося подолати обчислювальні проблеми, оскільки ті потребували тільки трьох перерахунків баєсової мережі на кожний параметр розподілів. Проте справжнім про-

ривом у вирішенні проблеми оцінювання чутливості стали результати Уффе К'єрульфа і Лінди ван дер Гааг [22]. У методі цих авторів апостеріорна гранична ймовірність пов'язана з простою математичною функцією з досліджуваним параметром, що значно спрощує процедуру оцінювання чутливості. Вона вимагає лише одного поширення назовні досліджуваного вузла для дерева BN і певного поширення вглиб для обчислення апостеріорних значень.

Для вихідних імовірностей треба розглянути апостеріорні граничні ймовірності у вигляді:

$$y = p(a|e),$$

де a — значення змінної A, а e позначає наявні апостеріорні значення. Тоді кожне значення ймовірності довільного параметра мережі має наступний вигляд: $x = p(b_i|F)$, де b_i — значення змінної з B, а F — довільна комбінація значень множини батьківських вузлів $\pi = p_a(B)$ з B. Тоді можна записати $p(a|e)(x)$ для позначення функції, що виражає апостеріорну маргінальну функцію $p(a|e)$ через параметр x.

Надалі ми будемо вважати, що в аналізі чутливості при зміні параметра $x = p(b_i|F)$, кожна з інших імовірностей $p(b_j|F)$ відповідно зміниться шляхом масштабування на співвідношення між множиною ймовірностей, які залишилися. Можна зробити припущення, що конкретний вигляд функцій $p(a|e)(x)$ представлено в (1):

$$p(b_i|\pi)(x) = \begin{cases} x & if\ j = i \\ p(b_i|\pi)\ \frac{1-x}{1-p(b_i|\pi)} & otherwise, \end{cases} \quad with\ p(b_i|\pi) < 1. \tag{2}$$

З урахуванням припущення про коваріацію, як зазначено вище, функція $y(x)$, що отримана за допомогою аналізу чутливості, є фактором двох лінійних функцій від x. Це і є ключовий момент, який дозволяє значно спростити процедуру обчислення та застосувати її для випадку багатофакторного оцінювання. На основі цього можна реалізувати алгоритм обрахування чутливості для обраних вузлів баєсової мережі. Такого роду алгоритм реалізований у деяких прикладних пакетах, зокрема в GeNIe Academic Version, за допомогою якого і було проведено оцінки чутливості баєсової моделі для раннього виявлення загроз.

Далі проведемо оцінювання чутливості моделі оцінювання кіберзагроз для представленої баєсової мережі у критичних вузлах. Результати розрахунків у критичних вузлах показано в табл. 2.

Таблиця 2. Оцінка чутливості в критичних вузлах

Критичний вузол	Значення чутливості
Захоплення Web Server	0,495
Успішна атака на File Server	0,331
Виконання шкідливого коду на робочій станції	0,234
Інсталяція Trojan Horse	0,426

Результати розрахунку оцінки факторів, що впливають на чутливість параметрів у критичних вузлах подано у табл. 2. Дані результати свідчать про низьку

чутливість баєсової моделі, що говорить про її перспективність в оцінюванні рівня загроз. Результати аналізу чутливості природнім чином переносяться на аналіз невизначеності для дослідження робастності моделі байєсівської мережі. Питання про робастність баєсових мереж у загальному вигляді залишається відкритим та потребує дослідження моделей баєсових мереж для конкретних предметних областей, зокрема і для раннього оцінювання кіберзагроз для енергетичної інфраструктури. Однак результати оцінювання в критичних вузлах дають підстави стверджувати, що баєсова модель кібератаки на енергосистему має властивості робастності.

Висновки

Представлені у даній статті результати досліджень дають підставу для висновків, що проблема зниження невизначеності під час кібератак на енергооб'єкт може бути вирішена в рамках методології використання баєсових мереж. У першу чергу, це стосується раннього виявлення загроз. Характерна особливість моделі полягає в тому, що апріорні розподіли ймовірностей будуються за допомогою баз даних CVE із застосуванням метрик CVSS. І хоча це суттєво підвищує достовірність моделі, проте водночас залишає питання щодо її точності та чутливості. Побудована баєсова мережа для раннього виявлення загроз продемонструвала належну релевантність для оцінки загроз на різних кроках кібератаки для певних базових сценаріїв. Цей факт дає підставу щодо використання даного підходу для кількісного оцінювання загроз. Крім того, в рамках статті проведено дослідження такого важливого питання як чутливість моделі від похибок апріорних параметрів. Практичні розрахунки за методом К'єрульфа та ван дер Гааг показали низьку чутливість моделі до похибок апріорних параметрів. У підсумку можна стверджувати, що використання баєсової мережі є перспективним підходом для застосування в оцінюванні загроз кібератак на ІТ-мережі енергооб'єктів.

Слід зазначити, що ми вважаємо даний підхід перспективним не тільки для раннього виявлення кіберзагроз, але і для всього циклу кібератаки. Що і буде предметом подальших досліджень.

1. Aljohani T.M. Cyberattacks on Energy Infrastructures as Modern War Weapons — Part I: Analysis and Motives. *IEEE Technology and Society Magazine*. June 2024. Vol. 43, No. 2. P. 59–69. doi: 10.1109/MTS.2024.3395688.

2. Stouffer K., Pillitteri V., Lightman S., Abrams M., and. Hahn A. NIST sp 800–82 rev 2, guide to industrial control systems (ICS) security. Technical report, NIST, 2015.

3. NSTB assessments summary report: Common industrial control system cyber security weaknesses. Technical report, Idaho National Laboratory, Gaithersburg, MD, United States, 2010.

4. Marcel Frigault and Lingyu Wang. Measuring network security using bayesian network-based attack graphs. In STPSA'08, 2008. В роботі [Liu, F.; Zhang, S.; Ma, W.; Qu, J. Research on Attack Detection of Cyber Physical Systems Based on Improved Support Vector Machine. *Mathematics.* 2022. **10**. 2713].

5. Wang W., Yang S., Hu F., Stanley H.E., He S., Shi M. An Approach for Cascading Effects within Critical Infrastructure Systems. *Phys. A Stat. Mech. Its Appl*. 2018. **510**. P. 164–177.

6. Abdelmalak M., Venkataramanan V., Macwan R. A Survey of Cyber-Physical Power System Modeling Methods for Future Energy Systems. *IEEE Access*. 2022. **10**. P. 99875–99896.

7. Davide Cerotti and Daniele Codetta Raiteri and Giovanna Dondossola and Lavinia Egidi and Giuliana Franceschinis and Luigi Portinale and Roberta Terruggia. A Bayesian Network Approach for

the Interpretation of Cyber Attacks to Power Systems. 2019. Italian Conference on Cybersecurity. URL: https://api.semanticscholar.org/CorpusID:59615894

8. Гальчинський Л., Личик В. Метрики оцінки кібервідмовостійкості (аналітичне оглядове дослідження). *Інформаційні технології та суспільство*. 2023. **2**(8). С. 27–33. https://doi.org/10.32689/maup.it.2023.2.3.

9. Личик В.В. Адаптивний підхід до реагування на порушення кіберстійкості об'єктів критичної інфраструктури. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво.* 2025. **59**. С. 16–28. https://doi.org/10.36910/6775-2524-0560-2025-59-02.

10. Jensen F.V., and Nielsen T.D. Bayesian Networks and Decision Graphs (2nd ed.). Springer, 2007.

11. Chen L., Zhang T., Ma Y., Li Y., Wang C., He C., Lv Z., Li N. A Bayesian–Attack–Graph–Based Security Assessment Method for Power Systems. *Electronics*. 2024. **13**. 2628. https://doi.org/10.3390/electronics13132628.

12. Jiali Wang, & Martin Neil, & Norman Fenton A Bayesian Network Approach for Cybersecurity Risk Assessment Implementing and Extending the FAIR Model. *Computers & Security*. 2019. 89. 101659. 10.1016/j.cose.2019.101659.

13. Davide Cerotti and Daniele Codetta Raiteri and Giovanna Dondossola and Lavinia Egidi and Giuliana Franceschinis and Luigi Portinale and Roberta Terruggia. A Bayesian Network Approach for the Interpretation of Cyber Attacks to Power Systems. 2019, Italian Conference on Cybersecurity. https://api.semanticscholar.org/CorpusID:59615894 P. Xie, J.H. Li, X. Ou, P. Liu, and R. Levy. Using Bayesian Networks for cybersecurity analysis. *In IEEE/IFIP DSN*. 2010. P. 211–220.

14. Zhang S., and Song S. A novel attack graph posterior inference model based on bayesian network. *Journal of Information Security*.2011. **2**(1). P. 8–27.

15. Schiffman M. Common Vulnerability Scoring System (CVSS). URL: http://www.first.org/cvss/cvss–guide.html.

16. Liu Y., and Man H. Network vulnerability assessment using Bayesian networks. In *Proceedings of the SPIE Conference on Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security*. 2005. Vol. 5812. P. 61–71. Orlando, FL.

17. Poolsappasit N., Dewri R., and Ray I. Dynamic Security Risk Management Using Bayesian Attack Graphs. *IEEE Transactions on Dependable and Secure Computing*. Jan.–Feb. 2012. Vol. 9, No. 1. P. 61–74. doi: 10.1109.

18. CISA. 2024. CISA Red Team's Operations Against a Federal Civilian Executive Branch Organization Highlights the Necessity of Defense–in–Depth. URL: https://www.cisa.gov/news–events/cybersecurity–advisories/aa24–193a.

19. Xie, Peng & Li, Jason & Ou, Xinming & Liu, Peng & Levy, Renato. Using Bayesian Networks for Cyber Security Analysis. Proceedings of the International Conference on Dependable Systems and Networks. 2010. P. 211–220. 10.1109/DSN.2010.5544924.

20. Castillo E., Gutierrez J.M., and Hadi A.S. Sensitivity analysis in discrete Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*. Part A: Systems and Humans. July 1997. Vol. 27, No. 4. P. 412–423. doi: 10.1109/3468.594909.

21. Laskey K.B. Sensitivity Analysis for Probability Assessments in Bayesian Networks. *IEEE Transactions on Systems, Man and Cybernetics*. 1995. **25**(6). P. 901−909.

22. Uffe Kjærulff and Linda C. van der Gaag. Making sensitivity analysis computationally efficient. In Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI). 2000. P. 317–325. San Francisco, California, Morgan Kaufmann Publishers.

Надійшла до редакції 24.08.2025