Models of Fuzzy Identification of Cyber Incidents in Information and Communication Systems by Intelligent SIEM Systems

Ihor Subach [1,2], Dmytro Mogylevych [1], Artem Mykytiuk [1], Volodymyr Kubrak [1]
and Vitalii Fesokha [2]

[1] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Verkhnoklyuchova str., 4, Kyiv, 03056, Ukraine*
[2] *Heroes of Kruty Military Institute of Telecommunications and Information Technologies, Kyivska str., 45/1, Kyiv, 01011, Ukraine*

Abstract

The issue of detecting cyber incidents that occur during the functioning of information and communication systems, based on the use of a SIEM system in the cyber protection circuit, as the basis for its design, is considered. This study concludes that the use of the latest information technologies, such as artificial intelligence, intelligent data analysis, big data processing, and machine learning, is necessary to improve cyber incident detection. These technologies make it possible to increase the effectiveness of the SIEM system by minimizing human involvement in solving cyber incident response tasks. The analysis of methods used to detect cyber incidents has revealed their major limitations, particularly their inability to consistently provide the desired outcome in scenarios where information about such incidents is incomplete or inaccurate. To address this issue, a new approach to detect cyber incidents is proposed. This approach is based on using knowledge delivery models within the knowledge base that provide information about the signs of cyber incidents occurring in information and communication systems, as well as the various types of cyber incidents. This knowledge will be used to identify cyber incidents by collecting and processing expert information, utilizing the theory of fuzzy sets. Additionally, the proposed approach includes the formalization of the causal relationships between the variables "signs of cyber incidents" and "types of cyber incidents." This is achieved by describing these relationships in natural language using the theory of fuzzy sets and linguistic variables. The models proposed in this study can improve the effectiveness of cyber incident detection by addressing the limitations of current methods.

Keywords

Cybersecurity, cyber incident, SIEM system, fuzzy sets, fuzzy production rules

1. Introduction

The use of a proactive SIEM system [1, 2], such as Security Information and Event Management, is essential for building an effective cyber protection system.

The SIEM system plays a crucial role in proactively managing cyber incidents in the protection circuit. System can predict future security events based on information about past security events within the information and communication system (ICS). This is made possible through the use of appropriate models and methods for detecting cyber incidents.

A new functional model for an intelligent and proactive SIEM system has been proposed in [2], which comprises subsystems for data collection and primary processing from multiple sources, data

XXII International Scientific and Practical Conference "Information Technologies and Security (ITS-2022)", November 16, 2022, Kyiv, Ukraine
EMAIL: igor_subach@ukr.net (I. Subach); mogilev11@ukr.net (D. Mogylevych); mukuta8888@gmail.com (A. Mykytiuk); volodymir.kubrak@ukr.net (V. Kubrak); vitaliifesokha@gmail.com (V. Fesokha)
ORCID: 0000-0002-9344-713X (I. Subach); 0000-0002-4323-0709 (D. Mogylevych); 0000-0002-8307-9978 (A. Mykytiuk); 0000-0001-8877-5289 (V. Kubrak); 0000-0001-6612-1970 (V. Fesokha)

management, data analysis, decision-making, and implementation. In [3], a method for the rational selection of a SIEM system for constructing a cyber security situational center has been discussed.

The effectiveness of the system developed based on the proposed model is heavily reliant on the utilization of the latest information technologies, including Artificial Intelligence (AI), Data Mining (DM), Big Data, Machine Learning (ML), and others. These technologies enable the system to minimize human involvement in responding to cyber incidents occurring within the ICS, thereby increasing the efficiency and accuracy of its decisions [4].

Analysis of recent research and publications confirms the feasibility of using these technologies for detecting and promptly responding to cyber incidents.

The works [5, 6] address the use of data mining (DM) technology for detecting cyber incidents in SIEM systems. Specifically, this study analyzes the efficiency of using DM methods, such as association rules, classification, clustering, prediction, and sequential patterns, for implementing the primary functions of a SIEM system.

The authors of [7] examine the application of association rules in intrusion detection systems (IDS), with a focus on how these rules are formed from big data.

The study [8] presents a classification model for cyber incidents based on rules with weighted attributes.

The work [9] focuses on detecting DoS attack patterns by analyzing HTTP, HTTPS, and FTP traffic.

The issues of detecting unknown cyber attacks by IDS-systems, which are built on the basis of statistical methods, are considered in [10, 11].

The work [10] is devoted to the application of hybrid methods of detecting cyberattacks based on associative rules and ant algorithms. The authors prove the effectiveness of this approach in comparison with known ones by conducting experimental studies with the NSL-KDD data set.

In [12] a simulation model was proposed for evaluating the effectiveness of cyber attack detection systems.

The study [13] focuses on increasing the effectiveness of detecting network attacks in IDS Snort by applying DM techniques. The effectiveness of the results is experimentally evaluated using the MIT-DARPA 1999 data set.

The authors of [14] propose using a self-learning SIEM system to analyze network events and detect abnormal patterns using associative rules.

Despite numerous scientific publications on the application of DM and other modern information technologies in SIEM systems to detect cyber incidents, knowledge organization in the knowledge bases of intelligent SIEM systems remains largely unexplored.

This study aims to develop a model for incorporating knowledge about cyber incidents into the knowledge base of an intelligent SIEM system.

2. Statement of the problem of cyber incidents identification

It is widely acknowledged that the effectiveness of any intelligent system depends largely on the quality of its knowledge base. In [15], the matter of organizing knowledge in the knowledge bases of intelligent systems is discussed. However, research suggests that the most commonly used methods for detecting cyber incidents in SIEM systems are rule-oriented [16] approaches that rely on classical production rules. Therefore, a production model of knowledge delivery is used to represent knowledge in the SIEM system's knowledge base.

However, when information about cyber incidents during ICS operation is incomplete or inaccurate, these methods may not produce the desired results, making them inefficient. To address this issue, it is recommended to utilize models and techniques based on fuzzy set theory for fuzzy inference [16-21].

According to this, the model for detecting (recognizing) cyber incidents by a SIEM system can be represented as [16]:

$$MF = \langle KF,\ O_i,\ RF,\ C \rangle, \tag{1}$$

where *KF* – fuzzy classifier;

$RF = \{RF_i\}$ – a set of fuzzy rules for recognizing cyber incidents:

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

Based on the works [22-25], the task of recognizing cyber incidents can be considered as the task of their identification, and its solution is to find a mapping:

$$O^* = (o_1^*,\ o_2^*,\ \ldots,\ o_n^*) \rightarrow c_j \in C = (c_1, c_2,\ \ldots,\ c_m), \tag{2}$$

where O^* –a set of signs of a cyber incident;

C – a set of possible cyber incidents.

The area of change of signs of a cyber incident $o_i \in [\underline{o_i}, \overline{o_i}]$, $i = \overline{1, n}$, and the initial value of the identification result $c_j \in [\underline{c_j}, \overline{c_j}]$ are considered to be known. Accordingly, $\underline{o_i}(\overline{o_i})$– the lower (upper) value of the parameters of the cyber incident, $o_i,\ i = \overline{1, n}$; $\underline{c_j}, \overline{c_j}, j = \overline{1, m}$ – the lower (upper) value of the identification result.

In practice, the most commonly used methods for solving the problem (2) are parametric identification methods, such as the least squares method, maximum likelihood method, mean discrepancy method, stochastic approximation method, and more [19].

The main drawbacks of these methods that complicate their use include [22, 23]:

- the lack of clear interpretation in "input-output" type models;
- the inability to work with qualitative input and output variables;
- the inability to incorporate expert experience regarding object structure in the form of "IF-THEN" logical statements.

To address these shortcomings, a different approach based on models and methods of knowledge engineering is needed, particularly models for delivering knowledge base (KB) about the signs of cyber incidents (SCI) and classes of cyber incidents (CCI) that occur during ICS operation. This approach involves collecting and processing expert information using fuzzy set theory.

The primary concept of this approach is to formalize causal relationships between "SCI-CCI" variables by describing them in natural language using the theory of fuzzy sets and linguistic variables. This enables the mathematical formalization of natural language statements to solve the problem of identifying cyber incidents in the statement of the problem (2).

To translate expert knowledge presented in "IF-THEN" linguistic statements into a mathematical model, the mathematical apparatus of membership functions (MF) is used to express the expert's degree of confidence that a certain value belongs to a fuzzy concept (term). In turn, methods of fuzzy logical inference make it possible to connect the FN of the signs of cyber incidents with the results of their identification, provided that there is a model of cyber incidents in the form of a set of fuzzy rules of the "IF–THEN" type – a fuzzy knowledge base (FKB).

The primary principles used in developing a mathematical model for identifying cyber incidents include [17, 18, 20, 22, 23]:

- the principle of linguistic state of the ICS, which considers the type of cyber incident and its signs as linguistic variables (LV) evaluated by qualitative terms;
- the principle of forming the structure of the dependence of the type of cyber incident on its signs by means of the FKB;
- the principle of the hierarchy of the FKB, according to which the reduction in the dimension of the FKB can be carried out by classifying input variables and building an "output tree" that defines a system of nested logical statements.

The third principle of the model allows for consideration of a wide range of indicators used in identifying cyber incidents. Its implementation is particularly significant when dealing with a complex model since adding new indicators and accumulating knowledge about them requires such an approach.

Therefore, the model for identifying (detecting) cyber incidents can be specified in the form of FKB, represented by a set of "IF-THEN" fuzzy rules that connect linguistic evaluations of the signs of cyber incidents with their identification results.

Based on the above, the problem (2) can be formulated as follows.

It is important to note that the formal solution to this problem requires the presence of dependence (3):

Let $O^* = (o_1^*,\ o_2^*,\ \dots,\ o_n^*)$ is a vector of fixed values of signs of a cyber incident, where $o_i^* \in [\underline{o_i}, \overline{o}_i]$, $i = \overline{1, n}$. Then the task of identification is to determine the type of cyber incident $c_j \in C$ based on the vector O^*. Note that a necessary condition for the formal solution of this problem is the presence of dependence (3):

$$c_j = \xi(o_1, o_2, \dots, o_n), \tag{3}$$

where $o_1, o_2, \dots, o_n$ – a set of values of signs of cyber incidents, c_j – the result of identification.

3. Models of fuzzy identification of cyber incidents

To solve the formulated problem, it is necessary to consider the input and output variables from (3) as linguistic variables defined on universal sets [22, 23]:

$$o_i = \left[\underline{o_i}, \overline{o_i}\right], c_j = \left[\underline{c_j}, \overline{c_j}\right]. \tag{4}$$

To evaluate a LV, it is quite appropriate to apply qualitative terms that make up term sets [22-25]:

$A_i = \{\alpha_i^1, \alpha_i^2, \dots, \alpha_i^{k_i}\}$ – term set of the variable $o_i, i = \overline{1, n}$, where α_i^k – k-linguistic term of the variable $o_i, k = \overline{1, k_i}, i = \overline{1, n}$;

$\Delta = \{\delta_1, \delta_2, \dots, \delta_m\}$ – term set of the variable c_j, where $\delta_j, j = \overline{1, m}$ – linguistic term of the variable c_j, m – number of possible classes of cyber incidents.

In general, cardinalities of term sets $A_i, i = \overline{1, n}$ can be different:

$$k_1 \neq k_2 \neq \dots \neq k_n. \tag{5}$$

Besides, term names $\alpha_i^1, \alpha_i^2, \dots, \alpha_i^{k_i}$ may differ for different linguistic variables $o_i, i = \overline{1, n}$.

Therefore, the linguistic terms $\alpha_i^k \in A_i, k = \overline{1, k_i}, i = \overline{1, n}$ and $\delta_j \in \Delta, j = \overline{1, m}$ can be considered as fuzzy sets defined on the universal sets o_i, c_j (5).

In turn, the fuzzy sets α_i^k and δ_j can be defined as follows [22, 23]:

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

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

where $\mu^{\alpha_i^k}(o_i)$ – MF of the value of the variable $o_i \in \left[\underline{o_i}, \overline{o_i}\right], i = \overline{1, n}$ of the term $\alpha_i^k \in \mathrm{A}_i, k = \overline{1, k_i}, i = \overline{1, n}$;

$\mu^{\delta_j}(c_j)$ – MF of the value of the variable $c_j = \left[\underline{c_j}, \overline{c_j}\right]$ of the term – the class of cyber incident $\delta_j \in \Delta, j = \overline{1, m}$.

Note that in expressions (6) and (7) the sign $\bigcup$ denotes the union of pairs $\mu(\omega)/\omega$.

Let L be the amount of data that links the input data – the characteristics of cyber incidents and the output value – the class of the cyber incident, and:

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

where l_j – the number of data received from experts that correspond to the output variable – the class of the cyber incident $\delta_j \in \Delta, j = \overline{1, m}$, m – the number of classes of cyber incidents, and in the general case: $l_1 \neq l_2 \neq \dots \neq l_m$.

Note that the number of data received from experts is much smaller than a complete enumeration of various combinations l_j of input signs of cyber incidents.

By numbering the combinations of these expert data, they can be presented in the form of a multidimensional table [22, 23]:

Table 1
Multidimensional table of signs of cyber incidents and their corresponding classes

The number of the input combination of cyber incident sign values	Signs of cyber incident						Class of cyber incident
	o$_1$	o$_2$	…	o$_i$	…	o$_n$	c
11	α_1^{11}	α_2^{11}	…	α_i^{11}	…	α_n^{11}	δ_1
12	α_1^{12}	α_2^{12}	…	α_i^{12}	…	α_n^{12}	
…	…	…	…	…	…	…	
1k$_1$	$\alpha_1^{1k_1}$	$\alpha_1^{1k_2}$	…	$\alpha_i^{1k_1}$	…	$\alpha_n^{1k_1}$	
…	…	…	…	…	…	…	…
j1	α_1^{j1}	α_2^{j1}	…	α_i^{j1}	…	α_n^{j1}	δ_j
j2	α_1^{j2}	α_2^{j2}	…	α_i^{j2}	…	α_n^{j2}	
…	…	…	…	…	…	…	
jk$_j$	$\alpha_1^{jk_j}$	$\alpha_2^{jk_j}$	…	$\alpha_i^{jk_j}$	…	$\alpha_n^{jk_j}$	
…	…	…	…	…	…	…	…
m$_1$	α_1^{m1}	α_2^{m1}	…	α_i^{m1}	…	α_n^{m1}	δ_m
m$_2$	α_1^{m2}	α_2^{m2}	…	α_i^{m2}	…	α_n^{m2}	
…	…	…	…		…		
mk$_m$	$\alpha_1^{mk_m}$	$\alpha_2^{mk_m}$	…	$\alpha_i^{mk_m}$	…	$\alpha_n^{mk_m}$	

Similarly to [22, 23], let us call this table a matrix of knowledge about cyber incidents. It has the following properties:

- dimension of the matrix: $(n+1)\times N$, where $(n+1)$ – the number of columns of the matrix, and $L = l_1 + l_2 + \ldots + l_m$ – the number of its rows;
- each row of the matrix is a combination of the input values of the signs of cyber incidents $o_i, i = \overline{1,n}$, which is assigned by the expert to one of its classes δ_j, with the first $1l$ rows corresponding to the class δ_1, and the last $1m$ rows – to the class δ_m;
- first *n* columns of the matrix correspond to the input values of the signs of cyber incidents $o_i, i = \overline{1,n}$, and $(n+1)$ column corresponds to the output value – the class of the cyber incident *c*.
- at the intersection of the *i*-column and the jk_j-row there is an element $\alpha_i^{jk_j}$, that corresponds to the linguistic assessment of the cyber incident sign *o$_i$* the row of the matrix jk_j, which belongs to the term set of the corresponding sign *o$_i$*: $\alpha_i^k \in \mathrm{A}_i, k = \overline{1,k_i}, i = \overline{1,n}$.

The matrix of knowledge about cyber incidents described above can be represented in the form of a system of fuzzy rules of "IF–THEN" type [22, 23], that connect the values of the input signs of cyber incidents $o_i, i=\overline{1,n}$ with one of the possible classes of cyber incidents $\delta_j \in \Delta, j=\overline{1,m}$:

$$\begin{gathered}
\text{IF}(o_i=\alpha_1^{11})\text{AND}(o_2=\alpha_2^{11})\text{AND}\ldots\text{AND}(o_n=\alpha_n^{11})\text{OR} \\
(o_i=\alpha_1^{12})\text{AND}(o_2=\alpha_2^{12})\text{AND}\ldots\text{AND}(o_n=\alpha_n^{12})\text{OR} \\
(o_i=\alpha_1^{1k_1})\text{AND}(o_2=\alpha_2^{1k_1})\text{AND}\ldots\text{AND}(o_n=\alpha_n^{1k_1})\text{THEN}(\text{c=}\delta_1),\ldots \\
\ldots,\ \text{IF}(o_i=\alpha_1^{j1})\text{AND}(o_2=\alpha_2^{j1})\text{AND}\ldots\text{AND}(o_n=\alpha_n^{j1})\text{OR} \\
(o_i=\alpha_1^{j2})\text{AND}(o_2=\alpha_2^{j2})\text{AND}\ldots\text{AND}(o_n=\alpha_n^{j2})\text{OR} \\
(o_i=\alpha_1^{k_j})\text{AND}(o_2=\alpha_2^{jk_j})\text{AND}\ldots\text{AND}(o_n=\alpha_n^{jk_j})\text{THEN}(\text{c=}\delta_j),\ldots \quad (9) \\
\ldots,\ \text{IF}(o_i=\alpha_1^{m1})\text{AND}(o_2=\alpha_2^{m1})\text{AND}\ldots\text{AND}(o_n=\alpha_n^{m1})\text{OR} \\
(o_i=\alpha_1^{m2})\text{AND}(o_2=\alpha_2^{m2})\text{AND}\ldots\text{AND}(o_n=\alpha_n^{m2})\text{OR} \\
(o_i=\alpha_1^{mk_m})\text{AND}(o_2=\alpha_2^{mk_m})\text{AND}\ldots\text{AND}(o_n=\alpha_n^{mk_m})\text{THEN}(\text{c=}\delta_m),
\end{gathered}$$

where α_i^{jk} – linguistic evaluation of the sign of cyber incident $o_i, i=\overline{1,n}$ in the k row of *j*-disjunction determined on the term set $\mathrm{A}_i=\{\alpha_i^1,\alpha_i^2,\ldots,\alpha_i^{k_i}\}$;

$\delta_j \in \Delta, j=\overline{1,m}$ – linguistic evaluation of the class of cyber incident, determined on the term set $\Delta=\{\delta_1,\delta_2,\ldots,\delta_m\}$.

Therefore, the expression (9) given in the form of a set of fuzzy rules of "IF–THEN" type, which are based on the matrix of knowledge about cyber incidents (Table 1), represents a model for the identification of cyber incidents by the SIEM system.

If the linguistic evaluations α_i^{jk} of the variables $o_1,o_2,\ldots,o_n$ and $\delta_j, j=\overline{1,m}$ from (9) are considered as fuzzy sets defined on universal sets $o_i=\left[\underline{o_i},\overline{o_i}\right], c=\left[\underline{c_j},\overline{c_j}\right], i=\overline{1,n}, j=\overline{1,m}$, then $\mu^{\alpha_i^{jk}}$ – MF (belongingness) of the sign of cyber incident $o_i=\left[\underline{o_i},\overline{o_i}\right]$ to the fuzzy term $\alpha_i^{jk}, i=\overline{1,n}, j=\overline{1,m}, k=\overline{1,k_j}$, and $\mu^{\delta_j}(o_1,o_2,\ldots,o_n)$ – to the vector of sign of a cyber incident, $O=\{o_1,o_2,\ldots,o_n\}, i=\overline{1,n}$, to the value of the initial assessment $c=\delta_j, j=\overline{1,m}$.

The relationship between them is determined through a fuzzy matrix of knowledge about cyber incidents (Table 1) and by replacing linguistic terms with their MF, as well as replacing logical operations AND or OR with operations $\wedge$ and $\vee$ can be represented as follows:

$$\begin{gathered}
\mu^{\delta_j}(o_1,o_2,\ldots,o_n)=\left[\mu^{\alpha_1^{j1}}(o_1)\wedge\mu^{\alpha_2^{j1}}(o_2)\wedge\ldots\wedge\mu^{\alpha_n^{j1}}(o_n)\right]\vee \\
\vee\left[\mu^{\alpha_1^{j2}}(o_1)\wedge\mu^{\alpha_2^{j2}}(o_2)\wedge\ldots\wedge\mu^{\alpha_n^{j2}}(o_n)\right]\vee\cdots \\
\cdots\vee\left[\mu^{\alpha_1^{jk_j}}(o_1)\wedge\mu^{\alpha_2^{jk_j}}(o_2)\wedge\ldots\wedge\mu^{\alpha_n^{jk_j}}(o_n)\right], j=\overline{1,m}
\end{gathered} \quad (10)$$

Then expression (10) can be represented as follows:

$$\mu^{\delta_j}(o_{1.}o_2,\ldots,o_n)=\bigvee_{k=1}^{k_j}\left[\bigwedge_{i=1}^{n}\mu^{\alpha_i^{jk}}(o_i)\right], j=\overline{1,m}. \quad (11)$$

Taking into account that in the theory of fuzzy sets, operations $\wedge$ and $\vee$ correspond to operations min and max [22, 23], then by transforming (11) we will obtain a model of identification of cyber incidents by SIEM systems (12):

$$\mu^{\delta_j}\left(o_i, o_2, \ldots, o_n\right) = \underset{k=\overline{1,k_j}}{\max} \left\{ \underset{i=\overline{1,n}}{\min} \left[\mu^{\alpha_i^{jk}}\left(o_i\right) \right] \right\}, j = \overline{1,m}. \tag{12}$$

This model can be the basis for the development of a rule-oriented method of detecting cyber incidents by a SIEM system.

4. Model of fuzzy identification of cyber incidents with weighted rules

The main drawback of the model (12) is that the expert's confidence in each "IF–THEN" rule, which is included in the fuzzy KB (9), may differ. This shortcoming can be eliminated by introducing the weight of the rule, which will characterize the significance of a particular rule during the identification of cyber incidents. Based on the works [24, 27, 28], the weight of a rule should be understood as a number in the interval [0, 1], which characterizes the subjective measure of the expert's confidence in a particular rule.

In this case, the multidimensional table of signs of cyber incidents and their corresponding classes (Table 1), taking into account the expert's confidence in a particular rule, which is specified using the weight of the rule, will take the following form (Table 2):

Table 2
Multidimensional table of signs of cyber incidents and their corresponding classes in accordance with the weight of the rules

The number of the input combination of cyber incident sign values	Signs of cyber incident						Weight of the rule	Class of cyber incident
	o₁	o₂	…	oᵢ	…	oₙ	ω	c
11	α_1^{11}	α_2^{11}	…	α_i^{11}	…	α_n^{11}	ω_{11}	δ_1
12	α_1^{12}	α_2^{12}	…	α_i^{12}	…	α_n^{12}	ω_{12}	
…	…	…	…	…	…	…		
1k₁	$\alpha_1^{1k_1}$	$\alpha_1^{1k_2}$	…	$\alpha_i^{1k_1}$	…	$\alpha_n^{1k_1}$	ω_{1k_1}	
…	…	…	…	…	…	…		…
j1	α_1^{j1}	α_2^{j1}	…	α_i^{j1}	…	α_n^{j1}	ω_{j1}	δ_j
j2	α_1^{j2}	α_2^{j2}	…	α_i^{j2}	…	α_n^{j2}	ω_{j2}	
…	…	…	…	…	…	…	…	
jkⱼ	$\alpha_1^{jk_j}$	$\alpha_2^{jk_j}$	…	$\alpha_i^{jk_j}$	…	$\alpha_n^{jk_j}$	ω_{jk_j}	
…	…	…	…	…	…	…	…	…
m₁	α_1^{m1}	α_2^{m1}	…	α_i^{m1}	…	α_n^{m1}	ω_{m1}	δ_m
m₂	α_1^{m2}	α_2^{m2}	…	α_i^{m2}	…	α_n^{m2}	ω_{m2}	
…	…	…	…	…	…	…	…	
mkₘ	$\alpha_1^{mk_i}$	$\alpha_2^{mk_m}$	…	$\alpha_i^{mk_m}$	…	$\alpha_n^{mk_m}$	ω_{mk_m}	

Thereby, taking into account the weight of the rules, the FKB, which is represented by a set of fuzzy "IF–THEN" rules that link linguistic evaluations of signs of cyber incidents with the results of their identification, will take the following form:

IF($o_i = \alpha_1^{11}$)AND($o_2 = \alpha_2^{11}$)AND…AND($o_n = \alpha_n^{11}$) with weight w_{11} OR
($o_i = \alpha_1^{12}$)AND($o_2 = \alpha_2^{12}$)AND…AND($o_n = \alpha_n^{12}$) with weight w_{12} OR
($o_i = \alpha_1^{1k_1}$)AND($o_2 = \alpha_2^{1k_1}$)AND…AND($o_n = \alpha_n^{1k_1}$) with weight w_{1k_1}
THEN(c=δ_1), …
…, IF($o_i = \alpha_1^{j1}$)AND($o_2 = \alpha_2^{j1}$)AND…AND($o_n = \alpha_n^{j1}$) with weight w_{j1} OR
($o_i = \alpha_1^{j2}$)AND($o_2 = \alpha_2^{j2}$)AND…AND($o_n = \alpha_n^{j2}$) з вагою OR
($o_i = \alpha_1^{jk_j}$)AND($o_2 = \alpha_2^{jk_j}$)AND…AND($o_n = \alpha_n^{jk_j}$) з вагою w_{jk_j}
THEN(c=δ_j), … (13)
…, IF($o_i = \alpha_1^{m1}$)AND($o_2 = \alpha_2^{m1}$)AND…AND($o_n = \alpha_n^{m1}$) with weight w_{m1} OR
($o_i = \alpha_1^{m2}$)AND($o_2 = \alpha_2^{m2}$)AND…AND($o_n = \alpha_n^{m2}$) with weight w_{m2}
OR
($o_i = \alpha_1^{mk_m}$)AND($o_2 = \alpha_2^{mk_m}$)AND…AND($o_n = \alpha_n^{mk_m}$) with weight w_{mk_m}
THEN(c=δ_m),

where α_i^{jk} – linguistic evaluation of the sign of cyber incident $o_i, i=\overline{1,n}$ in the k row of *j*-disjunction determined on the term set $\mathrm{A}_i = \{\alpha_i^1, \alpha_i^2, \dots, \alpha_i^{k_i}\}$;

$\delta_j \in \Delta, j=\overline{1,m}$ – linguistic evaluation of the class of cyber incident determined on the term set $\Delta = \{\delta_1, \delta_2, \dots, \delta_m\}$;

w_{jk} – weight of the rule.

Taking into account the weight of the rules, the fuzzy knowledge base (13) can be represented by a modified system of fuzzy equations (11) as follows:

$$\mu^{\delta_j}(o_{1.}o_2,\dots,o_n) = \overset{k_j}{\underset{k=1}{\vee}}\left\{\omega_{jk}\left[\overset{n}{\underset{i=1}{\wedge}}\mu^{\alpha_i^{jk}}(o_i)\right]\right\}, j=\overline{1,m} \tag{14}$$

By replacing the operations $\wedge$ and $\vee$ with min та max operations, which correspond to them, we will obtain a modified model of fuzzy identification of cyber incidents by SIEM systems (15) with weighted rules:

$$\mu^{\delta_j}(o_i,o_2,\dots,o_n) = \underset{k=\overline{1,k_j}}{\max}\left\{\omega_{jk}\underset{i=\overline{1,n}}{\min}\left\{\mu^{\alpha_i^{jk}}(o_i)\right\}\right\}, j=\overline{1,m}. \tag{15}$$

The given model makes it possible to eliminate the shortcomings present in model (12) by taking into account the expert's confidence in a particular rule.

5. Conclusion

The models proposed for cyber incident identification by a SIEM system differ from existing models in that they utilize fuzzy rules to identify incidents that occur during the operation of an information and communication system. As a result, it becomes possible to mitigate the shortcomings that arise from incomplete and inaccurate information in the identification process.

Potential areas for future research include developing rule-oriented methods for detecting cyber incidents using the proposed models in the SIEM system, along with techniques for generating fuzzy production rules to apply within these methods.

6. References

[1] I. Subach, V. Fesokha, N. Fesokha, Analysis of existing intrusion prevention solutions in information and telecommunication networks, opened on the basis of publicly available licenses, Information Technology and Security Vol. 5 Iss. 1 (2017) 29–41. doi:10.20535/2411-1031.2017.5.1.120554.

[2] I. Subach, V. Kubrak, A. Mykytiuk, Architecture and functional model of a promising proactive intelligent system SIEM-system for cyber protection of critical infrastructure, Information Technology and Security Vol. 7 Iss. 2 (2019) 208–215. doi:10.20535/2411-1031.2019.7.2.190570.

[3] I. Subach, V. Kubrak, A. Mykytiuk, Methodology of rational choice of security incident management system for building operational security center, CEUR Workshop Proceedings (CEUR-WS.org) Vol. 2577 (2019) 11–20. doi:10.5281/zenodo.7027782.

[4] I. Subach, B. Gerasimov, Quality indicators of information support and their impact on the effectiveness of decision support systems, Bulletin of Taras Shevchenko National University of Kiev Vol. 20 (2008) 18–25.

[5] A. R. Zope, A. Vidhate, N. Harale, Data Mining Approach in Security Information and EventManagement, International Journal of Future Computer and Communication Vol. 2 Iss. 2 (2013) 80–84. doi:10.7763/IJFCC.2013.V2.126

[6] F. Salo, M. Injadat, A. Nassif, A. Shami, A. Essex, Data Mining Techniques in Intrusion Detection Systems: A Systematic Literature Review, in Proc. IEEEAccess Vol. 6 (2018) 56046–56058. doi:10.1109/ACCESS.2018.2872784.

[7] D. Selvamani, V. Selvi, Association Rule Mining for Intrusion Detection System: A Survey, Asian Journal of Engineering and Applied Technology Vol. 8 Iss. 1 (2019) 20–24. doi:10.51983/ajeat-2019.8.1.1065.

[8] L. Mehrotra, P. S. Saxena, N. V. Doohan, A Data Classification Model: For Effective Classification of Intrusion in an Intrusion Detection System Based on Decision Tree Learning Algorithm, Information and Communication Technology for Sustainable Development vol 9 (2018) 61–66. doi:10.1007/978-981-10-3932-4_7.

[9] HC. Chen, SS. Kuo, DoS Attack Pattern Mining Based on Association Rule Approach for Web Server, International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. Innovative Mobile and Internet Services in Ubiquitous Computing Vol. 773 (2018) 527–536. doi:10.1007/978-3-319-93554-6_50.

[10] L. Mehrotra, P. S. Saxena, An Assessment Report on: Statistics-Based and Signature-Based Intrusion Detection Techniques, Information and Communication Technology Springer. Advances in Intelligent Systems and Computing Vol. 625 (2018) 321–327. doi:10.1007/978-981-10-5508-9_31.

[11] R. Shanmugavadivu, N. Nagarajan, Network Intrusion Detection System using Fuzzy Logic, Indian Journal of Computer Science and Engineering (IJCSE) Vol. 2 Iss. 1.(2011) 101–111.

[12] C. Gupta, A. Sinhal, R. Kamble, An Enhanced Associative Ant Colony Optimization Technique-based Intrusion Detection, Artificial Intelligence and Evolutionary Algorithms in Engineering Systems. Advances in Intelligent Systems and Computing Vol. 325 (2015) 541–553. doi:10.1007/978-81-322-2135-7_58.

[13] I. Subach, D. Mogylevych, A. Mykytiuk, V. Kubrak, V. Fesokha, Simulation Model of a Fuzzy Cyber Attack Detection System, CEUR Workshop Proceedings (CEUR-WS.org) Vol. 3421 (2021) 92–101. doi: 10.5281/zenodo.7247964.

[14] N. Khamphakdee, N. Benjamas, S. Saiyod, Improving Intrusion Detection System Based on Snort Rules for Network Probe Attacks Detection with Association Rules Technique of Data Mining, Journal of ICT Research and Applications, Vol. 8, Iss. 3, (2015) 234–250. doi:10.5614/itbj.ict.res.appl.2015.8.3.4.

[15] R. R. Tiwari, A. K. Singh, V. Singh, Self-Learning SIEM System Using Association Rule Mining, Journal of Advanced Database Management & Systems Vol. 2 Iss. 2, (2015) 10–23.

[16] I. Subach, B. Gerasimov, E. Nikiforov, Models of knowledge delivery for use in decision support systems, Scientific and technical information Vol. 1 (2005) 7–11.

[17] I. Subach, V. Kubrak, A. Mykytiuk, V. Korotaev, Rule-oriented method of cyber incidents detection by SIEM based on fuzzy logical inference, CEUR Workshop Proceedings (CEUR-WS.org) Vol. 2859 (2020) 210–219. doi:10.5281/zenodo.7123656.

[18] L. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Moscow, Mir, 1976.

[19] A.N. Borisov, O.A. Krumberg, I.P. Fedorov, Decision-making based on fuzzy models: examples of use, Riga, Zinatne, 1990.

[20] Y. Zaichenko, Operations Research: Fuzzy Optimization, Kiev, High school, 1991.

[21] A. Piegat, Fuzzy Modeling and Control, Physica-Verlag, Heidelberg, 2001.

[22] M. Dodonov, N. Dodonova, Automated detection system of insider attacks using fuzzy logic, CEUR Workshop Proceedings (CEUR-WS.org) Vol. 1490 (2015) 376–380.

[23] A. Rothstein, Medical diagnostics on fuzzy logic, Vinnytsia, Continent-PRIM, 1996.

[24] A. Rothstein, Intelligent identification technologies: fuzzy sets, genetic algorithms, neural networks, Vinnytsia, UNIVERSUM, 1999.

[25] O. Rothstein, G. Chernovolyk, E. Laryushkin, Method of constructing membership functions of fuzzy sets. Bulletin of VPI, Vol. 3 (1996) 72–75.

[26] Y. Mityushkin, B. Mokin, O. Rothstein, Soft Computing: identification of patterns of fuzzy knowledge bases: a monograph. Vinnytsia, UNIVERSUM-Vinnytsia, 2002.

[27] Rotshtein A.P., Katelnikov D.I. Identification of Nonlinear Objects by Fuzzy Knowledge Bases. Cybernetics and System Analysis Iss. 5 (1998) 53–61.

[28] Rotshtein A. Design and Tuning of Fuzzy Rule-Based Systems for Medical Diagnosis. In N.-H. Teodorescu (ed): Fuzzy and Neuro – Fuzzy Systems in Medicine (1998) 243–289.