# Models and algorithms for analyzing information risks during the security audit of personal data information system

Yuliia Kostiuk[1,†], Pavlo Skladannyi[1,*,†], Volodymyr Sokolov[1,†], Hennadii Hulak[1,†] and Nataliia Korshun[1,†]

[1] *Borys Grinchenko Kyiv Metropolitan University, Bulvarno-Kudryavska Str., 18/2, Kyiv, 04053, Ukraine*

**Abstract**

Security audits of information systems, including systems containing personal data, are a necessary step to ensure an adequate level of security in the face of growing cyber threats. Given the complexity and diversity of risks that may affect the confidentiality, integrity, and availability of information, effective tools for assessing and formulating recommendations are key to optimizing resources and strengthening security. The proposed risk analysis model uses methods that allow you to focus on identifying hidden threats, which significantly increases the efficiency of auditors. The intelligent system used for personal data security audits determines the level of security, assesses possible risks, and suggests measures to reduce potential threats. This approach contributes to the continuous improvement of the security and reliability of information systems in the context of the dynamic evolution of technologies and cyber threats.

**Keywords**

information security, personal data, risk assessment, fuzzy logic, model, decision support system, security audit, risk management

## 1. Introduction

Modern businesses are heavily dependent on information technology, making information system security and risk management strategically important issues. Information is becoming a key business asset, encompassing both static resources (databases, hardware configurations) and dynamic data processing processes. At the same time, the introduction of the latest technologies is accompanied by an increase in the threats of unauthorized access, privacy violations, and system failures [1–3].

Legal requirements, such as the GDPR in the European Union [4] and the Law of Ukraine “On Personal Data Protection” [5], which oblige organizations to conduct regular audits and assess potential threats, play an important role. In addition, the rapid development of technology, including the introduction of cloud services, the Internet of Things, and artificial intelligence, creates new vulnerabilities that must be taken into account when analyzing information security. Unaccounted-for risks can lead to significant financial losses, fines, and loss of user confidence [6]. Effective risk management contributes to the continuity of business processes, which guarantees the stable operation of information systems [7]. Particular attention should be paid to the protection of personal data at critical infrastructure facilities [8, 9], as their compromise can have large-scale consequences for national security, the economy and society. A reliable cybersecurity system for such facilities should include multi-level control mechanisms, regular threat monitoring, and the use of advanced encryption and authentication technologies.

*CH&CMiGIN’24: Third International Conference on Cyber Hygiene & Conflict Management in Global Information Networks, January 24–27, 2024, Kyiv, Ukraine*

[*] Corresponding author.

[†] These authors contributed equally.

y.kostiuk@kubg.edu.ua (Y. Kostiuk); p.skladannyi@kubg.edu.ua (P. Skladannyi); v.sokolov@kubg.edu.ua (V. Sokolov); h.hulak@kubg.edu.ua (H. Hulak); n.korshun@kubg.edu.ua (N. Korshun)

0000-0001-5423-0985 (Y. Kostiuk); 0000-0002-7775-6039 (P. Skladannyi); 0000-0002-9349-7946 (V. Sokolov); 0000-0001-9131-9233(H. Hulak); 0000-0003-2908-970X (N. Korshun)

## 2. Related work

An effective security audit of personal data information systems is an integral part of information security management. It allows you to assess the security of assets, analyze risks, and propose measures to improve information protection. However, the specifics of personal data make such an audit difficult to conduct, as data loss may be detected with a delay and the consequences are difficult to determine. There is often uncertainty in making decisions about the level of security and compliance with regulatory requirements [10–16].

An analysis of the literature shows the need for further research in the field of personal data protection. It is necessary to develop models, algorithms, and risk analysis tools to ensure prompt and effective decision-making during the security audit of personal data information systems [17–24]. They should take into account the specifics of the systems and current regulatory requirements. In general, the main purpose of the security system is to ensure the stable operation of the enterprise, prevent threats, protect legitimate interests, and prevent information loss. It should include classification of information, forecasting of threats, creation of response mechanisms, and increase of cost-effectiveness of protective measures.

## 3. Problem statement

Risk is a combination of the probability of an event and its consequences. In most cases, the meaning of risk $R$ is defined as follows:

$$R = P_i C_i, \tag{1}$$

where $P_i$ is the probability of successful implementation of the$i$ threat, $C_i$ is an assessment of the damage caused by the impact of an information security incident in the event of successful implementation of the $i$ threat. A threat is understood as a potential cause of an incident that can cause damage, so various methods are used to assess risks, including a basic vulnerability score, a full overlap protection model, an average risk value, and a loss probability function.

Information system security auditors should have access to effective tools that allow not only assessing the level of security but also formulating reasonable recommendations and developing countermeasures to improve it [13–15, 24–26]. This provides a favorable environment for in-depth analysis and complex tasks, which is important for increasing the professional efficiency of auditors. Risks to information systems, which can be caused by both technical errors and malicious acts, threaten basic security properties such as confidentiality, integrity, and availability of data. Many threats arise from insufficient adaptation of the infrastructure to changes in the external environment or internal inconsistencies in security measures, making risk analysis a crucial step in establishing the optimal level of protection, taking into account the allowable costs.

## 4. Proposed solution

In the case of information security risk analysis, a risk can be understood as any distribution $P$ belonging to a set of various distributions $R$ which is the set of outcomes

$$S \times D \longrightarrow R, \tag{2}$$

where $S$ is the set of system states; $D$ is the set of different decisions.

According to the definition of event probability, the process of risk analysis can be viewed as obtaining expert estimates of the frequency of realization of the threat $Y$ in information systems over a given period. In the absence of statistical data or the emergence of new threats, such as an anti-virus storm, the Bayesian approach is effectively used to quantify the factors that determine the level of security, because, in statistical problems, several probabilistic hypotheses are often considered, which change with the receipt of new information. The main advantage of the Bayesian approach is the ability to dynamically adjust probabilities that reflect the degree of confidence in various threat models based on current data, which allows you to obtain a posteriori estimates of the probability of information security incidents, track the receipt of new statistical data, determine the interdependencies between risk factors, and formulate logical conclusions with a physical interpretation of changes in the structure of the problem's dependencies.

The basis of the Bayesian approach is Bayes' theorem, the main theorem of probability theory, which allows us to determine the probability that a certain event has occurred based on statistical data. In the case of analyzing the level of security of information system resources, we consider a random variable $Y$, which has a probability density $p(y|\theta)$ with parameters $\theta$. Based on the obtained statistical data, we can conclude about another random variable $\theta$, which has a probability distribution $\pi(\theta)$. Then, according to the Bayes formula:

$$p(\theta|y) = \frac{p(y|\theta)\, p(\theta)}{p(y)}. \tag{3}$$

The main characteristics of the security of personal data information system resources are the following tuple of indicators: the ability to ensure confidentiality ($C$), integrity ($I$), and availability ($A$) of information in the face of possible threats.

In the event of a threat of a certain type to the studied or similar resource equipped with the same information security systems and no violations of confidentiality, integrity, and availability, it is possible to calculate a posteriori probability of hypotheses for a particular data source.

When collecting facts, the probabilities of hypotheses will increase if the facts support them or decrease if the facts refute them. If three indicators are obtained simultaneously, provided they are independent, the appropriate formula is used:

$$P(\theta_i|C,I,A) = \frac{P(C|\theta_i)\; P(I|\theta_i)P(A|\theta_i)P(\theta_i)}{\sum_{i=1}^{3}(C|\theta_i)\; P(I|\theta_i)P(A|\theta_i)P(\theta_i)}. \tag{4}$$

It is worth noting that it is obvious that if the results of the experiment indicate that the information security system did not provide a tuple of security indicators when exposed to a threat, then opposite scenarios should be considered. This means that it is necessary to carefully analyze the factors that contributed to the ineffectiveness of security measures and identify contradictions between the planned and implemented security measures. The use of risk analysis models and algorithms can help identify and resolve problems that arise when ensuring the security of personal data information systems:

$$P(\overline{C,I,A}|\theta_i) = 1 - P(C,I,A|\theta_i). \tag{5}$$

The use of the Bayesian approach to analyzing information risks during the security audit of personal data information systems is relevant due to the ability of this method to quantify the probability of threats, taking into account new information and the specifics of personal data processing. The application of Bayes' theorem allows us to systematize data on risks and vulnerabilities, evaluate the effectiveness of security measures, and determine optimal protection strategies [10, 11, 16, 19, 27, 28]. Particular attention should be paid to the confidentiality, integrity, and availability of personal data when identifying risks and developing protection strategies. Risk assessment, which takes into account the likelihood of a threat and potential losses, helps to increase the effectiveness of the information security system and helps to identify its weaknesses. Risk assessment $R$ is defined as follows:

$$R = \sum_i P_i C_i^y, \tag{6}$$

where $P_i$ is the probability of successful realization of the $i$th threat, $C_i^y$ is the damage estimate in case of successful realization of the $i$th threat, $i = 1 \dots n$ is the number of possible threats.

Despite the importance of risk analysis, there are difficulties in using expert judgment, as it complicates the interpretation of the results. It is advisable to develop methods based on quantitative expert assessments obtained in monetary terms to ensure objectivity and convenience. The solution to these problems determines the relevance of further research in the field of risk analysis and management in information security.

## 5. Risk analysis and management

The concept of ensuring the information security of personal data systems emphasizes that audit is a key component of the personal data security management cycle. There are three main types of audit: active, expert, and compliance audits, to ensure its correctness, an integrated approach with a combination of these types, as well as an analysis of security risks, is required [12–15, 24, 25, 28, 29]. Particular attention is paid to the study of decision support methods during the audit and the review of tools for automating the processes of verification of personal data information systems. The importance of the intellectual support of an expert with the use of data mining technologies to improve the efficiency of the audit is emphasized.

Risk ($R$) is considered by the vast majority of experts to be a complex value that involves the existence of such factors as threats, vulnerabilities, and the damage itself, and is expressed using a formula:

$$R = \lambda P_T P_V(z), \tag{7}$$

where $\lambda$ is the amount of damage (loss) in the event of an information asset security breach; $P_T$ is the probability of a threat occurring; $P_V(z)$ is a function describing the probability of a threat to an information asset depending on the cost of providing protective measures, where $z$ is the cost of providing information asset protection in monetary terms.

The amount of damage is determined exclusively by the protected information, the probability of a threat is a fixed value, and the probability of a threat being realized can be reduced by investing ($z$) in the information security of the asset [12, 19, 25]. There is a tendency to reduce the likelihood of a threat with increasing investment, but assessing the likelihood of threats, in particular, unauthorized access to protected information becomes difficult when taking into account the intentional actions of people. Increasing funding for information security reduces the likelihood of a threat exponentially [20, 26–29].

Then the function and graph of the dependence (Figure 1) of the probability of a threat on information security costs will look like this:

$$P_T(z) = te^{-\varphi \cdot z}, \quad \begin{Bmatrix} \forall(t) \in [0; 1], \\ \forall(z) \in R, \end{Bmatrix}, \tag{8}$$

where $t$ is the probability of a threat; $\varphi$ is the correction factor for information security costs; $z$ is the cost of protecting an information asset in monetary terms.

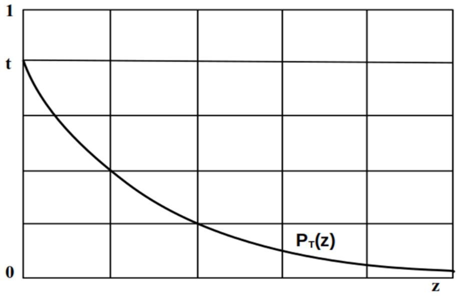


**Figure 1:** Dependence of the probability of a threat ($P_T$) on investments in information security ($z$).

The experience of information attacks shows that increased investment in information security for an organization’s valuable assets reduces the likelihood of a threat and increases the likelihood of its implementation. This is confirmed by the facts of hacking of government websites, law enforcement databases, and security systems of banks and corporations by professional hackers and attackers who have extensive experience and use advanced technologies to find weaknesses in protection [14, 15, 24–26, 29]. The proposed function and graph of dependence (Figure 2) of the probability of threat realization on information security costs will be in the form of a quadratic function:

$$P_V(z) = \frac{\alpha-\beta}{z_0^2}(z-z_0)^2+\beta, \quad \left\{\begin{matrix} P_V(z) \in [\alpha;\beta], \\ \forall \alpha, \beta \in [0;1], \\ \alpha \geq \beta, \\ \forall z \in R. \end{matrix}\right\}. \tag{9}$$

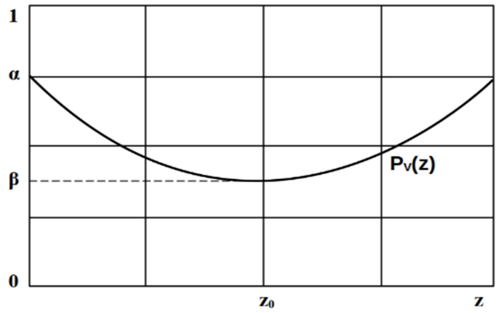


**Figure 2:** Risk management dependence of the probability of realization of the threat $(P_V)$ on investments in information security $(z)$.

The$z$ parameter indicates the amount of money allocated to protect the information asset. The parameters $\alpha$ and $\beta$ are the upper and lower limits of the probability of a threat being realized, determined by expert estimates. It is important to note that the inability to specify a lower bound on the probability of a threat materializing is a drawback of the known models. Even with significant investments in security, the probability of damage cannot be reduced to zero, as current experience shows. An example is the probability of various events such as accidents or disasters.

Determining the amount of money for protective measures at which the value of the probability function of the threat reaches the lower bound $\beta$ presents certain difficulties. The Gordon-Loeb economic model confirms that the optimal level of investment does not exceed $\frac{1}{e} \approx 36.8\%$ of the total losses in the event of an information asset security breach, which has been experimentally confirmed empirically [1–3, 17]. Given the lack of statistical data, and based on intuitive considerations, it seems appropriate to link the value of $z_0$ to a value equal to $\frac{1}{e} \approx 36.8\%$ of the amount of damage due to a breach of information security of an asset. Thus, the function of the dependence of the probability of a threat realization on information security costs will be as follows:

$$P_V(z) = \frac{\alpha-\beta}{\left(\frac{\lambda}{e}\right)^2}\left(z-\frac{\lambda}{e}\right)^2+\beta, \tag{10}$$

where $\lambda$ is the amount of losses in the event of an information asset security breach.

It should be noted that the study considers several alternative classes of functions of dependence on the probability of threat realization on information security costs. The choice of quadratic dependence in this work is due to the difficulty of determining additional parameters (coefficients) in alternative functions and their interpretation for providing expert assessments.

The task of finding the optimal level of investment in information security will then look like this:

$$C = \lambda P_T(z)P_V(z) + z = \lambda t e^{-\varphi \cdot z}\left(\frac{\alpha-\beta}{\left(\frac{\lambda}{e}\right)^2}\left(z-\frac{\lambda}{e}\right)^2+\beta\right) + z \longrightarrow \min. \tag{11}$$

The development of a threat model includes the stages of identifying types of information assets for risk assessment, determining the environment and sources of threats for each type of asset, which depends on the organization's information security needs, and the ratio of the cost of protection to the risk. The process of risk identification is cyclical: first, a general threat scheme is formed, an expert cost estimate is made, and then significant risk factors are identified for which detailed models are developed. Threats from people require a high level of detail, in particular for internal and external perpetrators, taking into account their motives and typical actions. The risk of a security breach is determined based on expert assessments of the likelihood of a threat, its realization, and the severity of the consequences. The optimal level of investment in security is determined through nonlinear programming, in particular with the help of the Mathcad software product, which provides a convenient interface for calculations and analysis [3, 11–14, 16, 18, 19, 25–28]. Analysis of possible losses from threats to integrity, availability, and confidentiality, including fines for violation of regulations and the cost of information recovery, is a key step in the risk assessment process:

$$S = S_{vst} + Y + S_2, \tag{12}$$

where $S_{vst}$ is the cost of recovery for the threat group, which is determined by the formula:

$$S_{vst} = \sum S_1, \tag{13}$$

where $S_1$ is the cost of recovery from the threat; $S_2$ is the cost of compensation to personal data subjects; $S_2 = Y_4 \cdot n, n$is the number of lawsuits from personal data subjects; $Y$ is the maximum fine calculated by the formula:

$$Y = \{Y_1, Y_2, Y_3, Y_4\}, \tag{14}$$

where $Y_1 = \{Y_{11}, Y_{12}, Y_{13}, Y_{14}, Y_{15}\}$.

The probability of realization of the identified and assessed threats is assessed by an expert method:

$$V_{ry} = \sum_{i=1}^{n} k_i\, \alpha_i, \tag{15}$$

where $k_i$ is the answer to the$i$ question of the questionnaire; $\alpha_i$ is the importance coefficient determined by the expert method and satisfying the condition

$$\sum_{i=1}^{n} \alpha_i = 1. \tag{16}$$

The degree of criticality of threat groups is proposed to be determined based on the calculated value of the risk factor

$$K_d = \frac{b}{N_y}, \tag{17}$$

where $K_d$ is the risk coefficient; $b$ is the consequences of the threat (asset value); $N_y$ is the number of threats in the group; $d$ is the risk measure.

The next step is to compare indicators for threat groups. Alternatives to $\alpha_i$ are compared for each indicator and threat groups are ranked:

$$\alpha_i = \{b, c, d\}. \tag{18}$$

The critical group of threats is determined by the risk coefficient, for this purpose, the assessments of alternatives are compared:

$$\varphi(\alpha_i)\varphi(\alpha_i) = \{K_d\}. \tag{19}$$

To assess the security risk of personal data information systems, an object-oriented assessment model can be used, based on the use of the Petri net apparatus, which is distinguished by the developed rules for triggering transitions and allows taking into account the degree of risk to personal data information systems, as well as the probability of realization and repulsion of personal data security threats. The method for formalizing a mathematical object-oriented risk assessment model based on colored Petri nets is as follows:

$$R =< P,\ V_{ry}, T, I, O >, \tag{20}$$

where $P$ is the set of Petri net states; $V_{ry}$ is the set of probabilities of threats realization; $T$ is the set of transitions that determine the rules for changing the network states; $I$ is the input positions (set of parameters of threats and countermeasures); $O$ is the output positions (set of residual risk values). The advantage of the model is the ability to implement the following features: a probabilistic network allows you to take into account both personal data risks and countermeasures to counter them by setting the probabilities of transitions. The Petri net helps to identify features related to personal data risks and countermeasures and ensures the implementation of a risk mitigation mechanism when implementing countermeasures. In addition, the network is dynamic, since at each cycle of calculation of the mathematical model, it is adapted to the changing properties of the personal data protection system [30].

To minimize the risks to personal data information systems, methods for identifying countermeasures, comparing them, calculating the probability of eliminating threats, and formulating an efficiency theorem have been proposed and implemented. When calculating a set of countermeasures that reduce the residual risk to zero, their effectiveness approaches one [12–14, 16, 18, 25]. The formed set of countermeasures ($V$) can be represented as a tuple

$$V =< K_d, T, Su >, \tag{21}$$

where $K_d$ is the value of the residual risk achieved by applying the formed set of countermeasures; $T$ is the ratio of neutralized threats to the total number of threats; $Su$ is the cost of the formed set of countermeasures. Thus, the modeling task is reduced to finding a set of countermeasures that has a residual risk of $K_d \longrightarrow 0, T \longrightarrow 1$, with an acceptable value of the total cost of this set of countermeasures.

The analysis and management of personal data information security risks include various methods, including tabular methods, fuzzy logic, game theory, and intruder modeling [12, 14–16, 19, 24, 25, 29]. Evaluation of the effectiveness of the protection system often uses economic indicators, but the complexity of the source data complicates the calculations. Risk analysis tools, such as COBRA, CRAMM, DS Office, FAIR, and OpenFAIR, are used to assess threats but have their limitations and requirements for integration with other systems. The choice of tool depends on the needs of the organization. The development of a new method includes collecting statistics on incidents, assessing specific threats, and using economic indicators to respond quickly to changes. Identifying vulnerabilities and system weaknesses is a key step in assessing the security of information systems, especially those containing personal data. In the context of risk management, risk analysis becomes an integral part of the methodology, which includes the identification of information system components, potential threats, and vulnerabilities (Figure 3). The complexity of information systems requires a systematic approach that includes several stages to ensure consistency and adequacy of security factors [21–24, 26–29].

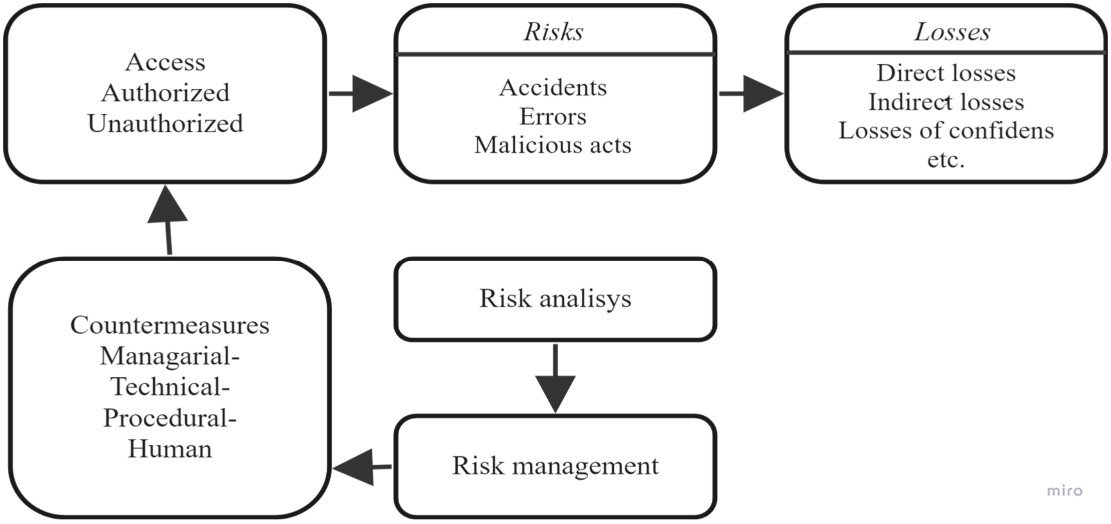


**Figure 3:** Risk management.

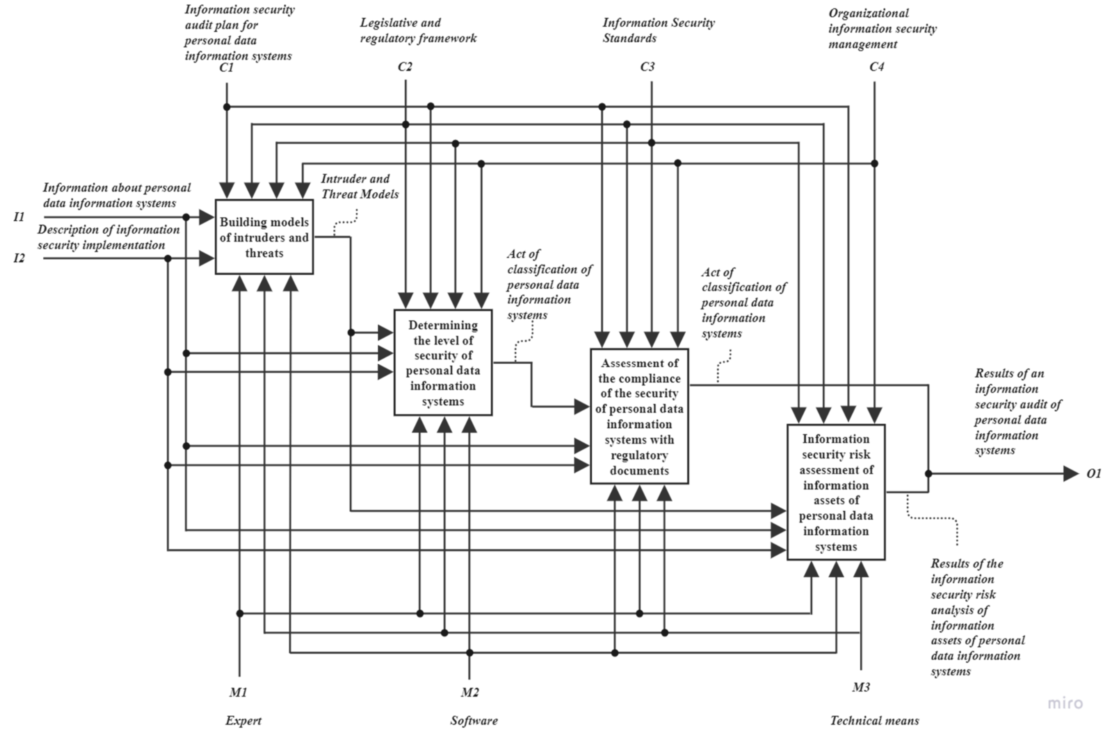
The risk management methodology should be universal so that it can be applied in different areas and ideas can be organized in a structured way. Traditional probabilistic methods of risk analysis are becoming ineffective due to the lack of sufficient statistical data. Risk management includes measures to minimize the consequences, such as avoidance, probability reduction, protection, risk transfer (e.g., through insurance), threat detection, and recovery. This ensures the validity of decisions and the resilience of information systems in the face of technological change and cyber threats.

## 6. Modeling of the information security audit process of personal data systems

Modeling the stages of an information security audit of personal data systems involves a comprehensive approach to assessing and improving measures to protect confidential information. The first step is to analyze information assets, determine their importance, identify threats and vulnerabilities, and develop models of offenders and threats. The system security assessment includes verification of compliance with standards and the effectiveness of security measures. The final stage is the development of recommendations for security modernization, implementation of new technologies, and monitoring of changes in threats. Modeling these stages helps to maintain a high level of personal information security and reduce the risks associated with personal data processing [14, 15, 20–22, 24, 29].

The process of building a systemic model of personal data protection based on IDEF technologies includes several stages. The first step is to create a model that reflects the relationship between the security audit of personal data information systems and the organization's activities. The second stage is the decomposition of the model for a more detailed analysis of the audit aspects. The third stage involves the formation of an ontology to unify terms in the field of personal data protection, which is important for standardizing the process. The last stage involves optimizing the model and ontology to increase efficiency and clarity [11–13, 17, 19].

First, a model of the relationship between the audit of information security of personal data systems and the main activities of the organization is built, which is then decomposed. Figure 4 shows a functional model that reflects the main stages of conducting an information security audit of personal data systems.



Information security audit plan for personal data information systems
C1
Legislative and regulatory framework
C2
Information Security Standards
C3
Organizational information security management
C4
Information about personal data information systems
I1
Description of information security implementation
I2
Building models of intruders and threats
Intruder and Threat Models
Determining the level of security of personal data information systems
Act of classification of personal data information systems
Assessment of the compliance of the security of personal data information systems with regulatory documents
Act of classification of personal data information systems
Information security risk assessment of information assets of personal data information systems
Results of an information security audit of personal data information systems
O1
Results of the information security risk analysis of information assets of personal data information systems
M1
Expert
M2
Software
M3
Technical means
miro

**Figure 4:** Functional model of security audit of personal data information system.

The analysis showed that the stages of "Assessment of compliance of personal data systems security with the requirements of regulatory documents" and "Assessment of risks to personal data systems information security" are labor-intensive and require expert support. The author proposes an ontology for systematizing the subject area of information security audit of personal data systems, including tasks, models, and decision-making methods, as well as a database of precedents for specific situations. The security profile of personal data information systems assesses compliance with security requirements through the evaluation indicators determined by experts for the level of protection [12–15, 20–22, 25–28]. A methodology for assessing the security of personal data information systems based on the creation of a security profile is proposed. To assess the compliance of the level of security of personal data information systems with the requirements of the regulatory framework, 15 group indicators of security of personal data information systems, designated as $M_i$, where $(i = 1,2, \dots ,15)$ (Table 1), were identified. Group indicators determine the measures to ensure the security of personal data by the requirements of regulatory documents on personal data protection.

**Table 1**
Group Indicators of Security of Personal Data Information Systems

| Designation | Name of group security indicators of personal data information systems |
|---|---|
| $M_1$ | Ensuring the security of personal data using identification and authentication of access subjects and objects |
| ... | ... |
| $M_1$ | Incident detection and response |
| $M_1$ | Configuration management of personal data information systems and personal data protection system |

The group indicators include private indicators of security of personal data information systems, designated as $M_{ij}$, the numerical estimates of $EV_{M_{ij}}$ of which form the overall estimates of group indicators $EV_{M_{ij}}$. The assessment of private security indicators of personal data information systems $M_{ij}$ is carried out by experts using questionnaires that take into account the requirements for each level of personal data security.

If the assessment of private indicators, the fulfillment of which is mandatory, results in a "no" answer, the score is $EV_{M_{ij}} = 0$; if the answer is "partially," then $EV_{M_{ij}} = 0{,}5$; if the answer is "yes," then $EV_{M_{ij}} = 1$. In the case of assessing private indicators, the fulfillment of which is optional, a "yes" answer leads to a score of $EV_{M_{ij}} = 1$; a "no" answer leads to the fact that the partial indicator is defined as unassessed and is not taken into account in the formation of the assessment results.

The final normalized score for the group indicator $EV_{M_{i\text{result}}}$ is calculated using the following formula:

$$EV_{M_{i\text{result}}} = \omega_i \frac{\sum EV_{M_{ij}}}{EV_{M_{i\max}}}, i = 1 \div 15, j = 1 \div N_i, \tag{22}$$

where $N_i$ is the number of private security indicators of personal data information systems representing the group indicator; $M_i EV_{M_i}$ is the value of the group indicator assessment; $EV_{M_{i\max}}$ is the maximum possible value of the quantitative assessment; $EV_{M_i}\omega_i$ is the significance coefficient $(0 \leq \omega_i \leq 1)$, which is assigned by an expert to adapt the personal data protection system to the conditions of their processing and the technical means used.

Based on the final scores of the group indicators $EV_{M_{i\text{result}}}$, a normalized score $\overline{EV}$ is calculated, which reflects the degree of compliance with the requirements of the regulatory framework for personal data protection:

$$\overline{EV} = \frac{\sum EV_{M_{i\text{result}}}}{EV_{\max}}, i = 1 \div 15, \tag{23}$$

where $EV_{\max}$ is the maximum possible value of the quantitative assessment $EV$. Since the maximum value of each of the 15 group indicators is 1, $EV_{\max} = 15$.

In this case, the method of expert assessments can be used not only to assess the degree of compliance with certain requirements for personal data protection but also to compare different options for building a personal data protection system [20–23, 27, 28]. The levels of compliance of the security of personal data information systems with the requirements of the regulatory framework are determined by expert assessments, taking into account the fact that the personal data operator can adapt the personal data protection system to the processing conditions and technical means used (Figure 5):

- $0 \leq \overline{EV} < 0{,}75$ is an unacceptable level of security in personal data information systems.
- $0{,}75 \leq \overline{EV} < 0{,}95$ is moderated security, but countermeasures are required.
- $0{,}95 \leq \overline{EV} \leq 1$ is the value of the security level that meets the requirements of regulatory documents on personal data protection.

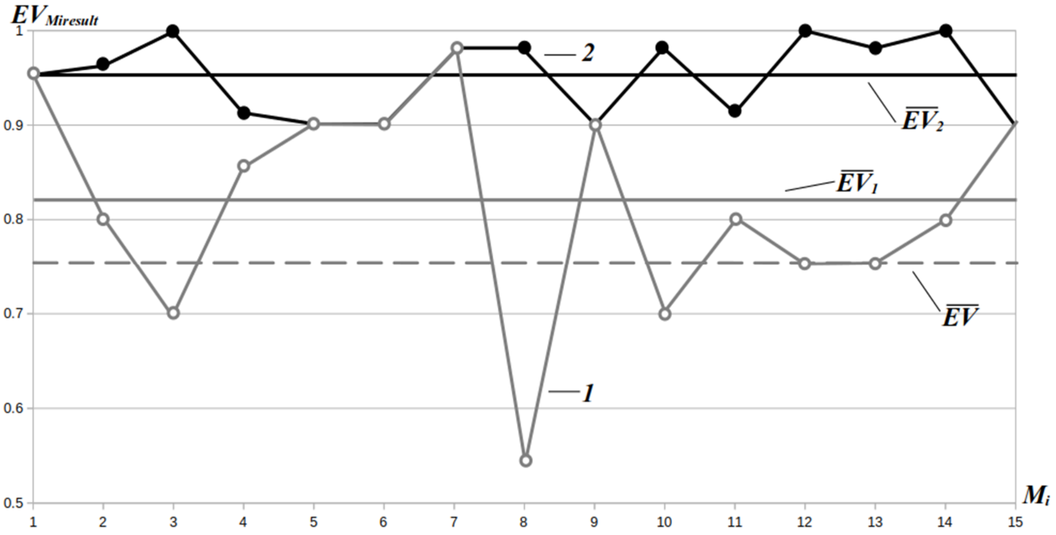


**Figure 5:** Diagram of assessment of compliance of the level of security of personal data information systems with the requirements of the regulatory framework.

Presentation of the results in the form of the above diagram allows for a visual assessment of the state of personal data protection. The results of the assessment reveal the inconsistency of the security of certain assets of personal data information systems with regulatory requirements, and then an assessment of information security risks caused by this inconsistency is made.

## 7. Decision-making algorithm for assessing personal data information security risks using data mining

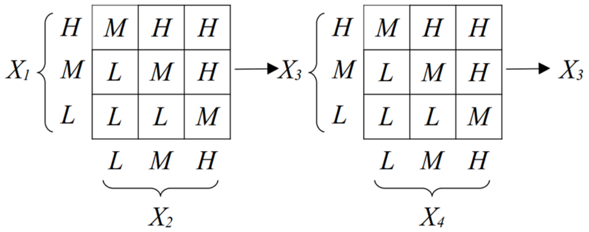
The use of fuzzy logic for information security risk assessment is driven by the need to adapt to uncertainty and the dynamics of cyber threats, which allows for the consideration of fuzzy or ambiguous concepts and expert experience. Methods for assessing information security risks include tabular methods, fuzzy logic, game theory, and a combination of expert and mathematical approaches. They allow risk assessment in the absence of accurate data and take into account social and psychological aspects. It is important to use fuzzy logic to formalize qualitative information, which provides flexibility in forecasting and risk assessment [10, 14–18, 21, 24].

Effective control of information system security audits using audit logs includes a personal data risk assessment algorithm that uses data mining. The assessment begins with the identification of the object and analysis of threats and vulnerabilities of systems, after which the likelihood and impact of threats on the confidentiality, integrity, and availability of data are assessed. Next, the overall risk is calculated, which allows you to make decisions on threat mitigation measures. An important part is the use of machine learning to detect anomalies in data processing and continuous security monitoring. A fuzzy logic-based information security risk assessment model supports experts in making decisions under uncertainty [20–24, 28]. In this context, the term "information security risk" $R_i$ is defined as the expected potential damage resulting from the impact of a threat due to the presence of vulnerabilities on an information asset:

$$R_i = \frac{C_{\text{damage/risk},i}}{C_{\text{information assent},i}}. \tag{24}$$

Consider information security risk in a context where $C_{\text{damage/risk},i}$ represents the damage caused by the impact of a threat on an information asset; $C_{\text{information assent},i}$ determines the value of this asset. The parameter $R_i$ which takes a value in the range from 0 to 1, determines the information security risk and can be expressed as a percentage ($0 \leq R_i \leq 100\%$) relative to the value of the asset $C_{\text{information assent},i}$.

The decomposition of the fuzzy inference system suggests dividing it into two relatively independent parts that solve the problem of assessing two indicators—the probability of successful threat implementation and the risk (potential damage) from the impact of the threat on personal data. Additionally, a set of fuzzy productive rules for risk assessment is proposed, obtained by converting the deterministic rules built in the ontology into fuzzy rules, such as: "IF Threat_Probability($X_1$) is high and Vulnerability($X_2$) is high, THEN Threat_Probability($X_3$) is high," ..., "IF Threat_Probability($X_3$) is high AND Information_Asset_Value($X_4$) is high, THEN Risk($X_5$) is high" (Figure 6).



| $X_1$ \ $X_2$ | L | M | H |
|---|---|---|---|
| H | M | H | H |
| M | L | M | H |
| L | L | L | M |

| $X_3$ \ $X_4$ | L | M | H |
|---|---|---|---|
| H | M | H | H |
| M | L | M | H |
| L | L | L | M |

**Figure 6:** A set of rules for risk assessment based on fuzzy logic.

The fuzzy logic method was used to analyze information security risks during the audit of the personal data system. Each term set $(L, M, H)$ for variables $(X_1, X_{2,} X_3, X_4, R)$ was assigned a "center of gravity," which allowed us to build a training set for a fuzzy neural network designed to assess information security risks to personal data system assets. For this purpose, a modular fuzzy neural network implemented in MATLAB using the ANFIS adaptive neuro-fuzzy inference system was used. This approach allows us to effectively assess information security risks in the face of uncertainty and system complexity (Figure 7).

In Figure 7 $\mu_L, \mu_M, \mu_H$ are the values of membership functions for input variables $X_1, X_{2,} X_4$ and output variables $X_3, X_5$; LOM-1 (logic output modules), LOM-2 are logic output modules; DM-1 (defasification modules), DM-2 are defasification modules; $\bar{X}_{3,1}, \bar{X}_{3,2}, \bar{X}_{3,3}$ are numerical values corresponding to term sets for variable;$X_3, \bar{X}_{5,1}, \bar{X}_{5,2}, \bar{X}_{5,3}$ are numerical values corresponding to term sets for variable $X_5$.

Figure 8 shows the possible options for decisions regarding the construction of information security systems: obtaining a rational solution (i.e., a solution that satisfies the given constraints on information security risk and the allocated costs for information security), accepting the risk, or increasing the budget.

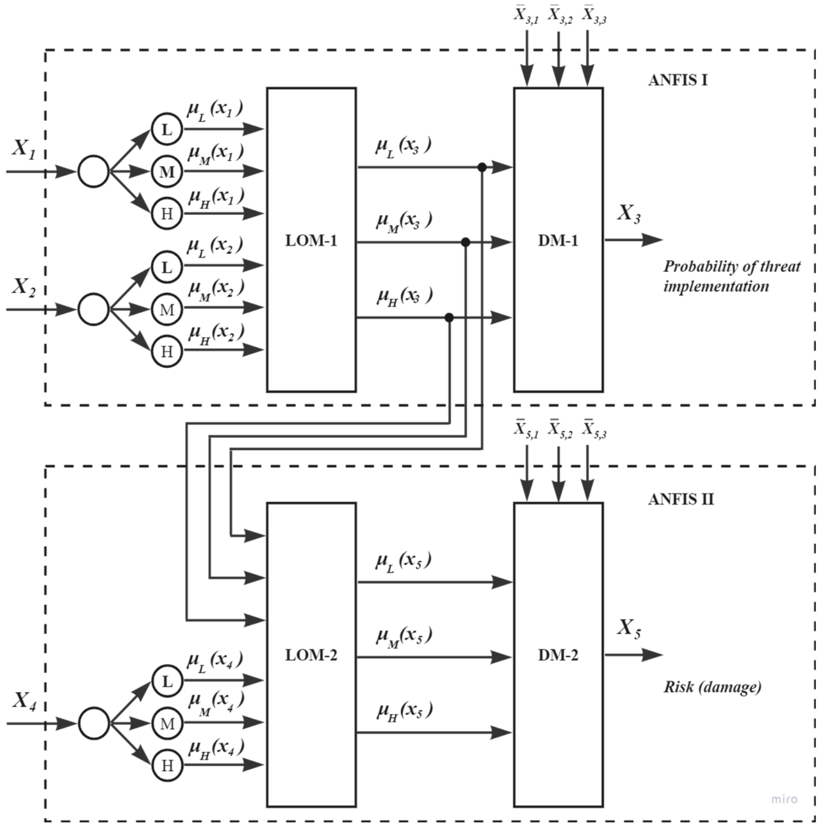


**Figure 7:** Structure of a modular neural network.

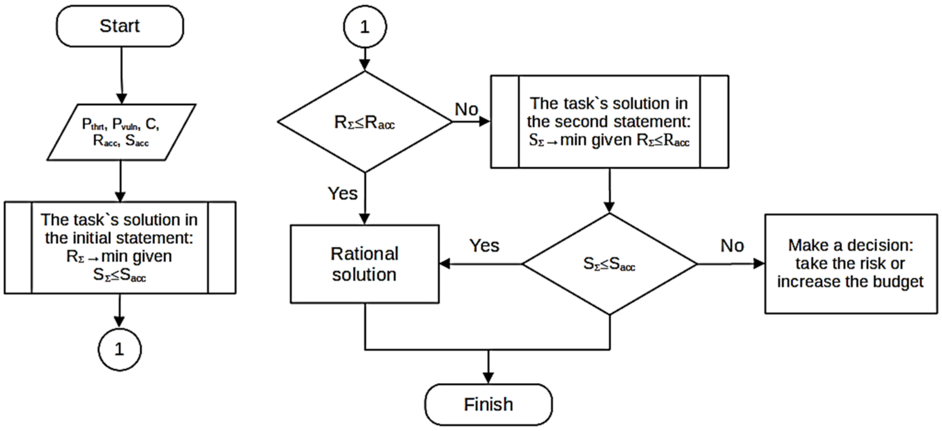

Start
Pthrt, Pvuln, C, Racc, Sacc
The task`s solution in the initial statement: RΣ→min given SΣ≤Sacc
1
1
RΣ≤Racc
No
Yes
The task`s solution in the second statement: SΣ→min given RΣ≤Racc
Rational solution
Yes
SΣ≤Sacc
No
Make a decision: take the risk or increase the budget
Finish

**Figure 8:** Algorithm for assessing the effectiveness of the personal data protection system.

The proposed modular neural network model has advantages, in particular, the ability to learn and reduce the number of fuzzy rules due to the division into modules. An algorithm for evaluating the effectiveness of personal data protection based on the Clements-Hoffman scheme with full overlap has been developed, which allows comparing and optimizing protection systems according to the criteria "information security risks is the cost of protection costs". The assessment can be carried out in two problem formulations:

- Minimizing the risk of information security $R_\Sigma$ in the presence of a limit on the total cost of information protection $R_\Sigma(S) \longrightarrow \min$ at $S_\Sigma = \Sigma_{q=1}^{z} S_q \leq S_{\text{acceptable}}$.
- Minimization of costs $S_\Sigma$ for the creation of protective barriers while limiting the total risk $R_\Sigma$ (at $R_\Sigma(S) \leq R_{\text{acceptable}}$) where $S_{\text{acceptable}}$ is the maximum allowable cost of creating

protective barriers; $R_{\text{acceptable}}$ is the maximum permissible value of the total risk $R_{\Sigma}$, i.e., the potential damage from the impact of threats; $S_q$ is the cost of creating a barrier $S_q^*$; $0 \leq S_q \leq 1$; $0 \leq R_{\Sigma}(S) \leq 1$; $0 \leq R_{\text{acceptable}} \leq 1$; $0 \leq S_{\Sigma} \leq 1$; $0 \leq S_{\text{acceptable}} \leq 1$.

## 8. Prototype of an intelligent decision support system for auditing the information security of a personal data system

A prototype of an intelligent decision support system for auditing the information security of personal data systems combines technical solutions and algorithmic approaches, using artificial intelligence and machine learning to analyze security. The system provides automatic detection and classification of threats and vulnerabilities, development of protection strategies, as well as recommendations for effective security measures, taking into account the specifics of the information system. An important component is real-time security monitoring for rapid response to new threats, which allows taking into account the dynamics of threats and modern technological challenges (Figure 9) [21, 23, 30–34].

\

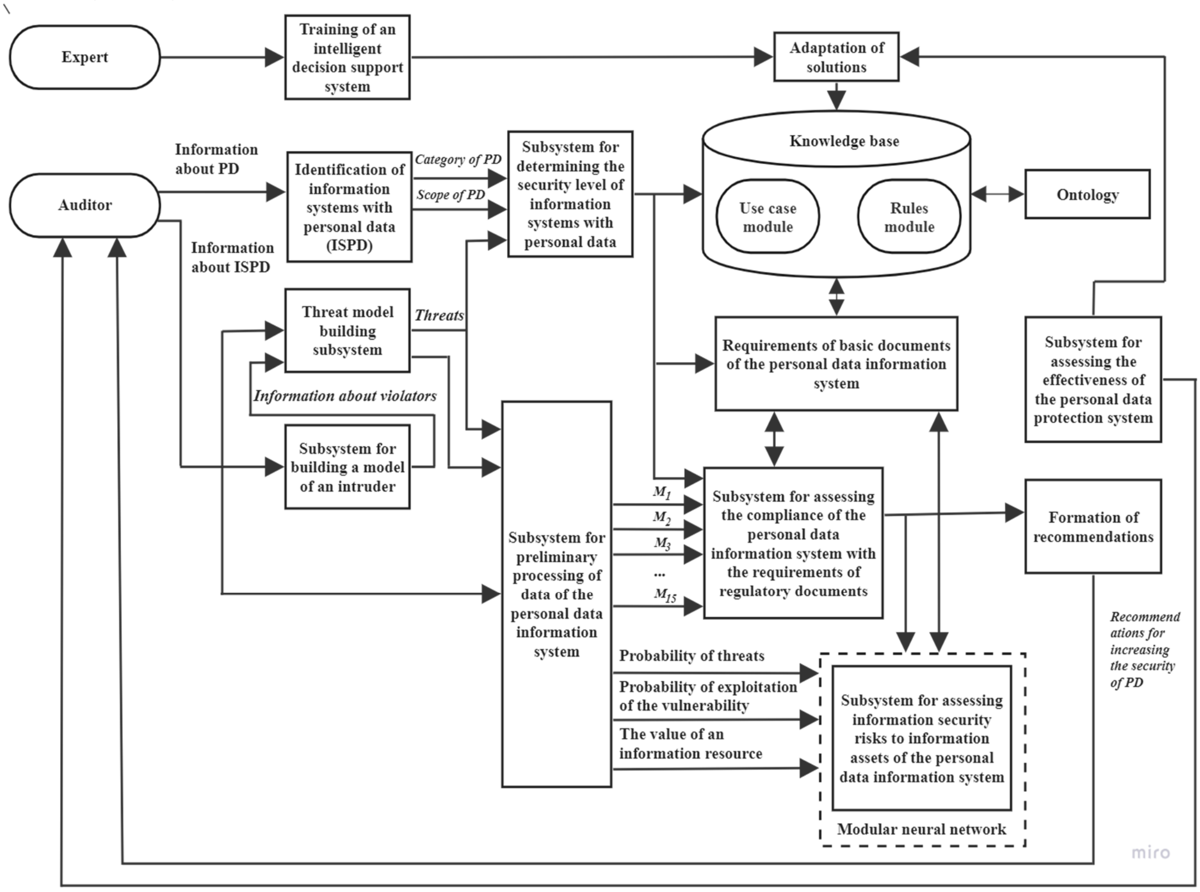

Expert
Training of an intelligent decision support system
Adaptation of solutions
Information about PD
Auditor
Information about ISPD
Identification of information systems with personal data (ISPD)
Category of PD
Scope of PD
Subsystem for determining the security level of information systems with personal data
Knowledge base
Use case module
Rules module
Ontology
Threat model building subsystem
Threats
Information about violators
Subsystem for building a model of an intruder
Requirements of basic documents of the personal data information system
Subsystem for assessing the effectiveness of the personal data protection system
Subsystem for preliminary processing of data of the personal data information system
M1
M2
M3
...
M15
Subsystem for assessing the compliance of the personal data information system with the requirements of regulatory documents
Formation of recommendations
Recommendations for increasing the security of PD
Probability of threats
Probability of exploitation of the vulnerability
The value of an information resource
Subsystem for assessing information security risks to information assets of the personal data information system
Modular neural network
miro

**Figure 9:** Architecture of an intelligent decision support system for auditing the information security of a personal data system.

The software of intelligent decision support systems for auditing the information security of systems with personal data implements the main modules through the MATLAB graphical interface, with the ability to enter data through the Excel tabular interface, which organizes access to data via COM/DCOM for the Python computing core and the neural network unit for analyzing information security risks. The system analyzes information systems, builds a model of an intruder and threats, and assesses the level of security and compliance with regulatory requirements, as well as risks to information assets. This makes it possible to formulate recommendations for the modernization of personal data protection, which helps to reduce the risks from threats to personal data information security [20–23, 27, 28, 31].

## 9. Mechanism for improving the accuracy of the information security risk model built using fuzzy logic

To improve the accuracy of the information security risk assessment model built using fuzzy logic, training is performed, during which the model parameters are iteratively changed to minimize deviations between the logical conclusion and experimental data [26–28]. This includes changing the weights and parameters of membership functions. It is important to keep in mind that transparency must be maintained during model training to ensure meaningful interpretation.

The parameters of the membership function can be determined by. The Gaussian function (21) (where $a$ is the mathematical expectation and $\sigma^2$ is the variance) has the form:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-a)^2}{2\sigma^2}} \tag{25}$$

before using it as a membership function, we modify (22) so that the "height" of a vertex (its ordinate) always remains equal to 1

$$c = e^{-k\cdot(g+a)^2}. \tag{26}$$

The two coefficients introduced for this purpose ($k$ and $a$) graphically represent the "width" of the membership function (coefficient $k$), which is analogous to the "variance" parameter, and the coordinate of the membership function vertex on the abscissa axis (coefficient $a$), which is analogous to the "mathematical expectation" parameter. This modification of the Gaussian function is due to the convenience of programming and a special restriction on the constancy of the ordinate of the vertices of the membership function. In the modified Gaussian function, the coordinates of the abscissa axis are determined by the variable $g, a$, and the coordinates of the ordinate axis are determined by the variable $c$.

To correct the shape of the membership function, two parameters of the Gaussian function are changed: the variance and the mathematical expectation. In the process of modeling in Fuzzy Logic, increasing the "width" of the membership function of input variables (variance for Gaussian curves) while reducing the variance of the output variable gives a smoother, more uniform surface [13–16, 21, 25, 29]. The Fuzzy Logic interface allows you to build a three-dimensional image of the "surface of the fuzzy inference system" and a graph of the dependence of the output variable on the input variables, which helps to control the quality of the inference mechanism, where a smooth and monotonous graph indicates the sufficiency and consistency of the rules. The use of the "center of gravity" method during defuzzification leads to a narrower range of output values, which means that the risk level will never reach maximum or minimum values. For a model using fuzzy logic, it is advisable to introduce a correction factor to eliminate the effect of narrowing the range. In this case, the object model will look like the one shown in Figure 10. This correction function stretches the output variable to the normalized value of the risk variable 0 ÷ 1 relative to the average value.

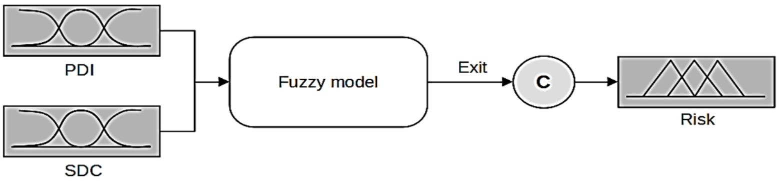


**Figure 10:** Object model using the correction factor.

Obtaining the parameters of the membership function in the information security risk assessment model using fuzzy logic is reduced to an optimization problem, where the criterion is the minimum area between the target and the obtained function, and the optimization parameters are the variance and mathematical expectation of the Gaussian function. Since the surface of the model has local minima and maxima that reduce accuracy, optimization allows you to obtain coefficients that make the model look monotonic. For multidimensional optimization, it is advisable to use the Gauss-Seidel method. The program algorithm shown in Figure 11, includes the creation of a model for assessing the risk of information security breach using fuzzy logic and automatic optimization of this model, as well as analysis and graphical display of the model in the form of a three-dimensional surface.

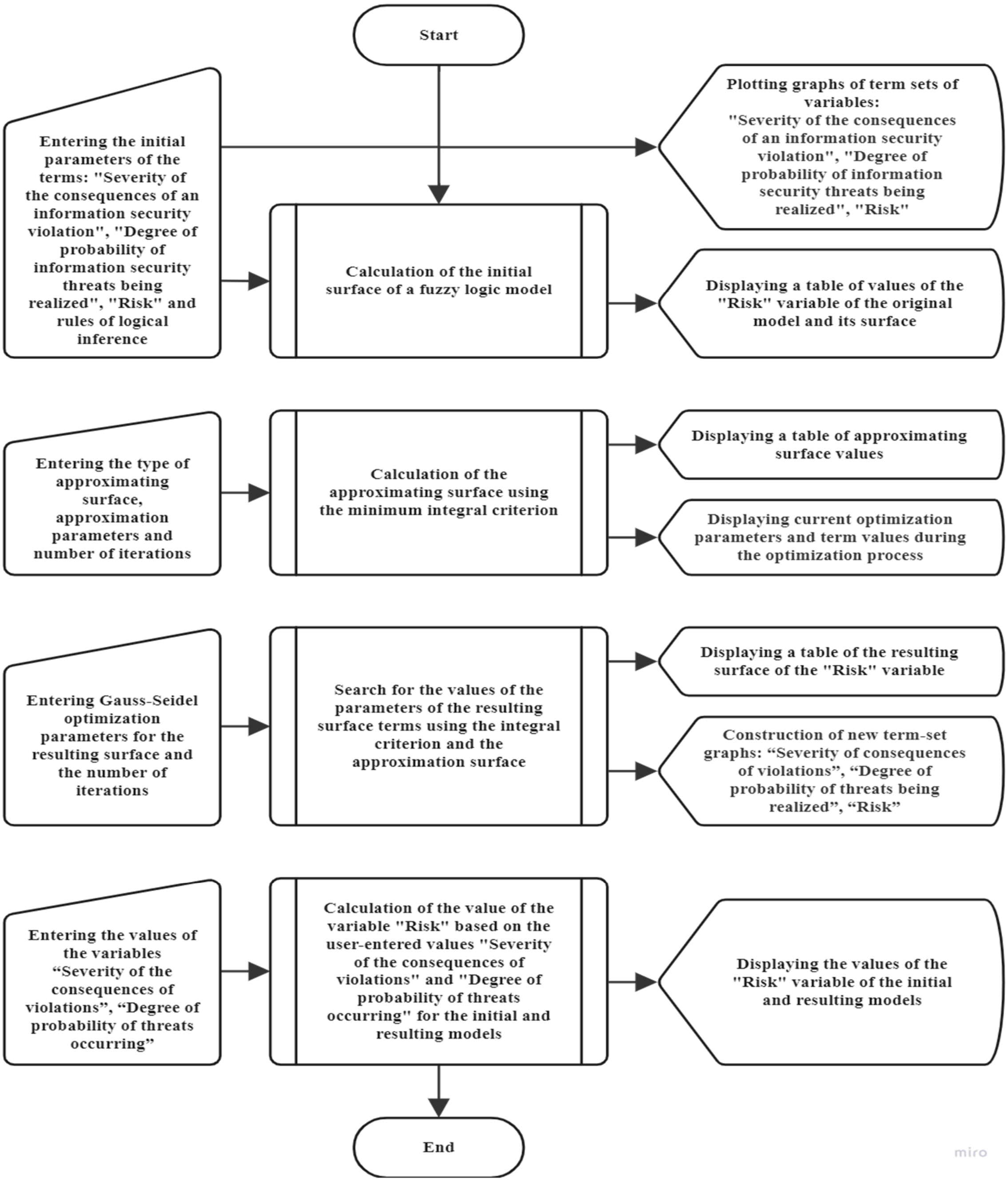

Start
Entering the initial parameters of the terms: "Severity of the consequences of an information security violation", "Degree of probability of information security threats being realized", "Risk" and rules of logical inference
Calculation of the initial surface of a fuzzy logic model
Plotting graphs of term sets of variables: "Severity of the consequences of an information security violation", "Degree of probability of information security threats being realized", "Risk"
Displaying a table of values of the "Risk" variable of the original model and its surface
Entering the type of approximating surface, approximation parameters and number of iterations
Calculation of the approximating surface using the minimum integral criterion
Displaying a table of approximating surface values
Displaying current optimization parameters and term values during the optimization process
Entering Gauss-Seidel optimization parameters for the resulting surface and the number of iterations
Search for the values of the parameters of the resulting surface terms using the integral criterion and the approximation surface
Displaying a table of the resulting surface of the "Risk" variable
Construction of new term-set graphs: "Severity of consequences of violations", "Degree of probability of threats being realized", "Risk"
Entering the values of the variables "Severity of the consequences of violations", "Degree of probability of threats occurring"
Calculation of the value of the variable "Risk" based on the user-entered values "Severity of the consequences of violations" and "Degree of probability of threats occurring" for the initial and resulting models
Displaying the values of the "Risk" variable of the initial and resulting models
End
miro

**Figure 11:** Algorithm of the information security risk assessment model using fuzzy logic.

The information security risk assessment toolkit includes dialog interaction procedures that simplify decision-making and methodological techniques for preparing information and obtaining results. The program module also solves the problem of narrowing the range of initial values and includes an adaptive learning mechanism.

## 8. Conclusions

Taking into account the multidimensionality and component heterogeneity of information technologies and systems, as well as the complexity of harmful effects, necessitates the use of probabilistic models to assess information risks. Within this methodology, risk analysis becomes the main tool for determining the level of security of information systems, since security is defined as a state in which risks do not exceed the acceptable level. Statistical monitoring of attacks and their consequences is becoming necessary for enterprises of all sizes, with a special emphasis on strategic risk management. The use of risk analysis models and algorithms during the security audit of personal data information systems, in particular through the integration of audit methods and fuzzy logic, creates an effective toolkit for risk assessment and management. This approach ensures flexibility and adaptability of the assessment, maintaining a high level of accuracy and transparency in the process. The inclusion of an adaptive learning mechanism allows for improving risk management methods, increasing the efficiency and reliability of information systems.

An important aspect is the cyber defense of critical facilities and infrastructures, where the consequences of attacks can be catastrophic. Risk analysis and risk management in such infrastructures using probabilistic models opens up new opportunities for forecasting and improving the efficiency of information systems. The use of intelligent systems to support decision-making in audits allows not only to determine the level of security and assess compliance with regulations but also to provide recommendations for modernizing personal data protection, reducing the risk of dangerous threats to information security.

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References

[1] H. Танака, Vulnerability and effects of Information security investment: A firm level empirical analysis of Japan, Forum on financial Information systems and cyber security, 2005.

[2] D. Landoll, The security risk assessment handbook, 2021. doi: 10.1201/9781003090441.

[3] V. Visintine, An introduction to information risk assessment, 2009.

[4] European Union, Regulation (EU) 2016/679 of the European Parliament and of the council, Official Journal of the European Union, 2016, 119/1–119/88. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

[5] Verkhovna Rada of Ukraine, On protection of personal data, 2297-VI, 2025. URL: https://zakon.rada.gov.ua/laws/show/en/2297-17#Text.

[6] O. Mykhaylova, et al., Mobile application as a critical infrastructure cyberattack surface, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems II, CPITS-II, vol. 3550 (2023) 29–43.

[7] A. Zahynei, et al., Method for calculating the residual resource of fog node elements of distributed information systems of critical infrastructure facilities, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems 3654 (2024) 432–439.

[8] S. Shevchenko, et al., Protection of information in telecommunication medical systems based on a risk-oriented approach, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, vol. 3421 (2023) 158–167.

[9] S. Shevchenko, et al., Information security risk management using cognitive modeling, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3550 (2023) 297–305.

[10] C. J. Alberts, et al., Operationally critical threat, asset and vulnerability evaluation, 2018.

[11] C. F. Endorf, Measuring ROI on security, Information security management handbook, edited by H. F. Tipton, M. Krauze, 6th ed., part 1, sect. 1.1, ch. 12 (2017) 133–137.

[12] J. Cebula, L. Young, A taxonomy of operational cyber security risks, 2010.

[13] W. Stallings, Effective cybersecurity: Understanding and using standards and best practices, Addison-Wesley, 2019.

[14] Y. Kostiuk, H. Heidarov, The role of tokens and session management in information security systems to counter cross-site attacks, Science and Technology Today 5(33) (2024) 1216–1231. doi: 10.52058/2786-6025-2024-5(33)-1216-1231.

[15] G. Wangen, Quantifying and analyzing information security risk from incident data, Graphical Models for Security (2019) 129–154. doi: 10.1007/978-3-030-36537-0_7.

[16] Y. Kostiuk, et al., Integrated protection strategies and adaptive resource distribution for secure video streaming over a Bluetooth network, in: Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3826 (2024) 129–138.

[17] L. Spedding, A. Rose, Business risk management handbook: A sustainable approach, 2018.

[18] K. Henry, Risk management and analysis, Information security management handbook, edited by H. F. Tipton, M. Krauze, 6th ed., part 1, sect.1.4, ch. 28 (2017) 321–329.

[19] M. Garnaeva, Kaspersky security bulletin 2015, Overall statistics for 2015 (2015).

[20] O. Kryvoruchko, et al., Methodology for developing an information system for internal audit support, in: Proceeding of IEEE 4th International Conference on Smart Information Systems and Technologies (SIST) (2024) 106–110. doi: 10.1109/sist61555.2024.10629532.

[21] S. Honchar, A. Onyskova, Relevance of the subjective component in cybersecurity risk assessment, Theoretical and Empirical Scientific Research: Concept and Trends 2 (2020) 22–23. doi: 10.36074/24.07.2020.v2.07.

[22] O. Kryvoruchko, et al., Analysis of technical indicators of efficiency and quality of intelligent systems, Journal of Theoretical and Applied Information Technology 101(24) (2023) 812–813.

[23] Y. Kostiuk, et al., Research of Methods of Control and Management of the Quality of Butter on the Basis of the Neural Network, in: Proceeding of International Conference on Smart Information Systems and Technologies (SIST) (2022) 1–6. doi: 10.1109/sist54437.2022.9945764.

[24] C. A. Wilhelmsen, T. Lee, Ostrom. Risk assessment: tools, techniques, and their applications, John Wiley & Sons, 2019.

[25] R. A. Caralli, Introducing OCTAVE Allegro: Improving the information security risk assessment process, 2008.

[26] United Kingdom Central Computer and Telecommunication Agency, CRAMM user guide, Risk analysis and management method, 2001.

[27] Y. Kostiuk, et al., Information and intelligent forecasting systems based on the methods of neural network theory, in: IEEE International Conference on Smart Information Systems and Technologies (SIST) (2023) 168–173. doi: 10.1109/sist58284.2023.10223499.

[28] B. Engelmann, R. Rauhmeier, The Basel II risk parameters, Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-16114-8.

[29] O. Skitsko, et al., Threats and risks of the use of artificial intelligence, Cybersecurity: Education, Science, Technique 1(25) (2023) 6–18. doi: 10.28925/2663-4023.2023.22.618.

[30] R. Syrotynskyi, et al., Methodology of Network infrastructure analysis as part of migration to zero-trust architecture, Cyber Security and Data Protection 3800 (2024) 97–105.

[31] Y. Chunxiao, W. Zhongfu, F. Yunqing, An attribute-based delegation model and its extension, Journal of Research and Practice in Information Technology 38(1) (2006) 220–234.

[32] O. Solomentsev, et al., Data processing through the lifecycle of aviation radio equipment, in: Proceedings of IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), IEEE, Lviv, Ukraine, 2022, pp. 146–151. doi: 10.1109/CSIT56902.2022.10000844.

[33] M. Zaliskyi, et al., Heteroskedasticity analysis during operational data processing of radio electronic systems, in: S. Shukla, A. Unal, J. Varghese Kureethara, D.K. Mishra, D.S. Han (Eds.), Data science and security, volume 290 of Lecture Notes in Networks and Systems, Springer, Singapore, 2021, pp. 168–175. doi: 10.1007/978-981-16-4486-3_18.

[34] I. Ostroumov, et al., A probability estimation of aircraft departures and arrivals delays, In: O. Gervasi, et al. (Eds.), Computational Science and Its Applications – ICCSA 2021. ICCSA 2021, volume 12950 of Lecture Notes in Computer Science, Springer, Cham, 2021, pp. 363–377. doi: 10.1007/978-3-030-86960-1_26.