Development of an Analytical Data Processing System for Monitoring Information Security of an Informatization Object's Information Support's Structure Models

Valerii Sizov [0000-0002-4844-4714] and Aleksei Kirov [0000-0002-8424-3071]

Plekhanov Russian University of Economics, Stremyanny lane 36, 117997 Moscow, Russia

Sizov.VA@rea.ru, Kirov.AD@rea.ru

Abstract. The work is aimed at improving the efficiency of information security management of an informatization object based on optimizing the structure of information support for an automated information security monitoring data processing system.

It is considered, that a distributed information security monitoring data processing system built on the basis of a universal information security event management system. The technical support of this system is a local computer network, which includes information security tools. Information security tools and network nodes are the main sources of information about information security events. Taking into account the large volumes of data on information security (IS) events in the analytical data processing system (ADPS), it is necessary to optimize the ADPS information support's structure, taking into account the structure and technical characteristics of the LAN.

The article proposes mathematical models for optimizing the structure of information support for an automated information security monitoring data processing system according to the maximin's criterion of information support for ADPS usefulness and the criterion for the maximum relevance of information support distributed over LAN nodes.

All problems are reduced to typical problems of integer mathematical programming, for the solution of which classical well-known methods can be used.

The proposed approach makes it possible to increase the efficiency of the procedure for identifying information security incidents by organizing a rational exchange of information in an automated data processing system for information security of an informatization object's monitoring, taking into account the usefulness of analytical data processing procedures.

Keywords: Information Security, Security Information Event Management, Information Support, Structure Models, Theory of Utility.

1. Introduction

Currently, in the field of information security (IS) management of modern objects of informatization, a group of tasks for managing IS incidents is distinguished, which

Proceedings of the of the XXIII International Conference "Enterprise Engineering and Knowledge Management" (EEKM 2020), Moscow, Russia, December 8-9, 2020.

includes the following main tasks: monitoring IS events of informatization objects and identifying IS incidents; registration of information security incidents; analysis of information security incidents; informing the administration about all cases of information security violations; collection of evidence and evidence for response to incidents of information security and others. From a practical point of view, one of the most effective approaches to creating information security monitoring is the SIEM-systems usage [1-4].

SIEM class solutions provide management of information and security events, implementing the functions of collecting and storing, processing and analyzing registered security events in order to identify and analyze incidents, as well as check the compliance of the IS management system with existing requirements and standards [5]. At the same time, most SIEM solutions include components of an analytical data processing system (ADPS) for monitoring information security of an informatization object, for example, a technology for identifying dependencies between individual information security events, used indicators of the state of the protected infrastructure, etc.

However, the main disadvantage of such systems is the relatively long period required for them to analyze the data and make a decision about whether this information security event, or their combination, is an information security incident or not [6-8].

This drawback is based on the contradiction between the distributed nature of information sources about an information security event (as a rule, these sources are information security tools integrated in a local area network (LAN)) and a centralized way of making decisions on actions with information security incidents.

To resolve this contradiction it is necessary, on the one hand, to provide the decision-making process with the most complete information, and on the other hand, this information must be relevant. Considering the large volumes of data on IS events in ADPS, it is necessary to optimize the structure of information support for ADPS, considering the structure and technical characteristics of the LAN. Therefore, the task of developing models of the structure of information support of ADPS for monitoring the information security of an object of informatization is relevant.

2. Statement of developing information support structure models for an analytical data processing system for monitoring the information security of an informatization object using the theory of utility

In this work, it is assumed that ADPS for information security monitoring, operating in the information security management system (ISMS) of an informatization object, built on the basis of a LAN, is used as an organizational and technical form of information security processes management, including information security monitoring processes occurring in real time. Such systems place increased demands on the efficiency of data processing in real operating conditions. For example, a fast response ISMS for monitoring the information security of an object related

to a critical information infrastructure impose increased requirements both on the safety of the data used in these systems and on the promptness of their processing. The solution to the problem of effective data processing in such systems is relevant and requires taking into account a number of factors, such as: distributed data processing, strict time constraints for obtaining a response to a request when making a decision, large amounts of data for analytical research and identification of information security incidents, etc.

Output information is understood as information obtained because of the performance of data analytics functions and issued to the object of its activity, users, or other systems. The quality of output information in ADPS for monitoring the information security of an informatization object is understood as a set of information properties that determine its suitability to meet the needs of a security officer in the timely identification and investigation of information security incidents.

One of the main systemic methods for improving the quality of output information in ADPS for monitoring the information security of an informatization object, aimed at improving the probabilistic and temporal characteristics of the functioning of systems, is computer modeling.

The features of the ADPS functioning for monitoring the information security of an informatization object based on a LAN allow, when solving problems of improving the quality of output information using computer modeling, in addition to traditional means of determining the optimal points from the corporate information system information security level's point of view storage points of information elements, use the methods of theory utility, allowing to evaluate the useful effect of placing information elements in certain computing nodes [9]. They take into account the property of information reliability not only as an error-free property (that is, the correspondence of the data received by the consumer with the data generated at the source), but also as a property of relevance (maintaining a sufficient degree of compliance with the real state of accounting objects at the time of using this information).

The essence of the application of the method based on the theory of utility and computer modeling to determine the computational nodes that are the storage location for information elements to optimize the use of these nodes and the distribution of information elements among the nodes is that, in most cases, information security incident management tasks make it possible to pre-select the necessary datasets (intermediate arrays) and distribute them across the LAN nodes for immediate use in the future, i.e. it is necessary to define in advance the information exchange of data on IS events in ADPS. The information support of ADPS for monitoring the informatization object information security can include both the data stating the informatization object information security state itself, obtained directly from information security services, as well as their copies and / or prehistories received in LAN nodes in places where they are used by ADPS for monitoring the informatization object information security[10-11].

Thus, in some cases it is allowed

$$[X(t) - X(t-\tau)] \leq e,\ e > 0,\ \forall \tau, \tau \in [0, t] \quad (1)$$

where X (t), X (t- T) are the corresponding values of any parameter of the information element at time t and (t- T), S is the absolute limit of the permissible deviations of the real degree of utility of the item from its expected degree of utility. The fulfillment of inequality (1) determines the usefulness of the information element when it is processed at the LAN node on the interval T. In this case, the degree of usefulness of the information support of ADPS is understood as the probability of the relevance of all its constituent elements on the interval T.

One of the main tasks of synthesizing the structure of information support of ADPS for monitoring the information security of an informatization object, solved at the stage of predesign analysis, is the task of determining the optimal content of ADPS information support and its placement on LAN nodes.

Let J - the number of LAN nodes, I - the number of information elements of the system; a; is the relative utility of the i-th information element (the degree of utility determined for a given time period T), t-j, is the time of transmission of the i-th information element from the j-th to the j'-th LAN node, b; is the value of the i-th information element's volume, W = (Ajj) is a matrix of interconnections of information sources (LAN nodes) and information elements,

$$\lambda_{ij} = \begin{cases} 1, \text{if the source of i} - \text{th information element} \\ \text{is j} - \text{th LAN node,} \\ 0 \text{ otherwise.} \end{cases} \quad (2)$$

P = (Pij) is the marginal utility matrix, where P,j are the weighted estimates of the utility for the user, obtained from placing each additional i- th information element in the j-th LAN node,

$P_{ij} \in [0,1]$, $\forall i$, i = 1,I,$\forall j$, j = 1,J.

The ways of constructing the matrix P are determined by the specific conditions of using ADPS for monitoring the information security of an informatization object. In particular, the estimate of the marginal utility for the user as a result of placing each additional i-th information element in the j-th LAN node can be made up of objective estimates (for example, the time of transmission of the i-th information element to the j-th node from other LAN nodes) and subjective assessment of the probability of obtaining a positive marginal utility in relation to the need to transfer the i-th information element to the j-th node of the LAN [12].

The usefulness of the i-th information item is determined with taking into account the quality of the analytical procedures performed in the j-th node using the i-th information item. Often, the quality of the procedures contradicts the required probabilistic and temporal characteristics of the information security monitoring process and identifying information security incidents. For example, processing procedures, procedures of neural and neuro-fuzzy data processing technologies, etc. [13] Therefore, the value of the positive marginal utility of placing a copy and

/ or the prehistory of an information element about an information security event of an informatization object in a specific LAN node, including information security services, is mainly influenced by the following factors:

The effectiveness of the procedure for processing data on the state of information security, performed in this LAN node;

The required completeness and relevance of the initial data for identifying information security incidents.

Then, using the following variables:

$$x_{ij} = \begin{cases} 1, \ if\ i-th\ information\ element \\ is\ placed\ in\ j-th\ LAN\ node, \\ 0, otherwise. \end{cases} \quad (3)$$

The degree of ADPS information support's usefulness for monitoring the information security of an informatization object $E(x_{ij})$ can be determined by the following formula:

$$E(x_{ij}) = 1 - \prod_{i=1}^{I}(1 - a_i \sum_{j=1}^{J} \sum_{j'=1}^{J} x_{ij}\lambda_{ij'}t_{jj'}^{i}). \quad (4)$$

In cases where geographically distributed systems consist of homogeneous elements based on the degree of risk of information security incidents, it is advisable to use the maximin criterion of the usefulness of information security of an information security object as the main criterion for synthesizing an ASOD for monitoring the information security of an informatization object for monitoring the information security of an informatization object, for complex, complex systems with a low degree of centralization of management, it is advisable to use the criterion of the maximum relevance of information support distributed over LAN nodes.

The task of designing the information support of ADPS for monitoring the information security of an informatization object structure according to the first criterion is as follows.

To find:

$$\max \min_j \sum_{i=1}^{I} P_{ij}X_{ij},, \quad (5)$$

with restrictions:

- on the degree of ADPS for monitoring the information security of an object of informatization information support's usefulness:

$$1 - \prod_{i=1}^{I}\left(1 - a_i \sum_{j=1}^{J} \sum_{j'=1}^{J} x_{ij}\lambda_{ij}t_{ij'}^{i}\right) \geq E^* \quad (6)$$

where E* is the minimum permissible degree of ADPS for monitoring the information security of an object of informatization information support's usefulness;

for the relative time of information elements transfer between LAN nodes;

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{j'=1}^{J} x_{ij}a_i\lambda_{ij'}t_{jj'}^{i} \leq T^* \quad (7)$$

where T^* is the maximum permissible relative time of transmission of information elements (on the interval τ) between LAN nodes;

- on the amount of memory of the *j*-th LAN node;

$$\sum_{i=1}^{I} x_{ij}\, b_i \le B_j^*, \forall j, j = \overline{1J} \tag{8}$$

where B_j^* is the maximum allowable memory size of the j-th node;

- for the absence of duplication of the information element in the LAN nodes:

$$\sum_{j=1}^{J} x_{ij} = 1, \forall i, i = \overline{1,I} \tag{9}$$

Statement 1. The solution to problem (1) - (5) is admissible for $E * \Longrightarrow 1$ and the fulfillment of constraints (6) - (9).

Evidence. Let

$$\phi_i = \sum_{j=1}^{J} \sum_{j'=1}^{J} X_{ij} a_i \lambda_{ij'} t_{jj'}^{i},$$

Then constraints (6), (7) will have the following form:

$$\begin{cases} 1 - \prod_{i=1}^{I} (1 - \phi_i) \le E^* \\ \sum_{i=1}^{I} \phi_i \le T^* \end{cases} \tag{10}$$

Taking the logarithm of the first inequality of system (10) and expanding the logarithm function in a power series, we successively obtain:

$$\begin{cases} \sum_{i=1}^{I} \ln(1 - \phi) \le \ln (1 - E^*) \\ \sum_{i=1}^{I} \phi_i \le T^* \end{cases} \quad \begin{cases} \sum_{i=1}^{I} \phi_i \le \ln(1 - E^*) - \chi \\ \sum_{i=1}^{I} \phi_i \le T^* \end{cases}$$

where χ is the remainder of the series.

Obviously, $E^* \Longrightarrow 1$, constraint (7) is more stringent. The proposed analysis of the rigidity of constraints can be used to reduce the dimension of tasks with specific initial data, especially for information security monitoring systems for complex objects of informatization of large dimensions and intensive exchange of information.

Modern ADPS for monitoring the information security of an informatization object, operating on a LAN basis, is critical to the amount of information transmitted through communication channels. Therefore, the following problem of determining the ADPS for monitoring the information security of an informatization object information support structure by the criterion of uniform usefulness of its components can be considered the most urgent.

To find:

$$\max_{X_{ij}} \min_{j} \sum_{i=1}^{I} P_{ij} X_{ij}$$

subject to constraint (7).

This problem is reduced to the problem of maximization by introducing additional variables y = {0,1}. It looks like this.

To find:

$$\max_{\{x_{ij}y_j\}} \sum_{j=1}^{J} y_j \sum_{i=1}^{I} P_{ij}X_{ij}$$

with restrictions:

$\sum_{j=1}^{J} y_j \sum_{i=1}^{I} P_{ij}X_{ij} < \sum_{i=1}^{I} P_{ij}X_{ij} \forall j, j = \overline{1,J}$,

$\sum_{j=1}^{J} y_j = 1$,

$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{j'=1}^{J} x_{ij} a_i \lambda_{ij'} t_{jj'}^{i} \le T^*$.

Let

$t_{jj'}^{i} = \text{const}, \forall j, j = \overline{1,J}, \forall j', j' = \overline{1,J}, j \ne j^* \forall i, i = \overline{1,I}$.

Then the solution to the problem of determining the optimal content of the information support of the ADPS for monitoring the information security of the informatization object components and their placement in the LAN nodes is reduced to solving M multidimensional knapsack problems, which are formulated as follows.

To find:

$$\sum_{i=1}^{I} P_{ij^*} X_{ij^*}, \forall j^*, j^* = \overline{1,J}$$

with restrictions:

$\sum_{j=1}^{J} \sum_{i=1}^{I} x_{ij} a_i \le D, D = \frac{T^*}{t_{jj'}^{i}}$,

$\sum_{i=1}^{I} (p_{ij^*} + a_i) + \sum_{i=1}^{I} (a_i - p_{ij^*}) x_{ij'} + \sum_{j=1}^{J} \sum_{i=1}^{I} x_{ij} a_i \le D, \forall j', j' = \overline{1,J}, j' \ne j^*$.

The result of solving the problem of developing the information support of the analytical data processing system for monitoring the information security of the informatization object's structure is the optimal according to the given criteria (including the general and / or marginal utility), the composition of the information support components of ADPS and their placement on the LAN nodes.

3. Conclusion

Thus, in this article, the problem of determining a rational structure of ADPS

for monitoring the information security of an informatization object's information support based on modern SIEM systems is considered.

The formalization and solution of this problem are based on the methods of utility theory and operations research, which allow using computer modeling to determine the optimal composition of the ADPS information support components and their distribution among LAN nodes from the point of view of the general usefulness of information, taking into account the analytical information technologies used in these nodes for identifying information security incidents.

In general, this approach makes it possible to increase the efficiency of the procedure for identifying information security incidents by organizing a rational exchange of information between LAN nodes (information protection means), taking into account the characteristics of analytical data processing procedures.

References

1. Sizov V.A., Kirov A.D. Problems of SIEM-systems implementation in the practice of information security management of economic entities. Open education, 2020
2. Lee J., Kim Y.S., Kim J.H., Kim I.K. Toward the SIEM architecture for cloud-based security services. *Communications and Network Security IEEE Conference*, 2017. DOI: 10.1109/CNS.2017.8228696
3. G.G., El-Barboni M, Debar H. New Types of Alert Correlation for Security Information and Event Management Systems. *New Technologies, Mobility and Security IFIP International Conference*, 2016. DOI: 10.1109/NTMS.2016.7792462.
4. Kavanagh M., Rochford O. Magic Quadrant for Security Information and Event Management. *Gartner technical report*. 2015. 15 p.
5. Markov A.S., Tsirlov V.L. *Strukturnoye soderzhaniye trebovaniy informatsionnoy bezopasnosti. Monitoring pravoprimeneniya*. 2017. № 1(22). P. 53-61. DOI: 10.21681/2412-81632017-1-53-61.
6. Nabil M., Soukainat S., Lakbabi A., Ghizlane O. SIEM selection criteria for an efficient contextual security. In: *2017 International Symposium on Networks, Computers and Communications (ISNCC)*, 2017. DOI: 10.1109/ISNCC.2017.8072035.
7. Kotenko I.V., Fedorchenko A.V., Sayenko I.B., Kushnerevich A.G. *Tekhnologii bolshikh dannykh dlya korrelyatsii sobytiy bezopasnosti na osnove ucheta tipov svyazey. Voprosy kiberbezopasnosti.* 2017. № 5 (24). P. 2-16. DOI: 10.21681/2311-3456-2017-5-2-16.
8. Fedorchenko A.V., Levshun D.S., Chechulin A.A., Kotenko I.V. *Analiz metodov korrelyatsii sobytiy bezopasnosti v SIEM-sistemakh. Part 2. Trudy SPIIRAN*. 2016. no. 49. P. 208-225. DOI: 10.15622/sp.49.11.
9. Kirsanov K.K. The theory of utility in the period of conceptual provisions change. *Bulletin of Eurasian Science*. 2015. No. 2 (27).
10. Sizov V.A. Models and methods of virtual-restorative data backup of automated information-control systems in emergency situations. *Journal of Automation and Telemechanics*, No. 7, 1998 p. 176-184.
11. Sizov V.A. Development of models for increasing the efficiency of data safety in a distributed computing environment based on dynamic data backup. *Collection of articles of the XXI International Scientific and Practical Conference "Advances in Science and Technology"*. Moscow, Actuality.RF Publ., 2019. pp. 96-100. URL: http://xn--80aa3afkgvdfe5he.xn--p1ai/AST-21 originalmaket N.pdf

12. Uralsky N.B., Sizov V.A. Development of a modified genetic algorithm for solving the problem of parallelizing multiplication of high-dimensional matrices in distributed data processing systems. In: *IX International Scientific and Practical Conference "Innovative Development of the Russian Economy": in 6 volumes*. Moscow: FGBOU VO "PRUE im. G. V. Plekhanov ", 2016.
13. Mikryukov A.A., Babash A.V., Sizov V.A. Classification of events in information security systems based on neural network technologies. *Open Education*. 2019; 23(1):57- 63. DOI: https://doi.org/10.21686/1818-4243-2019-1-57-63