Fuzzy Logic Data Protection Management

Alexander Bolshakov, Anastasia Zhila
Moscow Technical University of Communications and Informatics
Moscow, Russia
as.bolshakov57@mail.ru, ai.zhila@yandex.ru

***Abstract*—This article discusses the problem of information security management in computer systems and describes the process of developing an algorithm that allows to determine measures to protect personal data. The organizational and technical measures formulated by the FSTEC are used as measures.**

I. INTRODUCTION

Recently, many articles have been aimed at information security of personal data. Measures to ensure personal data using the basic threat model are formulated by FSTEC [1,4]. However, unfortunately, there are no methods, incl. Using a certain mathematical apparatus that establishes the relationship between the threats under consideration and measures of information security. This paper proposed to fill this gap by establishing such a connection using fuzzy logic.

To ensure the required level of security of personal data in computer systems, it is advisable to create an information security management system that would help to choose protective measures of the required level of security for certain input data.

According to the authors, the apparatus of fuzzy inference can be used as a mathematical apparatus for describing such situation, which is one of the directions of the modern theory of decision making under conditions of an indefinite relationship between input and output parameters. The main advantage of using this mathematical approach in modeling a control system is the description of conditions and methods for solving problems in conditions of uncertainty in a language close to natural [2]. Thus, exploring existing and potential problems in poorly structured systems in the process of modeling a set of causal relationships.

So far, there has been a lot of research on computer system security using fuzzy rules. In this paper, a new approach to preventing attacks using a fuzzy expert system is developed. The fuzzy system proposed in this study provides valuable information to system administrators to improve the achievement of computer system security. This work can be adapted to different attack scenarios on the computer system.

II. PROTECTION MANAGEMENT SYSTEM

The information security management system in the researching personal data protection model can be represented in the form of a functional diagram presented in Fig. 1, in which information security will be managed according to the rules of fuzzy logic.

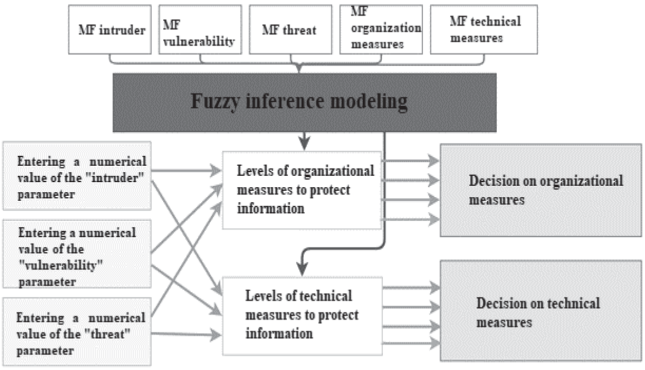

MF intruder
MF vulnerability
MF threat
MF organization measures
MF technical measures
Fuzzy inference modeling
Entering a numerical value of the "intruder" parameter
Entering a numerical value of the "vulnerability" parameter
Entering a numerical value of the "threat" parameter
Levels of organizational measures to protect information
Levels of technical measures to protect information
Decision on organizational measures
Decision on technical measures

Fig. 1. Functional diagram of the developing information security management system

The algorithm developed in this work is a set of rules based on fuzzy logic that connects the input parameters that characterize the implementation of threats to the security of personal data, and the output parameters of the model in the form of organizational and technical protection measures. This model identifies the impact of information security threats on decision-making on the choice of measures to protect personal data in a computer system.

The numerical values of the input parameters characterize the implementation of information security threats, defining a specific type of linguistic variable from the set belonging to each of the parameters. The numerical values of the output parameters correspond to the levels of information protection measures obtained using fuzzy inference modeling. Based on the respective levels of the output variables, a decision is made about what measures should be taken to neutralize the simulated threat scenarios.

III. DERIVATION OF THE PARAMETERS OF THE FUZZY OUTPUT MODEL

To identify the impact of information security threats on the choice of measures to protect personal data using fuzzy logic, it is necessary to determine threats to the security of personal data for a specific type of ISPD.

The combination of conditions and factors that create a potential threat to the information security of personal data is formed taking into account the characteristics of the information system containing the protected information and the characteristics of the threat sources. Threats modeling will be carried out in accordance with the FSTEC basic threat model [1].

In practice, when compiling private models of threats and security for ISPD, the identified threats presented in Fig. 2 are relevant.

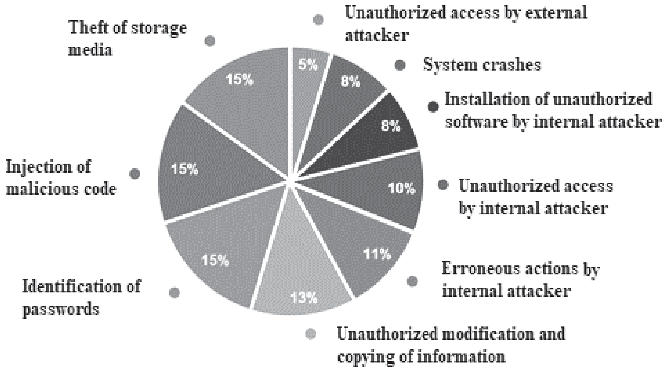


Fig. 2. The result of the analysis of current ISPD threats

To build an information security model, it is necessary to have an idea of the mechanisms violating the properties of information in the ISPD. Namely, to disclose the content of the "chain": "threat" = "intruder" - "vulnerability" - "protected resource" + "information security incident (Table I) [3].

TABLE I. DESCRIPTION OF THE THREATS OF IS OF THE DEVELOPED MODEL

Intruder	IS threat	Vulnerability
External intruder	Revealing passwords	Simple password in the system
	Unauthorized access to information	Lack of access control
	Unauthorized modification, copying of information	Lack of differentiation of rights
	Theft of information carriers	Lack of recording of information carriers
	Malicious code injection	Lack of anti-virus protection
	System crashes	Lack of backup
ISPD personnel	Erroneous actions	Lack of control over user actions
	Unauthorized modification, copying of information	Lack of differentiation of rights
	Unauthorized access to information	Lack of access control
	Installation of inconsistent software	Lack of control over user actions
	System crashes	Lack of backup
	Revealing passwords	Lack of access control
	Theft of information carriers	Lack of recording of information carriers

According to the requirements of FSTEC of February 18, 2013 No. 21 [4], we will make up a table of organizational and technical measures to protect personal data that ought to be followed (Table II).

TABLE II. DESCRIPTION OF TECHNICAL AND ORGANIZATIONAL PROTECTION MEASURES

Organizational measures	Technical measures
Identification and authentication of subjects of access and objects of access (IAF)	Protection of technical means (ZTS)
Ensuring the integrity of the information system and personal data (OTsL)	Restriction of the software environment (OPS)
Personal data security control (ANZ)	Access control of subjects of access to access objects (UPD)
	Antivirus protection (AVZ)
	Ensuring the availability of personal data (ODT)
	Protection of machine media (ZNI)
	Information System Configuration Management (UCF)

IV. SIMULATION OF FUZZY OUTPUT

When building a model based on fuzzy logic, each input and output parameter is a linguistic variable, the values of which are words of a natural language. This set of meanings is a term set of a linguistic variable. The elements of this set are terms that are formalized by a fuzzy set using the membership function in the scale [0,1], ie the degree of belonging to the set [5]. Despite the fact that fuzzy systems can have membership functions of an arbitrary structure, from a practical point of view, functions of a triangular type are most popular [6,7].

The main stages of fuzzy inference are related to the process of forming classification conclusions [8]:

1. Formation of certain variables.

The variables are described in the range of real numbers from 0 to 1. Then we have the following clear initial input variables:

$X_1 \epsilon [0,1]$ – Intruder;
$X_2 \epsilon [0,1]$ – IS threat;
$X_3 \epsilon [0,1]$ – Vulnerability.

Therefore, at the input of the model there is an initial clear vector $\{X_1, X_2, X_3\} \epsilon [0,1] \times [0,1] \times [0,1]$. The set of possible values for this vector is a 3-dimensional cube with edge = 1.

At the output of the model, certain variables should be formed:

$Y_1 \epsilon [0,1]$ – organizational measures;
$Y_2 \epsilon [0,1]$ – technical measures.

Therefore, at the output of the model there is an initial clear vector $\{Y_1, Y_2\} \epsilon [0,1] \times [0,1]$. The set of possible values for this vector is a 2-dimensional square with edge = 1.

2. *Fuzzification of input and output variables (formation of fuzzy linguistic variables).*

As a term-set of the variable X_1, we will use the set T_1= *(*external attacker, internal attacker*)* = *(*$\mathrm{T}_{1.1}, \mathrm{T}_{1.2}$*)* with membership functions, respectively $\mu_{1.1}(\mathrm{X}_1) \in [0,1]$, $\mu_{1.2}$ $(\mathrm{X}_1) \in [0,1]$.

Based on the data obtained (in Fig. 2), the percentage of the frequency of occurrence of threats from the total number associated with an internal attacker is 29%, while those associated with an external attacker are 20%, and 51% of threats are associated with both an external attacker and an internal. Thus, the input variable "Intruder" in the developed fuzzy logic model will have the form shown in Fig. 3.

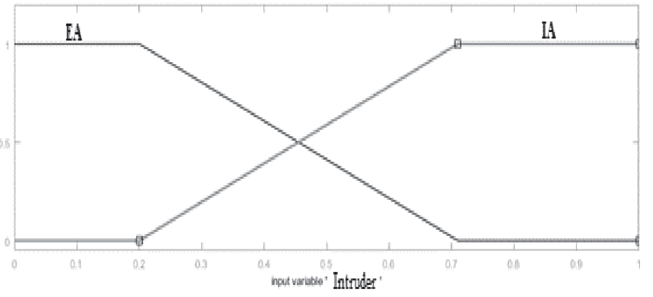


Fig. 3. Description of the linguistic variable "Intruder"

As a term set of the variable X_2, we will use the set T_2 = (system failures, erroneous actions, installation of inconsistent software, injection of malicious code, identification of passwords, unauthorized access to information, unauthorized modification and copying of information, theft of storage media) = *(*$\mathrm{T}_{2.1}, \mathrm{T}_{2.2}, \mathrm{T}_{2.3}, \mathrm{T}_{2.4}, \mathrm{T}_{2.5}, \mathrm{T}_{2.6}, \mathrm{T}_{2.7}, \mathrm{T}_{2.8}$ *)* with membership functions, respectively $\mu_{2.1}(\mathrm{X}_2) \in [0,1], \mu_{2.2}(\mathrm{X}_2) \in [0,1], \mu_{2.3}(\mathrm{X}_2) \in [0,1], \mu_{2.4}(\mathrm{X}_2) \in [0,1, \mu_{2.5}(\mathrm{X}_2) \in [0,1]$, $\mu_{2.6}(\mathrm{X}_2) \in [0,1]$, $\mu_{2.7}(\mathrm{X}_2) \in [0,1]$, $\mu_{2.8}(\mathrm{X}_2) \in [0,1]$.

Based on the data obtained, namely the percentage of the frequency of occurrence of threats from the total number, the input variable "IS threat" in the developed fuzzy logic model will have the form shown in Fig. 4.

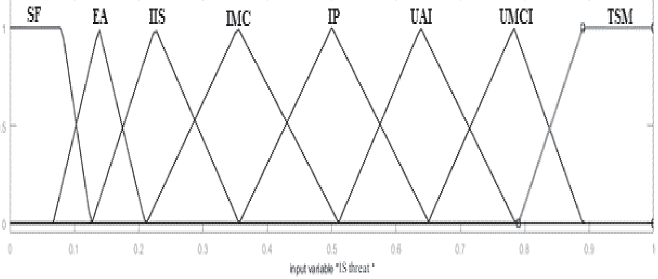


Fig. 4. Description of the linguistic variable "Information Security threat"

As a term-set of variable X_3, we will use the set T_3 = (no anti-virus protection, no backup, no media accounting, simple password in the system, no access control, no differentiation of rights, no control over user actions) = $(\mathrm{T}_{3.1}, \mathrm{T}_{3.2}, \mathrm{T}_{3.3}, \mathrm{T}_{3.4}, \mathrm{T}_{3.5}, \mathrm{T}_{3.6}, \mathrm{T}_{3.7},)$ with membership functions : $\mu_{3.1}(\mathrm{X}_3) \in [0,1], \mu_{3.2}(\mathrm{X}_3) \in [0,1], \mu_{3.3}(\mathrm{X}_3) \in [0,1], \mu_{3.4}(\mathrm{X}_3) \in [0,1]$, $\mu_{3.5}(\mathrm{X}_3) \in [0,1]$, $\mu_{3.6}(\mathrm{X}_3) \in [0,1], \mu_{3.7}(\mathrm{X}_3) \in [0,1]$.

Since there is no data on the severity of vulnerabilities, they are evenly distributed. Thus, the input variable "Vulnerability" in the developed fuzzy logic model will have the form shown in Fig. 5.

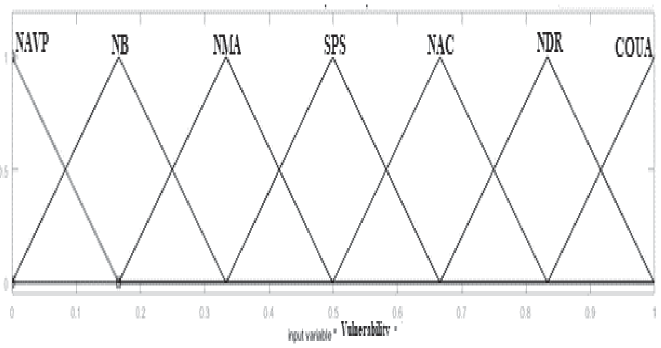


Fig. 5. Description of the linguistic variable "Vulnerability"

As a term-set of the output variable Y_1, we will use the set T_{Y_1}= (identification and authentication of subjects and objects of access, control of the security of personal data, ensuring the integrity of the information system and personal data) = $(T_{Y_{1.1}}, T_{Y_{1.2}}, T_{Y_{1.3}})$ with membership functions : $\mu_{Y_{1.1}}(Y_1) \in [0,1]$, $\mu_{Y_{1.2}}(Y_1) \in [0,1]$, $\mu_{Y_{1.3}}(Y_1) \in [0,1]$.

Since there is no data on the effectiveness of the measures taken in relation to threats, vulnerabilities and the intruder, we will assume that they are evenly distributed. Thus, the output variable "Organizational measures" in the developed fuzzy logic model will have the form shown in Fig. 6.

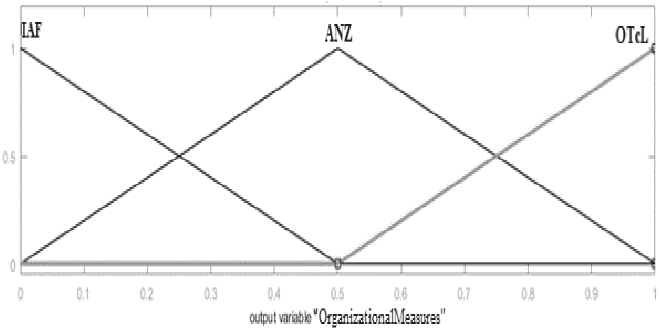


Fig. 6. Description of the linguistic variable "Organizational measures"

As a term-set of the output variable Y_2, we will use the set T_{Y_2} = (ensuring the availability of personal data, controlling access of subjects of access to access objects, managing the configuration of the information system, protecting hardware, limiting the software environment, antivirus protection, protecting machine media) = $(T_{Y_{2.1}}, T_{Y_{2.2}}, T_{Y_{2.3}}, T_{Y_{2.4}}, T_{Y_{2.5}}, T_{Y_{2.6}}, T_{Y_{2.7}})$ with membership functions : $\mu_{Y_{2.1}}(Y_2) \in [0,1]$, $\mu_{Y_{2.2}}(Y_2) \in [0,1]$, $\mu_{Y_{2.3}}(Y_2) \in$

$[0,1], \mu_{Y_{2.4}}(Y_2) \in [0,1], \mu_{Y_{2.5}}(Y_2) \in [0,1]$, $\mu_{Y_{2.6}}(Y_2) \in [0,1]$, $\mu_{Y_{2.7}}(Y_2) \in [0,1]$.

Since there is no data on the effectiveness of the measures taken in relation to threats, vulnerabilities and the intruder, we will assume that they are evenly distributed. Thus, the output variable "Technical measures" in the developed fuzzy logic model will have the form shown in Fig. 7.

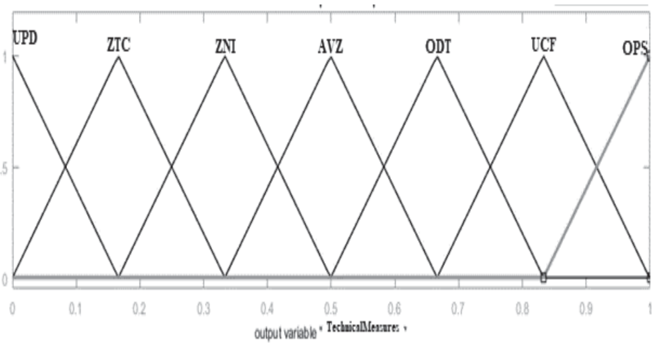


Fig. 7. Description of the linguistic variable "Technical measures"

3. Formation of the rule base of the fuzzy inference system.

Fuzzy rules take the form of sentences of the form: IF "..." AND "..." THEN "...", the conditional part of which is an expression of fuzzy logic over the linguistic values of the selected criteria and the relationship between them and constitute a construction: Rule 1: If "Condition A1" and “Condition B1”, then “Corollary C1”. This model of a logical inference system is based on a process of reasoning similar to human reasoning [7]. The input variables describe the conditions for its applicability, and the conclusion of the rule determines the membership functions of the values of the output linguistic variables.

To define the system's rule base, organizational and technical measures have to be defined. To select and evaluate the effectiveness of the selected protection measures, the method of expert assessments can be used, which refines the choice depending on the characteristics of specific ISPD. The proposed list of measures to eliminate each of the threats is advised by the requirement of FSTEC order No. 21. Let's form 54 rules of fuzzy inference with the corresponding term-sets.

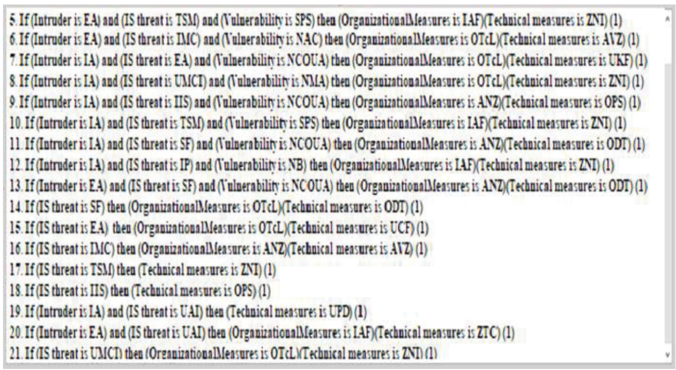

5. If (Intruder is EA) and (IS threat is TSM) and (Vulnerability is SPS) then (OrganizationalMeasures is IAF)(Technical measures is ZNI) (1)
6. If (Intruder is EA) and (IS threat is IMC) and (Vulnerability is NAC) then (OrganizationalMeasures is OTcL)(Technical measures is AVZ) (1)
7. If (Intruder is IA) and (IS threat is EA) and (Vulnerability is NCOUA) then (OrganizationalMeasures is OTcL)(Technical measures is UKF) (1)
8. If (Intruder is IA) and (IS threat is UMCI) and (Vulnerability is NMA) then (OrganizationalMeasures is OTcL)(Technical measures is ZNI) (1)
9. If (Intruder is IA) and (IS threat is IIS) and (Vulnerability is NCOUA) then (OrganizationalMeasures is ANZ)(Technical measures is OPS) (1)
10. If (Intruder is IA) and (IS threat is TSM) and (Vulnerability is SPS) then (OrganizationalMeasures is IAF)(Technical measures is ZNI) (1)
11. If (Intruder is IA) and (IS threat is SF) and (Vulnerability is NCOUA) then (OrganizationalMeasures is ANZ)(Technical measures is ODT) (1)
12. If (Intruder is IA) and (IS threat is IP) and (Vulnerability is NB) then (OrganizationalMeasures is IAF)(Technical measures is ZNI) (1)
13. If (Intruder is EA) and (IS threat is SF) and (Vulnerability is NCOUA) then (OrganizationalMeasures is ANZ)(Technical measures is ODT) (1)
14. If (IS threat is SF) then (OrganizationalMeasures is OTcL)(Technical measures is ODT) (1)
15. If (IS threat is EA) then (OrganizationalMeasures is OTcL)(Technical measures is UCF) (1)
16. If (IS threat is IMC) then (OrganizationalMeasures is ANZ)(Technical measures is AVZ) (1)
17. If (IS threat is TSM) then (Technical measures is ZNI) (1)
18. If (IS threat is IIS) then (Technical measures is OPS) (1)
19. If (Intruder is IA) and (IS threat is UAI) then (Technical measures is UPD) (1)
20. If (Intruder is EA) and (IS threat is UAI) then (OrganizationalMeasures is IAF)(Technical measures is ZTC) (1)
21. If (IS threat is UMCI) then (OrganizationalMeasures is OTcL)(Technical measures is ZNI) (1)

Fig. 8. Formed rules of fuzzy inference

Fig. 9 shows the implementation of the rules graphically.

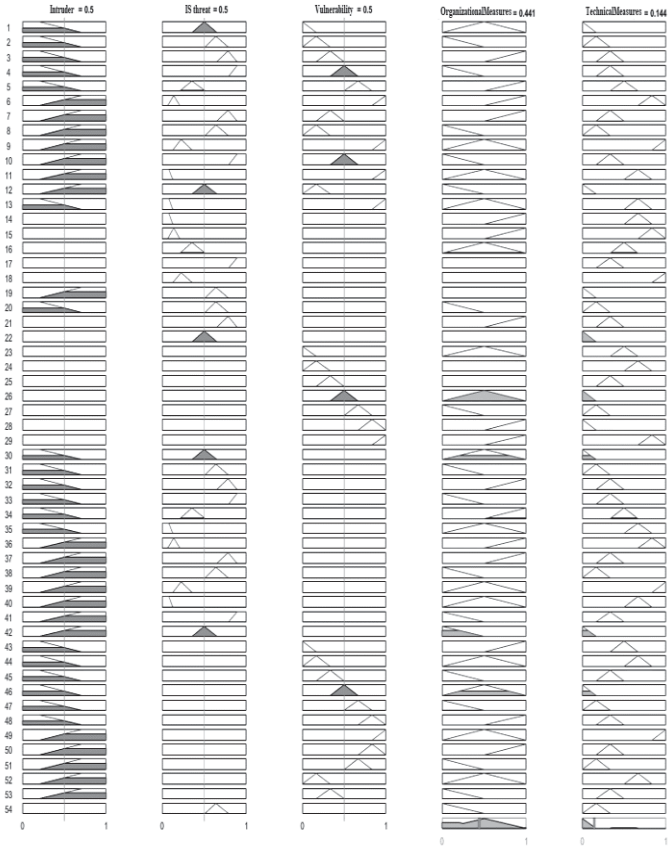


Fig. 9. A graphical representation of the fuzzy inference rule

4. Aggregation of subconditions in fuzzy production rules.

Determination of the degree of truthfulness of the conditions for each rule of the fuzzy inference system and determination of the cut-off level for the left side of each of the rules by the formula:

$$alfa_i = min_i(A_{ik}(X_k)),$$

where A_{ik} − the degree of truth of fuzzy statements;
X_k − fuzzy element.

5. Activation of subconclusions in fuzzy production rules.

Next, the truncated membership functions are found by the formula:

$$B_i^*(y) = min_i\left(alfa_i, B_i(y)\right),$$

where $B_i^*(y)$ − activated membership function;
$alfa_i$ − degree of truth of the i-th subconclusion;
$B_i(y)$ − term membership function.

6. Accumulation of conclusions of fuzzy production rules.

The union of the obtained truncated functions by the maximum composition of fuzzy sets [7]:

$$MF(y) = max_i\,(B_i^*(y)),$$

where $MF(y)$ − the membership function of the final fuzzy set.

7. Defuzzification (transformation of fuzzy sets into a specific value of the output variables at the output)

At this stage, the clear meaning of the output variables is determined - the meaning of organizational and technical protection measures. The value is determined using the centroid method - determining the center of gravity of the resulting curve to determine the maximum degree of compliance [9]:

$$R = \frac{\int_0^1 R * \mu_Z(R)dR}{\int_0^1 \mu_Z(R)\, dR},$$

where $R -$ clear value of the output variable;
$\mu_Z(R) -$ output variable membership function.

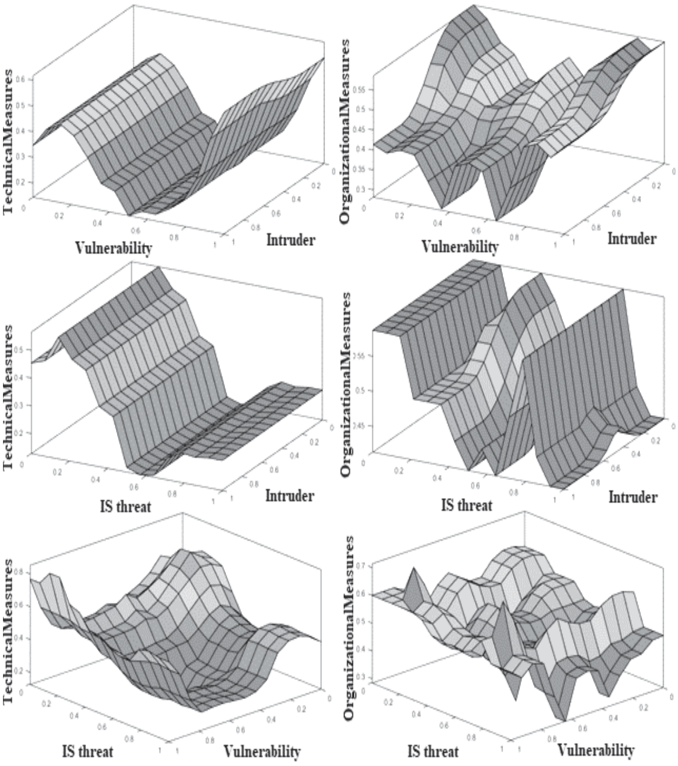


Fig. 10. Planes of centroid values

The planes obtained with fuzzy inference show the relationship between the input and output parameters of the model.

To make a decision on information protection measures, it is advisable to determine the boundary values of the output parameters, which will indicate the level of need for such measures. The decision-making thresholds can vary based on the characteristics and the required level of protection of the ISPD, and the composition of such measures is determined by the subject of information security. In this paper, three thresholds are proposed, depending on which a decision can be made:

- if $R \in [0,0.37]$ – **green zone,** then the decision on the implementation of the received measures to neutralize the threat may not be taken;
- if $R \in [0.37,0.64]$ – **yellow zone,** then the decision on the implementation of the received measures to neutralize the threat must be taken into account;
- if $R \in [0.64,1]$ – **red zone,** then the decision on the implementation of the received measures to neutralize the threat must be carried out.

Decision thresholds are shown in Fig 11.

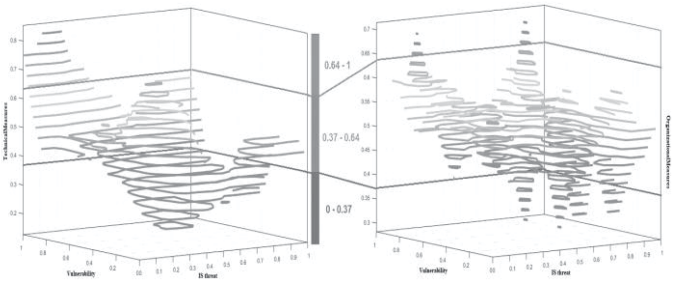


Fig. 11. Thresholds for decision making

The vector of input variables of the model determines a specific point on the plane, the position of which determines the decision on the choice of information protection measures based on the specified thresholds.

V. INFLUENCE OF CHANGE IN INPUT VARIABLES ON THE COMPLEX OF PROTECTIVE MEASURES TO PROVIDE IS

Since for a specific ISPD the probabilistic characteristics of threats, vulnerabilities and intruders with their potential capabilities usually have individual characteristics, this article studies the response of the proposed model to a change in input characteristics.

To identify the influence of the input variables on the output values of the constructed model, it was assumed that the terms of the membership functions of the input parameters were uniformly distributed (Fig. 12).

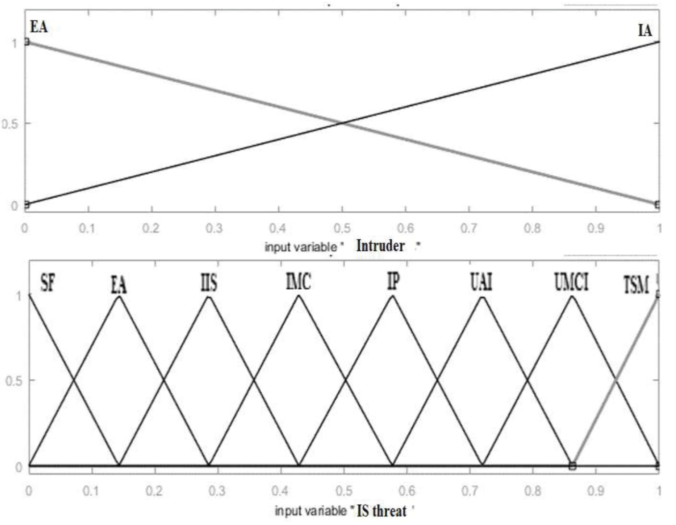


Fig. 12. Modified membership functions

Comparing the results of modeling the planes of centroid values indicated in Fig. 13, where plane a) is built on the basis of the distribution of threats according to Fig. 4, and plane b) is based on the distribution of threats according to Fig. 12, the following conclusions can be drawn.

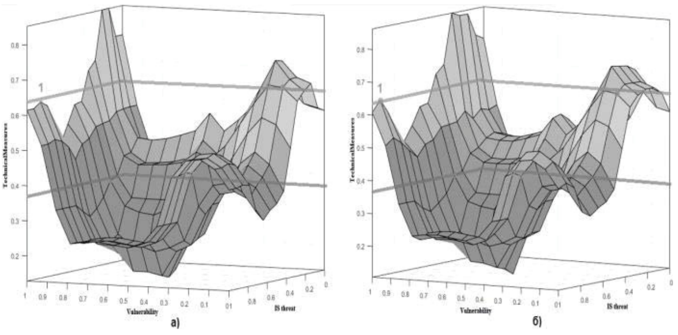
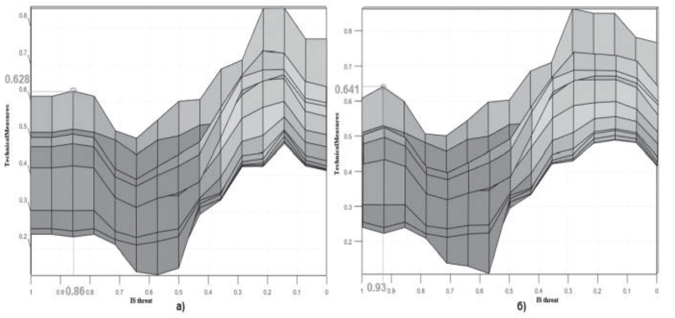


Fig. 13. Illustrations explaining the thresholds for decision-making on the measures of RFI, for various input variables

As can be seen from Figure 13, the indicated points on the planes have different levels of decision making for the same values of the input variables. But since the indicated values of the output parameters are on the border of the adopted decision-making thresholds, the decisions on the choice of the necessary information protection measures can differ significantly. Let's consider point 1 in more detail and define the value of the "IS threat" parameter (Fig 14).



Fig, 14. Value of the variable "Information security threat"

Based on Figures 13 and 14, such a difference in decisions regarding technical measures is likely in the scenario when the numerical value of the input parameter "IS threat" has the values shown in Fig. 15. The numerical value of the input parameter "Vulnerability" is equal to 1.

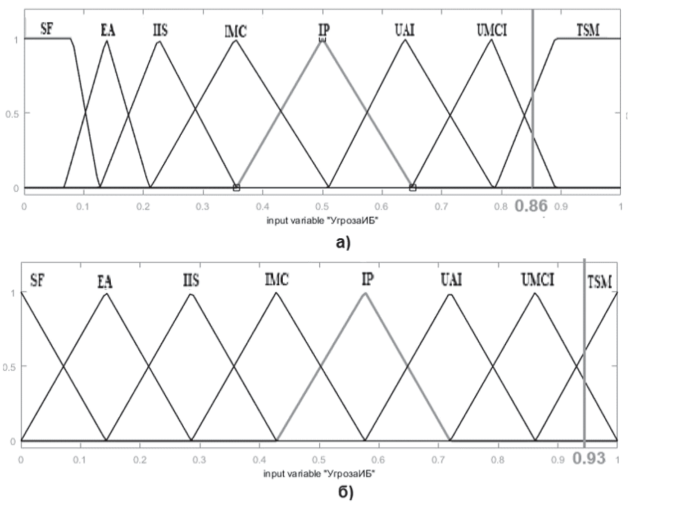


Fig. 15. Comparison of values on accessory functions

Describing this scenario, the numerical value of the input parameter “IS threat” belongs rather to the interval of the term “Theft of media” rather than the term “Unauthorized modification and copying”. The value of the input parameter "Vulnerability" belongs to the term "Lack of user control". The value of the "Intruder" input parameter belongs equally to the terms "External intruder" and the term "Internal intruder".

In case of the location of the terms specified in option a), the numerical value of the output parameter "Technical measures" is in the yellow threshold of decision-making, while when the terms of option b) are located, the level of measures is in the range of the red threshold. The value of the output parameter "Technical measures" in this case belongs to the interval of the term "ODT". The numerical values of the “Organizational Measures” output parameter are 0.489 and 0.549, which belongs to the “ANZ” term interval. These values are in one decision-making threshold - yellow.

Thus, the given example of changing the membership functions of input variables indicates the need for an adequate description of input variables, on the one hand, and the response of the model's output variables to input variables in order to take effective measures to protect information, on the other.

VI. CONCLUSION

The model developed on the basis of this algorithm using fuzzy inference identifies the influence of information security threats on decision-making on the choice of measures to protect personal data in a computer system, and, based on the obtained planes, it allows to determine the managerial decision on the choice of measures depending on the values of the selected parameters.

REFERENCES

[1] FSTEC official website, Basic model of threats to the security of personal data when processing them in personal data information systems, Web: https://fstec.ru/component/attachments/download/289.

[2] 11 Zadeh L.A., Kacprzyk, J. Fuzzy Logic for the Management of Uncertainty. – NY: John Wiley. 1992.

[3] E. K. Baranova, A.V. Babash. Modeling of information security systems. Workshop: textbook, Moscow RIOR INFRA-M, 2018. - 150 p.

[4] FSTEC official website, FSTEC Order No. 21 "On approval of the composition and content of organizational and technical measures to ensure the security of personal data during their processing in Personal data Information systems", Web: https://fstec.ru/component/attachments/download/561.

[5] Zadeh, L.A., “Fuzzy sets”, Information and control, vol.8, pp. 338-353, 1965

[6] A. S. Bolshakov, E. A. Rogatneva. information security risk assessment using fuzzy logic algorithms.Telecommunications and information technologies. 2018. Vol. 5. No. 2. pp. 142-147.

[7] A. S. Bolshakov, E. A. Rogatneva. The application of fuzzy logic for the management of information risk. Information society technologies. Materials of the XIII International Industrial Scientific and Technical Conference. 2019. pp. 331-335.

[8] Azhmukhamedov I. M., Solving problems of ensuring information security on the basis of system analysis and fuzzy cognitive modeling, Monograph, 2012, p. 344.

[9] R. Belfer, D. Kalyuzhnyy, D.Tarasova, Analysis of dependence of risk level of safety of communication networks on expert data during calculations with the use of a model of the illegible sets// Cybersecurity issues №1(2) – 2014, с 36.