КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№1 (9), 2020
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2020.9.4558
УДК 004.275:004.7-049.5

Лахно Валерій Анатолійович
доктор технічних наук, професор, завідувач кафедри комп'ютерних систем і мереж НУБіП України
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID: 0000-0001-9695-4543
valss21@ukr.net

Гусєв Борис Семенович
кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем і мереж НУБіП України
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID: 0000-0003-1658-7822
gusevbs@nubip.edu.ua

Блозва Андрій Ігорович
кандидат педагогічних наук, доцент кафедри комп'ютерних систем і мереж НУБіП України
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID: 0000-0002-4377-0916
andriy.blozva@nubip.edu.ua

Касаткін Дмитро Юрійович
кандидат педагогічних наук, доцент, доцент кафедри комп'ютерних систем і мереж НУБіП України
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID: 0000-0002-2642-8908
d.kasatkin@nubip.edu.ua

Осипова Тетяна Юріївна
кандидат педагогічних наук, доцент, кафедри комп'ютерних систем і мереж НУБіП України
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID ID 0000-0002-9199-3436
t_osipova@nubip.edu.ua

КЛАСТЕРИЗАЦІЯ ОЗНАК МЕРЕЖЕВИХ АТАК В ЗАДАЧАХ АНАЛІЗУ ЗАХИЩЕНОСТІ ІНФОРМАЦІЇ

Анотація. У статті запропоновано алгоритм з елементами самонавчання для систем виявлення вторгнень, а також вдосконалена методика кластеризації, що фіксується системою даних, які стосуються подій інформаційної безпеки. Запропоновані підходи відрізняються від відомих, застосуванням ентропійного підходу, що дозволяє представляти дані як однорідні групи, причому кожна така група (або кластер) може відповідати заздалегідь заданим параметрам. Запропоновані рішення стосуються можливостей оцінювання динамічних залежностей між кластерами, що характеризують аналізовані класи вторгнень. В ході досліджень було встановлено, що в разі прояву нових ознак подій інформаційної безпеки, змінюється і відповідна шкала, що описує відстані між кластерами. Для перевірки працездатності та адекватності запропонованих рішень було проведено обчислювальний експеримент. В ході обчислювального експерименту встановлено, що покрокове обчислення параметрів інформативних ознак мережевих атак, дозволяє сформувати досить інформативні кластерні структури даних, що володіють характерними атрибутами. Ці атрибути в подальшому стають основою для бази знань інтелектуальних систем виявлення мережевих атак. Розраховані динамічні залежності між кластерами, дозволяють досить точно визначати множину подій інформаційної безпеки, які можуть стати вихідними даними для подальшої автоматичної оцінки ступеня поточних загроз, зафіксованих системами виявлення атак. Представлені в статті методика і алгоритм кластеризації ознак мережевих атак, на наш погляд, є більш простими для програмної реалізації, ніж існуючі аналоги.

Ключові слова: кібербезпека, об'єкт інформатизації, події інформаційної безпеки, ознаки, кластеризація, алгоритм.

1. ВСТУП

Зростаюча роль факторів інформаційної та кібербезпеки (далі ІБ і КрБ) в загальній системі національної безпеки (НБ) держав, що реалізують політику цифровізації економік, зафіксована у відповідних стратегіях і доктринах багатьох країн, наприклад, США, держав Європейського союзу та ін. [1]-[3]. Зростання кількості кіберзагроз, а також кількості і складності деструктивних, в тому числі цілеспрямованих впливів на інформаційно-телекомунікаційні засоби і інформаційні системи (далі по тексту об'єкти інформатизації - ОБІ) робить тренд на посилення політики ІБ і КрБ одним з пріоритетних у розвитку сучасного цифрового суспільства. Зауважимо, що політика протидії застосуванню потенціалу засобів і методів кібератак в сучасному високотехнологічному світі стає одним із пріоритетних завдань, які потребують вирішення для багатьох держав, що роблять акцент на розвиток інформаційних технологій (ІТ).

Керуючись парадигмою забезпечення ІБ і КрБ ОБІ, одним з ключових постає завдання щодо протистояння руйнівним комп'ютерним атакам. Останнє можна забезпечити, застосовуючи технічні засоби, в тому числі, засоби виявлення вторгнень (ЗВВ).

В даному контексті, однією з підзадач, що вирішуються ЗВВ для захисту ОБІ, є аналіз подій, пов'язаних з ІБ в ОБІ (далі ПІБ). В результаті такого аналізу повинні бути сформовані рекомендації і відкоригована політика ІБ. При цьому виникає природне протиріччя між великою кількістю реєстрованих мережевих подій в ЗВВ і необхідністю оперативно коригувати локальну політику ІБ для конкретного ОБІ. Моделі, які застосовуються в даний час, методи і засоби аналізу ПІБ, з використанням систем управління базами даних (СУБД), не завжди дають бажаний результат [4]. Особливо це очевидно при збільшенні масштабів ОБІ [5].

Багатьма компаніями [2], [3] і незалежними дослідниками [4], [5] ведуться роботи в сегменті ІБ і КрБ ОБІ, пов'язані з вивченням можливості розв’язання, зазначеної вище суперечності. І як один з варіантів вирішення, видається варіант застосування методів інтелектуального аналізу даних (МІАД), кластеризації і методів машинного навчання в задачах обробки великих масивів інформації в ЗВВ.

Постановка проблеми. Таким чином, ми вважаємо, що актуальність і тематика нашого дослідження, викликані необхідністю додаткового вивчення можливостей МІАД в ЗВВ, є актуальними. Особливо це стає очевидним, якщо розглядати завдання забезпечення ІБ і КрБ ОБІ в контексті розробки високоефективного інструментарію, що дозволяє аналізувати мережеві ЗВВ.

Аналіз останніх досліджень і публікацій. Дослідження, пов'язані із застосуванням МІАД в ЗВВ на думку деяких авторів досить перспективні. Особливо це завдання актуальне для зниження кількості помилкових спрацьовувань в ЗВВ.

В [6] описаний метод, який базується на традиційному байєсовському підході в поєднанні з бустінгом (підсиленням) [7]. Але, за зауваженням авторів, їм не вдалося подолати проблему бустінга (підсилення), пов'язану з необхідністю перенавчання.

У [8] розглянуто можливість комбінування МІАД з методами нечіткої логіки (НЛ), і асоціативних правил. Дослідження в цьому напрямі авторами ще не завершені. У [9] пропонувалося об'єднати переваги генетичних алгоритмів (ГА), НЛ і клітинних автоматів. В роботі [9] авторами відзначалося скорочення часу на навчання ЗВВ. Однак, не була вирішена проблема помилкових спрацьовувань. В [10] розглянуто задачу виявлення вторгнень для рідкісних класів атак. Авторами наголошується, що

пріоритетом в їх дослідженні є можливість застосування розробленого алгоритму в режимі реального часу. В [11] було запропоновано тренування ЗВВ реалізовувати на основі навчальних правил і методів кластеризації. Роботи в цьому напрямку ведуться. В [12] були розглянуті принципи використання НЛ і ГА в ЗВВ. Робота в цьому напрямку триває. У [13] досліджувалась можливість спільного застосування методів опорних векторів і дерев рішень в задачах виявлення атак. Автори прийшли до таких висновків: 1) метод дерев рішень більш кращий для виявлення атак класів Probe, U2R, R2L, однак метод опорних векторів краще працює в ситуаціях виявлення DoS/DDoS атак.

Однак, як показав аналіз розглянутих досліджень, в цілому розробленість проблематики застосування МІАД для ЗВВ залишається досить невисокою [14], [15]. Вищезазначене обумовлює релевантність обраної нами тематики дослідження, спрямованого на розвиток методологічного і алгоритмічного забезпечення при дослідженні статистики ПІБ.

Мета статті. Вдосконалення методів і моделей аналітичної обробки даних для систем виявлення вторгнень на об'єктах інформатизації.

Для досягнення мети досліджень необхідно вирішити такі завдання:

- розробити алгоритм і вдосконалити методику кластеризації даних з елементами самонавчання і з урахуванням моніторингу несуперечності часу фіксації конкретних подій ІБ;
- розробити методику, що дозволяє оцінювати рівні кіберзагроз для об'єкта інформатизації, керуючись експертизою і ретроспективним аналізом даних, пов'язаних з ПІБ.

2. ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Для аналізу даних, пов'язаних з ПІБ для ОБІ великого масштабу, дані необхідно розбити на однорідні групи - кластери [16]. В процесі кластеризації, як було показано в наших попередніх дослідженнях [16], [17], в якості критерію результативності навчання системи підтримки прийняття рішень в задачах детектування аномалій або атак, можна застосувати ентропійну міру Шеннона та інформаційно-дистанційний критерій Кульбака - Лейблера.

3. МЕТОДИКА ДОСЛІДЖЕННЯ

В якості критеріїв для обробки подій, пов'язаних з ІБ для ОБІ, виступають такі мінімальні величини: кількість елементів в однорідній (однотипній) групі даних (кластері); значення близькості елементів в групі; значення динамічної залежності, яка описує несуперечливість за часом фіксації ПІБ для різних кластерів.

Кожен кластер відповідає утвореній статистичній сутності. Мережеву активність можна уявити, як низку подій, зафіксованих в часі. Отже, можливо попарно порівнювати отримані кластери для визначення ступеня несуперечності з часом ПІБ, які увійшли в різні кластери. На наступному етапі оцінювання загроз для ІБ і КрБ різних ОБІ виконується експертиза і аналіз ретроспективних даних, що стосуються ПІБ. На заключному етапі виконується проектування попереднього досвіду і накопиченої інформації в базах знань, наприклад, експертної системи або СППР [18]-[21] для синтезу поточної оцінки, що характеризує рівень аналізованих загроз.

Як приклад розглянемо такі основні ознаки кіберзагроз для ОБІ: 1) номер сигнатури; 2) і 3) IP-адреси джерела і цілі атаки, відповідно; 4) і 5) номер порту джерела і атаки, відповідно.

Кластеризація виконувалася в кілька етапів:

Етап 1. Визначається найбільш інформативна ознака (далі - ІО:) для дослідження множини елементів:

$$IS_i = \frac{IS_i^1}{IS_i^2}, \quad (1)$$

де

$$IS_i^1 = \sum_{j \neq i} \left(\begin{array}{l} \sum_{n=1}^{s_i} p_{i(n)} \log_2 p_{i(n)} + \sum_{m=1}^{s_j} p_{j(m)} \log_2 p_{j(m)} - \\ - \sum_{n=1}^{s_i} \sum_{m=1}^{s_j} \left(p_{i(n),j(m)} \log_2 p_{i(n),j(m)} \right) \end{array} \right);$$

$$IS_i^2 = \sum_{j \neq i} \left(\sum_{n=1}^{s_i} \sum_{m=1}^{s_j} \left(p_{i(n),j(m)} \log_2 p_{i(n),j(m)} \right) \right);$$

i, j – номери ознак; $n = 1,..., s_i$, $m = 1,..., s_j$ – номери можливих значень для ознак i, j, відповідно; $p_{i1},..., p_{i(s_i)}$, $p_{j1},..., p_{j(s_j)}$ – ймовірності виникнення відповідних значень в ознаках i, j, відповідно; $p_{i(n),j(m)}$ – ймовірність одночасного виникнення в елементі значень з номерами n, m для ознак i, j, відповідно; $1 \le n \le s_i; 1 \le m \le s_j$.

Тоді будемо вважати, що ознака, яка володіє найвищим значенням, буде містити найбільшу кількість інформації. Відповідно, будемо вважати цей показник загрози, аномалії або атаки найбільш інформативним.

Етап 2. Визначається найбільш інформативне значення (далі - ІЗ) серед множини ІО:

$$IS_{i(n)} = \frac{IS_{i(n)}^1}{IS_{i(n)}^2}, \quad (2)$$

де $IS_{i(n)}^1 = \sum_{j \neq i} (A - B);$

$$A = \left(p_{i(n)} \log_2 p_{i(n)} + p_{\overline{i(n)}} \log_2 p_{\overline{i(n)}} + p_{i(n),j} \log_2 p_{i(n),j} + p_{\overline{i(n)},j} \log_2 p_{\overline{i(n)},j} \right);$$

$$B = \left(\begin{array}{l} \sum_{m=1}^{s_j^{i(n)}} \left(p_{i(n),j(m)} \log_2 p_{i(n),j(m)} \right) - \\ - \sum_{m=1}^{s_j^{i(n)}} \left(p_{\overline{i(n)},j(m)} \log_2 p_{\overline{i(n)},j(m)} \right) \end{array} \right) - p_{\overline{i(n)},j} \log_2 p_{\overline{i(n)},j};$$

$$IS_{i(in)}^2 = \sum_{j \neq i} \left(\begin{array}{l} \sum_{m=1}^{s_j^{i(n)}} \left(p_{i(n),j(m)} \log_2 p_{i(n),j(m)} \right) + \\ + \sum_{m=1}^{s_j^{i(n)}} \left(p_{\overline{i(n)},j(m)} \log_2 p_{\overline{i(n)},j(m)} \right) \end{array} \right) + p_{\overline{i(n)},j} \log_2 p_{\overline{i(n)},j};$$

$IS_{i(n)}$ – параметр, що визначає наскільки ознака є інформативною і відповідає значенню n; $p_{i(n)}$ – ймовірність виникнення елемента зі значенням n в ознаці i; $p_{i(n),j}$ – ймовірність виникнення елемента з довільним значенням для ознаки i, (вважаємо, що дана поява значення n вже хоча б один раз зафіксована раніше для i); $p_{\overline{i(n)},j}$ –

ймовірність виникнення елемента з довільним значенням для ознаки i, (вважаємо, що дана поява значення n жодного разу раніше не зафіксована для i); $p_{\overline{i(n)},j(m)}$ – ймовірність виникнення елемента, що має значення m для j, але відрізняється від значення m для i; $s_j^{i(n)}$ – число комбінацій для значень n ознак з номерами i, j.

Етап 3. Здійснюється вибірка елементів початкової множини, для яких величина ІО відповідає ІЗ.

Етап 4. Визначається параметр однорідності або однотипності для отриманих множин (ho):

$$ho = \frac{\sum_{j=1}^{k} z_{\max(j)}}{z \cdot k}, \tag{3}$$

де z – кількість елементів для досліджуваної множини; кількість ознак для конкретного елемента; $z_{\max(j)}$ – максимальна кількість елементів з ідентичними значеннями ознаки з номером j.

Етап 5. Перевіряється параметр однорідності. Якщо даний параметр менше порогової величини, то отриману множину переміщуємо в список початкових множин та ще раз проробляємо етапи 1-5.

Етап 6. Реалізується перевірка кількості елементів у множині. Якщо їх кількість менша граничної величини, то ці елементи вважаються статистично малозначущими. Малозначущі елементи переміщуємо в множину даних, які відносяться до некластеризованних.

Етап 7. Ті групи елементів, які підходять обом граничним значенням, і відповідають підсумковим кластерам для яких формується шаблон. Даний шаблон зберігається в базі даних (БД) або БЗ СППР або ЕС.

Далі визначається параметр, що характеризує несуперечливість інформації про час фіксації конкретних подій ІБ. Причому, цей параметр обчислюється попарно для всіх кластерів:

$$CO_{XY} = \frac{h_{XY}}{H_{XY}}, \tag{4}$$

де $h_{XY} = \sum_{i=1}^{z+1} p_i \cdot \log_2 p_i + \sum_{j=1}^{k+1} p_j \cdot \log_2 p_j - \sum_{i=1}^{z+k+1} p_r \cdot \log_2 p_r$ – перетин показників ентропії для двох ознак в кластері (див. рис. 1).

$H_{XY} = \sum_{r=1}^{z+k+1} p_r \cdot \log_2 p_r$ – ентропія для пари елементів в порівнюваних кластерах XY;

$p_i = dt_i / T, \quad p_j = dt_j / T, \quad p_r = dt_r / T$;

$dt_i\left(dt_j, dt_r\right)$ – тимчасові інтервали з номерами $(i(j,r))$, для яких зафіксовані ПІБ кластерів $(X(Y, XY))$;

T – часовий проміжок для якого виконується поточний аналіз; z, k – число ПІБ в кластерах X, Y, відповідно.

Можливі кілька варіантів настання ПІБ для кластерів: 1) події відбуваються синхронно, тоді $CO_{XY} = 1$; 2) події розділені максимальними інтервалами часу, тоді $CO_{XY} \to 0$; 3) проміжні значення для CO_{XY}.

ISSN 2663 - 4023

Крім розрахунку та відповідного аналізу показника ентропії для порівнюваних кластерів, можливий додатковий аналіз ПІБ. Подібний підхід може базуватися на ретроспективному аналізі ПІБ. При цьому поточні ПІБ порівнюються з тими, які вже були підлягали експертній оцінці. Тоді можлива більш деталізована оцінка ступеня реальних загроз для конкретного ОБІ, який є об'єктом захисту. Застосування запропонованого алгоритму та методики, є платформою подальшої розробки програмних продуктів для автоматизації аналізу досліджуваних ПІБ і оцінки ступеня їх небезпеки. Хоча, слід зазначити, що подібна оцінка буде носити імовірнісний характер. Це пов'язано з тим, що подібна оцінка є лише проекцією попереднього досвіду розслідування ПІБ і зіставлення його з поточними загрозами.

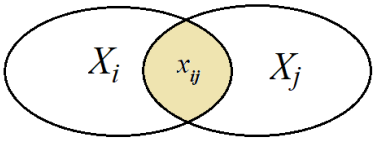


Рис. 1. Перетин показників ентропії для двох ознак в кластері

Оцінювати ступінь небезпеки поточних загроз і ПІБ за допомогою розробленого програмного продукту для аналізу подій пропонується на основі таких критеріїв: 1) ступеня дослідження ПІБ (Ψ); 2) ступеня несуперечності даних, отриманих в процесі досліджень ПІБ (Ω).

Критерії визначаються так:

$$\Psi = \begin{cases} 1, \quad if \quad Z_v \neq 0, Z_{\bar{v}} = 0; \\ \left(Z_k \cdot HO_{\bar{k}}\right) \Big/ \left(Z_s \cdot \left(HO_{\bar{k}} + HO_k\right)\right), \quad if \quad Z_k \neq 0, Z_{\bar{k}} \neq 0; \\ 0, \quad if \quad Z_k = 0, Z_{\bar{k}} \neq 0, \end{cases} \tag{5}$$

де $Z_s = Z_{\bar{k}} + Z_k$, $Z_{\bar{k}}$, Z_k – сумарне число, а також кількість немаркованих і маркованих елементів для шаблону кластера, який піддається аналізу, відповідно; $HO_{\bar{k}}, HO_k$ – показник однотипності немаркованих і маркованих елементів для шаблону кластера, який піддається аналізу.

$$\Omega = ER \cdot \left(1 + P_{k^+} \cdot \log_2 P_{k^+} + P_{k^-} \cdot \log_2 P_{k^-}\right), \tag{6}$$

де $P_{k^+} = z_{k^+} \big/ Z$, $P_{k^+} = z_{k^-} \big/ Z$ – ймовірності позитивного або негативного висновку, відповідно; z_{k^+}, z_{k^-} – кількості елементів з позитивною (негативною) оцінкою в процесі аналізу ПІБ; Z – загальна кількість елементів ПІБ, які оцінюються в ході експертизи ER:

$$ER = \begin{cases} 1, \quad if \quad z_{k^+} \geq z_{k^-}; \\ -1, \quad if \quad z_{k^+} < z_{k^-}. \end{cases} \tag{7}$$

Запропонований алгоритм, показаний на рис. 2.

ISSN 2663 - 4023

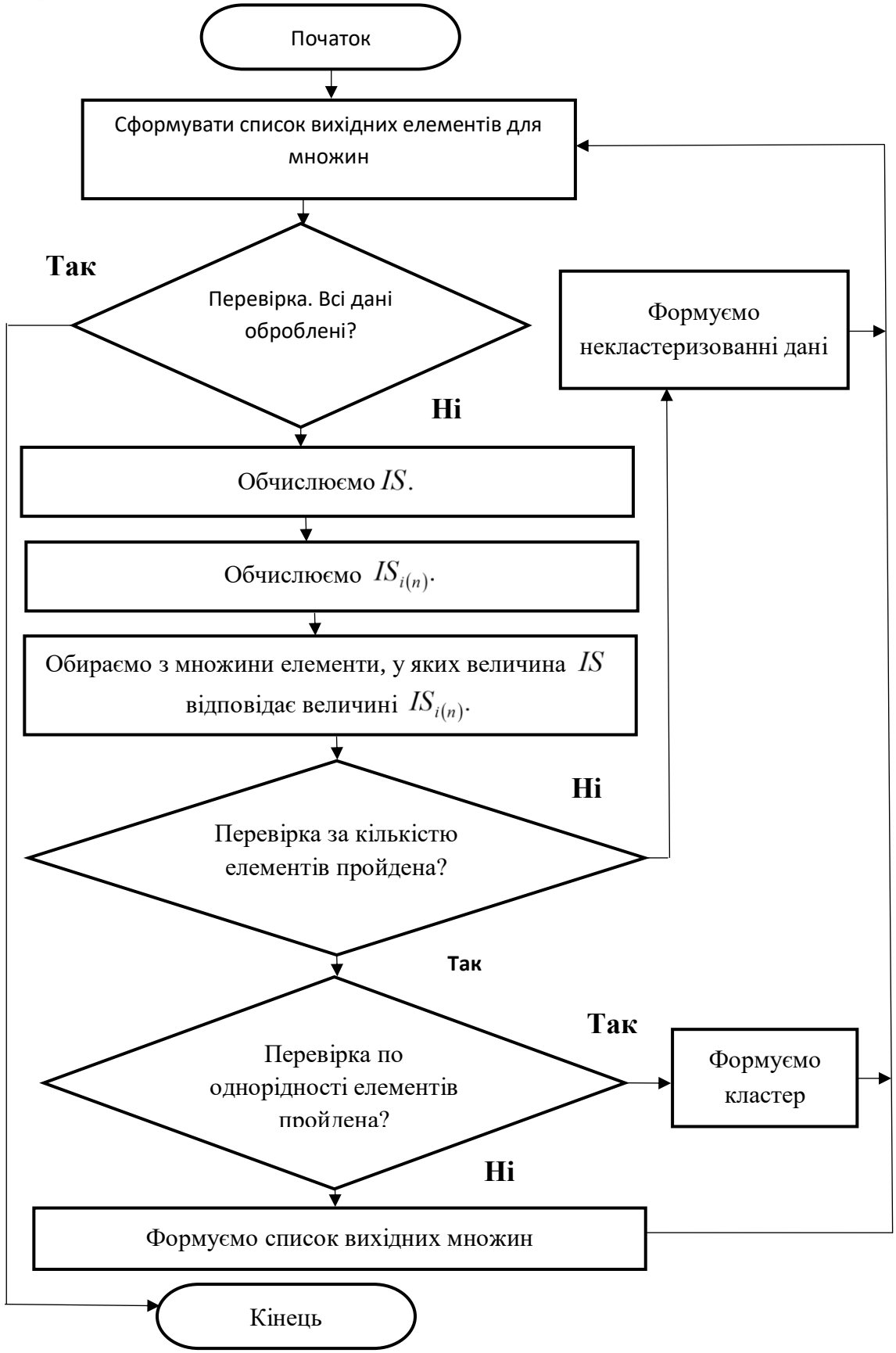


Рис. 2. Блок-схема алгоритму формування кластерів для аналізу ПІБ

Для того щоб підтвердити правильність запропонованих алгоритмів був виконаний обчислювальний експеримент.

Кількість вихідних елементів для реалізації алгоритму кластеризації - 350. В якості вихідних даних прийняті мережеві ПІБ, які фіксувалися за допомогою СОВ Snort [1], [4], [21]. IP-адреси генерувалися як випадкові дані.

Результати були визначені для таких параметрів:

$Z_{\min}$ – min число елементів в аналізованому класі;

$HO_{\min}$ – min порогова величина для показника однотипності елементів в аналізованому класі.

В ході обчислювального експерименту прийнято: $Z_{\min} = 10; \quad HO_{\min} = 80.$

Результати експерименту для досліджуваних ознак, показані в таблиці 1.

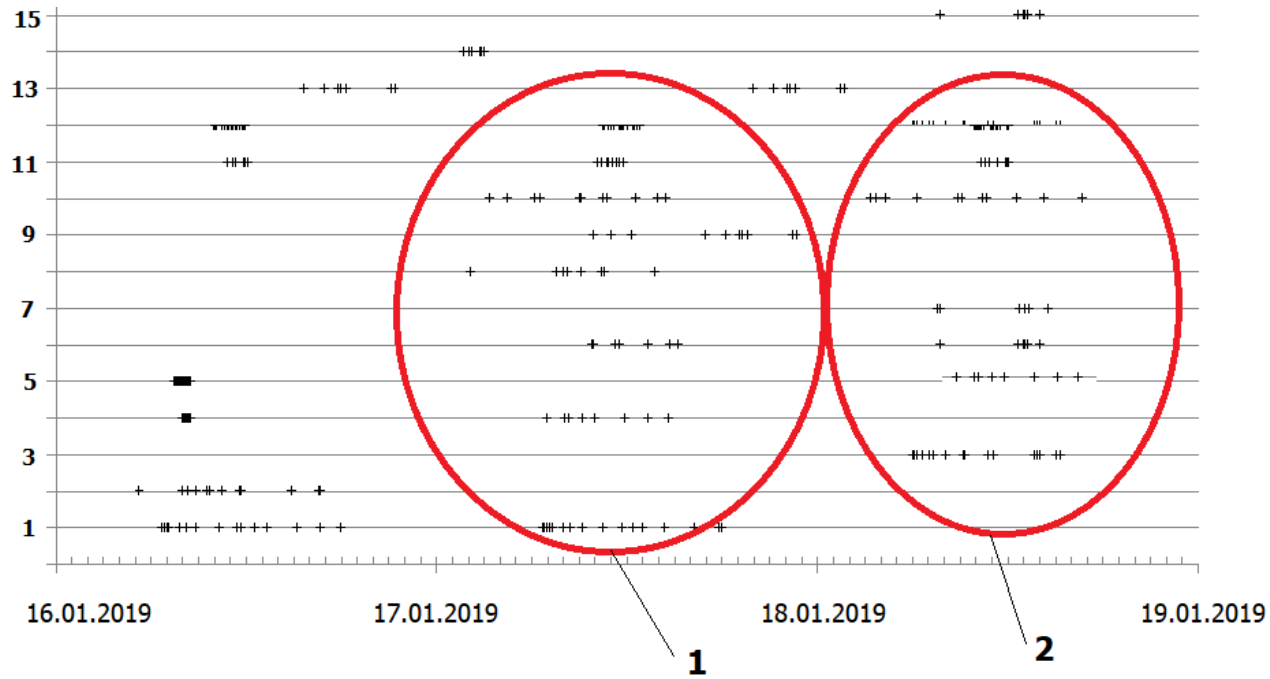
Таблиця 1

Кластеризація ознак ПІБ

Номер шаблону	**Аналізовані ознаки**					**Кількість елементів**
	Номер сигнатури	IP-адреси джерела	IP-адреси цілі атаки	номер порту джерела	номер порту цілі атаки	
	№1	№2	№3	№4	№5	
1.	101	36.254.254.80	-	62200	443	32
2.	152	-	195.213.95.121	62200	25	13
3.	195	205.102.45.132	-	53	62200	16
4.	106	202.205.12.28	-	-	62200	34
5.	117	207.219.38.34	180.209.188.171	80	62200	28
6.	115	-	200.231.16.154	80	62200	16
7.	115	-	200.231.16.179	80	62200	11
8.	148	-	40.101.36.112	80	62200	13
9.	162	-	197.67.214.141	80	62200	13
10.	175	201.23.211.208	-	80	62200	22
11.	186	-	198.81.213.31	8080	62200	29
12.	186	-	198.81.213.31	80	62200	65
13.	190	192.23.211.208	-	80	62200	21
14.	190	192.25.201.162	197.81.213.31	8080	62200	19
15.	192	-	200.202.16.179	80	62200	18

Графік, що відображає і узагальнює інформацію, подану в табличній формі (див. табл. 1), показаний на рис. 3. На графіку по осі ординат показані відповідні номери шаблонів для обчислених кластерів аналізованих ознак загроз для ОБІ. На осі абсцис вказані часові інтервали збору даних для подальшої кластеризації ознак і ретроспективного аналізу загроз для ОБІ.

ISSN 2663 - 4023



1, 2 – кластери ПІБ, і відповідні шаблони, які пов'язані з потенційними загрозами для ОБІ

Рис. 3. Кластеризація ПІБ і отримання шаблонів аналізу загроз для об'єкту інформатизації

Графік дозволяє оцінити зміни в отриманих кластерах динамічно. Події для аналізованих 15 кластерів, корелювалися за часом, що дало можливість робити висновок про ступінь загроз, пов'язаних з аналізованими ПІБ для конкретних ОБІ.

4. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Аналіз результатів обчислювального експерименту дозволяє сформулювати такі висновки:

- покрокове обчислення параметрів інформативних ознак і значень (ІО, ІЗ), а також однотипності множин (етапи 1-4 процедури кластеризації) для контрольної вибірки, дозволило сформувати досить інформативні кластерні структури даних, що володіють характерними атрибутами;
- розраховані динамічні залежності між кластерами, дозволяють досить точно визначати множину подій інформаційної безпеки, які можуть стати вихідними даними для подальшої автоматичної оцінки ступеня поточних загроз, зафіксованих ПІБ для ОБІ.

У порівнянні з результатами наших попередніх досліджень [17]-[20], а також близьких концептуально досліджень інших авторів [6], [8], [10], представлені в даній роботі методика та алгоритм кластеризації ознак ПІБ, на наш погляд, є більш простими для програмної реалізації. До того ж, всі наведені в статті математичні залежності базуються на неухильних математичних висновках.

На поточному етапі досліджень ми обмежилися порівняно невеликою вибіркою вихідних даних для обчислювального експерименту, що є певним недоліком роботи.

Тому логічним продовженням досліджень в даному напрямку стане збільшення кількості первинних даних для експерименту і розробка на основі запропонованих моделей та алгоритмів програмних продуктів для автоматизованого аналізу ПІБ.

Робота виконана в рамках грантового фінансування проекту AP05132723 «Розробка адаптивних експертних систем в області кібербезпеки критично важливих об'єктів інформатизації».

5. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Розроблено алгоритм з елементами самонавчання ЗВВ і вдосконалена методика кластеризації, фіксованих системою даних, які стосуються подій інформаційної безпеки, що відрізняються від відомих, застосуванням ентропійного підходу, що дозволяє представляти дані як однорідні групи, причому кожна така група (або кластер) може відповідати заздалегідь заданим параметрам.

Розроблено алгоритм і методика оцінювання динамічних залежностей між кластерами, що характеризують аналізовані класи вторгнень, і відрізняються від відомих моделлю, яка характеризує ступінь несуперечності в часі подій, пов'язаних з ІБ. При цьому встановлено, що в разі прояву нових ознак ПІБ, змінюється відповідна шкала, що описує відстані між кластерами.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Johanson, D. (2013). The evolving US cybersecurity doctrine, Security Index: *A Russian Journal on International Security*, 19(4), 37–50.

[2] Harknett, R. J., & Stever, J. A. (2011). The new policy world of cybersecurity, *Public Administration Review*, 71(3), 455–460.

[3] Newmeyer, K. P. (2015). Elements of national cybersecurity strategy for developing nations, *National Cybersecurity Institute Journal*, 1(3), 9–19.

[4] Bass, T. (2000). Intrusion detection systems and multisensor data fusion. *Communications of the ACM*, 43(4), 99-105. https://doi.org/10.1145 / 332051.332079

[5] Lakhno, V., Kazmirchuk, S., Kovalenko, Y., Myrutenko, L., & Zhmurko, T. (2016). Design of adaptive system of detection of cyber-attacks, based on the model of logical procedures and the coverage matrices of features, *Eastern-European Journal of Enterprise Technologies*, (3 (9)), pp. 30–38. https://doi.org/10.15587/1729-4061.2016.71769

[6] Rahman, C. M., Farid, D. M., & Rahman, M. Z. (2011). Adaptive intrusion detection based on boosting and naïve Bayesian classifier, Vol., 24, No.3, pp. 12-19. https://doi.org/10.5120/2932-3883

[7] Jyothsna, V. V. R. P. V., Prasad, V. R., & Prasad, K. M. (2011). A review of anomaly based intrusion detection systems, *International Journal of Computer Applications*, 28(7), 26-35.

[8] Harshna, N. K. (2014). Fuzzy Data Mining Based Intrusion Detection System Using Genetic Algorithm, *International Journal of Advanced Research in Computer and Communication Engineering*, 3(1), 5021–5028.

[9] Sree, P. K., & Babu, I. R. (2008, December). Investigating Cellular Automata Based Network Intrusion Detection System for Fixed Networks (NIDWCA), *In Advanced Computer Theory and Engineering*, 2008. ICACTE'08. International Conference on (pp. 153-156). IEEE.

[10] Dokas, P., Ertoz, L., Kumar, V., Lazarevic, A., Srivastava, J., & Tan, P. N. (2002, November). Data mining for network intrusion detection. In Proc. NSF Workshop on Next Generation Data Mining (pp. 21–30).

[11] Chan, P. K., Mahoney, M. V., & Arshad, M. H. (2005). Learning rules and clusters for anomaly detection in network traffic. In *Managing Cyber Threats* (pp. 81-99). Springer, Boston, MA.

[12] Borgohain, R. (2012). Fugeids: Fuzzy genetic paradigms in intrusion detection systems. arXiv preprint arXiv:1204.6416.

[13] Peddabachigari, S., Abraham, A., & Thomas, J. (2004). Intrusion detection systems using decision trees and support vector machines. *International Journal of Applied Science and Computations*, USA, 11(3), 118-134.

[14] Pan, S., Morris, T., & Adhikari, U. (2015). Developing a hybrid intrusion detection system using data mining for power systems. *IEEE Transactions on Smart Grid*, 6(6), pp. 3104–3113. https://doi.org/10.1109/TSG.2015.2409775

[15] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2), pp. 1153–1176.

[16] Lakhno, V., Tkach, Y., Petrenko, T., Zaitsev, S., & Bazylevych, V. (2016). Development of adaptive expert system of information security using a procedure of clustering the attributes of anomalies and cyber attacks. *Eastern-European Journal of Enterprise Technologies*, (6 (9)), pp. 32–44. https://doi.org/10.15587/1729-4061.2016.85600

[17] Lakhno, V.A., Kravchuk, P. U., Pleskach, V. L., etc. (2017). Applying the functional effectiveness information index in cybersecurity adaptive expert system of information and communication transport systems, *Journal of Theoretical and Applied Information Technology*, Vol. 95, No 8, pp. 1705–1714.

[18] Akhmetov, B., Lakhno, V., Akhmetov, B., & Alimseitova, Z. (2018, September). Development of Sectoral Intellectualized Expert Systems and Decision Making Support Systems in Cybersecurity. In *Proceedings of the Computational Methods in Systems and Software* (pp. 162–171). Springer, Cham.

[19] Akhmetov B., Kydyralina, L., etc. (2018). Model for a computer decision support system on mutual investment in the cybersecurity of educational institutions, *International journal of mechanical engineering & technology* (IJMET), Vol. 9, Iss. 10, pp. 1114–1122.

[20] Lakhno V.A., Petrov, A.S., Petrov, A.A. (2017). Development of a support system for managing the cyber security of information and communication environment of transport, *Advances in Intelligent Systems and Computing* / Information Systems Architecture and Technology: Proceedings of 38th International Conference on Information Systems Architecture and Technology – ISAT 2017 Part II Editors: Świątek, Jerzy, Borzemski, Leszek, Wilimowska, Zofia (Eds.) , pp. 113–127.

[21] Akhmetov, B.B. etc. (2017). Designing a decision support system for the weakly formalized problems in the provision of cybersecurity, *Eastern-European Journal of Enterprise Technologies/ Information and controlling system*, Vol. 1/2, Iss. 85, pp. 4–15.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (9), 2020
ISSN 2663 - 4023

Valerii A. Lakhno
Dr. Tech. Sc., Professor, Head of the Department of Computer System and Networks
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID: 0000-0001-9695-4543
valss21@ukr.net

Borys S. Husiev
Ph.D, Docent, Associate Professor at the Department of Computer System and Networks
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID: 0000-0003-1658-7822
gusevbs@nubip.edu.ua

Andrii I. Blozva
Ph.D., Associate Professor at the Department of Computer System and Networks
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID: 0000-0002-4377-0916
andriy.blozva@nubip.edu.ua

Dmytro Y. Kasatkin
Ph.D., Docent, Associate Professor at the Department of Computer System and Networks
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID: 0000-0002-2642-8908
d.kasatkin@nubip.edu.ua

Tetiana Y. Osypova
Ph.D, Associate Professor at the Department of Computer System and Networks
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID: 0000-0002-9199-3436
t_osipova@nubip.edu.ua

CLUSTERING NETWORK ATTACK FEATURES IN INFORMATION SECURITY ANALYSIS TASKS

Abstract. The paper proposes an algorithm with self-learning elements for intrusion detection systems, as well as an improved clustering technique which is recorded by the data system concerning information security events. The proposed approaches differ from those known using an entropy approach allowing data to be presented as homogeneous groups, moreover, each such group (or cluster) may correspond to predetermined parameters. The proposed solutions relate to the possibilities of assessing dynamic dependencies between clusters characterizing the analysed classes of invasions. The studies have found that in case of manifestation of new signs of information security events, the corresponding scale changes and describes the distances between clusters. A computational experiment was conducted to verify the operability and adequacy of the proposed solutions. During the computational experiment, it has been found that step-by-step calculation of parameters of informative characteristics of network attacks allows to form sufficiently informative cluster structures of data having characteristic attributes. These attributes further become the basis for the knowledge base of intelligent network attack detection systems. Dynamic dependencies between clusters are calculated allowing for a sufficiently accurate definition of the many information security events that can become the source data for further automatic assessment of current threats extent detected by attack detection systems. The methodology and algorithm presented in the paper for clustering the signs of network attacks, in our opinion it is simpler for software implementation than existing analogues.

Keywords: cybersecurity, informatization object, information security events, signs, clustering, algorithm.

REFERENCES

[1] Johanson, D. (2013). The evolving US cybersecurity doctrine, Security Index: *A Russian Journal on International Security*, 19(4), 37–50.
[2] Harknett, R. J., & Stever, J. A. (2011). The new policy world of cybersecurity, *Public Administration Review*, 71(3), 455–460.
[3] Newmeyer, K. P. (2015). Elements of national cybersecurity strategy for developing nations, *National Cybersecurity Institute Journal*, 1(3), 9–19.
[4] Bass, T. (2000). Intrusion detection systems and multisensor data fusion. *Communications of the ACM*, 43(4), 99-105. https://doi.org/10.1145 / 332051.332079
[5] Lakhno, V., Kazmirchuk, S., Kovalenko, Y., Myrutenko, L., & Zhmurko, T. (2016). Design of adaptive system of detection of cyber-attacks, based on the model of logical procedures and the coverage matrices of features, *Eastern-European Journal of Enterprise Technologies*, (3 (9)), pp. 30–38. https://doi.org/10.15587/1729-4061.2016.71769
[6] Rahman, C. M., Farid, D. M., & Rahman, M. Z. (2011). Adaptive intrusion detection based on boosting and naïve Bayesian classifier, Vol., 24, No.3, pp. 12-19. https://doi.org/10.5120/2932-3883
[7] Jyothsna, V. V. R. P. V., Prasad, V. R., & Prasad, K. M. (2011). A review of anomaly based intrusion detection systems, *International Journal of Computer Applications*, 28(7), 26-35.
[8] Harshna, N. K. (2014). Fuzzy Data Mining Based Intrusion Detection System Using Genetic Algorithm, *International Journal of Advanced Research in Computer and Communication Engineering*, 3(1), 5021–5028.
[9] Sree, P. K., & Babu, I. R. (2008, December). Investigating Cellular Automata Based Network Intrusion Detection System for Fixed Networks (NIDWCA), *In Advanced Computer Theory and Engineering*, 2008. ICACTE'08. International Conference on (pp. 153-156). IEEE.
[10] Dokas, P., Ertoz, L., Kumar, V., Lazarevic, A., Srivastava, J., & Tan, P. N. (2002, November). Data mining for network intrusion detection. In Proc. NSF Workshop on Next Generation Data Mining (pp. 21–30).
[11] Chan, P. K., Mahoney, M. V., & Arshad, M. H. (2005). Learning rules and clusters for anomaly detection in network traffic. In *Managing Cyber Threats* (pp. 81-99). Springer, Boston, MA.
[12] Borgohain, R. (2012). Fugeids: Fuzzy genetic paradigms in intrusion detection systems. arXiv preprint arXiv:1204.6416.
[13] Peddabachigari, S., Abraham, A., & Thomas, J. (2004). Intrusion detection systems using decision trees and support vector machines. *International Journal of Applied Science and Computations*, USA, 11(3), 118-134.
[14] Pan, S., Morris, T., & Adhikari, U. (2015). Developing a hybrid intrusion detection system using data mining for power systems. *IEEE Transactions on Smart Grid*, 6(6), pp. 3104–3113. https://doi.org/10.1109/TSG.2015.2409775
[15] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2), pp. 1153–1176.
[16] Lakhno, V., Tkach, Y., Petrenko, T., Zaitsev, S., & Bazylevych, V. (2016). Development of adaptive expert system of information security using a procedure of clustering the attributes of anomalies and cyber attacks. *Eastern-European Journal of Enterprise Technologies*, (6 (9)), pp. 32–44. https://doi.org/10.15587/1729-4061.2016.85600
[17] Lakhno, V.A., Kravchuk, P. U., Pleskach, V. L., etc. (2017). Applying the functional effectiveness information index in cybersecurity adaptive expert system of information and communication transport systems, *Journal of Theoretical and Applied Information Technology*, Vol. 95, No 8, pp. 1705–1714.
[18] Akhmetov, B., Lakhno, V., Akhmetov, B., & Alimseitova, Z. (2018, September). Development of Sectoral Intellectualized Expert Systems and Decision Making Support Systems in Cybersecurity. In *Proceedings of the Computational Methods in Systems and Software* (pp. 162–171). Springer, Cham.
[19] Akhmetov B., Kydyralina, L., etc. (2018). Model for a computer decision support system on mutual investment in the cybersecurity of educational institutions, *International journal of mechanical engineering & technology* (IJMET), Vol. 9, Iss. 10, pp. 1114–1122.
[20] Lakhno V.A., Petrov, A.S., Petrov, A.A. (2017). Development of a support system for managing the cyber security of information and communication environment of transport, *Advances in Intelligent Systems and Computing* / Information Systems Architecture and Technology: Proceedings of 38th International Conference on Information Systems Architecture and Technology – ISAT 2017 Part II Editors: Świątek, Jerzy, Borzemski, Leszek, Wilimowska, Zofia (Eds.) , pp. 113–127.

[21] Akhmetov, B.B. etc. (2017). Designing a decision support system for the weakly formalized problems in the provision of cybersecurity, *Eastern-European Journal of Enterprise Technologies/ Information and controlling system*, Vol. 1/2, Iss. 85, pp. 4–15.