КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4(20), 2023
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2023.20.8192
УДК 004.056.53

Субач Ігор Юрійович
доктор технічних наук, доцент, завідувач кафедри
Інститут спеціального зв'язку та захисту інформації Національного технічного університету України Київський політехнічний інститут імені Ігоря Сікорського, Київ, Україна
ORCID ID: 0000 – 0002 – 9344 – 713X
igor_subach@ukr.net

Кубрак Володимир Олександрович
аспірант
Інститут спеціального зв'язку та захисту інформації Національного технічного університету України Київський політехнічний інститут імені Ігоря Сікорського, Київ, Україна
ORCID ID: 0000 – 0001 – 8877 – 5289
volodymir.kubrak@ukr.net

МОДЕЛЬ ІДЕНТИФІКАЦІЇ КІБЕРІНЦИДЕНТІВ SIEM-СИСТЕМОЮ ДЛЯ ЗАХИСТУ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМ

Анотація. У статті представлено модель ідентифікації кіберінцидентів SIEM-системою, які відбуваються в ході функціонування інформаційно-комунікаційних систем (ІКС). Наведено перелік задач, які виконує SIEM-система в контурі захисту ІКС та механізмів, що складають її основу та які, у свою чергу, є складовими загального процесу кореляції подій, що відбуваються в ІКС. Проведено аналіз методів процесу кореляції, направлених на видалення, об'єднання та зв'язування даних про події в ІКС з встановленням її причинності та пріоритетності. Зроблено висновок про неефективність застосування існуючих методів в умовах неповноти та неточності інформації про кіберінциденти. Проаналізовано кортежну модель розпізнавання кіберінцидентів та для усунення її недоліків запропоновано удосконалену модель, що ґрунтується на теорії нечітких множин та лінгвістичних термів. Запропонована нова постановка задачі розпізнавання кіберінцидентів, яка зводиться до їхньої ідентифікації. Проаналізовано методи її рішення та виділено низку суттєвих їхніх недоліків, які утруднюють їх використання на практиці. Запропоновано підхід до рішення сформульованої задачі ідентифікації кіберінцидентів SIEM-системою на основі формування нечіткої бази знань SIEM-системи про їхні ознаки на основі збору експертної інформації та її подальшої обробки шляхом застосування теорії нечітких множин. Сформульовано основні принципи, які мають бути використаними під час розробки математичної моделі ідентифікації кіберінцидентів SIEM-системою. Запропонована модель нечіткої бази знань про кіберінцидентіи у вигляді багатовимірної таблиці з ознаками кіберінцидентів, представлених лінгвістичними термами та класами, що їм відповідають. Наведено представлення нечіткої бази (матриці) знань у вигляді системи нечітких правил виду "ЯКЩО-ТО" та на їх основі, шляхом застосування операцій min та max, запропоновано модель ідентифікації кіберінцидентів SIEM-системою. Зроблено висновок про доцільність застосування представленої в роботі моделі для захисту інформаційно-комунікаційних систем в умовах неповноти та неточності інформації про кіберінциденти, що виникають в ході їхнього функціонування.

Ключові слова: інформаційно-комунікаційа система; кіберзахист; кіберінцидент; SIEM; багатопараметрична ідентифікація; теорія нечітких множин; база знань.

ВСТУП

Постановка проблеми. Аналіз задач моніторингу інформаційно-комунікаційних систем (ІКС) та методів підвищення ефективності їх функціонування показує [1], що задача кіберзахисту їх є однією з найважливіших та актуальних на сьогоднішній день. Так, відповідно до даних, наданих Оперативним центром реагування на кіберінциденти Державного центру кіберзахисту Державної служби спеціального зв'язку та захисту

інформації України, кількість детектованих подій інформаційної безпеки системою виявлення вразливостей і реагування на кіберінциденти та кібератаки під час первинного аналізу, у 2022 році склала – 181 млн.

Аналіз показує, що основою побудови ефективної системи кіберзахисту ІКС має бути застосування проактивної SIEM-системи [2].

Застосування в контурі захисту ІКС SIEM-системи дозволяє ефективно здійснювати проактивне управління кіберінцидентами. Суть такого управління полягає у тому, що використовуючи інформацію про події безпеки, які вже відбулися в ІКС, здійснюється прогнозування подій безпеки в майбутньому [2]. Це стає можливим завдяки розробці та застосуванню відповідних моделей і методів виявлення та розпізнавання кіберінцидентів.

Аналіз останніх досліджень і публікацій. Функціональну модель проактивної SIEM-системи розглянуто в роботах [2, 3]. Відповідно до задач, які виконує дана система (збір, обробку та аналіз подій безпеки, що поступають до неї з множини різнорідних розподілених джерел), основу її функціонування складають такі механізми, як: нормалізація, фільтрація, класифікація, агрегація, кореляція, пріоритезація та аналіз подій і кіберінцидентів та їхніх наслідків, а також генерація різноманітних звітів, повідомлень і візуального представлення даних для оперативного та обґрунтованого прийняття рішень [2].

У деяких джерелах дані механізми розглядаються, як етапи загального процесу, який має назву – процес кореляції [4-6]. Йому відводиться особливе місце в роботі SIEM-системи, оскільки його призначенням є виявлення кібератак, шкідливої активності, порушень політики безпеки та інші [4]. Це забезпечується завдяки вирішенню широкого спектру задач, які він охоплює: визначення потенційних взаємозв'язків між різнорідною інформацією безпеки; групування низькорівневих подій безпеки до подій безпеки більш високого рівня; виявлення потенційних кіберінцидентів на основі аналізу поведінки різних об'єктів інфраструктури та інші.

Технологічно, у складі SIEM-системи, метод кореляції включає послідовність дій над даними, яка направлена на виявлення певним способом ознак видалення, об'єднання та зв'язування інформації, що обробляється, а також встановлення її причинності та пріоритетності [4-6]. Дані ознаки називають кореляційними ознаками.

Для цього, на різних етапах процесу кореляції застосовуються велике різноманіття методів [7-11], таких як: метод на основі кінцевих автоматів, який застосовується для ідентифікації небезпечних станів системи; правило-орієнтований метод, який ґрунтується на правилах, що мають зрозумілі синтаксис та семантику; метод міркувань на основі прецедентів; метод баєсових мереж, який застосовується на етапі багатокрокової кореляції подій, аналізу збитків та пріоритезації; штучні нейронні мережі, які, також, застосовуються для кореляції подій, аналізу збитків і пріоритезації та інші.

Аналіз показує, що найбільш розповсюдженим методом є правило-орієнтований метод, проте в наслідок того, що він ґрунтуються на класичних продукційних правилах, які не завжди в умовах неповноти та неточності інформації про кіберінциденти дають очікуваний результат, застосування його є не завжди ефективним.

Мета статті. Метою статті є висвітлення підходів щодо підвищення ефективності процесу ідентифікації (розпізнавання) кіберінцидентів SIEM-системою, що виникають в ході функціонування інформаційно-комунікаційної системи в умовах неповноти та неточності інформації про них, шляхом розробки моделі, яка ґрунтується досягненнях теорії нечітких множин та лінгвістичних термів.

ISSN 2663 - 4023

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Виклад основного матеріалу дослідження. Будь-який кіберінцидент характеризується множиною інформаційних ознак, на основі яких, у свою чергу, він може бути розпізнаним.

Нехай $O = \{o_i\} \mid i = \overline{1,n}$ – множина інформаційних ознак кіберінцидентів, які відбуваються в системі та представляються за допомогою множини $C = \left\{ C_j \mid C_j = (o_{j1}, o_{j2}, \ldots, o_{jm}) \right\}$, $j = \overline{1,J}$, де $o_{ji} \in O$ – інформаційні ознаки, що асоціюються з кіберінцидентом C_j.

Тоді модель розпізнавання кіберінцидентів можна представити за допомогою кортежу [3]:

$$M = \langle K, O_i, R, C \rangle, \quad (1)$$

де K – класифікатор ознак;

$o_i \in O$ – множина ознак кіберінцидентів, що спостерігаються;

$R = \{R_i\}$ – множина правил розпізнавання кіберінцидентів;

C – кіберінцидент.

Процес розпізнавання кіберінцидентів здійснюється на основі правил, зазвичай продукційних:

$$R_1 : (K, O_i),\ R_2 : (K, O_i),\ \ldots,\ R_l : (K, O_i) \to C.$$

Проте, у традиційних продукційних системах, правила є класичними продукціями, які не у повній мірі відповідають умовам неповноти та неточності інформації про кіберінциденти, що виникають в ході функціонування інформаційно-комунікаційних систем. Для цього, як правило, застосовуються методи та моделі теорії нечітких множин на нечіткого логічного виводу [12-18].

З урахуванням наведеного, модель (1) може бути удосконаленою та представленою у наступному виді:

$$MF = \langle KF, O_i, RF, C \rangle, \quad (2)$$

де KF – нечіткий класифікатор;

$RF = \{RF_i\}$ – множина нечітких правил розпізнавання кіберінцидентів:

$$RF_1 : (K, O_v),\ RF_2 : (K, O_v),\ \ldots,\ RF_l : (K, O_v) \to C$$

З іншого боку, ґрунтуючись на роботах [13, 14] задача розпізнавання кіберінцидентів може розглядатися як задача їхньої ідентифікації, рішення якої полягає у знаходженні відображення:

$$O^* = (o_1^*, o_2^*, \ldots, o_n^*) \to c_j \in C = (c_1, c_2, \ldots, c_m), \quad (3)$$

де O^* – множина ознак кіберінцидента;

C – множина можливих кіберінцидентів.

ISSN 2663 - 4023

Область зміни ознак кіберінцидентів $o_i \in \left[\underline{o}_i ,\ \overline{o}_i \right],\ i = \overline{1,n},$ і вихідного значення результату ідентифікації $c_j \in \left[\underline{c}_j ,\ \overline{c}_j \right]$ вважаються відомими. Відповідно, $\underline{o}_i \left(\overline{o}_i \right)$– нижнє (верхнє) значення параметрів кіберінцидентів, $o_i, i = \overline{1,n},\ \underline{c}_j, \left(\overline{c}_j \right)$ – нижнє (верхнє) значення результату ідентифікації c_j.

Для вирішення на практиці задачі (3) найбільш широкого розповсюдження набули методи параметричної ідентифікації. Для їхнього застосування в якості апріорної інформації необхідно мати рівняння моделі об’єкту. У залежності від критерію ідентифікації або алгоритму, який застосовується для обчислення невідомих параметрів, представниками даної групи методів є: метод найменших квадратів, метод максимальної правдоподібності, метод середніх нев’язок, метод стахостичної апроксимації та ін. [15]. Проте, вони мають низку суттєвих недоліків, які утруднюють їх використання [13, 14]:

- моделі об’єктів типу “входи-вихід”, як правило не мають явної інтерпретації;
- відсутня можливість роботи з вхідними та вихідними змінними якісного типу;
- відсутня можливість використання досвіду експерта про структуру об’єкту, що формалізується у вигляді логічних висловлювань типу “ЯКЩО-ТО”.

Таким чином, наведені методи не в повній мірі пристосовані для вирішення сформульованої задачі ідентифікації кіберінцидентів у постановці (3).

З урахуванням наведеного, можна зробити висновок про доцільність формування бази знань [19] про ознаки кіберінцидентів (ОК), що виникають в ході функціонування ІКС та типи кіберінцидентів (ТК), для їхньої ідентифікації на основі збору експертної інформації та подальшої її обробки за допомогою теорії нечітких множин та лінгвістичних змінних.

Для цього необхідно формалізувати причинно-наслідкові зв’язки між змінними “ОК-ТК”, шляхом опису цих зв’язків природною мовою із застосуванням теорії нечітких множин та лінгвістичних змінних [13, 14, 17, 18]. Це дозволить здійснити математичну формалізацію природно-мовних висловлювань щодо їхнього застосування для вирішення задачі ідентифікації кіберінцидентів у постановці задачі (3).

Для перетворення експертних знань, які представлені у вигляді лінгвістичних висловлювань типу “ЯКЩО-ТО” до математичної моделі, необхідно застосувати математичний апарат функцій належності (ФН). Саме вони характеризують степінь впевненості експерта у тому, що деяке значення належить нечіткому поняттю (терму).

Наприклад, для ознаки кіберінциденту “кількість з’єднань до хосту, встановлених віддаленою стороною, що використовують один сервіс”, представленою множиною лінгвістичних змінних {“М – мала [-1, 500, 1000]”, “НС– нижче середньої [1000, 5000, 9999]”, “С – середня [10000, 60000, 100000]”, “ВС – вище середньої [125000, 600000, 1000000]”, “В – велика [1100000, 1800000, 2500000]”, ДВ – дуже велика [2400000, 3500000, 5250000]”} на універсумі [-1, 5250000], функція належності може мати наступний вид (рис. 1):

ISSN 2663 - 4023

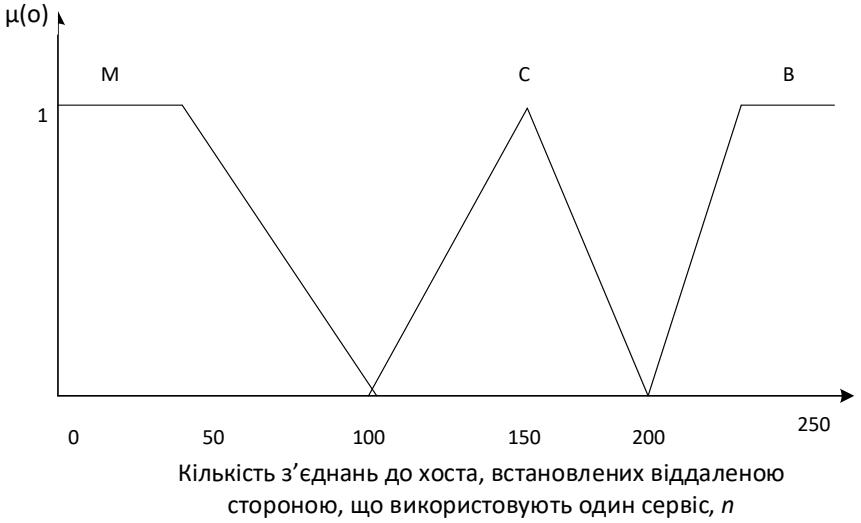

Рис. 1. Приклад графічного представлення функції належності для ознаки кіберінциденту "кількість з'єднань до хоста, встановлених віддаленою стороною, що використовують один сервіс"

А для ознаки кіберінциденту "відсоток з'єднань з помилкою типу REJ для даної служби приймача", заданою множиною {"M – мала [-0.5, 0, 0.6]", "B – велика [0.5, 0.8, 1.2]"} на універсумі [-0.5, 1.2], функція належності може мати вид, наведений на рис. 2.

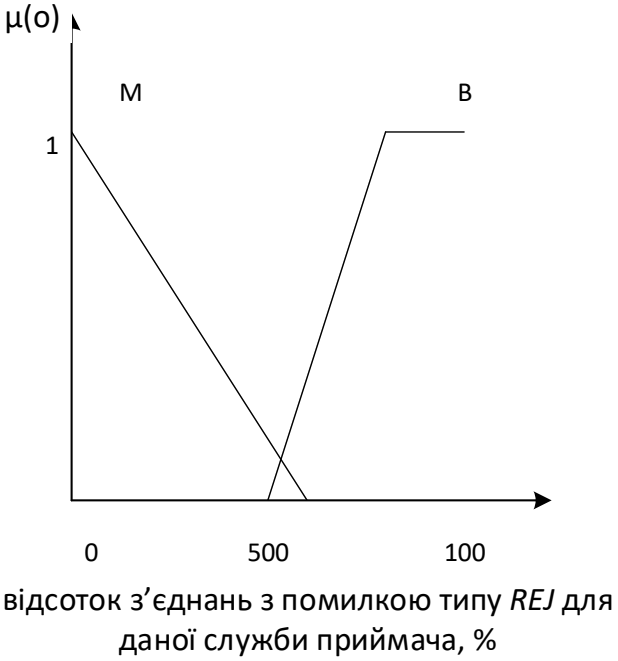

Рис. 2. Приклад графічного представлення функції належності для ознаки "відсоток з'єднань з помилкою типу REJ для даної служби приймача"

У свою чергу, методи нечіткого логічного виводу дозволяють зв'язати ФН ознак кіберінцидентів з результатами їхньої ідентифікації, при умові, що існує модель кіберінцедентів у вигляді множини нечітких правил типу "ЯКЩО-ТО" – нечіткої бази знань (НБЗ).

Аналіз літератури [13-18] та досвіду рішення на практиці задач ідентифікації, дозволяє сформулювати наступні принципи, які мають бути використаними під час

розробки математичної моделі ідентифікації (розпізнавання) кіберінцидентів, які виникають під час функціонування ІКС:

- принцип лінгвістичності стану ІКС, суть якого полягає у тому, що тип кіберінциденту та його ознаки, розглядаються як лінгвістичні змінні (ЛЗ), які оцінюються якісними термами;

- принцип формування залежності типу кіберінциденту від його ознак у вигляді НБЗ;

- принцип ієрархічності НБЗ, відповідно до якого, зменшення розмірності НБЗ може бути здійсненим шляхом проведення класифікації вхідних змінних та побудові "дерева виводу", яке визначає систему вкладених одне в одне логічних висловлювань.

За рахунок реалізації в моделі саме третього принципу, можна враховувати велику кількість ознак кіберінцидентів, які використовуються в процесі їхньої ідентифікації (розпізнавання). Проте, аналіз наукових публікацій [20] показує, що під час побудови дерева рішень для ідентифікації кіберінцидентів, необхідно намагатися, щоб число аргументів у кожному вузлу дерева було: 7±2.

Особливої важливості це набуває в умовах ускладнення моделі (додавання нових ознак) по мірі накопичення знань про них (донавчання) у НБЗ.

Таким чином, модель ідентифікації (розпізнавання) кіберінцидентів може бути заданою у вигляді сукупності нечітких правил "ЯКЩО-ТО", що зв'язують лінгвістичні оцінки ознак кіберінцидентів з результатами їхньої ідентифікації.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Нехай $O^* = \left(o_1^*, o_2^*, \ldots, o_n^*\right)$ - вектор фіксованих значень ознак кіберінцидентів, де $o_i^* \in \left[\underline{o_i},\ \overline{o_i}\right]$, $i = \overline{1,n}$. Тоді задача ідентифікації полягає в тому, щоб на основі вектору O^* визначити тип кіберінциденту $c_j \in C$. Тоді, необхідною умовою для формального рішення цієї задачі є наявність залежності (4):

$$c = \xi(o_1, o_2, \ldots, o_n), \tag{4}$$

де $o_1, o_2, \ldots, o_n$ – набір значень ознак кіберінцидентів, с – результат ідентифікації.

Для цього, необхідно розглянути вхідні і вихідні змінні з (4), як лінгвістичні змінні, які задані на універсальних множинах [13, 14, 17, 18]:

$$o_i = \left[\underline{o_i}, \overline{o_i}\right], c_j = \left[\underline{c_j}, \overline{c_j}\right]. \tag{5}$$

Для оцінки ЛЗ (5) цілком доцільно застосувати якісні терми, які складають терм-множини [13, 14, 17, 18]:

$\mathrm{A}_i = \left\{\alpha_i^1, \alpha_i^2, \ldots, \alpha_i^{k_i}\right\}$ – терм-множина змінної $o_i, i = \overline{1,n}$, де α_i^k – k-й лінгвістичний терм змінної $o_i, k = \overline{1,k_i}, i = \overline{1,n}$;

$\Delta = \left\{\delta_1, \delta_2, \ldots, \delta_m\right\}$ – терм-множина змінної c, де $\delta_j, j = \overline{1,m}$ – лінгвістичний терм змінної c, m – число можливих класів кіберінцидентів.

У загальному випадку, потужності терм-множин $\mathrm{A}_i, i = \overline{1,n}$ можуть бути різними. Тоді справедливо:

$$k_1 \neq k_2 \neq \ldots \neq k_n. \tag{6}$$

Також, назви термів $\alpha_i^1, \alpha_i^2, \ldots, \alpha_i^{k_i}$ можуть відрізнятися для різних лінгвістичних змінних $o_i, i = \overline{1,n}$.

Отже, лінгвістичні терми $\alpha_i^k \in \mathrm{A}_i, k = \overline{1,k_i}, i = \overline{1,n}$ та $\delta_j \in \Delta, j = \overline{1,m}$ можна розглядати як нечіткі множини, які задані на універсальних множинах o_i, c_j (5).

У свою чергу, нечіткі множини α_i^k та δ_j можна визначити наступним чином [13, 14]:

$$\alpha_i^k = \int_{\underline{o_i}}^{\overline{o_i}} \mu^{\alpha_i^k}(o_i)/o_i, \tag{6}$$

$$\delta_j = \int_{\underline{c}}^{\overline{c}} \mu^{\delta_j}(c)/c, \tag{7}$$

де $\mu^{\alpha_i^k}(o_i)$ – ФН значення змінної $o_i \in \left[\underline{o_i}, \overline{o_i}\right], i = \overline{1,n}$ терму $\alpha_i^k \in \mathrm{A}_i, k = \overline{1,k_i}, i = \overline{1,n}$; μ

$\mu^{\delta_j}(c)$ – ФН значення змінної $c = \left[\underline{c_i}, \overline{c_i}\right]$ терму – класу кіберінциденту $\delta_j \in \Delta, j = \overline{1,m}$.

Зауважимо, що у виразах (6) та (7) знак інтегралу позначає об'єднання пар $\mu(\omega)/\omega$.

Нехай L – кількість даних, які зв'язують вхідні дані – ознаки кіберінцидентів та вихідне значення – клас кіберінциденту, причому:

$$L = l_1 + l_2 + \ldots + l_m, \tag{8}$$

де l_j – число даних, що були отримані від експертів та які відповідають вихідній змінній – класу кіберінциденту $\delta_j \in \Delta, j = \overline{1,m}$, m – число класів кіберінцидентів, причому у загальному випадку: $l_1 \neq l_2 \neq \ldots \neq l_m$.

Зауважимо, що число даних, які були отримані від експертів є набагато меншою повного перебору різних поєднань l_j вхідних ознак кіберінцидентів.

Пронумерувати комбінації цих експертних даних можна представити у вигляді багатовимірної таблиці [13, 14]:

За аналогією [17, 18] назвемо дану таблицю матрицею знань про кіберінциденти. Вона має наступні властивості:

- розмірність даної матриці: $(n+1) \times N$, де $(n+1)$ – число стовпчиків матриці, а $L = l_1 + l_2 + \ldots + l_m$ – число її рядків;

- кожний рядок матриці є комбінацією вхідних значень ознак кіберінцидентів $o_i, i = \overline{1,n}$, яка віднесена експертом до одного з його класів δ_j, причому перші l_1 рядків відповідають класу δ_1, а останні l_m рядків – класу δ_m.

- перші n стовчиків матриці відповідають вхідним значенням ознак кіберінцидентів $o_i, i = \overline{1,n}$, а $(n+1)$-ий стовпчик відповідає вихідному значенню – класу кіберінциденту c.

- на перетині i-го стовпчика та jk_j-го рядку знаходиться елемент $\alpha_i^{jk_j}$, який відповідає лінгвістичній оцінці ознаки кіберінциденту o_i у рядку матриці jk_j, яка належить терм-множині відповідної ознаки o_i: $\alpha_i^k \in \mathrm{A}_i, k = \overline{1,k_i}, i = \overline{1,n}$.

Таблиця 1

Багатовимірна таблиця ознак кіберінцидентів і класів, що їм відповідають

Номер вхідної комбінації значень ознак кіберінцидентів	Ознаки кіберінцидентів						Клас кіберінциденту
	o_1	o_2	…	o_i	…	o_n	c
11	α_1^{11}	α_2^{11}	…	α_i^{11}	…	α_n^{11}	δ_1
12	α_1^{12}	α_2^{12}	…	α_i^{12}	…	α_n^{12}	
…	…	…	…	…	…	…	
$1k_1$	$\alpha_1^{1k_1}$	$\alpha_1^{1k_2}$	…	$\alpha_i^{1k_1}$	…	$\alpha_n^{1k_1}$	
…	…	…	…	…	…	…	…
$j1$	α_1^{j1}	α_2^{j1}	…	α_i^{j1}	…	α_n^{j1}	δ_j
$j2$	α_1^{j2}	α_2^{j2}	…	α_i^{j2}	…	α_n^{j2}	
…	…	…	…	…	…	…	
jk_j	$\alpha_1^{jk_j}$	$\alpha_2^{jk_j}$	…	$\alpha_i^{jk_j}$	…	$\alpha_n^{jk_j}$	
…	…	…	…	…	…	…	…
m_1	α_1^{m1}	α_2^{m1}	…	α_i^{m1}	…	α_n^{m1}	δ_m
m_2	α_1^{m2}	α_2^{m2}	…	α_i^{m2}	…	α_n^{m2}	
…	…	…	…		…		
mk_m	$\alpha_1^{mk_m}$	$\alpha_2^{mk_m}$	…	$\alpha_i^{mk_m}$	…	$\alpha_n^{mk_m}$	

Легко побачити, що описана вище матриця знань про кіберінциденти, може бути представленою у вигляді системи нечітких правил виду “ЯКЩО-ТО” [13, 14, 17, 18], які зв’язують значення вхідних ознак кіберінцидентів $o_i, i = \overline{1,n}$, з одним із можливих класів кіберінцидентів $\delta_j \in \Delta, j = \overline{1,m}$:

$$
\begin{aligned}
&\text{ЯКЩО}(o_i = \alpha_1^{11})\text{ТА}(o_2 = \alpha_2^{11})\text{ТА}\ldots\text{ТА}(o_n = \alpha_n^{11})\text{АБО}\\
&\quad(o_i = \alpha_1^{12})\text{ТА}(o_2 = \alpha_2^{12})\text{ТА}\ldots\text{ТА}(o_n = \alpha_n^{12})\text{АБО}\\
&\quad(o_i = \alpha_1^{1k_1})\text{ТА}(o_2 = \alpha_2^{1k_1})\text{ТА}\ldots\text{ТА}(o_n = \alpha_n^{1k_1})\text{ТО}(\text{c}=\delta_1),\ldots\\
&\ldots,\text{ЯКЩО}(o_i = \alpha_1^{j1})\text{ТА}(o_2 = \alpha_2^{j1})\text{ТА}\ldots\text{ТА}(o_n = \alpha_n^{j1})\text{АБО}\\
&\quad(o_i = \alpha_1^{j2})\text{ТА}(o_2 = \alpha_2^{j2})\text{ТА}\ldots\text{ТА}(o_n = \alpha_n^{j2})\text{АБО}\\
&\quad(o_i = \alpha_1^{k_j})\text{ТА}(o_2 = \alpha_2^{jk_j})\text{ТА}\ldots\text{ТА}(o_n = \alpha_n^{jk_j})\text{ТО}(\text{c}=\delta_j),\ldots\\
&\ldots,\text{ЯКЩО}(o_i = \alpha_1^{m1})\text{ТА}(o_2 = \alpha_2^{m1})\text{ТА}\ldots\text{ТА}(o_n = \alpha_n^{m1})\text{АБО}
\end{aligned}
\qquad (9)
$$

$$(o_i = \alpha_1^{m2})\text{ТА}(o_2 = \alpha_2^{m2})\text{ТА}\ldots\text{ТА}(o_n = \alpha_n^{m2})\text{АБО}$$
$$(o_i = \alpha_1^{mk_m})\text{ТА}(o_2 = \alpha_2^{mk_m})\text{ТА}\ldots\text{ТА}(o_n = \alpha_n^{mk_m})\text{ТО(с=}\delta_m\text{)},$$

де $\alpha_i^{jk_j}$ – лінгвістична оцінка ознаки кіберінциденту $o_i, i=\overline{1,n}$ у рядку k_j j-ої диз'юнкції, що визначається на терм-множині $\text{A}_i = \{\alpha_i^1, \alpha_i^2, \ldots, \alpha_i^{k_i}\}$;

$\delta_j \in \Delta, j=\overline{1,m}$ – лінгвістична оцінка класу кіберінциденту, що визначається на терм-множині $\Delta = \{\delta_1, \delta_2, \ldots, \delta_m\}$.

Отже , вираз (3.5), заданий у вигляді сукупності нечітких правил виду "ЯКЩО-ТО", який ґрунтуються на матриці знань про кіберінциденти (табл. 1) представляє собою модель ідентифікації (розпізнавання) кіберінцидентів SIEM-системою.

Якщо лінгвістичні оцінки α_i^{jk} змінних $o_1, o_2, \ldots, o_n$ та δ_j, $j=\overline{1,m}$ з (9) розглянути, як нечіткі множини, що визначені на універсальних множинах $o_i = \left[\underline{o_i}, \overline{o_i}\right], c_o = \left[\underline{c_j}, \overline{c_j}\right], i=\overline{1,n}, j=\overline{1,m}$, то $\mu^{\alpha_i^{jk}}$ – функції належності ознаки кіберінциденту $o_i = \left[\underline{o_i}, \overline{o_i}\right]$ нечіткому терму $\alpha_i^{jk}, i=\overline{1,n}, j=\overline{1,m}, k=\overline{1,k_j}$, а $\mu^{\delta_j}(o_1, o_2, \ldots, o_n)$ – ФН вектору ознак кіберінцидентів $O = \{o_1, o_2, \ldots, o_n\}, i=\overline{1,n}$, значенню вихідної оцінки $c=\delta_j$, $j=\overline{1,m}$.

Зв'язок між ними визначається через нечітку матрицю знань про кіберінциденти та шляхом заміни лінгвістичних термів на їхні ФН, а також заміни логічних операцій ТА чи АБО на операції $\wedge$ та $\vee$ може бути представленим у наступному вигляді:

$$\begin{aligned}\mu^{\delta_j}(o_1, o_2, \ldots, o_n) &= \mu^{\alpha_1^{j1}}(o_1) \wedge \mu^{\alpha_2^{j1}}(o_2) \wedge \ldots \wedge \mu^{\alpha_n^{j1}}(o_n) \vee \\ &\vee \mu^{\alpha_1^{j2}}(o_1) \wedge \mu^{\alpha_2^{j2}}(o_2) \wedge \ldots \wedge \mu^{\alpha_n^{j2}}(o_n) \vee \\ &\vee \mu^{\alpha_1^{jk_j}}(o_1) \wedge \mu^{\alpha_2^{jk_j}}(o_2) \wedge \ldots \wedge \mu^{\alpha_n^{jk_j}}(o_n)\end{aligned} \quad (10)$$

Тоді вираз (3.9) може бути представленим наступним чином:

$$\mu^{\delta_j}(o_i) = \bigvee_{k=1}^{k_j}\left[\bigwedge_{i=1}^{n} \mu^{\alpha_i^{jk}}(o_i)\right], j=\overline{1,m}. \quad (11)$$

Враховуючи, що в теорії нечітких множин операції $\wedge$ та $\vee$ можна представити операціями min та max [13, 14], то шляхом подібного перетворення виразу (11), отримаємо модель ідентифікації кіберінцидентів SIEM-системою (12):

$$\mu^{\delta_j}(o_i) = \max_{k=\overline{1,k_j}}\left\{\min_{i=\overline{1,n}}\left[\mu^{\alpha_i^{jk}}(o_i)\right]\right\}. \quad (12)$$

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Таким чином, запропонована в роботі модель ідентифікації кіберінцидентів SIEM-системою, що виникають в ході функціонування інформаційно-комунікаційних систем, застосовує нечіткі правила виду "ЯКЩО-ТО", які у сукупності складають нечітку базу знань SIEM-системи про кіберінциденти. Застосування даної моделі на практиці дозволяє усунути неповноту та неточність інформації про кіберінциденти у процесі їх ідентифікації.

Напрямом подальших досліджень є розробка методу виявлення кіберінцидентів SIEM-системою, в основу якого покладена запропонована модель та обґрунтування вибору показників ефективності [24] для оцінки ефекту від впровадження запропонованих рішень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1 Герасимов, Б.М., Субач, І.Ю., Хусаінов, П.В., Міщенко, В.О. (2008) Аналіз задач моніторингу інформаційних мереж та методів підвищення ефективності їх функціонування. *Сучасні інформаційні технології у сфері безпеки та оборони*, *3*(3), 24−27.

2 Субач, І., Кубрак, В., Микитюк, А. (2019) Архітектура та функціональна модель перспективної проактивної інтелектуальної системи SIEM-системи для кберзахисту об'єктів критичної інфраструктури. *Information Technology and Security*, *7*(2), 208–215. https://doi.org/10.20535/2411-1031.2019.7.2.190570.

3 Самохвалов, Ю., Толюпа, С. (2017). Кореляция событий в SIEM-системах на основе немонотонного вывода. *Захист інформації*, *19*(1), 5-9.

4 Jakobson, G., Weissman M. (1993). Alarm correlation. IEEE Network, *7*(6), 52 59.

5 Tiffany, M. (2002). A survey of event correlation techniques and related topics. http://www.tiffman.com/netman/netman.html.

6 Sadoddin, R., Ghorbani, A. (2006). Alert Correlation Survey: Framework and Techniques, In *Proceedings of International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST`06*), *October, 2006. Article no. 37*. (*pp. 1–10*).

7 Borkar, P. (2018). SIEM Rules or Models for Threat Detection? *Exabeam.* https://www.exabeam.com/siem/siem-threat-detection-rules-or-models/.

8 Salo, F., Injadat, M., Nassif, A., Shami, A., Essex, (2018). A Data Mining Techniques in Intrusion Detection Systems: A Systematic Literature Review, In *Proc. IEEEAccess, September 2018, 6,(pp. 56046–56058).*

9 Muller, A. (2009). Event Correlation Engine. Master`s Thesis. *Swiss Federal Institute of Technology Zurich.*

10 Hanemann, A., Marcu, P. (2008). Algorithm Design and Application of ServiceOriented Event Correlation, In *Proceedings of Conference BDIM 2008, 3rd IEEE/IFIP International Workshop on Business-Driven IT Management. (pp. 61–70).*

11 Elshoush, H., Osman, I.M. (2011). Alert correlation in collaborative intelligent intrusion detection systems. A survey. *Applied Soft Computing*, 4349–4365.

12 Zadeh, L. (1976). *The concept of a linguistic variable and its application to approximate decision making*. Mir.

13 Rothstein, A.P. (1996). *Medical Diagnostics on Fuzzy Logic*. Continent-PRIM.

14 Rothstein, A.P. (1999). *Intelligent Identification Technologies: Fuzzy Sets, Genetic Algorithms, Neural Networks.* UNIVERSUM.

15 Zaichenko, Y.P. (1991). *Operations Research: Fuzzy Optimisation.* Vyshcha Shkola.

16 Borisov, A.N., Krumberg, O.A., Fedorov, I.P. (1990). *Decision-Making on the Basis of Fuzzy Models: Examples of Use*. Zinatne.

17 Fesokha, V., Subach, I., Kubrak, V., Mykytiuk, A., Korotaiev, S. (2020). Zero-day polymorphic cyberattacks detection using fuzzy inference system. *Austrian Journal of Technical and Natural Sciences*, 5-6, 8-13. https://doi.org/10.29013/AJT-20-5.6-8-13.

18 Субач, І, Здоренко, Ю., Фесьоха, В. (2018). Методика виявлення кібератак типу JS(HTML)/Scrinject на основі застосування математичного апарату теорії нечітких множин. *Збірник наукових праць Військового інституту телекомунікацій та інформатизації імені Героїв Крут*, *4*, 125–131.

19 Герасимов, Б, Субач, І., Нікіфоров, Є. (2005). Моделі надання знань для використання в системах підтримки прийняття рішень. *Науково-технічна інформація*, *1*, 7 – 11.

20 Kalnish, V. (2019). Monitoring psychophysiological functions of operators in the process of their work activity. *Ukrainian journal of occupational healt, 15*(3), 204-215. https://doi.org/10.33573/ujoh2019.03.204.

21 Субач, І., Герасимов, Б. (2008). Показники якості інформаційного забезпечення та їх вплив на ефективність застосування ІСППР. *Вісник Національного університету ім. Тараса Шевченка*, *20*, 27–29.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (20), 2023
ISSN 2663 - 4023

Ihor Subach
doctor of technical science, associate professor, head of department
Institute of special communications and information security National technical university of Ukraine
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
ORCID ID: 0000 – 0002 – 9344 – 713X
igor_subach@ukr.net

Volodymyr Kubrak
postgraduate student
Institute of special communications and information security National technical university of Ukraine
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
ORCID ID: 0000 – 0001 – 8877 – 5289
volodymir.kubrak@ukr.net

MODEL OF CYBER INCIDENT IDENTIFICATION BY SIEM FOR PROTECTION OF INFORMATION AND COMMUNICATION SYSTEMS

Abstract. The article presents a model for identifying cyber incidents by a SIEM system that occur in the course of operation of information and communication systems (ICS). A list of tasks performed by the SIEM system in the ICS protection circuit and the mechanisms that form its basis, which, in turn, are components of the general process of correlation of events occurring in the ICS, is given. The methods of the correlation process aimed at removing, combining and linking data on events in the ICS with the establishment of its causality and priority are analyzed. It is concluded that the existing methods are ineffective in the context of incomplete and inaccurate information about cyber incidents. The tuple model for recognizing cyber incidents is analyzed and an improved model based on the theory of fuzzy sets and linguistic terms is proposed to eliminate its shortcomings. A new formulation of the problem of recognizing cyber incidents is proposed, which is reduced to their identification. The methods for solving it are analyzed and a number of their significant shortcomings are identified, which make it difficult to use them in practice. An approach to solving the formulated problem of identifying cyber incidents by a SIEM system is proposed on the basis of forming a fuzzy knowledge base of the SIEM system about their features based on the collection of expert information and its further processing by applying the theory of fuzzy sets. The basic principles that should be used when developing a mathematical model for identifying cyber incidents by a SIEM system are formulated. A model of a fuzzy knowledge base of cyber incidents is proposed in the form of a multidimensional table with the features of cyber incidents represented by linguistic terms and classes that correspond to them. A representation of the fuzzy knowledge base (matrix) in the form of a system of fuzzy rules of the "IF-THEN" type is presented, and on their basis, by applying the min and max operations, a model for identifying cyber incidents by a SIEM system is proposed. It is concluded that it is expedient to use the model presented in the paper to protect information and communication systems in the conditions of incomplete and inaccurate information about cyber incidents arising in the course of their operation.

Keywords: information and communication system; cyber defense; cyber incident; SIEM; multi-parameter identification; fuzzy set theory; knowledge base.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1 Herasymov, B.M., Subach, I.Iu., Khusainov, P.V., Mishchenko, V.O. (2008) Analiz zadach monitorynhu informatsiinykh merezh ta metodiv pidvyshchennia efektyvnosti yikh funktsionuvannia. Suchasni informatsiini tekhnolohii u sferi bezpeky ta oborony, 3(3), 24–27.

2 Subach, I., Kubrak, V., Mykytiuk, A. (2019) Arkhitektura ta funktsionalna model perspektyvnoi proaktyvnoi intelektualnoi systemy SIEM-systemy dlia kberzakhystu obiektiv krytychnoi infrastruktury. Information Technology and Security, 7(2), 208-215. https://doi.org/10.20535/2411-1031.2019.7.2.190570.

3 Samokhvalov, Yu., Toliupa, C. (2017). Koreliatsyia sobытyi v SIEM-systemakh na osnove nemonotonnoho vыvoda. Zakhyst informatsii, 19(1), 5-9.

4 Jakobson, G., Weissman M. (1993). Alarm correlation. IEEE Network, 7(6), 52 59.

5 Tiffany, M. (2002). A survey of event correlation techniques and related topics. http://www.tiffman.com/netman/netman.html.
6 Sadoddin, R., Ghorbani, A. (2006). Alert Correlation Survey: Framework and Techniques, In Proceedings of International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST`06), October, 2006. Article no. 37. (pp. 1–10).
7 Borkar, P. (2018). SIEM Rules or Models for Threat Detection? Exabeam. https://www.exabeam.com/siem/siem-threat-detection-rules-or-models/.
8 Salo, F., Injadat, M., Nassif, A., Shami, A., Essex, (2018). A Data Mining Techniques in Intrusion Detection Systems: A Systematic Literature Review, In Proc. IEEEAccess, September 2018, 6,(pp. 56046–56058).
9 Muller, A. (2009). Event Correlation Engine. Master`s Thesis. Swiss Federal Institute of Technology Zurich.
10 Hanemann, A., Marcu, P. (2008). Algorithm Design and Application of ServiceOriented Event Correlation, In Proceedings of Conference BDIM 2008, 3rd IEEE/IFIP International Workshop on Business-Driven IT Management. (pp. 61–70).
11 Elshoush, H., Osman, I.M. (2011). Alert correlation in collaborative intelligent intrusion detection systems. A survey. Applied Soft Computing, 4349–4365.
12 Zadeh, L. (1976). The concept of a linguistic variable and its application to approximate decision making. Mir.
13 Rothstein, A.P. (1996). Medical Diagnostics on Fuzzy Logic. Continent-PRIM.
14 Rothstein, A.P. (1999). Intelligent Identification Technologies: Fuzzy Sets, Genetic Algorithms, Neural Networks. UNIVERSUM.
15 Zaichenko, Y.P. (1991). Operations Research: Fuzzy Optimisation. Vyshcha Shkola.
16 Borisov, A.N., Krumberg, O.A., Fedorov, I.P. (1990). Decision-Making on the Basis of Fuzzy Models: Examples of Use. Zinatne.
17 Fesokha, V., Subach, I., Kubrak, V., Mykytiuk, A., Korotaiev, S. (2020). Zero-day polymorphic cyberattacks detection using fuzzy inference system. Austrian Journal of Technical and Natural Sciences, 5-6, 8-13. https://doi.org/10.29013/AJT-20-5.6-8-13.
18 Subach, I, Zdorenko, Yu., Fesokha, V. (2018). Metodyka vyiavlennia kiberatak typu JS(HTML)/Scrinject na osnovi zastosuvannia matematychnoho aparatu teorii nechitkykh mnozhyn. Zbirnyk naukovykh prats Viiskovoho instytutu telekomunikatsii ta informatyzatsii imeni Heroiv Krut, 4, 125–131.
19 Herasymov, B, Subach, I., Nikiforov, Ye. (2005). Modeli nadannia znan dlia vykorystannia v systemakh pidtrymky pryiniattia rishen. Naukovo-tekhnichna informatsiia, 1, 7 – 11.
20 Kalnish, V. (2019). Monitoring psychophysiological functions of operators in the process of their work activity. Ukrainian journal of occupational healt, 15(3), 204-215. https://doi.org/10.33573/ujoh2019.03.204.
21 Subach, I., Herasymov, B. (2008). Pokaznyky yakosti informatsiinoho zabezpechennia ta yikh vplyv na efektyvnist zastosuvannia ISPPR. Visnyk Natsionalnoho universytetu im. Tarasa Shevchenka, 20, 27–29.