INFORMATION SECURITY RISK MANAGEMENT

DOI 10.20535/2411-1031.2023.11.1.283575
UDC 004.056:323.1

IHOR SUBACH,
ARTEM MYKYTIUK

METHODOLOGY OF FORMATION OF FUZZY ASSOCIATIVE RULES WITH WEIGHTED ATTRIBUTES FROM SIEM DATABASE FOR DETECTION OF CYBER INCIDENTS IN SPECIAL INFORMATION AND COMMUNICATION SYSTEMS

The article presents the method of forming associative rules from the database of the SIEM system for detecting cyber incidents, which is based on the theory of fuzzy sets and methods of data mining. On the basis of the conducted analysis, a conclusion was made about the expediency of detecting cyber incidents in special information and communication systems (SICS) by applying rule-oriented methods. The necessity of applying data mining technologies, in particular, methods of forming associative rules to supplement the knowledge base (KB) of the SIEM system with the aim of improving its characteristics in the process of detecting cyber incidents, is substantiated. For the effective application of cyber incident detection models built on the basis of the theory of fuzzy sets, the use of fuzzy associative rule search methods is proposed, which allow processing heterogeneous data about cyber incidents and are transparent for perception. The mathematical apparatus for forming fuzzy associative rules is considered and examples of its application are given. In order to increase the effectiveness of the methods of searching for fuzzy associative rules from the database of the SIEM it is proposed to use weighting coefficients of attributes that characterize the degree of manifestation of their importance in the fuzzy rule. A formal formulation of the problem of forming fuzzy associative rules with weighted attributes and which are used for the identification of cyber incidents is given. A scheme of their formation and application for identification of cyber incidents is proposed. The method of forming fuzzy associative rules with weighted attributes from the database of the SIEM is given. The problem of determining the weighting coefficients of the relative importance of SIEM system DB attributes is formulated and a method for its solution is proposed. The formulation of the problem of finding sets of elements that have a weighted fuzzy support of at least the given one and are used to form fuzzy associative rules with weighted attributes is given. Methods for its solution are proposed.

Key words: cyber protection, cyber incident, SIEM, theory of fuzzy sets, data mining, associative rules.

Statement of the problem. A significant increase in the number of cyber attacks requires new efforts to create effective cyber protection systems for special information and communication systems (SICS).

The basis of the modern cyber protection system is the SIEM system, and the main tasks it must solve [1] are the following: collection of data on cyber incidents from a number of disparate distributed sources, their processing and analysis; real-time or near-real-time detection of cyber incidents and security policy violations; assistance in the investigation of cyber incidents; formation of effective solutions regarding cyber protection of SICS; formation of reporting documents, visualization of system status and other.

The most widespread methods used in SIEM systems to detect (identify) cyber incidents that occur during the operation of SICS are rule-oriented methods, which are based on the mechanism of logical deduction based on production rules found in the knowledge base (KB) of the intelligent SIEM system.

Analysis of the latest studies and publications. The models of fuzzy identification of cyber incidents by IDS/IPS systems that occur in SICS are proposed in [2]. However, it is a well-known fact that the effectiveness of the functioning of any intelligent system is primarily determined by the power of its KB.

A large number of publications [3] - [10] are devoted to the issue of filling (uploading) production rules into the KB of cyber protection systems (SIEM, IDS, etc.). However, they are mainly based on the use of expert knowledge, which has significant shortcomings:

– human psychology is such that in order to confirm their point of view experts often select “convenient” data and underestimate “undesirable” data, leaving it without necessary attention;

– if a set of factors is very large, there is a need to analyze a large number of situations, but an expert cannot qualitatively predict changes in the state of a non-linear object, which is influenced by many factors with known dynamic characteristics;

– unmanageability of physiological processes, for example, forgetting leads to the limited use of expert’s basic information resource – accumulated experience, etc.

On the other hand, modern information technologies such as Data Mining (DM), Big Data (BD), Machine Learning (ML) and other have been actively used to solve the issues.

It is the use of the above technologies that allows to make the systems of detection of cyber incidents more flexible and adaptive in the process of rapidly changing conditions of their functioning.

Currently, there are a large number of DM methods used in cyber protection systems [11] - [13], but according to the cyber incident identification models proposed in [2], [16], in order to form a KB about cyber incidents, it is quite appropriate to apply the methods of finding fuzzy associative rules that allow to process heterogeneous data and are transparent to perceive [2], [4], [10], [14], [18] - [21].

Presentation of the main material of the study. Let the database (DB) of *DBI* cyber incidents, information about which is accumulated in the SIEM system, consists of records that include attributes $O = \{o_1, o_2, \ldots, o_n\}$, where $o_j, j = \overline{1,n}$ – a numeric or categorical attribute that contains a sign of a cyber incident. For any entry $t \in DBI$, $t\ [o_j]$ has some value *v*, that takes an attribute o_j in entry *t*.

Table 1 – A fragment of the SIEM system DB on cyber incidents [15]

TID	*Boot_level*, %	*Request, n*	…	SICS cyber incident class
1	10	1	…	Normal condition
2	85	3	…	Normal condition
3	65	30	…	*JS (HTML)/ScrInject*
4	69	35	…	*JS (HTML)/ScrInject*
5	89	40	…	*JS (HTML)/ScrInject*

Then each numeric attribute $o_j \in O$ can be given on the domain $dom(o_j) = [\underline{v}, \bar{v}] \subseteq R$.

Accordingly, $l_{o_j k}, k = \overline{1, K_{o_j}}$ – a set of linguistic terms that relate to the numeric attribute $o_j \in O$ where each linguistic term $l_{o_j k}$ can be represented by a fuzzy set $L_{o_j k}$, which is given on the domain $dom(o_j)$ with the membership function $\mu_{L_{o_j k}}(v) : dom(o_j) \to [0,1]$.

The degree of membership of any value *v* to some linguistic term $l_{o_j k}$ is characterized by the membership function $\mu_{L_{o_j k}}(v)$.

In accordance with the data contained in Table 1, Fig. 1 shows the examples of setting the sets of linguistic terms $l_{number\ of\ requests} = \{\text{small (S), medium (M), large (L)}\}$ for numeric attribute o_j – “number of requests to Internet resources for updates” and $l_{boot\ level} = \{\text{low (L), below average (bA), average (A), above average (aA), high (H)}\}$ for numeric attribute o_p – “OS workload” to identify the cyber incident associated with *JS (HTML)/ScrInject* cyber attack [15].

Accordingly, for any categorical attribute $o_j \in O$, which is defined on the domain $dom(o_j) = \{v_1, v_2, ..., v_m\}$, a set of linguistic terms $l_{o_jk}, k = \overline{1, K_{o_j}}$ is set, where each linguistic l_{o_jk} is represented by a fuzzy set L_{o_jk}.

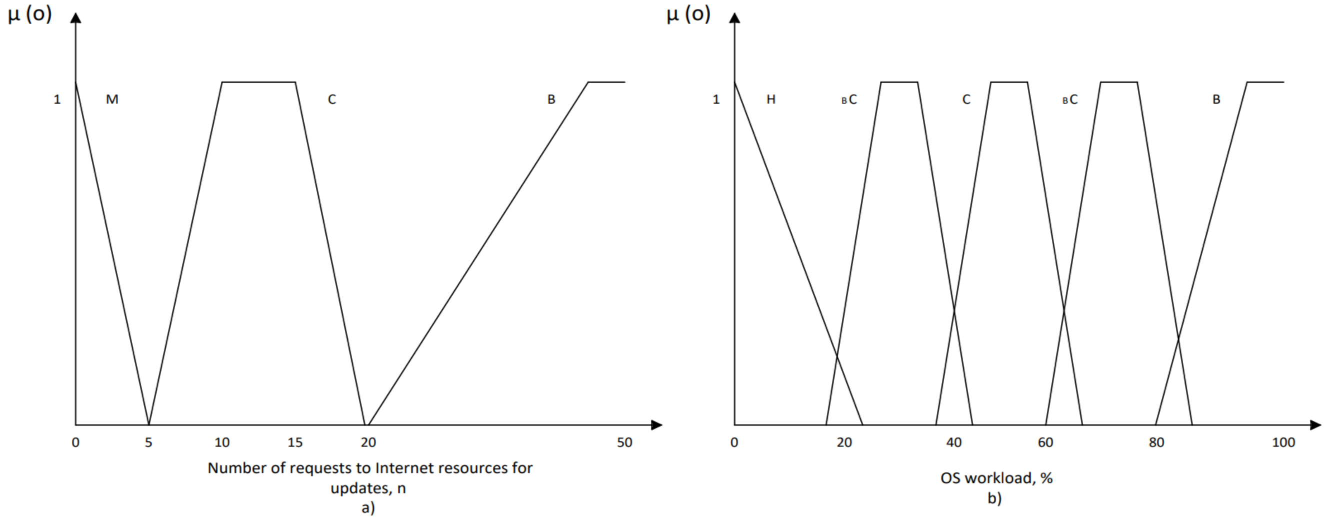


Figure 1 – Linguistic terms for: a) attribute Request – “number of requests to Internet resources for updates”; b) attribute Boot_level – “OS workload”

So, the set of attributes O DBI DB is represented using a set of linguistic terms: $l = \{l_{o_jk} \mid j = \overline{1, n},\ k = \overline{1, K_{o_j}}\}$, where K_{o_j} – the number of linguistic attribute terms o_j, and the set of linguistic terms, in turn, is presented with the help of fuzzy sets: $L = \{L_{o_jk} \mid j = \overline{1, n},\ k = \overline{1, K_{o_j}}\}$.

An example of representing a SIEM system DB attributes by linguistic terms in accordance with the data contained Table 1 and shown in Fig. 1. The example is given in Table 2.

Table 2 – An example of representing the attributes of a SIEM system DB containing a sign of a cyber incident using a set of linguistic terms

TID	*Boot_Level*	*Request*	…	*Normal*	*JS (HTML)/ScrInject*
1	L	S	…	H	L
2	H	S	…	H	L
3	aA	L	…	L	H
4	aA	L	…	L	H
5	H	L	…	L	H

Table 3 – An example of representing the attributes of a SIEM system DB containing a cyber incident sign by membership functions

TID	*Boot_Level*	*Request*	…	*Normal*	*JS (HTML)/ScrInject*
1	0.75	0.95	…	1.0	0.0
2	0.7	0.80	…	1.0	0.0
3	0.65	0.65	…	0.0	1.0
4	1.0	0.70	…	0.0	1.0
5	0.99	0.95	…	0.0	1.0

Then a fuzzy associative rule takes the form [18]:

$$\left\langle X=\left\{ o_1,o_2,...,o_p \right\}, A=\left\{ l_{o_1k_{i_1}}, l_{o_2k_{i_2}},...,l_{o_pk_{i_p}} \right\}\right\rangle \Rightarrow \left\langle Y=\left\{ o_{p+1},o_{p+2},...,o_n \right\}, B=\left\{ l_{o_{p+1}q_{o_{p+1}}}, l_{o_{P+2}q_{o_{p+2}}},...,l_{o_nq_{o_n}} \right\}\right\rangle, \quad (1)$$

where sets of elements $X \subset O, Y \subset O, X \bigcap Y = \varnothing$;
$l_{o_jk_{o_j}} \in l, l_{o_pq_{o_p}} \in l, j \neq p$ – linguistic terms that are given on fuzzy sets $L_{o_jk_{o_j}} \in L$ and $L_{o_pq_{o_p}} \in L$ correspondingly.

Examples of fuzzy associative rules, according to the data given in Tables 1-3:
<*Boot_level* = aA>, <*Request* = L> ⇒<*JS (HTML)/ScrInject* = H>
<*Boot_level* = L>, <*Request* = S> ⇒<*Normal* = H>

These rules can be interpreted as follows:

If the OS workload is above average and the number of accesses to Internet resources is large, then a cyber incident related to JS (HTML)/ScrInject cyber attack occurs.

If the OS workload is not high and the number of accesses to Internet resources is small, then SICS is functioning in the normal mode.

The fuzzy support of a set of elements <*X*, *A*> is denoted as $FS_{<X,A>}$ and is calculated as follows [18]:

$$FS_{<X,A>} = \frac{\sum\limits_{t_h \in D^F} \prod\limits_{o_j \in X} \mu_{L_{o_j} \in A}(t_h[o_j])}{|D^F|}, \quad (2)$$

where D^F – transformed SIEM system DB on cyber incidents, which is obtained from the original one using fuzzy sets specified by the cybersecurity analyst $L = \{L_{o_jk} \mid j=\overline{1,n},\ k=\overline{1,K_{o_j}}\}$.

For the data given in Tables 1-3, the fuzzy support for element sets will be as follows:
<*Boot_level*,L> =0.75/5=0.2; <*Boot_level*,aA> = (0.65+1.0)/5 = 0.33;
<*Boot_level*,H> = (0.7+0.99)/5 = 0.34; <*Request*,S> = (0.95+0.80)/5=0.35;
<*Request*,L> = (0.65+0.70+0.95)/5=0.46; <*JS(HTML)/ScrInject*,H> =(1.0+1.0+1.0)/5=0.6;…
{<*Boot_level*,L>, <*Request*,*S*>} = (0.75·0.95)/5 = 0.14;
{<*Boot_level*,H>, <*Request*,S>} = (0.70·0.80)/5 = 0.11; {<*Boot_level*,aA>, <*Request*,L>} = ((0.65·0.65)+(1.0·0.70))/5 = 0.23; {<*Boot_level*,H>, <*Request*,L>} = (0.99·0.95)/5=0.19;…
{<*Boot_level*,aA>, <*Request*,L>,<*JS(HTML)/ScrInject*,H>} = ((0.65·0.65·1.0)+(1.0·0.70·1.0))/5 = 0.23;…

The fuzzy probability of rule (1) is denoted by $FC_{<<X,A>,<Y,B>>}$ and is calculated as follows:

$$FC_{<<X,A>,<Y,B>>} = \frac{\sum_{t_h \in D^F} \prod_{o_j \in Z} \mu_{L_{o_j} \in C}(t_h[o_j])}{\sum_{t_h \in D^F} \prod_{o_j \in X} \mu_{L_{o_j} \in A}(t_h[o_j])}, \tag{3}$$

where $Z = X \bigcup Y,\ C = A \bigcup B,\ X \bigcap Y = \varnothing,\ A \bigcap B = \varnothing.$

For the data given in Tables 1–3, the fuzzy probability of rules may be as follows:
<{*Boot_level* =aA>, <*Request*=L}> ⇒ <{*JS (HTML)/ScrInject* = H}> = ((0.65·0.65·1.0)+(1.0·0.70 ·1.0))/((0.65·0.65)+(1.0·0.70))=0.23/0.23=1;
<{*Boot_level*,L>, <*Request,S*}>⇒<{*Normal* = H}> = ((0.75·0.95·1.0)/ (0.75·0.95)) = 1.0.

Obviously, both rules are interesting due to the fact that they have a fuzzy probability of 1.0 or 100%.

The weighted fuzzy support of the set of elements <*X,A*> denoted as $WFS_{<X,A>}$ and is calculated as follows [18]:

$$WFS_{<X,A>} = \frac{\sum_{t_h \in D^F} \prod_{o_j \in X} \mu_{L_{o_j} \in A}((t_h[o_j]) \cdot \omega(o_j.l_{o_jk}))}{| D^F |}, \tag{4}$$

where $\omega(o_j.l_{o_jk})$ – the weighted factor that characterizes the relative importance of the pair $< o_j.l_{o_jk} >$, $o_j \in I$, $j = \overline{1,n}$, $l_{o_jk} \in l$, $k = \overline{1,K_{o_j}}$ and is a degree of manifestation of the integral property “importance of the attribute”.

Table 4 – An example of representing the attributes of a SIEM system DB containing a cyber incident sign by membership functions and weighted factors

TID	*Boot_Level/w*	*Request/w*	…	*Normal/w*	*JS (HTML)/ScrInject/w*
1	0.75/0.6	0.95/0.8	…	1.0/0.5	0.0/1.0
2	0.7/0.6	0.80/0.8	…	1.0/0.5	0.0/1.0
3	0.65/0.6	0.65/0.8	…	0.0/0.5	1.0/1.0
4	1.0/0.6	0.70/0.8	…	0.0/0.5	1.0/1.0
5	0.99/0.6	0.95/0.8	…	0.0/0.5	1.0/1.0

For the data given in Table 4, the weighted fuzzy support of the set of elements will be as follows:
<*Boot_level*,L> =0.75·0.6/5=0.1; <*Boot_level*,aA> = (0.65·0.6+1.0·0.6)/5 = 0.2;
<*Boot_level*,H> = (0.7·0.6+0.99·0.6)/5 = 0.2; <*Request*,S> = (0.95·0.8+0.80·0.8)/5=0.28;
<*Request*,L> = (0.65·0.8+0.70·0.8+0.95·0.8)/5=0.37; <*JS(HTML)/ScrInject*,H> =(1.0·1.0+1.0·1.0+ 1.0·1.0)/5=0.6;…{<*Boot_level*,L>, <*Request,S*>} = ((0.75·0.6)·(0.95·0.8))/5 = 0.07;
{<*Boot_level*,H>, <*Request*,S>} = ((0.70·0.6)·(0.80·0.8))/5 = 0.05; {<*Boot_level*,aA>, <*Request*,L>} = ((0.65·0.6)·(0.65·0.8))+((1.0·0.6)·(0.70·0.8))/5 = 0.29; {<*Boot_level*,H>, <*Request*,L>} = ((0.99·0.6)·(0.95·0.8))/5=0.27;…
{<*Boot_level*,aA>, <*Request*,L>,<*JS(HTML)/ScrInject*,H>} = (((0.65·0.6)·(0.70·0.8)·(1.0·1.0))+((1.0·0.6)·(0.70·0.8)·(1.0·1.0)))/5 = 0.11;…

The weighted fuzzy probability of rule (1) is denoted as $WFC_{<<X,A>,<Y,B>>}$ and is calculated as follows [18]:

$$FC_{<<X,A>,<Y,B>} = \frac{\sum\limits_{t_h \in D^F} \prod\limits_{o_j \in Z} \mu_{L_{o_j} \in C}((t_h[o_j]) \cdot \omega(o_j.l_{o_jk}))}{\sum\limits_{t_h \in D^F} \prod\limits_{o_j \in X} \mu_{L_{o_j} \in A}((t_h[o_j]) \cdot \omega(o_j.l_{o_jk}))}, \qquad (5)$$

where $Z = X \bigcup Y,\ C = A \bigcup B,\ X \bigcap Y = \varnothing,\ A \bigcap B = \varnothing.$

For the data given in Table 4, the weighted fuzzy probability of rules may be as follows:
<{*Boot_level* =aA>, <*Request*=L}> ⇒ <{*JS (HTML)/ScrInject* = H}> = (((0.65·0.6)·(0.65·0.8)·(1. 0·1.0))+((1.0·0.6)(0.70·0.8)(1.0·1.0))/((0.65·0.6)·(0.65·0.8))+((1.0·0.6)·(0.70·08)))=0.54/0.54=1;
<{*Boot_level*,L>, <*Request,S*}> ⇒<{*Normal* = H}> = ((0.75·0.6)·(0.95·0.8)·(1.0·0.5))/ (0.75·0.6)·(0.95·0.8)) = 0.5.

Analysis of the obtained values of fuzzy and weighted fuzzy support and probability for the given rules shows that in the case of the second rule, its weighted fuzzy probability, due to the introduced weight for the normal condition of the SICS, equal to 0.5, has significantly decreased, which clearly makes this rule uninteresting.

A set of elements $< X, A >$ is considered a frequent element set if its weighted fuzzy support is greater than or equal to the minimum support value specified by the security analyst:

$$WFS_{\langle X,A,\omega \rangle} \geq FS_{\min}, \qquad (6)$$

where $FS_{\min}$ – minimum support value specified by the security analyst.

Let us assume that a fuzzy associative rule in the form (1) is promising (interesting) if $\langle\langle X, A\rangle, \langle Y, B\rangle\rangle$ is a frequent set of elements, and the weighted fuzzy probability $WFC_{\langle\langle X,A\rangle,\langle Y,B\rangle\rangle}$ of the rule is greater than or equal to the minimum probability value specified by the security analyst:

$$WFC_{\langle\langle X,A\rangle,\langle Y,B\rangle\rangle} \geq FC_{\min}, \qquad (7)$$

where $FC_{\min}$ – minimum probability value specified by the user.

Statement of the problem of searching for fuzzy association rules. On this basis, the problem of forming fuzzy associative rules containing weighted attributes and which are used to identify cyber incidents can be formulated as follows.

Given: The *SIEM* system *D* DB on cyber incidents containing relational tables (relationships) defined by the schema

$R(O_1, O_2, \ldots, O_m)$, where *R* – name of the table;

$O_1, O_2, \ldots, O_m$ – names of the attributes included in it and containing the values of signs of cyber incidents;

$D_1, D_2, ..., D_m$ – domains on which attribute values are set $O_1, O_2, \ldots, O_m$;

O_{V_k} – set of values

$$O_k \times D_k \,|\, k = \overline{1, m};$$

pair $\langle x_k, v_{k_i} \rangle \in O_{V_k}$ denotes an attribute x_k, corresponding to the value v_{k_i}, set on the domain

$$D_k : \langle x_k, v_{k_i} \rangle \in O_k \times D_k \mid k = \overline{1, m};\ i = \overline{1, n}.$$

Required: to find all fuzzy association rules of the form (1) with weighted attributes, for which the fuzzy support and fuzzy probability must be no less than the limit values pre-specified by the security analyst, which are called the minimum support $FS_{\min}$ (6) and the minimum probability $FC_{\min}$ (7):

$$M = \left\{ \langle X, A \rangle \Rightarrow \langle Y, B \rangle \ \middle|\, WFC_{(\langle X,A\rangle,\langle Y,B\rangle)} \geq FC_{\min}, WFS_{\langle X,A \rangle} \geq FS_{\min} \right\} \qquad (8)$$

Limitations and assumptions: we believe that the SIEM system DB containing data on cyber incidents is built on the basis of a relational data model.

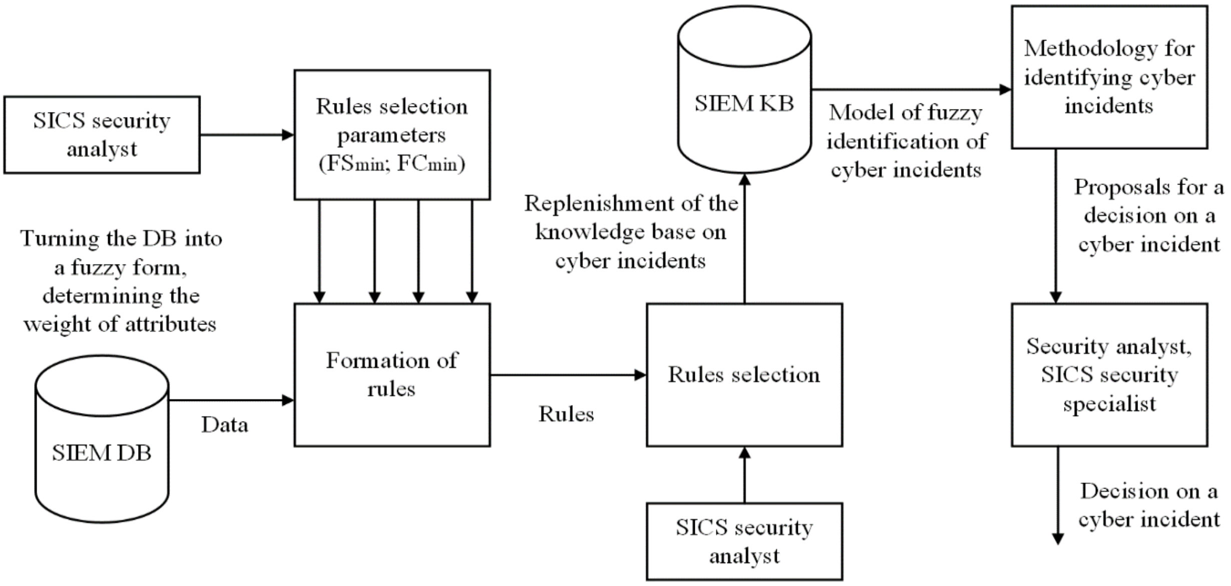
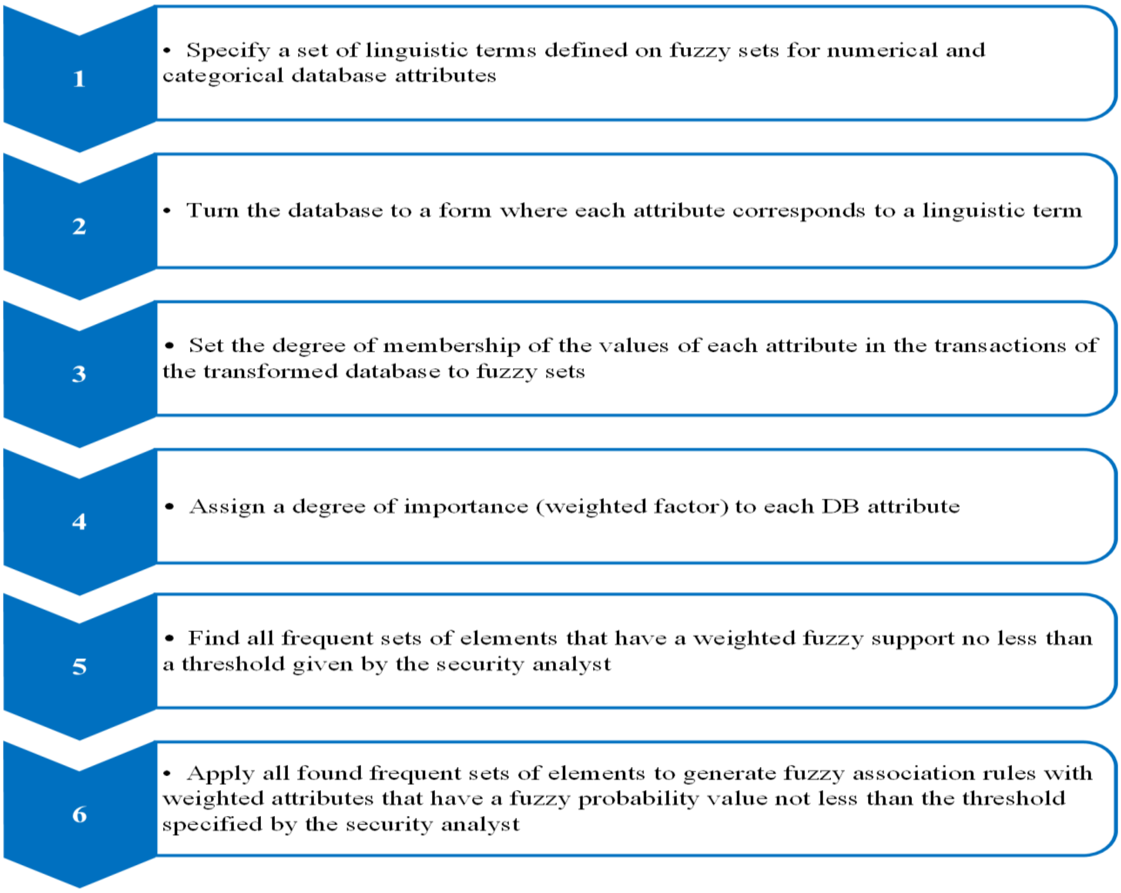

SICS security analyst
Rules selection parameters (FSmin; FCmin)
SIEM KB
Model of fuzzy identification of cyber incidents
Methodology for identifying cyber incidents
Turning the DB into a fuzzy form, determining the weight of attributes
Replenishment of the knowledge base on cyber incidents
Proposals for a decision on a cyber incident
Formation of rules
Rules selection
Rules
Security analyst, SICS security specialist
SIEM DB
Data
SICS security analyst
Decision on a cyber incident

Figure 2 – Scheme for the formation and application of fuzzy associative rules for cyber incidents identification

The general scheme of the formation of fuzzy associative rules for cyber incidents identification by the SIEM system, which occur during the operation of the SICS, is shown in Fig. 2.

Methodology of formation of fuzzy associative rules with weighted attributes. Thus, to solve problem (7), it is necessary to perform the following sequence of actions (Fig. 3).


1 • Specify a set of linguistic terms defined on fuzzy sets for numerical and categorical database attributes

2 • Turn the database to a form where each attribute corresponds to a linguistic term

3 • Set the degree of membership of the values of each attribute in the transactions of the transformed database to fuzzy sets

4 • Assign a degree of importance (weighted factor) to each DB attribute

5 • Find all frequent sets of elements that have a weighted fuzzy support no less than a threshold given by the security analyst

6 • Apply all found frequent sets of elements to generate fuzzy association rules with weighted attributes that have a fuzzy probability value not less than the threshold specified by the security analyst

Figure 3 – The methodology of formation of fuzzy associative rules with weighted attributes for cyber incidents identification by the SIEM system

At the first step, for each numerical or categorical attribute $o_j \in I$, a set of linguistic terms is specified $l_{o_j k_{i_j}} \in l$, which are defined on fuzzy sets $L_{o_j k_{i_j}} \in L$ with a membership function $\mu_{L_{o_j k_{i_j}}}(t[o_j])$.

At the second step, transform the DB scheme $D = \{t_1, t_2, ..., t_H\}$, where $t_h, h = \overline{1, H}$ – transactions consisting of attributes $O = \{o_1, o_2, ..., o_n\}$, into the form of D^F, where each transaction $t_h^F, h = \overline{1, H}$ includes attributes $\left\langle o_j \cdot l_{o_j k_{o_j}} \right\rangle \in O^F$, $k_{o_j} = \overline{1, K_{o_j}}$, where K_{o_j} – number of linguistic terms of the attribute o_j, and has the following form:

$$O^T = \left\{ \left\{ \left\langle o_1. \cdot l_{o_1 1} \right\rangle, \left\langle o_1 \cdot l_{o_1 2} \right\rangle, ..., \left\langle o_1 \cdot l_{o_1 K_{o_1}} \right\rangle \right\}, \left\{ \left\langle o_2 \cdot l_{o_2 1} \right\rangle, \left\langle o_2 \cdot l_{o_2 2} \right\rangle, ..., \right. \right.$$
$$\left. \left. \left\langle o_2 \cdot l_{o_2 K_{o_2}} \right\rangle \right\}, ..., \left\{ \left\langle o_n \cdot l_{o_n 1} \right\rangle, \left\langle o_n \cdot l_{o_n 2} \right\rangle, ..., \left\langle o_n \cdot l_{o_n K_{o_n}} \right\rangle \right\} \right\}. \tag{9}$$

At the next step, the degree of membership of the value v of each attribute o_j of the record $t \in D$ to fuzzy sets $L_{o_j k_{i_{oj}}} \in L$ is established:

$$\mu_{L_{o_j}}(t[o_j]) = \sum_{k_{o_j}=1}^{K_{o_j}} \mu_{L_{o_j k_{o_j}}}(t[o_j]), \tag{10}$$

and assign the resulting values to the attributes. $\left\langle o_j . l_{o_j k_{o_j}} \right\rangle$ of the records $t_h^T, h = \overline{1, H}$ of database D^F .

At the fourth step, each attribute of the transformed database is assigned a weighted factor characterizing the degree of importance of the attribute in the transaction.

The problem of determining the weighted factors of the relative importance of the SIEM system DB attributes is formulated as follows.

Given: a database containing transactions as follows:

$$O^T = \left\{ \left\{ \left\langle o_1 . l_{o_1 1} \right\rangle, \left\langle o_1 . l_{o_2 2} \right\rangle, ..., \left\langle o_1 . l_{o_1 K_{o_1}} \right\rangle \right\}, \left\{ \left\langle o_2 . l_{o_2 1} \right\rangle, \left\langle o_2 . l_{o_2 2} \right\rangle, ..., \right. \right.$$
$$\left. \left. \left\langle o_2 . l_{o_2 K_{o_2}} \right\rangle \right\}, ..., \left\{ \left\langle o_n . l_{o_n 1} \right\rangle, \left\langle o_n . l_{o_n 2} \right\rangle, ..., \left\langle o_n . l_{o_n K_{o_n}} \right\rangle \right\} \right\}, \tag{11}$$

where $\left\langle o_j . l_{o_j k_{o_j}} \right\rangle \in O^T$, $k_{o_j} = \overline{1, K_{o_j}}$, where K_{o_j} – the number of linguistic terms of the attribute, o_j and in the general $K_{o_1} \neq K_{o_2} \neq ... \neq K_{o_n}$.

Required: for each pair of attributes $o_j . l_{o_j k_{i_j}}, o_p . l_{o_p k_{o_p}} \in O^T$; $j, p = \overline{1, n}$; $l_{o_j k_{o_j}}, l_{o_p k_{o_p}} \in l$; $k_{o_j} = \overline{1, K_{o_j}}; k_{o_p} = \overline{1, K_{o_p}}$ determine the degree of preference $a_{o_j k_{o_j}, o_p k_{o_p}}; j \neq p$ and build a membership function to the fuzzy set “important attribute”, characterizing the “intensity of importance” and representing the condition for the dominance of the DB D^F attributes on the set O^T .

Analysis of the task and the nature of the activities of Security Operational Center officials shows that a method of the following type can be applied to solve it:

$$\left\langle d_1, p_3, e_2, v_1, N, L \right\rangle, \tag{12}$$

where d_1 – individual survey method;

p_3 – data collection procedure – comparison of pairs of objects;

e_2 – typology of expert information – paired or sequential comparison;

v_1 – the nature of the comparison or evaluation – the comparison is carried out according to the degree of severity of a certain characteristic;

N – interpretation of expert information – considering the results of comparison or evaluation as deterministic;

L – simplicity and ease of implementation.

The specified requirements are met by a method based on the idea of distributing degrees of membership of elements of a universal set according to their ranks or a method of pairwise comparisons based on rank assessments [8].

At the fifth step, all sets of elements that have weighted fuzzy support $WFS_{\langle X,A\rangle}$ no less than a user-specified threshold FS_{min} are found:

$$\forall\langle X,A\rangle : WFS_{\langle X,A\rangle} \geq FS_{min},$$

where $WFS_{\langle X,A\rangle}$ – the amount of fuzzy support of a set of elements <*X*,*A*>.

The formal statement of this problem is as follows:

Given: D^F – transformed DB with attributes representing characteristics and classes of cyber incidents, and which, in turn, are represented by linguistic terms, and their values are represented by membership functions;

FIS_k – a set of frequently occurring k-element sets;

CIS_k – a set of k-element candidate sets;

O^T – finite set of elements of the transformed DB D^F;

ASE_k – an auxiliary set of *k*-element sets;

FS_{min} – a threshold of minimum fuzzy support of a set of elements;

Required: find all frequent sets of elements for which the weighted fuzzy support value is not less than the threshold value: $WFS_{\langle X,A,\omega\rangle} \geq FS_{\min}$.

The solution to this problem can be implemented by applying one of the methods given in [9] - [12].

At the last step, all found frequent data elements from the set FIS_k are used to generate fuzzy association rules with weighted attributes of the form (1), having a weighted fuzzy probability value $WFC_{\langle\langle X,A\rangle,\langle Y,B\rangle\rangle}$ not less than the user-specified threshold FC_{min}:

$$\forall\langle X,A\rangle \Rightarrow \langle Y,B\rangle : WFC_{\langle\langle X,A\rangle,\langle Y,B\rangle\rangle} \geq FC_{min},$$

where $WFC_{\langle\langle X,A\rangle,\langle Y,B\rangle\rangle}$ – the value of the weighted fuzzy probability of the rule $\langle\langle X,A\rangle \Rightarrow \langle Y,B\rangle\rangle$.

Conclusions. The application of the method for the formation of fuzzy association rules with weighted attributes from the SIEM system DB for detecting cyber incidents that occur during the operation of the SICS proposed in the paper allows to increase the efficiency of the use of SIEM systems in practice by improving their characteristics such as adaptability to new types of cyber incidents and the accuracy of their detection.

REFERENCES

[1] I. Subach, A. Mykytiuk, and V. Kubrak, “Architecture and functional model of a perspective proactive intellectual SIEM for cyber protection of objects of critical infrastructure”, *Information technology and security*, vol. 7, no. 2, pp. 208-215, 2019, doi: https://doi.org/10.20535/2411-1031.2019.7.2.190570.

[2] I. Subach, and V. Fesokha, “Model of detection of anomalies in information and telecommunication networks of military management bodies on the basis of fuzzy sets and fuzzy logic output”, *Collection scientific works MITI*, vol. 3, pp. 158-164, 2017.

[3] I. Subach, B. Gerasimov, and E. Nikiforov, “Models of knowledge delivery for use in deision support systems”, *Scientific and technical information*, vol. 1, pp. 7-11, 2005.

[4] F. S. Tsai, “Network intrusion detectionusing association rules”, *International journal of recent trends in engineering*, vol. 2, no. 2, pp. 202-204, 2009.

[5] Y. Wang, I. Kim, G. Mbateng, and S.-Y. Ho, “A latent class modeling approach to detect network intrusion”, *Comput. Commun.*, vol. 30, no. 1, pp. 93-100, 2006, doi: https://doi.org/10.1016/j.comcom.2006.07.018.

[6] R. Shanmugavadivu, and N. Nagarajan, “Network intrusion detection system using fuzzy logic”, *Indian journal of computer science and engineering*, vol.2, no.1, pp. 101-111, 2011.

[7] N. Naidu, and D. R. V. Dharaskar, “An Effective Approach to Network Intrusion Detection System using Genetic Algorithm”, *Int. J. Comput. Appl.*, vol. 1, no. 3, pp. 26-32, 2010, doi: https://doi.org/10.5120/89-188.

[8] D. M. Farid, and M. Z. Rahman, “Anomaly network intrusion detection based on improved self adaptive bayesian algorithm”, *J. Comput.*, vol. 5, no. 1, 2010, doi: https://doi.org/10.4304/jcp.5.1.23-31.

[9] S.-J. Horng *et al.*, “A novel intrusion detection system based on hierarchical clustering and support vector machines”, *Expert Syst. with Appl.*, vol. 38, no. 1, pp. 306-313, 2011, doi: https://doi.org/10.1016/j.eswa.2010.06.066.

[10] H. Kabamba, “An evolution strategy approach toward ruleset generation for network intrusion detection systems (IDS)”, *International Journal of Soft Computing and Engineering*, vol. 2, iss. 5, pp. 1-5, 2012.

[11] I. Subach, V. Fesokha, and N. Fesokha, “Analysis of existing solutions for preventing invasion in information and telecommunication networks”, *Information technology and security*, vol. 5, no. 1, pp. 29-41, 2017, doi: https://doi.org/10.20535/2411-1031.2017.5.1.120554.

[12] T. Lappas, and K. Pelechrinis, “Data mining techniques for (network) intrusion detection systems”, 2007. [Online]. Available: https://www.researchgate.net/publication/228745997_Data_Mining_Techniques_for_Network_Intrusion_Detection_Systems. Accessed on: Feb. 21, 2023.

[13] A. Youssef, and A. Emam, “Network intrusion detection using data mining and network behaviour analysis”, *Int. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 6, pp. 87-98, 2011, doi: https://doi.org/10.5121/ijcsit.2011.3607.

[14] I. Subach, and V. Fesokha, “Model of detecting cybernetic attacks on information-telecommunication systems based on description of anomalies in their work by weighed fuzzy rules”, *Information technology and security*, vol. 5, no. 2, pp. 145-152, 2017, doi: https://doi.org/10.20535/2411-1031.2017.5.2.136984.

[15] I. Subach, Y. Zdorenko, and V. Fesokha, “Method for detecting cyber-attacks of the JS (HTML) / ScrInject type based on the use of the mathematical apparatus of the theory of fuzzy sets”, *Collection scientific works MITI*, vol. 4, pp. 125-131, 2018.

[16] I. Subach, A. Mykytiuk, S. Korotaev, and V. Kubrak, “Rule-oriented method of cyber incidents detection by SIEM based on fuzzy logical inference”, *Inf. technol. security, CEUR.*, vol. 2859, pp. 210-219, 2020, doi: https://doi.org/10.5281/zenodo.7123656.

[17] A. Rothstein, *Medical diagnostics on fuzzy logic*. Vinnytsia, Ukraine: Continent-PRIM, 1996.

[18] A. Gyenesei, “A fuzzy approach for mining quantitative association rules”, *Acta Cybernetica*, vol. 15, pp. 305-320, 2001.

[19] A. Gyenesei, “Fuzzy partitioning of quantitative attribute domains by a cluster goodness index”, *TUCS Technical Reports*, no. 368, Oct. 2000. [Online]. Available: https://www.researchgate.net/publication/2359339_Fuzzy_Partitioning_of_Quantitative_Attribute_Domains_by_a_Cluster_Goodness_Index. Accessed on: Feb. 13, 2023.

[20] Wai-Ho Au, and K. C. C. Chan, “An effective algorithm for discovering fuzzy rules in relational databases”, in *Proc. 1998 IEEE Int. Conf. Fuzzy Syst. IEEE World Congr. Comput. Intell.*, Anchorage, AK, USA, doi: https://doi.org/10.1109/fuzzy.1998.686309.

[21] K. C. C. Chan, and W.-H. Au, "Mining fuzzy association rules in a database containing relational and transactional data", in *Data Mining and Computational Intelligence*. Heidelberg, Germany: Physica-Verlag HD, 2001, pp. 95–114, doi: https://doi.org/10.1007/978-3-7908-1825-3_4.

The article was received 26.04.2023.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] І. Субач, А. Микитюк, та В. Кубрак, "Архітектура та функціональна модель перспективної проактивної інтелектуальної SIEM-системи для кіберзахисту об'єктів критичної інфраструктури", *Information technology and security,* vol. 7, no. 2, pp. 208-215, 2019, doi: https://doi.org/10.20535/2411-1031.2019.7.2.190570.

[2] І. Субач, та В. Фесоха, "Модель виявлення аномалій в інформаційно-телекомунікаційних мережах органів військового управління на основі нечітких множин та нечіткої логіки виводу", *Збірник наукових праць ВІТІ*, т. 3, с. 158-164, 2017.

[3] І. Субач, Б. Герасимов, та Є Нікіфоров, "Моделі надання знань для використання в системах підтримки прийняття рішень", *Науково-технічна інформація*, т. 1, с. 7-11, 2005.

[4] F. S. Tsai, "Network intrusion detectionusing association rules", *International journal of recent trends in engineering*, vol. 2, no. 2, pp. 202-204, 2009.

[5] Y. Wang, I. Kim, G. Mbateng, and S.-Y. Ho, "A latent class modeling approach to detect network intrusion", *Comput. Commun.*, vol. 30, no. 1, pp. 93-100, 2006, doi: https://doi.org/ 10.1016/j.comcom.2006.07.018.

[6] R. Shanmugavadivu, and N. Nagarajan, "Network intrusion detection system using fuzzy logic", *Indian journal of computer science and engineering*, vol.2, no.1, pp. 101-111, 2011.

[7] N. Naidu, and D. R. V. Dharaskar, "An Effective Approach to Network Intrusion Detection System using Genetic Algorithm", *Int. J. Comput. Appl.*, vol. 1, no. 3, pp. 26–32, 2010. doi: https://doi.org/10.5120/89-188.

[8] D. M. Farid, and M. Z. Rahman, "Anomaly network intrusion detection based on improved self adaptive bayesian algorithm", *J. Comput.*, vol. 5, no. 1, 2010, doi: https://doi.org/10.4304/ jcp.5.1.23-31.

[9] S.-J. Horng *et al.*, "A novel intrusion detection system based on hierarchical clustering and support vector machines", *Expert Syst. with Appl.*, vol. 38, no. 1, pp. 306-313, 2011, doi: https://doi.org/10.1016/j.eswa.2010.06.066.

[10] H. Kabamba, "An evolution strategy approach toward ruleset generation for network intrusion detection systems (IDS)", *International Journal of Soft Computing and Engineering*, vol. 2, iss. 5, pp. 1-5, 2012.

[11] І. Субач, В. Фесоха, та Н. Фесоха, "Аналіз існуючих рішень запобігання вторгненням в інформаційно-телекомунікаційні мережі", *Information technology and security*, vol. 5, no. 1, pp. 29-41, 2017, doi: https://doi.org/10.20535/2411-1031.2017.5.1.120554.

[12] T. Lappas, and K. Pelechrinis, "Data mining techniques for (network) intrusion detection systems", 2007. [Online]. Available: https://www.researchgate.net/publication/228745997_ Data_Mining_Techniques_for_Network_Intrusion_Detection_Systems. Accessed on: Feb. 21, 2023.

[13] A. Youssef, and A. Emam, "Network intrusion detection using data mining and network behaviour analysis", *Int. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 6, pp. 87-98, 2011, doi: https://doi.org/10.5121/ijcsit.2011.3607.

[14] І. Субач, та В. Фесоха, "Модель виявлення кібернетичних атак на інформаційно-телекомунікаційні системи на основі описання анамалій їх роботи зваженими нечіткими правилами", *Information technology and security*, vol. 5, no. 2, pp. 145-152, 2017, doi: https://doi.org/10.20535/2411-1031.2017.5.2.136984.

[15] І. Субач, Ю. Здоренко, та В. Фесоха, “Методика виявлення кібератак типу JS (HTML) / Scrinject на основі застосування математичного апарату теорії нечітких множин”, *Збірник наукових праць ВІТІ*, т. 4, с. 125-131, 2018.

[16] I. Subach, A. Mykytiuk, S. Korotaev, and V. Kubrak, “Rule-oriented method of cyber incidents detection by SIEM based on fuzzy logical inference”, *Inf. technol. security, CEUR.*, vol. 2859, pp. 210-219, 2020, doi: https://doi.org/10.5281/zenodo.7123656.

[17] A. Rothstein, *Medical diagnostics on fuzzy logic*. Vinnytsia, Ukraine: Continent-PRIM, 1996.

[18] A. Gyenesei, “A fuzzy approach for mining quantitative association rules”, *Acta Cybernetica*, vol. 15, pp. 305-320, 2001.

[19] A. Gyenesei, “Fuzzy partitioning of quantitative attribute domains by a cluster goodness index”, *TUCS Technical Reports*, no. 368, Oct. 2000. [Online]. Available: https://www.researchgate.net/publication/2359339_Fuzzy_Partitioning_of_Quantitative_Attribute_Domains_by_a_Cluster_Goodness_Index. Accessed on: Feb 13, 2023.

[20] Wai-Ho Au, and K. C. C. Chan, “An effective algorithm for discovering fuzzy rules in relational databases”, in *Proc. 1998 IEEE Int. Conf. Fuzzy Syst. IEEE World Congr. Comput. Intell.*, Anchorage, AK, USA, doi: https://doi.org/10.1109/fuzzy.1998.686309.

[21] K. C. C. Chan, and W.-H. Au, “Mining fuzzy association rules in a database containing relational and transactional data”, in *Data Mining and Computational Intelligence*. Heidelberg, Germany: Physica-Verlag HD, 2001, pp. 95-114. doi: https://doi.org/10.1007/978-3-7908-1825-3_4.

ІГОР СУБАЧ,
АРТЕМ МИКИТЮК

МЕТОДИКА ФОРМУВАННЯ НЕЧІТКИХ АСОЦІАТИВНИХ ПРАВИЛ ІЗ ЗВАЖЕНИМИ АТРИБУТАМИ З БАЗИ ДАНИХ SIEM-СИСТЕМИ ДЛЯ ВИЯВЛЕННЯ КІБЕРІНЦИДЕНТІВ В СПЕЦІАЛЬНИХ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМАХ

У статті представлено методику формування асоціативних правил з бази даних SIEM-системи для виявлення кіберінцидентів, яка ґрунтується на теорії нечітких множин та методах інтелектуального аналізу даних. На підставі проведеного аналізу зроблено висновок про доцільність виявлення кіберінцидентів в спеціальних інформаційно-комунікаційних системах (СІКС) шляхом застосовування правило-орієнтованих методів. Обґрунтовано необхідність застосування технологій інтелектуального аналізу даних (Data Mining), зокрема, методів формування асоціативних правил для поповнення ними бази знань (БЗ) SIEM-системи з метою покращення її характеристик в процесі виявлення кіберінцидентів. Для ефективного застосування моделей виявлення кіберінцидентів, побудованих на основі теорії нечітких множин, запропоновано використання методів пошуку нечітких асоціативних правил, які дозволяють обробляти різнорідні дані про кіберінциденти та є прозорими для сприйняття. Розглянуто математичний апарат для формування нечітких асоціативних правил та наведено приклади його застосування. Для підвищення ефективності методів пошуку нечітких асоціативних правил з БД SIEM-системи, запропоновано застосування вагових коефіцієнтів атрибутів, які характеризують степінь проявлення їхньої важливості у нечіткому правилі. Наведено формальну постановку задачі формування нечітких асоціативних правил із зваженими атрибутами та які застосовуються для ідентифікації кіберінцидентів. Запропоновано схему їх формування та застосування для ідентифікації кіберінцидентів. Наведено методику формування нечітких асоціативних правил із зваженими атрибутами з БД SIEM-системи. Сформульовано задачу визначення вагових коефіцієнтів відносної важливості атрибутів БД SIEM-системи та запропоновано метод для її вирішення. Наведено постановку задачі знаходження наборів елементів, що мають зважену нечітку підтримку не менше заданої, та використовуються для формування нечітких асоціативних правил із зваженими атрибутами. Запропоновано методи для її вирішення.

Ключові слова: кіберзахист, кіберінцидент**,** SIEM-система, теорія нечітких множин, інтелектуальний аналіз даних, асоціативні правила.

Subach Ihor, doctor of technical science, associate professor, head at the cybersecurity and application of information systems and technologies academic department, Institute of special communications and information security National technical university of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine, ORCID 0000-0002-9344-713X, igor_subach@ukr.net.

Mykytiuk Artem, postgraduate student, Institute of special communications and information security National technical university of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine, ORCID 0000-0002-8307-9978, mukuta8888@gmail.com.

Субач Ігор Юрійович, доктор технічних наук, доцент, завідувач кафедри кібербезпеки і застосування інформаційних систем і технологій, Інститут спеціального зв’язку та захисту інформації Національного технічного університету України “Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна.

Микитюк Артем В’ячеславович, аспірант, Інститут спеціального зв’язку та захисту інформації Національного технічного університету України “Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна.