DOI: 10.18372/2310-5461.48.15125

УДК 004.056.53

С. В. Толюпа, д-р техн. наук, проф.
Київський національний університет
імені Тараса Шевченка
orcid.org */0000-0002-1919-9174*
e-mail: tolupa@i.ua;

Р. С. Одарченко, д-р техн. наук, доц.
Національний авіаційний університет
orcid.org/0000-0002-7130-1375
e-mail: odarchenko.r.s@ukr.net;

І. І. Пархоменко, канд. техн. наук, доц.
Київський національний університет
імені Тараса Шевченка
orcid.org/0000-0001-6889-9284
e-mail: parkh08@ukr.net;

С. Ю. Даков, канд. техн. наук
Київський національний університет
імені Тараса Шевченка
orcid.org/0000-0001-9413-3709
e-mail: dacov@ukr.net;

ВИЯВЛЕННЯ АТАК В КОРПОРАТИВНІЙ МЕРЕЖІ ЗА ДОПОМОГОЮ ПРАВИЛ НЕЧІТКОЇ ЛОГІКИ

Вступ

Нечіткі системи виявлення мережевих вторгнень використовують множини нечітких правил для визначення ймовірності конкретних або загальних мережевих атак. Нечітка множина може бути сформовано для опису трафіку в конкретній мережі. Нечіткі набори правил асоціації використовуються для опису нормальних і аномальних класів. Належність записи певного класу визначається за допомогою відповідної метрики. Нечіткі асоціативні правила формуються на основі звичайних навчальних вибірок. Тестований зразок класифікується як нормальний, якщо згенерований сукупністю правил показник буде вище певного порогового значення. Зразки з більш низьким показником вважаються аномальними

Аналіз останніх досліджень і публікацій

В досліджених публікаціях [1; 2; 3; 4; 5] такий аналіз робиться на базі існуючих експертного або більш складного та ресурсо-затратного методу перебору, що ускладняє використання його на практиці, оскільки обладнання потребує в рази більше розрахункової потужності і має дуже складну структуру програмного забезпечення. Дана система вирішить ці питання і зробить можливість використання правил нечіткої логіки більш гнучким та доступним.

Постановка завдання

Основними цілями інтелектуального аналізу даних є пошук функціональних і логічних закономірностей в накопиченої інформації, побудові моделей і правил, які пояснюють знайдені аномалії і/або прогнозують розвиток деяких процесів, а також виявлення прихованих знань у вигляді кореляцій, тенденцій і взаємозв'язків, які аналітик не в змозі виявити й узагальнити самостійно.

Мета статті

Метою даної роботи розробити процес підбору та складання правил, синтез таблиці лінгвістичних правил (ТЛП) системи. А також моделювання системи з ітераційною корекцією лінгвістичних правил. За відсутності аналітичного опису об'єкта коригування правил проводиться безпосередньо після впровадження системи.

Основна частина

В експертній системі знання фахівців-експертів формалізуються у вигляді набору правил, що дозволяють приймати рішення в складних ситуаціях. Структуруванням знань експертів у вигляді бази знань займається аналітик (інженер по знаннях). Заснована на правилах експертна система складається з бази знань, механізму логічного висновку, блоку пояснення результатів і призначеного для користувача інтерфейсу. Загальна структура експертної системи (ЕС) наведена на рис. 1.

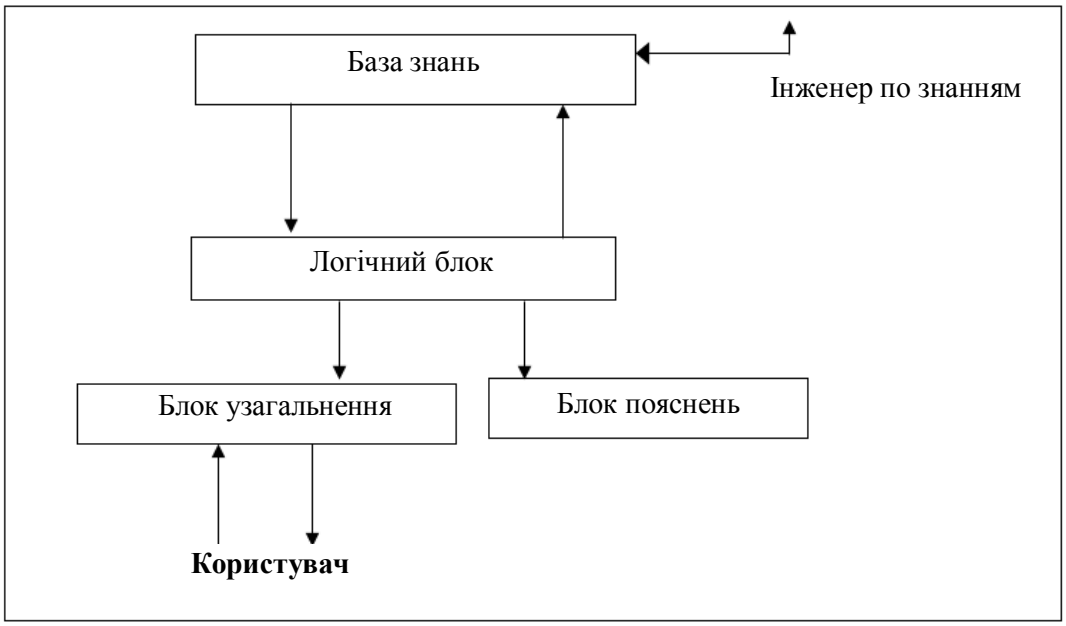

Рис. 1. Структурна схема експертної системи

Використання ЕС являє собою поширений метод виявлення атак, при якому інформація про атаки формулюється у вигляді правил. Ці правила можуть бути записані, наприклад, у вигляді послідовності дій або у вигляді сигнатури. При виконанні будь-якого з цих правил приймається рішення про наявність несанкціонованої діяльності [6]. Важливою перевагою такого підходу є практично повна відсутність помилкових тривог.

База даних (БД) експертної системи повинна містити сценарії більшості відомих на сьогодні атак. Для того щоб залишатися актуальними, експертні системи вимагають постійного оновлення БД, оскільки навіть невелика зміна вже відомої атаки може стати серйозною перешкодою для функціонування системи виявлення атак.

Іншим підходом є використання нечіткої логіки, яка дозволяє застосувати концепцію невизначеності в логічних висновках. Нечітка логіка дозволяє описувати правила в незавершеному, «розмитому» режимі на основі знань і ваг подій, що дозволяють припустити ймовірність атаки. В результаті можна працювати не з конкретними значеннями параметрів, а з їх якісними описами.

Для певної корпоративної мережі є характерні параметри трафіку, які можна визначити накопивши статистичну інформацію по поведінці мережі за довільний період роботи. Можливий підозрілий трафік в мережі може бути такий:

- що передається вузлами внутрішньої локальної мережі характерний для відповідей після успішного проведення атаки;
- командно-контрольних серверів ботмереж;
- що відноситься до протоколів та програм для миттєвого обміну повідомленнями;
- характерний для атак відмови в обслуговуванні;
- від вузлів зі списку Spamhaus Drop list;
- від вузлів, які відомі як джерела атак, на основі списку DShield;
- характерний для програм використання вразливостей (експлойтів);
- з використанням ICMP, характерний для проведення мережевих атак, наприклад, сканування портів;
- характерного для атак на сервіси IMAP;
- характерного для шкідливих програм (*malware*);
- характерного для мережевих черв'яків, що використовують протокол NetBIOS;
- який може суперечити політиці безпеки організації (наприклад, трафік VNC або використання анонімного доступу по протоколу FTP);
- програм сканування портів;
- характерного для атак на сервіси SMTP;
- характерного для атак на TFTP;
- характерного для троянських програм;
- характерного для атак на web-сервери;
- характерного для атак на основі ін'єкцій SQL (SQL-injection attacks);
- характерного для мережевих черв'яків;
- характерного для атак із застосуванням сигнатурних методів;
- характерного для аномалій у роботі протоколів і додатків;

- характерного для статистичних аномалій виявлення атак типу «відмова в обслуговуванні» (DoS);
- характерного при неправильно сформованих IP-пакетів, і відповідно з політикою, знищувати такий пакет, IP-сесію, до якої він належить або весь трафік від можливого джерела атаки.

Отже, застосування теорії нечітких множин для вирішення практичних задач передбачає як першого кроку формалізацію нечітких понять і відношень, які використовуються при описі елементів задачі управління, в даному випадку характеристик трафіку.

Методу, який можна застосувати, пропонується процедура побудови функцій належності $M_A(x)$ на основі кількісного парного порівняння степенів належності індивідуальним ОПР (особою, що приймає рішення (ОПР)). Результатом опитування ОПР є матриця $M = \|a_{ij}\|$ розмірністю n, де n — число точок, у яких порівнюються значення функцій.

Число m_{ij} показує, у скільки разів, на думку ОПР, $M_A(x_i)$ більше $M_A(x_j)$. При цьому кількість питань до ОПР становить $(n^2 - n)/2$. Значення функцій належності $M_A(x_1)$, …, $M_A(x_n)$ у точках x_1, …, x_n визначаються на основі рішення задачі

$$M\Phi^T = V_{\max}\Phi, \qquad (1)$$

де $\Phi = (\Phi_1,\ldots\Phi_n)$ — вектор довжиною n, $V_{\max}$ — максимальне число матриці M; T — символ транспонування.

Оскільки матриця M — позитивна по побудові, вирішення задачі (1) існує і є єдиним [7]. Остаточно отримаємо

$$M_A(x_i) = \frac{\Phi_i}{\sum_{i=1}^{n}\Phi_i} \qquad (2)$$

Звідси випливає, що $\sum_{i=1}^{n} M_A(x_i) = 1$..

Обчислення степенів належності по формулі (1) на основі вирішення задачі (1) витікає з таких міркувань [8]. M_0 — матриця, складена із відношень степенів належності, а $\Phi_0 = (M_A(x_1), M_A(x_2),\ldots,M_A(x_n))$.

$$M_0 = \begin{Vmatrix} \frac{M_A(x_1)}{M_A(x_1)} & \frac{M_A(x_1)}{M_A(x_2)} & \cdots & \frac{M_A(x_1)}{M_A(x_n)} \\ \frac{M_A(x_2)}{M_A(x_1)} & \frac{M_A(x_2)}{M_A(x_2)} & \cdots & \frac{M_A(x_2)}{M_A(x_n)} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{M_A(x_n)}{M_A(x_1)} & \frac{M_A(x_m)}{M_A(x_2)} & \cdots & \frac{M_A(x_n)}{M_A(x_n)} \end{Vmatrix}$$

Тоді очевидно, що $M_0\Phi\,\Phi_0^T = n\Phi_0$.

Оскільки M_0 — невід'ємна матриця, що дорівнює 1, то її власне число $V_{\max} = n$, а вектор Φ_0, складений із степенів належності — власний вектор.

Матриця M є апроксимацією матриці M_0, створеної на основі відповідей ОПР. Тому вектор степені належності і розраховується із виразу (1), величини a_{ij} інтерпритуються відповідно з таблицею.

Як зазначалося вище метод, що застосовується, використовує матрицю парних порівнянь елементів універсальної множини. Отже, нехай L — деяка властивість, яка розглядається як лінгвістичний терм $\tilde{L}$, являє собою сукупність пар [9] (див. таблицю)

$$\tilde{L} = \left\{ \frac{Ms(U_1)}{U_1}, \frac{Ms(U_2)}{U_2}, \ldots, \frac{Ms(U_n)}{U_n} \right\}.$$

Таблиця

a_{ij} значення	Зміст
1	$M_A(x_i)$ приблизно дорівнює $M_A(x_j)$
3	$M_A(x_i)$ несуттєво більше $M_A(x_j)$
5	$M_A(x_i)$ більше $M_A(x_j)$
7	$M_A(x_i)$ досить більше $M_A(x_j)$
9	$M_A(x_i)$ значно більше $M_A(x_j)$
2, 4, 6, 8	Значення проміжних по степені між перерахованими

де $\{U_1, U_2, \ldots, U_n\} = U$ — універсальна множина на якій задається нечітка множина $\tilde{L} \subset U$; $Ms(U_i)$ — степінь належності елементів $U_i \in U$ до нечіткої множини $\tilde{L}$.

Задача полягає в тому, щоб визначити значення $Ms(U_i)$ для всіх $i = \overline{1,n}$. Сукупність цих значень і буде складати невідому функцію належності.

Метод, який застосовується для рішення поставленого завдання, базується на ідеї розподілення степеня належності елементів універсальної множини відповідно з їх рангами.

У нашому випадку під рангом елемента $U_i \in V$ розуміється число $rs(U_i)$, яке характеризує значимість (чи вагомість) цього елемента у формуванні властивості, яка описується нечітким термом $\tilde{L}$. Припустимо також, що має місце правило: чим більше ранг елемента, тим більше степінь його належності.

У подальшому викладенні будемо позначати [10]:

$$rs(U_i) = r_i, \quad Ms(U_i) = M_i, \quad i = \overline{1,n}.$$

Тоді правило розподілення степенів належності задаємо у вигляді співвідношення:

$$\frac{M_1}{r_1} = \frac{M_2}{r_2} = ... = \frac{M_n}{r_n} \tag{3}$$

до якого долучається умова нормування

$$M_1 + M_2 + \ldots + M_n = 1. \tag{4}$$

Використовуючи співвідношення (3) легко визначити степінь належності всіх елементів універсальної множини через степінь належності опорного елемента. Якщо опорним є елемент $U_1 \in U$ є ступенем належності M_1, то

$$M_2 = \frac{r_2}{r_1} M_1,\ M_3 = \frac{r_3}{r_1} M_1, ...,\ M_n = \frac{r_n}{r_1} M_1. \tag{5}$$

Для опорного елемента $U_2 \in U$ з належністю M_2 отримаємо:

$$M_1 = \frac{r_1}{r_2} M_2,\ M_3 = \frac{r_3}{r_2} M_2, ...,\ M_n = \frac{r_n}{r_2} M_2. \tag{6}$$

І, нарешті, для опорного елемента $U_n \in U$ з належністю M_n маємо:

$$M_1 = \frac{r_1}{r_n} M_n,\ M_2 = \frac{r_2}{r_n} M_n, ...,\ M_{n-1} = \frac{r_{n-1}}{r_n} M_n. \tag{7}$$

Ураховуючи умови нормування виразу (4) і співвідношення (5)–(7) знаходимо

$$\left.\begin{aligned} M_1 &= \left(1 + \frac{r_2}{r_1} + \frac{r_3}{r_1} + ... + \frac{r_n}{r_1}\right)^{-1} \\ M_2 &= \left(\frac{r_1}{r_2} + 1 + \frac{r_3}{r_2} + ... + \frac{r_n}{r_2}\right)^{-1} \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ M_n &= \left(\frac{r_1}{r_n} + \frac{r_2}{r_n} + \frac{r_3}{r_n} + ... + 1\right)^{-1} \end{aligned}\right\}. \tag{8}$$

Отримання формули (8) дає можливість обрахувати степінь належності $Ms(U_i)$ елементів $U_i \in U$ до нечіткого терму $\tilde{L}$ двома незалежними шляхами [10]:

1) по абсолютній оцінці рівнів $r_i (i = \overline{f,n})$, які визначаються за методиками, запропонованими в теорії структурного аналізу систем. Для експертних оцінок рангів можливо скористатися дев'ятибальною шкалою (1 — найменший ранг, 9 — найвищий ранг).

2) за відносними оцінками рангів $r_i / r_j = a_{ij}$ $_{(i,j=\overline{1,n})}$, які створюють матрицю

$$A = \begin{bmatrix} 1 & \frac{r_2}{r_1} & \frac{r_3}{r_1} & ... & \frac{r_n}{r_1} \\ \frac{r_1}{r_2} & 1 & \frac{r_3}{r_2} & ... & \frac{r_n}{r_2} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \frac{r_1}{r_n} & \frac{r_2}{r_n} & \frac{r_3}{r_n} & ... & 1 \end{bmatrix} \tag{9}$$

Оскільки матриця (9) може бути інтегрована, як матриця парних порівнянь рангів, то для експертної оцінки елементів цієї матриці можливо скористатися шкалою Сааті [11]. У нашому випадку ця шкала формується відносно таблиці таким чином, за допомогою формул (8), експертні знання про ранги елементів чи їх парних порівнянь перетворюються у функцію належності нечіткого терма.

Для реалізації даного метода необхідно виконати такі дії [12]:

1. Задати лінгвістичну змінну *x*.
2. Визначити універсальну множину, на якій задається змінна.
3. Задати сукупність нечітких термів $\{L_1, L_2, \ldots, L_m\}$, які використовуються для оцінки змінної *x*.
4. Для кожного терму L_j ($j = \overline{1,m}$) сформувати матрицю (9).
5. Користуючись формулами (8), розрахувати елементи функції належності для кожного терма.

Визначення якісних факторів типів мережевого трафіку здійснюється за допомогою експертних оцінок.

Для мережевого трафіку можна виділити наступні позиції, які будуть визначені якісними експертними оцінками:

- кількість пакетів;
- тривалість захвату (сек);
- середня к-ть пакетів/сек;
- середній розмір пакета (байт);
- середня кількість байт/сек.

По кожній позиції застосовується шкала оцінок від 0 (мінімум) до 1 (максимум). Отже маючи кількісні значення для всіх показників mi та загальну кількість показників *n*, а також вагові коефіцієнти показників k_i, які приймають цілі

значення від 1 до 9 (де 9 — це абсолютне домінування показника над іншими, 1 — це мінімальний вплив показника на визначення кількісного значення якості одиниці обладнання) можливо визначити кількісне значення якості мережного трафіку R (головного чи допоміжного), яке міститься також у діапазоні від 0 до 1:

$$R = \frac{\sum_{i=1}^{n} k_i * m_i}{n + \sum_{i=1}^{n} (k_i - 1)} \qquad (10)$$

Узагальнений комплексний коефіцієнт, який визначає якість у межах від 0 до 1 і визначається за формулою:

$$Q = \frac{\sum_{i=1}^{l} h_i * R_i}{l + \sum_{i=1}^{l} (h_i - 1)} \qquad (11)$$

де l — це загальна кількість пакетів; h_i — відповідні вагові коефіцієнти різного типу трафіка, які теж мають діапазон від 1 до 9.

Лінгвістичні змінні y_1 та y_2, які мають терми: «низька», «середня», «висока» приймають відповідні значення залежно від коефіцієнта Q. Так лінгвістична змінна матиме значення «висока» у випадку коли Q > 0,8, «середня» 0,6 < = Q < = 0,8, в інших випадках низька.

Основу проєктування інтелектуальних нечітких систем складає конструювання «знань» із застосуванням методів подання та пошуку знань [13; 14]. В самій БЗ правила-продукції мають різний пріоритет.

Структурна схема нечіткої системи виявлення аномальних показників трафіку для сегменту мережі наведена на рис. 2, де f, $\overline{f}$ — параметри трафіку та їх похідні; f^*, $\dot{f}^*$ — лінгвістична форма змінних параметрів трафіку та їх похідних; y — поточні показники трафіку; q — оптимальні показники трафіку; e, $\dot{e}^*$ — розузгодження, та похідні розузгодження; U — нечіткі правила; e^*, $\dot{e}^*$ — лінгвістична форма розузгоджень та їх похідних; D_f, D_e — диференціатори відповідно по трафіку та розузгодженню; ТЛП — таблиця лінгвістичних правил.

Відповідно до схеми аналізатор здійснює діагностику та фільтрацію вхідних даних, фазифікатор переводить з числової в лінгвістичну форму відповідні дані.

Класифікатор аналізує отриману вхідну інформацію визначає відповідну ситуацію, по якій в базі знань, активізуючи певні продукційні правила.

Дефазифікатор переводить з лінгвістичної форми у цифрову і генерує відповідне правило.

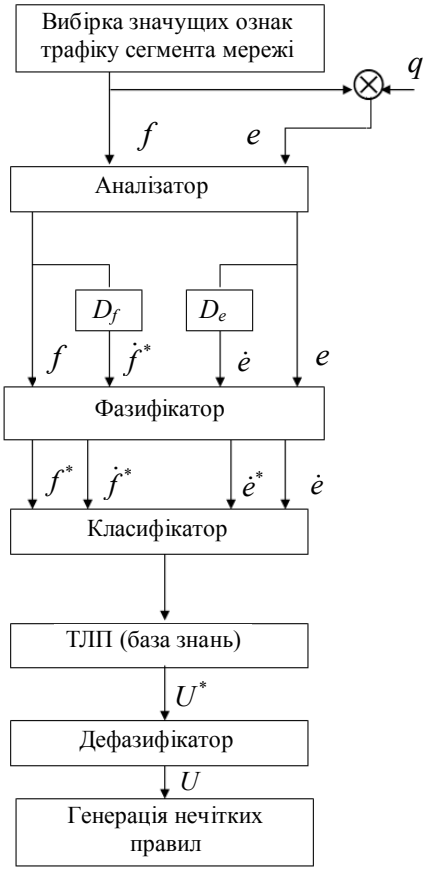

Рис. 2. Структурна схема нечіткої системи для виявлення аномального трафіку в сегменті мережі

Система реалізується в класі нечітких відображень вигляду [15]:

якщо f_i^* та якщо $\dot{f}_i^*$, то U_i^*;

якщо e_n^* та якщо e_m^*, то U_m.

Тобто в нечіткій системі розглядаються такі нечіткі підмножини [16]:

$$F_i(i = \overline{1, n_1});\ F_{ij}(j = \overline{1, n_2});\ E_e(e = \overline{1, n_3});$$

$$E_{eg} = (g = \overline{1, n_4});\ U_r = (r = \overline{1, n_5});\ U_{rk} = (k = \overline{1, n_6}),$$

де F_i, F_{ij} — універсальні підмножини змінного збурення, та похідної змінного збурення; E_e, E_{eg} — універсальні підмножини відхилення та похідної відхилення; U_r,U_{rk} — універсальні підмножини генерованих правил.

Відповідні нечіткі підмножини визначаються за допомогою лінгвістичної мови:

для змінних параметрів трафіку:

$F_{ij} \cong (f_m,\ \mu_i(f_m(t)))$, $i = 1,2,\ldots,n_1$),

де $f(t)$ — поточне значення трафіку; $\mu_i(f(t))$ — функція належності до нечіткої підмножини.

Аналогічно, для швидкості зміни (похідної) параметра трафіку.

$F_{ij} \cong (\dot{f}_m, \mu_i(e_m(t)))$, $j = 1,2,\ldots, n_2)$.

Похибки та швидкості її зміни

$E_e \cong (e_m, \mu_e(e_m(t))), e = 1,2,...,n_3;$

$E_{cq} \cong (\dot{e}_m, \mu_q(\dot{e}_m(t))$, $q = 1,2,...,n_4$.

Результат застосування нечітких правил:

$U_r \cong (U_2, \mu_2, (U_2(t)), \quad r = 1,2,...,n_5;$

$U_{rk} \cong (U_k, \mu_k, (U_k(t))), \quad k = 1,2,..,n_6$.

Висновки

Найбільш важливим моментом в процедурі синтезу нечіткої системи є підбір та складання правил, чи іншими словами, синтез таблиці лінгвістичних правил (ТЛП) системи [17]. Лінгвістичні правила системи евристично складаються розробником, який добре проінформований про технологічні особливості об'єкта. При аналітичному описі об'єкта проводиться машинне моделювання розроблюваної системи з ітераційною корекцією лінгвістичних правил. За відсутності аналітичного опису об'єкта коригування правил проводилось безпосередньо після впровадження системи.

ЛІТЕРАТУРА

1. **Motro A.,** Smet, P. Uncertainty Management in Information Systems: From Needs to Solutions. Springer, 1997. 464 p.

2. **Parsons S.** Current Approaches to Handling Imperfect Information in Data and Knowledge Bases. *Knowledge and Data Engineering IEEE*. 1996. Vol. 8. №3. P. 483–488.

3. **Sugeno M.,** Takagi T. Fuzzy Identification of Systems and It's Applications to Modeling and Control. *IEEE Trans. On Systems, Man, and Cybernetics*. 1985. №15. P. 116–132.

4. **Ishibuchi H.**, Nojima Y. Pattern Classification with Linguistic Rules. Fuzzy Sets and Their Extensions: Representation, Aggregation and Models Studies in Fuzziness and Soft Computing. 2008. Vol. 220. P. 377–395.

5. **Толюпа С. В.,** Штаненко С. С., Берестовенко Г. Класифікаційні ознаки систем виявлення атак та напрямки їх побудови. *Збірник наукових праць Військового інституту телекомунікацій та інформатизації імені Героїв Крут*. 2018. Вип. 3. С. 56–66.

6. **Пархоменко І. І.** Автоматизоване управління ділянкою очищення дифузійного соку на базі нечіткої логіки. *Автоматизація виробничих процесів*. 2001. №1(12). С. 36–44

7. **Zadeh L. A.** Fuzzy Sets. *Information and Control*. 1965. Vol.8. P. 338–353.

8. **Заде Л. А.** Понятие лингвистической переменной: и ее применение к принятию приближенных решений. М. : Мир, 1976. 167 с.

9. **Yang H.**, Xie F., Lu Y. Clustering and classification based anomaly detection. Fuzzy Systems and Knowledge Discovery. 2006. Vol. 4223. P. 1082–1091.

10. **Bhattacharyya D. K.,** Kalita J. K. Network Anomaly Detection. A Machine Learning Perspective. CRC Press, 2014. 364 p.

11. **Tajbakhsh A.,** Rahmati M., Mirzaei A. Intrusion detection using fuzzy association rules. *Applied Soft Computing*. 2009. Vol. 9. No. 2. P. 462.

12. **Takagi T.,** Sugeno M. Fuzzy Identification of Systems and Its Applications to Modeling and Control. *IEEE Transactions on Systems. Man and Cybernetics*. 1985. Vol. SMC-15. №1. Pp. 11.6–132.

13. **Popat D.,** Sherda H., Taniar D. Classification of Fuzzy Data in Database Management System. *Proceedings of 8th International KES Conference* (Wellington, New Zealand). 2004. P. 691–-697.

14. **Blanco I. J.,** Marin N., Martinez Cruz C., Vila M.A. About the Use of Ontologies for Fuzzy Knowledge Representation. Proceedings of the Joint 4th Conference of the *European Society for Fuzzy Logic and Technology* (Barcelona, Spain, 2005). 2005. P. 106–111.

15. **Гнатчук Є. Г.** Моделювання нечіткого логічного висновку процесу діагностування комп'ютерних засобів. *Вісник Вінницького політехнічного інституту*. 2005. №6 (63). С. 220–224.

16. **Ma Z.M.,** Yan, L. A Literature Overview of Fuzzy Database Models. J. Inf. Sci. Eng. 2008. №24. P. 189–202.

17. **Кравець П.,** Киркало Р. Системи прийняття рішень з нечіткою логікою. *Вісник Національного університету "Львівська політехніка"*. Львів. 2009. №650. С. 116–123.

Толюпа С. В., Одарченко Р. С., Пархоменко І. І., Даков С. Ю.

ВИЯВЛЕННЯ АТАК В КОРПОРАТИВНІЙ МЕРЕЖІ ЗА ДОПОМОГОЮ ПРАВИЛ НЕЧІТКОЇ ЛОГІКИ

Розглянуто задачу виявлення можливих атак на ресурси корпоративної мережі. Виконано аналіз підходів до виявлення порушень інформаційної безпеки з використанням теорії нечітких множин. Показано, що для підвищення ефективності виявлення ситуацій, пов'язаних з можливим вторгненням, необхідно використовувати сучасні технології інтелектуального аналізу з використанням правил і методів нечіткої логіки. Запропонована структурна схема нечіткої системи для виявлення аномального трафіку в сегменті

мережі. В експертній системі знання фахівців-експертів формалізуються у вигляді набору правил, що дозволяють приймати рішення в складних ситуаціях. Структуруванням знань експертів у вигляді бази знань займається аналітик (інженер по знаннях). Заснована на правилах експертна система складається з бази знань, механізму логічного висновку, блоку пояснення результатів і призначеного для користувача інтерфейсу. Для певної корпоративної мережі є характерні параметри трафіку, які можна визначити накопивши статистичну інформацію по поведінці мережі за довільний період роботи. Найбільш важливим моментом в процедурі синтезу нечіткої системи є підбір та складання правил, чи іншими словами, синтез таблиці лінгвістичних правил системи. Лінгвістичні правила системи евристично складаються розробником, який є добре проінформований про технологічні особливості об'єкта. При аналітичному описі об'єкта проводиться машинне моделювання розроблюваної системи з ітераційною корекцією лінгвістичних правил. При відсутності аналітичного опису об'єкта корегування правил проводилось безпосередньо після впровадження системи. Відповідно до схеми аналізатор здійснює діагностику та фільтрацію вхідних даних, фазифікатор переводить з числової в лінгвістичну форму відповідні дані. Класифікатор аналізує отриману вхідну інформацію визначає відповідну ситуацію, по якій в базі знань, активізуючи певні продукційні правила. Дефазифікатор переводить з лінгвістичної форми у цифрову і генерує відповідне правило.

Ключові слова: інформаційна безпека; вторгнення; корпоративна мережа; інтелектуальний аналіз даних; нечітка логіка; нечітка система.

Toliupa S., Odarchenko R., Parkhomenko I., Dakov S.

DETECTION OF ATTACKS IN THE CORPORATE NETWORK USING THE RULES OF FUZZY LOGIC

The problem of identifying possible attacks on corporate network resources is considered. An analysis of approaches to the detection of information security violations using fuzzy set theory is performed. It is shown that in order to increase the efficiency of detecting situations related to a possible invasion, it is necessary to use modern technologies of intellectual analysis using the rules and methods of fuzzy logic. A block diagram of a fuzzy system for detecting abnormal traffic in a network segment is proposed. In the expert system, the knowledge of experts is formalized in the form of a set of rules that allow you to make decisions in difficult situations. The analyst (knowledge engineer) is structuring the knowledge of experts in the form of a knowledge base. The rules-based expert system consists of a knowledge base, a inference mechanism, a result explanation unit and a user interface. For a particular corporate network there are characteristic traffic parameters that can be determined by accumulating statistical information on network behavior for any period of operation. The most important point in the procedure of fuzzy system synthesis is the selection and compilation of rules, or in other words, the synthesis of the table of linguistic rules of the system. The linguistic rules of the system are heuristically compiled by the developer, who is well informed about the technological features of the object. In the analytical description of the object, machine modeling of the developed system with iterative correction of linguistic rules is performed. In the absence of an analytical description of the object, the rules were adjusted immediately after the implementation of the system. According to the scheme, the analyzer performs diagnostics and filtering of the input data, the fafifier translates from numerical to linguistic form the corresponding data. The classifier analyzes the received input information determines the relevant situation in which the knowledge base, activating certain production rules. The defasifier translates from linguistic to digital form and generates a corresponding rule.

Keywords: information security; intrusion; corporate network; data mining; fuzzy logic; fuzzy system.

Толюпа С. В., Одарченко Р. С., Пархоменко І. І., Даков С. Ю.

ОБНАРУЖЕНИЕ АТАК В КОРПОРАТИВНОЙ СЕТИ С ПОМОЩЬЮ ПРАВИЛ НЕЧЕТКОЙ ЛОГИКИ

Рассмотрена задача выявления возможных атак на ресурсы корпоративной сети. Выполнен анализ подходов к выявлению нарушений информационной безопасности с использованием теории нечетких множеств. Показано, что для повышения эффективности выявления ситуаций, связанных с возможным вторжением, необходимо использовать современные технологии интеллектуального анализа с использованием правил и методов нечеткой логики. Предложенная структурная схема нечеткой системы для выявления аномального трафика в сегменте сети. В экспертной системе знания специалистов-экспертов формализуются в виде набора правил, позволяющих принимать решения в сложных ситуациях. Структурированием знаний экспертов в виде базы знаний занимается аналитик (инженер по знаниям). Основана на правилах экспертная система состоит из базы знаний, механизма логического вывода, блока объяснения результатов и пользовательского интерфейса. Для определенной корпоративной сети характерные параметры трафика, которые можно определить накопив статистическую информацию по поведению сети за произвольный период работы. Наиболее важным моментом в процедуре синтеза нечеткой системы является подбор и составление

правил, или другими словами, синтез таблицы лингвистических правил системы. Лингвистические правила системы эвристический состоят разработчиком, который хорошо проинформирован о технологических особенностях объекта. При аналитическом описании объекта проводится машинное моделирование разрабатываемой системы с итерационной коррекцией лингвистических правил. При отсутствии аналитического описания объекта корректировки правил проводилось непосредственно после внедрения системы. Согласно схеме анализатор осуществляет диагностику и фильтрацию входных данных, фазификатор переводит с числовой в лингвистическую форму соответствующие данные. Классификатор анализирует полученную входную информацию определяет соответствующую ситуацию, по которой в базе знаний, активизируя определенные продукционные правила. Дефазификатор переводит с лингвистической формы в цифровую и генерирует соответствующее правило.

Ключевые слова: информационная безопасность; вторжения; корпоративная сеть; интеллектуальный анализ данных; нечеткая логика; нечеткая система.

Стаття надійшла до редакції 16.11.2020 р.
Прийнято до друку 10.12.2020 р.