Rule-oriented method of cyber incidents detection by SIEM based on fuzzy logical inference

© Ihor Subach, © Volodymyr Kubrak, © Artem Mykytiuk, © Stanislav Korotayev

Institute of Special Communication and Information Protection of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine

igor_subach@ukr.net

Abstract. We consider the role of SIEM in the protection circuit of information and telecommunication system for proactive cyber incident management. We provide the main mechanisms of the process of correlation of events on the detection of cyber-attacks, malicious activity, and violations of security policy. We analyze identification methods of signs of deletion, integration, and connection of the processed information, as well as the establishment of its reasons and priorities. We outline the main disadvantages of the rule-oriented method. We propose the implementation of the model and method of cyber incident recognition under incompleteness or inaccuracy of information about the incidents based on the application of fuzzy set theory and fuzzy inference. We present the formal statement of the problem of cyber incident detection by the SIEM and propose its solution. The problem of incident identification is solved by finding a mapping between the set of signs of cyber incidents and the set of their possible classes. Graphical interpretation of the problem of cyber incident identification is presented and the main difficulties that arise during its solution are formulated. Emphasis is placed on the expediency of creating a subsystem of intelligent decision support in the SIEM, which should be based on the model of cyber incident identification based on fuzzy rules and fuzzy inference, where the causal relationship between a cyber incident and its features are described by an expert in natural language, and then formalized as a set of fuzzy logical rules. An algorithm for deciding on cyber incident identification is proposed. The data on the practical effectiveness of the proposed method is presented.

Keywords: cybersecurity, cyber defense, cyber-attack, cyber incident, SIEM, fuzzy set theory, tuple recognition model, rule-oriented method

1 Introduction

Building of an effective cyber defense system should be based on proactive Security Information and Event Management (SIEM) [1]. The use of SIEM in the protection circuit allows for effective proactive management of cyber incidents, based on automated mechanisms that use information about events that have already occurred in the system, predict future events that will occur in it, and adapt system protection parameters to its current status.

A cyber incident is an event or a series of adverse events that bear signs of a possible cyberattack, which threaten the security of electronic communications systems, process control systems, create a possibility of violation of the normal operation of such systems, including failure and/or blocking of the system and/or unauthorized management of its resources, and endanger the security of electronic information resources [2].

The architecture and functional model of the proactive SIEM were considered in [1]. According to the tasks performed by this system (collection, processing, and analysis of security events coming to it from many disparate distributed sources), the basis of its operation includes the following mechanisms: normalization, filtering, classification, aggregation, correlation, prioritization, and analysis of events and cyber incidents and their consequences, as well as generation of various reports, messages and visual presentation of data for prompt and informed decision-making [1]. The methodology of rational selection of SIEM for the construction of SOC (Security Operation Center) is given in [3].

In some sources [4, 5] these mechanisms are considered as stages of the general process, which is called the correlation process. It has a special place in the SIEM, as its purpose is to detect cyberattacks, malicious activity, security policy violations, and others [6]. This purpose is achieved by addressing a wide range of tasks that it covers: identifying potential relationships between disparate security information; grouping low-level events into higher-level events; detecting potential incidents based on analysis of the behavior of various infrastructure objects, and others.

Technologically, as part of the SIEM, the correlation method includes a sequence of actions on the data, which aims to identify, in a certain way, signs of deletion, integration, and linking of processed information, as well as establishing its causality and priority [4, 5]. These features are called correlation features.

To achieve these objectives, at different stages of the correlation process, a wide variety of methods are used [7, 8, 9, 10, 11], such as: the method based on finite machine states (finite state machines), which is used to identify dangerous states of the system; rule-oriented method, which is based on rules that have clear syntax and semantics; the method of reasoning based on precedents; Bayesian network method, which is used at the stage of multi-step event correlation, loss analysis, and prioritization; artificial neural networks, which are also used for event correlation, loss analysis, and prioritization, and others.

Analysis shows that the most common method is the rule-oriented method, but due to the fact that it is based on classical production rules, which do not always give the expected result in terms of incomplete and inaccurate information about cyber incidents, its application is not always effective.

Therefore, the task of developing models and methods for recognizing cyber incidents in conditions of incompleteness or inaccuracy of information about them is relevant.

The aim of the work is to develop a model and rule-oriented method of detecting cyber incidents by SIEM based on fuzzy inference.

2 Statement of the problem of cyber incident detection by SIEM

Any cyber incident is characterized by a set of information features, on the basis of which, in turn, it ~~can be~~ recognized.

Let $O = \{o_i\} \mid i = \overline{1,n}$ – the set of information features of cyber incidents that occur in the system and are represented by the set $C = \left\{ C_j \mid C_j = (o_{j1}, o_{j2}, \ldots, o_{jm}) \right\}, j = \overline{1,J},$ where information signs are associated with a cyber incident C_j.

Then the model of cyber incident recognition can be represented by a tuple [12]:

$$M = \langle K, O_i, R, C \rangle, \tag{1}$$

where K is a feature classifier; $o_i \in O$ – is a set of the observed features; $R = \{R_i\}$ is a set of cyber incidents recognition rules; C – a cyber incident.

The process of recognizing cyber incidents is carried out based on rules (usually, production rules):

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

However, in traditional production systems, the rules are classic products that do not fully meet the conditions of incompleteness and inaccuracy of information about cyber incidents that occur during operation of information and telecommunications systems. As a rule, for this purpose, methods, and models of fuzzy set theory on fuzzy inference are used [13].

3 The method of solving the problem

Based on above-listed considerations, model (1) can be developed and presented as follows:

$$MF = \langle KF, O_i, RF, C \rangle, \tag{2}$$

where KF is a fuzzy classifier, $RF = \{RF_i\}$ is a set of fuzzy cyber incident recognition rules:

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

On the other hand, based on the works [15, 16], the problem of recognizing cyber incidents can be considered as a problem of their identification, the solution of which is to find a mapping:

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

where $O^* -$ is a set of signs a cyber incident; a set of possible cyber incidents.

Range of change of signs of cyber incidents $o_i \in \left[\underline{o}_i, \overline{o}_i\right]$, $i = \overline{1,n}$, and the original value of the identification result $k \in \left[\underline{k}, \overline{k}\right]$ are considered known. Accordingly, $\underline{o}_i\left(\overline{o}_i\right)$ is the lower (upper) value of the cyber incidence parameters, o_i, $i = \overline{1,n}$, $\underline{k}, \left(\overline{k}\right)$ is the lower (upper) value of the identification result *k*.

Graphically, the problem of identifying cyber incidents can be represented as follows (see Fig. 1):

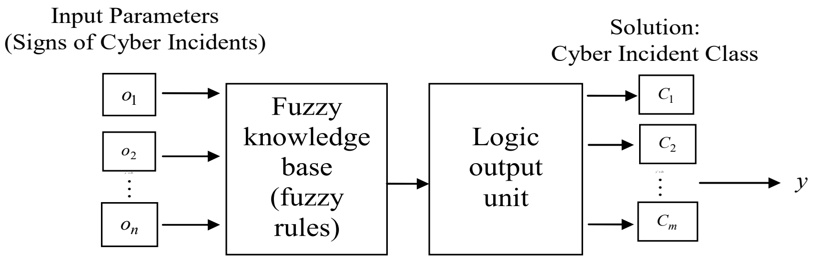

Fig. 1. Graphical interpretation of the problem of cyber incident identification.

In this case, the main difficulties that arise when solving the problem (3) are as follows:

first: for correct identification of a cyber incident, it is necessary to take into account a large number of heterogeneous parameters of the system (quantitative and qualitative), which, in turn, requires a highly qualified cybersecurity officer, as well as appropriate time;

second: the lack of analytical dependence between the cyber incident and its signs.

These difficulties confirm the expediency of creating an intelligent decision support subsystem within the SIEM. Its operation should be based on a model of cyber incident identification based on fuzzy rules and fuzzy inference.

At the same time, for its development, it is necessary to take into account the linguistic nature of the type of cyber incident (output variable) and its features (input variables). In turn, the causal relationship between a cyber incident and its signs must be described by an expert in plain language and then formalized as a set of fuzzy logical rules.

It should be noted that with a large number of signs of cyber incidents, it is advisable to build a tree of inference, which determines the order of embedding statements in each other.

Figure 2 shows the inference tree for the correlation rule [16]: if on one computer with the same IP address, seven user authentication attempts using different user IDs have failed within ten minutes, and a successful login of a user into the system from any computer with the same IP address has been successful, then this event must be addressed by a security officer.

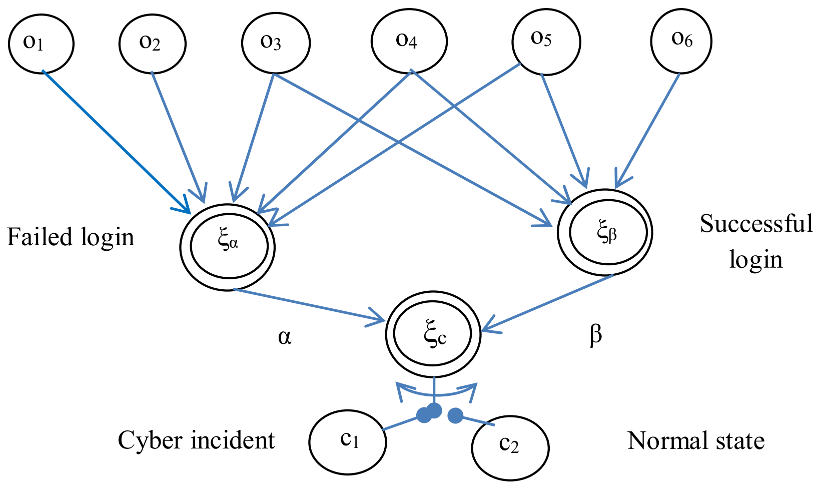

Fig. 2. Logical inference tree for the correlation rule.

Here through $o_1 \div o_6$ marked signs of a cyber incident (Table 1).

Table 1. Signs of a cyber incident.

Sign	The content of the sign	Type
O_1	The number of failed login attempts	numerical
O_2	The number of users IDs	numerical
O_3	The duration of login attempts	numerical
O_4	The number of IP-addresses involved during login	numerical
O_5	The number of computers involved during login	numerical
O_6	The number of successful login attempts	numerical

In turn, c_1 and c_2 indicate the type of event occurring in the system (Table 2).

Table 2. The type of event occurring in the system.

Event	Event content
α	Successful login
β	Failed login
C_1	Cyber incident
C_2	Successful login

The structure of the logical output tree corresponds to relations (4)-(6):

$$c = \xi_c(\alpha, \beta), \tag{4}$$

$$\alpha = \xi_\alpha(o_1, o_2, o_3, o_4, o_5), \tag{5}$$

$$\beta = \xi_\beta(o_3, o_4, o_5, o_6). \tag{6}$$

Table 3. Multidimensional matrix of knowledge about cyber incidents.

Sign	Value	Type	Cyber incident
		α	
O_1	L		
O_2	H		
O_3	L	H	
O_4	L		
O_5	aA		C_1
Sign	Value	β	
O_3	bA		
O_4	L	H	
O_5	H		
O_6	L		
Sign	Value	α	
O_1	aA		
O_2	aA		
O_3	bA	aA	
O_4	L		
O_5	A		
Sign	Value	β	C_1
O_3	bA		
O_4	L	aA	
O_5	aA		
O_6	L		

A single scale of qualitative terms $o_1 \div o_6, \alpha, \beta$ is used to estimate the values of linguistic variables: L - low; bA - below average; A - average; aA - above average; H - high. Each of these terms is given by the corresponding membership function.

From a formal point of view, the problem of cyber incident identification based on fuzzy rules and fuzzy inference corresponds to the mathematical model of object identification with a discrete output [14, 15]. Thus, to identify a cyber incident c_1, the ratio is as follows:

$$\mu^{c_1}(c) = \left[\mu^H(\alpha) \wedge \mu^H(\beta)\right] \vee \left[\mu^{aA}(\alpha) \wedge \mu^{aA}(\beta)\right], \tag{7}$$

where

$$\mu^H(\alpha) = \left[\mu^{\mathrm{H}}(o_1) \wedge \mu^{\mathrm{H}}(o_2) \wedge \mu^{\mathrm{L}}(o_3) \wedge \mu^{\mathrm{L}}(o_4) \wedge \mu^{\mathrm{aA}}(o_5)\right],$$

$$\mu^H(\beta) = \left[\mu^{\mathrm{bA}}(o_3) \wedge \mu^{\mathrm{L}}(o_4) \wedge \mu^{\mathrm{H}}(o_5) \wedge \mu^{\mathrm{L}}(o_6)\right];$$

$$\mu^{aA}(\alpha) = \left[\mu^{\mathrm{aA}}(o_1) \wedge \mu^{\mathrm{aA}}(o_2) \wedge \mu^{\mathrm{bA}}(o_3) \wedge \mu^{\mathrm{L}}(o_4) \wedge \mu^{\mathrm{A}}(o_5)\right],$$

$$\mu^{aA}(\beta) = \left[\mu^{\mathrm{bA}}(o_3) \wedge \mu^{\mathrm{L}}(o_4) \wedge \mu^{\mathrm{aA}}(o_5) \wedge \mu^{\mathrm{L}}(o_6)\right];$$

and $\mu(c), \mu(\alpha), \mu(\beta), \mu(o_i) -$ are corresponding membership functions.

These fuzzy logical equations allow us to make a decision in favor of identification of a cyber incident based on the following algorithm:

Step 1. The values of the signs of cyber incidents are recorded $O^* = \left(o_1^*, o_2^*, \ldots, o_6^*\right)$.

Step 2. The values of membership functions $\mu^k\left(o_i^*\right)$ are determined at fixed parameter values $o_i^*, i = \overline{1,6}; k = \{\mathrm{L, bA, A, aA, H}\}$.

Step 3. Based on logical equations (7), the values of membership functions $\mu^{c_j}\left(o_1^*, o_2^*, \ldots, o_6^*\right)$ are calculated by the vector of attributes $O^* = \left(o_1^*, o_2^*, \ldots, o_6^*\right)$ for all types of cyber incidents c_1, c_2. Logical operations AND $(\wedge)$ and OR $(\vee)$ on membership functions are replaced by operations min and max:

$$\mu^k\left(o_i^*\right) \wedge \mu^k\left(o_j^*\right) = \min\left[\mu^k\left(o_i^*\right), \mu^k\left(o_j^*\right)\right]; i \neq j, \tag{8}$$

$$\mu^k\left(o_i^*\right) \vee \mu^k\left(o_j^*\right) = \max\left[\mu^k\left(o_i^*\right), \mu^k\left(o_j^*\right)\right]; i \neq j, \tag{9}$$

Step 4. Choice of solution c_j^* (the type of cyber incident) provided:

$$\mu^{c_j}\left(o_1^*, o_2^*, \ldots, o_6^*\right) = \max\left[\quad \mu^{c_j}\left(o_1^*, o_2^*, \ldots, o_6^*\right)\right]. \tag{10}$$

It should be noted that the adequacy of this model and the effectiveness of the method of detecting cyber incidents, which is based on the proposed model, respectively, are determined by the quality of membership functions, through which linguistic estimates are quantified. Due to the fact that these membership functions are determined by experts, the adequacy of the fuzzy knowledge base will depend entirely on the qualifications of experts.

However, it should be noted that as a result of SIEM operation, statistics on cyber incidents will be collected, which makes it possible to assess the adequacy of the proposed model and the method developed on its basis.

Thus, it is quite expedient to perform additional training (system settings). This, in turn, will allow the identification of cyber incidents that were not previously identified by the system during its operation.

Comparative analysis of the proposed method showed that, in comparison with existing methods (the method of reference vectors, neural networks, k-nearest neighbors, the method based on immune systems), it can increase the accuracyof cyber incident detection (11) by 2-15 % (Table 4).

$$P = \frac{TP}{TP + FP}, \tag{11}$$

where P (precision) is the accuracy of cyber incident detection;

TP – the number of cyber incidents that are properly classified;

FP – the number of cyber incidents classified as a normal state of the system [17].

Table 4. Comparative analysis of the proposed method of cyber incidents detection.

Method	The accuracy of cyber incident detection	Δ
Method of reference vectors	0,83	+0,15
K-nearest neighbors	0,85	+0,13
Method based on immune systems	0,96	+0,02
Neural networks	0,86	+0,12
The proposed method	0,98	-

4 Conclusions

As a result of the conducted research, it is shown that the main role in a protection circuit of information and telecommunication system for proactive management of cyber incidents belongs to SIEM.

The results of the analysis indicate the feasibility of using a rule-oriented method to identify signs of deletion, aggregation, and linking of information processed, as well as to establish its causality and priority.

To increase the efficiency of the rule-oriented method of recognizing cyber incidents in conditions of incompleteness and inaccuracy of information about them, a model based on the theory of fuzzy sets and fuzzy inference is proposed. Based on the model, a rule-oriented method of cyber incident identification, based on mapping of the set of incident features to the set of possible classes of cyber incidents, and the algorithm for its implementation have been developed.

To implement the developed model and method, it is advisable to modify the structure of the SIEM-system by introducing an intelligent decision support subsystem, which should be based on the model of cyber incident identification based on fuzzy rules and fuzzy inference, where causal relationships of a cyber incident and its sighs are described by the expert in plain language and then formalized as a set of fuzzy logical rules.

The simulation results show that the proposed method allows us to increase the accuracy of cyber incident detection by 2-15%.

The obtained results can be used in practice for solving the problem of detecting cyber incidents by SIEM, which is part of the SOC software and hardware.

References

1. I. Subach, V. Kubrak, and A. Mykytiuk, "Architecture and functional model of a promising proactive intelligent system SIEM-system for cyber protection of critical infrastructure objects", Information Technology and Security, Vol 7., Iss. 2., 2019, pp. 208-215, DOI: 10.20535 / 2411-1031.2019.7.2.190570, Access mode: https://doi.org/10.20535/2411-1031.2019.7.2.190570
2. Law of Ukraine On the Basic Principles of Cyber Security of Ukraine: Official Publication: Vidomosti Verkhovnoi Rady, 2017, № 45, Art. 403.
3. I. Subach, V. Kubrak, and A. Mykytiuk, "Methodology of rational choice of security incident management system for building operational security center", CEUR Workshop Proceedings, 2019, 2577, p.p. 11-20, Режим доступу: http://ceur-ws.org/Vol-2577/paper2.pdf
4. A. Fedorchenko, D. Levshun, A. Chechulin, and I. Kotenko, "Analysis of methods for correlating security events in SIEM systems. Part 1 ", Proceedings of SPIIRAN, issue 4 (47), 2016, pp. 5-27, DOI: 10.15622 / sp.47.1.
5. A. Fedorchenko, D. Levshun, A. Chechulin, and I. Kotenko, "Analysis of methods for correlating security events in SIEM systems. Part 2 ", Proceedings of SPIIRAN, issue 6 (49), 2016, pp. 208-225, DOI: 10.15622 / sp.49.11.
6. Elshoush H.T., Osman I.M. Alert correlation in collaborative intelligent intrusion detection systems — A survey // Applied Soft Computing, 2011, pp. 4349–4365.
7. Muller A. Event Correlation Engine. Master`s Thesis. Swiss Federal Institute of Technology Zurich. 2009. 165 p.

8. Jakobson G., Weissman M.D. Alarm correlation // IEEE Network. 1993. no. 7(6). pp. 52 59.
9. Tiffany M. A survey of event correlation techniques and related topics. URL: http://www.tiffman.com/netman/netman.html (дата обращения: 26.04.2016).
10. Sadoddin R., Ghorbani A. Alert Correlation Survey: Framework and Techniques // Proceedings of 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST`06). 2006. Article no. 37.
11. Hanemann A., Marcu P. Algorithm Design and Application of Service-Oriented Event Correlation // Proceedings of Conference BDIM 2008, 3rd IEEE/IFIP International Workshop on Business-Driven IT Management. 2008. pp. 61–70.
12. Yu. Samokhvalov, and S. Tolyupa. "Correlation of events in SIEM-systems based on non-monotone inference", Information protection, .Volume 19, № 1, 2017, pp. 5-9.
13. L. Zade, The concept of a linguistic variable and its application to approximate decision making, Moscow, Russia: Mir, 1976.
14. A.P. Rothstein, Medical diagnostics on fuzzy logic, Vinnytsia, Ukraine: Continent-PRIM, 1996.
15. A.P. Rothstein, Intelligent identification technologies: fuzzy sets, genetic algorithms, neural networks, Vinnytsia, Ukraine: UNIVERSUM, 1999.
16. SIEM Rules or Models for Threat Detection? Exabeam, 2018.[Online]. Available: https://www.exabeam.com/siem/siem-threat-detection-rules-or-models/. Accessed on: November29, 2020.
17. F. Salo, M. Injadat, A. Nassif, A. Shami, and A. Essex, ''Data Mining Techniques in Intrusion Detection Systems: A Systematic Literature Review,'' in Proc. IEEE*Access*, September 2018, Vol. 6, pp. 56046–56058. DOI:10.1109/ACCESS.2018.2872784.