Structural and Analytical Models for Early APT-attacks Detection in Critical Infrastructure

Zhadyra Avkurova[a], Sergiy Gnatyuk[b,c,d], Bayan Abduraimova[a]

[a] *L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan*
[b] *National Aviation University, Kyiv, Ukraine*
[c] *State Scientific and Research Institute of Cybersecurity Technologies and Information Protection, Kyiv, Ukraine*
[d] *Yessenov University, Aktau, Kazakhstan*

Abstract

Modern information and communication technologies (ICT) are vulnerable to APT-attacks (advanced persistent threats) and other relevant threats. APT-attack is a stealthy threat actor, typically a nation-state or state-sponsored group, which gains unauthorized access to ICT and remains undetected for an extended period. Early detection of APT-attack is very important for ICT of critical infrastructure. But existed approaches don't allow to detect effectively in fuzzy environment (networks / cyberspace). In this paper, a method of linguistic terms using statistical data was used for structural and analytical models of parameters (both host and network parameters) as well as intruder model based on the defined host and networks parameters was developed. Based on this, logical rules can be developed to provide the functioning of IDS based on honeypot technology for APT-attacks detection and intruder type identification in ICT.

Keywords

APT-attack, Early Detection, Identification, Honeypot, Fuzzy Logic, Parameter, ICT.

1. Introduction

The development of information and communication technology (ICT) creates new types of threats to information security, among which the intruder in computer systems and networks (for example, APT-attacks or other negative influences) occupies a prominent place. To effectively counter this threat, IDS (intruder detection system) are being developed to detect and identify an intruder. Early detection is important and not simple task for security side. Typical IDS should perform the following main functions [1]:

- monitor and analyze the activity of ICS (information and communication system) users;
- capture system configurations and vulnerabilities;
- assess the integrity of critical system files and data files;
- recognize activity patterns that reflect known attacks;
- perform statistical analysis to detect abnormal behavior;
- recognize violations of security policy by the system user.

2. Related papers analysis

The IDS tasks can be divided into global and local. Global tasks – recognition of the violator (intruder) and legitimate user – the solution of this problem contains the following stages [2-3]: data collection, filtering, behavior classification – directly the process of recognizing the violator, report and response system. As can be seen from the main functions and tasks of IDS, one of the most important aspects of their functioning is not only the fixation of intrusion in ICS, but also its identification.

There are many studies related with APT-attacks early detection. In [4-5] the big data processing approach was proposed for APT-attacks detection. In [6-9] authors proposed malware and DoS-attacks detection system as well as game theory based approach for APT-attacks detection. Presented techniques have many advantages (indicators, correlation, high-speed and others), but they don't allow identifying intruders` category as well as don't give possibility to operate with fuzzy parameters. That is why, the ***main task*** of this study is creation the possibility for early APT-attacks detection using developed structural and analytical models as well as method of linguistic terms using statistical data.

3. Development of the structural and analytical models based on host and network parameters

3.1. Basic parameters for intruders identification

In the process of attack, the violator, acting on the system, changes certain parameters, creates or terminates its inherent processes, and so on. All these actions are reflected in the state of the system. Evaluating these parameters, you can detect the fact of intrusion into the system. The work of modern IDSs is based on this principle. Thus, the NIDES system performs audits of such processes as logging in, working with files and processes, administration and fixing errors and failures. Previous works describe the parameters by which the violator is identified by the developed system. These parameters (are host settings) include:

1) Host Parameters: Username at login, *UID;* Login time, *Tlog;* Frequency of login requests, *Nlog;* Time spent logging in, *TSlog;* Intensity of actions, *I;* Processor time / CPU usage, *CPU;* The amount of RAM load, *Muse;* Number of executable files, *NEF;* The type of files used in the attack, *AtEF;* Number of failures and errors, *NEr;* Process / file execution time, *RTPr/F;* Unusual processes, *UPr;* File transfer to the system, *TrFin;* Files changes, *ModF;copying* / transferring files from the system, *TrFout;* Pressing the keyboard keys, *KS.*

2) Network Settings – characteristics of *ARP-*, *IP-*, *ICMP-* and *TCP*-packages.

Since the process of detection and identification of the violator takes place in conditions of uncertainty, and some of the parameters of the IDS are unclear, the operation of such a system should be based on fuzzy logic. To identify the violator, we can use the logical-linguistic approach and the basic model of parameters, partially described in [10], which will be the basis for the construction of the developed IDS. For example, to detect the process of port scanning in section [11] used linguistic variables (LV) "Number of virtual channels" and "Age of virtual channels", and in section [12] LV "Number of simultaneous connections", "Query processing speed", "Delay between requests" and 'Number of packets with the same sender and recipient address"– to detect DDoS attacks and spoofing.

The process of detecting and identifying the violator requires determining the necessary parameters and their properties. In this regard, the main purpose of this work is to build models of standards required for the operation of IDS in a vaguely defined, poorly formalized environment.

3.2. Method of linguistic terms using statistical data

Consider the method of linguistic terms using statistical data (MLTS) [13], where as a measure of belonging of the element to the set is an estimates of the frequency of use of the concept, which is given by a fuzzy set to characterize the element. To do this, the values of the linguistic variable (LV) are placed on the universal scale [0; 1] $X=\{x_1, x_2, \dots, x_n\}$. The method is based on the condition that the same number of experiments falls into each interval of the scale, but this is usually not followed in practice. An empirical table is compiled in real conditions, in which experiments can be unevenly distributed over intervals. Some of them may not be involved, and then the data is processed using a matrix of prompts.

May it is necessary to estimate in values of LV deviations of the parameter $\Delta B \in [0, B]$ (where B is the maximum possible deviation), which characterizes the current measurements. Next for $n = 5$ determine the value of LV $\{x_1, x_2, x_3, x_4, x_5\}$. Interval $[0, B]$ and $\Delta B/B$ (estimated ratio) divided into k segments (for example, 5), on which the statistics characterizing frequency of use by the expert of the value of drugs for the display of the conclusions gathers. Then the data are entered into the table and processed to reduce the errors made during the experiment: the table is removed individual elements on the left side and on the right side of which there are zeros in the row. The tooltip matrix is a string whose elements are calculated by the formula:

$$k_j = \sum_{i=1}^{n} b_{ij} = \sum_{i=1}^{5} b_{ij},\ j = \overline{1,5}\ . \qquad (1)$$

Next, in the resulting row of the matrix, the maximum element is selected $k_{\max} = \max k_j$, and then all elements of the table are converted by expression

$$c_{ij} = b_{ij} k_{max} / k_j, \ i = \overline{1,5}; j = \overline{1,5}, \quad (2)$$

and for columns, where $k_j = 0$ the linear approximation is applied $c_{ij} = (c_{ij-1} + c_{ij+1})/2, \ i = \overline{1,5}; j = \overline{1,5}$. Next, calculate the value of MF (membership function) by the formula

$$\mu_{ij} = c_{ij} / c_{imax}, \ c_{imax} = \max_j c_{ij}, \ i = \overline{1,5}; j = \overline{1,5}. \quad (3)$$

The described method uses data from statistical studies. Their processing is quite time consuming, because to build a MF of one term it is necessary to conduct statistical studies of all terms of LV. We construct a model of standards of linguistic variables for fuzzy parameters of violator identification from the set of parameters (host and network). Model contains (4) as well as Table 1 and Table 2.

$$DIO = <UID, Tlog, Nlog, TSlog, I, CPU, MUse, NEF, AtEF, NEr, RTPr/F, UPr, TrFin, ModF, TrFout, KS, ARP, IP, ICMP, TCP>. \quad (4)$$

3.3. Models of intruders host and networks parameters

The system must monitor certain parameters of the IS (Table 1), record them and identify violator.

Table 1

Host parameters for violator identification and their characteristics

Parameter	Blur	Human				Bot	
		Misinformer	*Spammer*	*Cracker*	*Hacker*	*Spam-bot*	*Bot-hackers*
UID	-	+	-	+	+	-	+
Tlog	+	With a certain probability depending on the time of day (*)	-	*	*	-	*
Nlog	+	Above average	-	Above average	Above average	-	High
TSlog	+	Above average	-	Above average	Above average	-	Above average
I	+	Within the norm	Within the norm	Within the norm	Within the norm	Above the norm	Above the norm
CPU	+	Above the norm	Above the norm	Above the norm	Above the norm	Above the norm	Above the norm
MUse	+	Above the norm	Above the norm	Above the norm	Above the norm	Above the norm	Above the norm
NEF	+	Not within the norm	-	Not within the norm	Not within the norm	-	Not within the norm
AtEF	-	Scripts and PHP scripts	PHP scripts	Executable files	Scripts	PHP scripts	Scripts
NEr	+	Above the norm	Above the norm	Above the norm	Above the norm	Above the norm	Above the norm
RTPr/F	+	Differs from the typical time (**)	**	**	**	**	**
UPr	-	Present	Present	Present	Present	Present	Present
TrFin	-	Present	Present	Present	Mostly absent	Present	Mostly absent

ModF	-	Present	Absent	Present	Mostly present	Absent	Mostly present
TrFout	-	Absent	Absent	Mostly present	Present	Absent	Present
KS	-	It is fixed	It is fixed	It is fixed	It is fixed	It is not fixed	It is not fixed

Consider ***host parameters*** in more detail:

1) Login username (UID). In the database of VDS or honeypot, on the basis of which this system is built, a list of user names (logins) who are allowed to use IS resources (i.e. which are authorized) must be defined and stored. Any other usernames which is not included in this list are considered unauthorized and their appearance indicates unauthorized login. This parameter is clear because the appearance of an inherent login clearly indicates intruder. However, spammers, spam bots and search bots usually do not require authorization in the system and therefore the fact of their penetration into the system by this parameter is mostly impossible to determine.

2) System login time (Tlog). The parameter is based on the fact that the activity of IS and users of these systems depends on the time of day. Usually, more user activity is logged in during the day, less - at night, but other statistics are possible, which is determined by the mode of operation of the organization to which the IS belongs. The nature of this parameter is unclear, because it is impossible to unambiguously draw a conclusion about the illegal activity of intruder. Thus, in organizations with working hours from 08:00 to 16:00, the probability that the user who has logged in is the lowest is at 08:00 and increases over time, reaching a maximum in the hours after 16:00. However, it should be noted that in the concept of honeypot-technologies, this parameter loses some weight, as any activity on them is considered malicious.

3) Frequency of login requests (Nlog). It is clear that the highest frequency of login requests will be observed during attacks by bots (including hacker bots, as spammers do not require login). Intruder human is also marked by an increased frequency of requests due to attempts to circumvent the protection and the theoretical assumption that it does not have a legitimate login and password, so it will be forced to make at least a few attempts. Moreover, the greater the number of attempts, the more likely that the IS is really trying to enter intruder. It is clear that this parameter is also fuzzy.

4) Time spent logging in (TSlog). A parameter that is closely related to the previous one. The time spent by intruder is in most cases longer than the time spent by a legitimate user. But it is unclear because it does not allow unambiguous identification.

5) Intensity of actions (I). This means the number of any user actions, including login / logout, transfer, change, copy files, start / stop processes, etc. per unit time. The intensity may not differ in intruder human and in the legitimate user, but in bots it is much higher, so it is most important for the identification and differentiation of human-bot categories. Although a significant excess of the norm indicates the activity of unauthorized automatic systems - intruder (bots), but *I* is a fuzzy parameter, because the normal value of the intensity index is very difficult to determine.

6) Processor time / CPU usage (CPU). Since the number of active processes on honeypot-systems must be minimal, any increase in load is a sign of the activity of intruder in the system. In real IS, the probability that the activity is caused by the intruder is slightly lower, and, of course, the normal value of CPU time is higher. However, this parameter can still be effectively used to identify violations in intrusion detection systems and VDS. Since it is impossible to give an unambiguous answer about the IS for this parameter, primarily due to the possible activity of viruses, the CPU is a fuzzy parameter.

7) The amount of loaded RAM (MUse). Similar in content to the previous one and is also vague.

8) Number of executable files (NEF). It can also be included in the group of fuzzy parameters. The fact of actions of the malefactor on this parameter is defined by deviation from norm. Thus, in each organization, in accordance with the security policy and job responsibilities, each legitimate user can use certain files at a given time, and the simultaneous use of many files at once is virtually excluded. This makes it possible to detect both external and internal intruder, but with a certain probability.

9) The type of files used in the attack (AtEF). If you notice a recently modified or created file that identifies as a script, we are dealing with a hacker, a person who is highly computer literate and able to use the script to further break into systems. If the observed file is an executable file, then by definition "...the result of the cracker's work is ... modified ("cracked" or "broken") program with the required

functionality" we can claim that the person who broke the server protection is a cracker. And finally, the case when we find a PHP script, clearly speaks of a cracker running on the Internet. According to modern research, the largest number, among the considered categories of attackers, DDoS-diggers and spammers. The misinformer can use several types of files, mainly scripts and PHP scripts. Since the result of applying the parameter gives an unambiguous answer about the presence of intruder and its class, the parameter is clear.

10) Number of failures and errors (NEr). This parameter is unclear, as failures and errors can occur during the operation of both the authorized user and intruder. However, with frequent failures or errors, it can be concluded with a certain degree of probability that the system is attacked. This group includes a wide range of events from authorization errors to failures in certain processes or files. During active operation, regardless of its class and category, the frequency of failures will be slightly higher. It should also be noted that it is possible that when identifying intruder bot, this frequency will be even higher.

11) Process / file execution time (RTPr/F). Examining the statistics of IS of different enterprises and organizations, it is easy to see that depending on the specifics of the time spent on a particular operation is approximately the same for the same type of IS and their tasks. Honeypot-systems mainly run system processes, i.e. those that support the work of the honeypot itself, or administrator processes that run at a certain time for a certain period. Thus, when identifying such processes, we can conclude that the attack system intruder. Since this state of affairs can be caused by negligence of the employee, the conclusion is ambiguous and, accordingly, the parameter is unclear.

12) Unusual processes (UPr). According to the concepts of VDS and honeypot-systems in the IS should be constantly monitored processes. Thus, during the operation of the system can be formed so-called molds of the system, which record all the activity on the host or create lists of processes and their characteristics that have been started. In the case of an unusual process in the work of the IS, i.e. one that for a long time did not start at all or started a small number of times, our VDS immediately notes the fact of the emergence of intruder. Since in this case the probability of the correct raising of the alarm is almost equal to "1", the parameter can be classified as clear.

13) File transfer to the system (TrFin), files change (ModF), copy / transfer files from the system (TrFout) - it is a group of clear parameters. Any actions with files are inherent in each attack, but those actions that prevail during the attack, determine the class and category of intruder. For example, a spam attack usually indicates the transfer of files to the system, but their change or transfer from the system is mostly absent. Similarly identify other categories of intruder.

14) Pressing the keyboard keys (KS). This technology uses monitoring to detect attacks by pressing keyboard key. The main idea is that the sequence of user clicks sets the attack pattern. The disadvantage of this approach is the lack of a fairly reliable mechanism for intercepting the keyboard without the support of the operating system, as well as a large number of possible options for presenting the same attack. In addition, without a semantic analyzer of clicks, various command aliases can easily destroy this technology. Because it aims to analyze keystrokes, automated attacks that result from an attacker's programs may also not be detected. But this fact is the most useful for us in the process of identifying intruder and classifying it as a class of people or robots. This parameter is unambiguous, so it is also classified as clear.

Network part works with network traffic and detect attacks associated with low-level impact on network protocols, and can detect attacks on multiple network hosts. Network VDS is based on an intelligent traffic analyzer, which processes each frame of data passing through it, in order to search for prohibited signatures that indicate attacks. Network data, network traffic is received from a network adapter operating in a promiscuous mode (i.e. receiving all packets on the network).

Consider ***network parameters*** (with the characteristics of the TCP / IP protocols) in more detail:

ARP request is monitored by the following parameters: IP address of source; source hardware address; network interface that limits the ARP request.

IP-fragment: source address; receiver address; protocol field; offset field; length; header length; MF bit; identification.

ICMP message: source IP address; IP address of the receiver; ICMP field type; ICMP identifier; ICMP sequence number.

TCP-package: source IP address; IP address of the receiver; TCP source port; TCP receiver port; bits of the TCP code.

Table 2
Network parameters for intruder identification and their characteristics

Parameter	Blur	Human				Bot	
		Misinformer	*Spammer*	*Cracker*	*Hacker*	*Spam bot*	*Bot hackers*
ARP-request	-	Does not meet the allowed (***)	***	***	***	***	***
IP-fragment	-	***	***	***	***	***	***
ICMP-message	-	***	***	***	***	***	***
TCP-package	-	***	***	***	***	***	***

All these network parameters, provided the correct configuration of the interconnection policy, clearly indicate the attack, and therefore belong to the group of clear.

3.4. Structural and analytical models investigation

Login time, Tlog. This parameter is based on the fact that the activity of the ICS and users of this system depends on the time of receipt. Usually, the usual greater activity of users to log in is detected on the last day, less – at night. Still, other statistics are possible, determined by the mode of operation of the organization to which the ICS belongs. The nature of these parameters is unclear, due to which it is impossible to conclude the message's illegal activity unambiguously. Thus, in organizations working from 08.00 to 16.00, the probability of who is the user who logs in – the message is lowest at 08.00 and increases over time, reaching a maximum in the years after 16.00. However, it should be changed that in the concepts of honeypot-technology, this parameter loses weight, as any activity on them is considered criminal. Let's evaluate the LV "Level of legitimacy over time". Determine the value of the linguistic variable $\{x_1, x_2, x_3\}$, corresponding {legitimate, suspicious, illegitimate}. That is $T_{Tlog} = \bigcup_{i=1}^{3} T^i{}_{Tlog} = \{legitimate, suspicious, illegitimate\}$, we use statistics for B = 24 hours. It is advisable to divide the total interval into 4 intervals [00:00;06:00], [06:00;12:00], [12:00;18:00], [18:00;24:00].

Table 3
Data for LV *Tlog*

The value of LV	Interval			
	№1	№2	№3	№4
High	0	8	6	1
Middle	2	1	2	3
Low	6	1	1	4

Using expression (1), we define $k_j = \|8\,10\,9\,8\|$, where k_{max} = 10, and in accordance with (2) calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 0 & 8 & 6,66 & 1,25 \\ 2,5 & 1 & 2,22 & 3,75 \\ 7,5 & 1 & 1,11 & 5 \end{Vmatrix}.$$

Calculate the MF by formula (3):

$$\left\|\mu_{ij}\right\| = \begin{Vmatrix} 0 & 1 & 0,83 & 0,16 \\ 0,66 & 0,26 & 0,59 & 1 \\ 1 & 0,13 & 0,15 & 0,66 \end{Vmatrix}.$$

For $\bigcup_{i=1}^{3} \mu_{ij}$ accordingly, we find the evaluation relationship $\bigcup_{i=1}^{3} \Delta B_i / B = \{0,25;\ 0,5;\ 0,75;\ 1\}$, and we obtain the following fuzzy numbers:

$$L = \{0/0,25;\ 1/0,5;\ 0,83/0,75;\ 0,16/1\},$$

$$P = \{0,66/0,25;\ 0,26/0,5;\ 0,59/0,75;\ 1/1\},$$

$$N = \{1/0,25;\ 0,13/0,5;\ 0,15/0,75;\ 0,66/1\}.$$

Schedule MF terms LV *Tlog* is shown in Fig. 1.

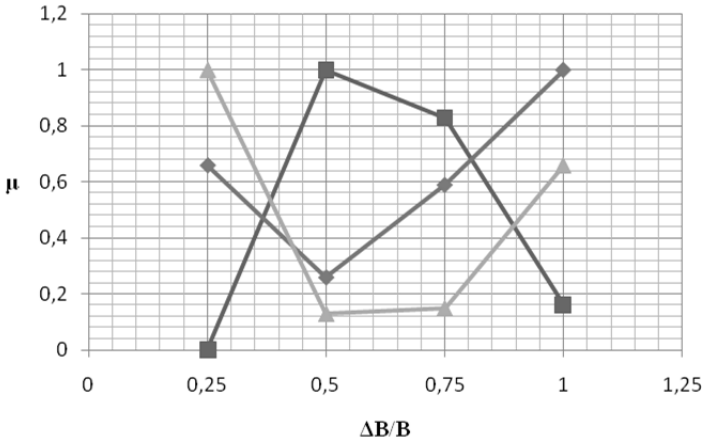

Figure 1: Linguistic standards of fuzzy numbers for *Tlog*

Frequency of login requests, Nlog. The highest frequency of login requests will be observed during system attacks by bots (including hacker bots, as spammers do not require login). The human offender is also marked by an increased frequency of requests due to attempts to circumvent the protection and the theoretical assumption that he does not have a legitimate login and password, so he will be forced to make at least a few attempts. And the greater the number of attempts, the more likely the violator is trying to enter the ICS. This parameter is also fuzzy.

Using (1) – (3), we can form structural and analytical models for analogically *Nlog* (Fig. 2, Table 4):

$$T_{Nlog} = \bigcup_{i=1}^{5} T^{i}{}_{Nlog} = \{\ low,\ below\ average,\ medium,\ above\ average,\ high\ \}.$$

Table 4
Data for LV *Nlog*

The value of LV	Interval				
	№1	№2	№3	№4	№5
Low	8	0	0	0	0
Below average	5	2	0	0	0
Average	1	6	4	0	0
Above average	0	2	8	1	0
High	0	0	1	6	6

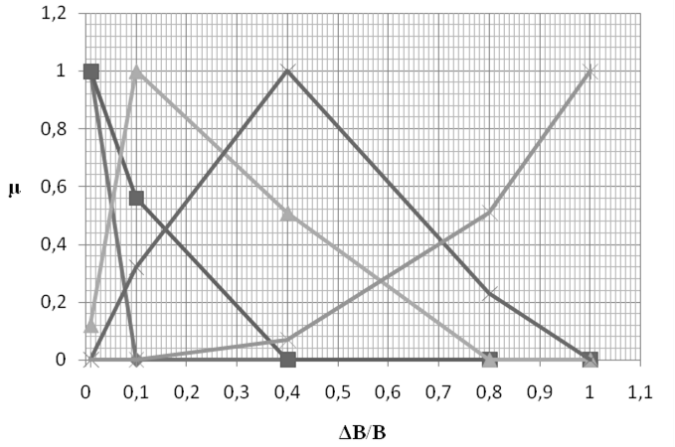

Figure 2: Linguistic standards of fuzzy numbers for *Nlog*

Time spent logging in, TSlog. A parameter that is closely related to the previous one. The time spent by the infringer is in most cases longer than the time spent by the legitimate user. But it is unclear because it does not allow unambiguous identification.

Using (1) – (3), we can form structural and analytical models for analogically *TSlog* (Fig. 3, Table 5):

$$T_{Slog} = \bigcup_{i=1}^{5} T^{i}_{Slog} = \{very\ small,\ small,\ medium,\ large,\ very\ large\}.$$

Table 5

Data for LV *TSlog*

The value of LV	Interval				
	№1	№2	№3	№4	№5
Very small	9	3	0	0	0
Small	5	10	1	0	0
Medium	1	7	5	0	0
Large	0	1	2	9	2
Very large	0	0	1	6	9

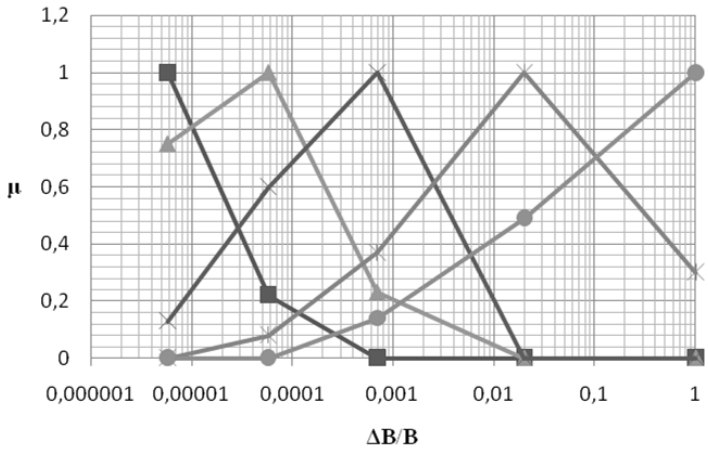

Figure 3: Linguistic standards of fuzzy numbers for *TSlog*

The intensity of actions, I. Here we mean the number of any user actions, including login/logout, transfer, change, copy files, start/stop processes, etc., per unit time. The intensity may not differ between the human violator and the legitimate user. Still, in bots, it is much higher, so it is essential to identify and delimitate human-bot categories. However, a significant excess of the norm indicates unauthorized automatic systems-violators (bots), a fuzzy parameter because the normal value of the intensity index is complicated to determine.

Using (1) – (3), we can form structural and analytical models for analogically *I* (Fig. 4, Table 6):

$$T_I = \bigcup_{i=1}^{3} T^i{}_I = \{low, medium, high\} .$$

Table 6
Data for LV *I*

The value of LV	Interval			
	№1	№2	№3	№4
Low	7	5	1	0
Middle	0	7	4	0
High	0	1	5	7

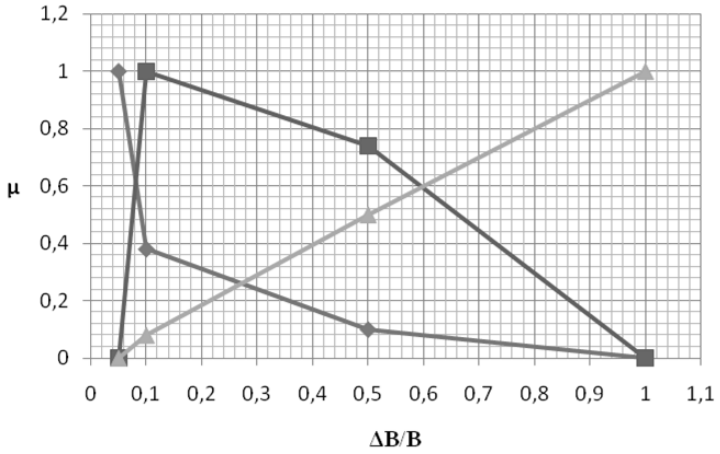

Figure 4: Linguistic standards of fuzzy numbers for *I*

Analogically, using (1) - (3), structural and analytical models for other parameters (*CPU, Muse, NEF, NEr, RTPr/F)* can be formed and presented.

3.5. Advantages of the proposed approach

The first part of research study contains structural and analytical models of the host and network parameters, MF as well as intruder model based on the defined parameters. In comparison with other approaches, these models give possibility to operate with fuzzy parameters (network and cyberspace are not determined environment) and create more flexible tools in the future. These results will be used for formation of the logical rules of IDS or other cyber threat detection system, which can be used for effective APT-attacks detection and intruder category identification.

4. Conclusions

In this paper, defined linguistic variables were introduced as well as structural and analytical models of parameters *Tlog, Nlog, TSlog, I, CPU, Muse, NEF, NEr, RTPr/F* were built. Also, for each described linguistic variables, MF were calculated and schedules of their terms were constructed. The formed standards are necessary for formation the system of logical rules allowing to provide functioning of IDS for APT-attacks detection and intruder category identification. Also, the intruder model based on the defined host and networks parameters was developed. These results can be used in sectors of critical infrastructure because APT-attacks are directed on them frequently.

The obtained results will be further used to build an IDS system (or other cyber threat detection system) based on honeypot technology. Next step of authors` research study is the rules system development for effective detection the fact of intrusion in ICS and identification of the person (category) of the intruder.

References

[1] M. Khosravi and B. T. Ladani, "Alerts Correlation and Causal Analysis for APT Based Cyber Attack Detection", in IEEE Access, Vol. 8, pp. 162642-162656, 2020, doi:10.1109/ACCESS.2020.3021499.

[2] Denning D.E. "An Intrusion-Detection Model", IEEE Transactions On Software Engineering, February 1987, Vol. SE-13, No. 2, pp. 222-232.

[3] Hu Z., Odarchenko R., Gnatyuk S., Zaliskyi M., Chaplits A., Bondar S., Borovik V. "Statistical techniques for detecting cyberattacks on computer networks based on an analysis of abnormal traffic behavior", International Journal of Computer Network and Information Security, Vol. 12, Issue 6, pp. 1-13, 2020.

[4] Y. Qi, R. Jiang, Y. Jia and A. Li, "An APT Attack Analysis Framework Based on Self-define Rules and Mapreduce", 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC), 2020, pp. 61-66, doi: 10.1109/DSC50466.2020.00017.

[5] D. Liu, H. Zhang, H. Yu, X. Liu, Y. Zhao and G. Lv, "Research and Application of APT Attack Defense and Detection Technology Based on Big Data Technology", 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), 2019, pp. 1-4, doi: 10.1109/ICEIEC.2019.8784483.

[6] X. Liu, L. Li, Z. Ma, X. Lin and J. Cao, "Design of APT Attack Defense System Based on Dynamic Deception", 2019 IEEE 5th International Conference on Computer and Communications (ICCC), 2019, pp. 1655-1659, doi: 10.1109/ICCC47050.2019.9064206.

[7] S. -P. Hong, C. -H. Lim and H. J. Lee, "APT attack response system through AM-HIDS", 2021 23rd International Conference on Advanced Communication Technology (ICACT), 2021, pp. 271-274, doi: 10.23919/ICACT51234.2021.9370749.

[8] Y. Su, "Research on APT attack based on game model", 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2020, pp. 295-299, doi: 10.1109/ITNEC48623.2020.9084845.

[9] M. Zaliskyi, R. Odarchenko, S. Gnatyuk, Yu. Petrova. A. Chaplits, "Method of traffic monitoring for DDoS attacks detection in e-health systems and networks", CEUR Workshop Proceedings, Vol. 2255, pp. 193-204, 2018.

[10] A. Paradise et al., "Creation and Management of Social Network Honeypots for Detecting Targeted Cyber Attacks", in IEEE Transactions on Computational Social Systems, vol. 4, No. 3, pp. 65-79, Sept. 2017.

[11] Svarovskiy S. "Approximation of membership functions for linguistic variables", Mathematical issues of data analysis, pp. 127-131, 1980.

[12] M. Zuzcak and P. Bujok, "Causal analysis of attacks against honeypots based on properties of countries", in IET Information Security, Vol. 13, No. 5, pp. 435-447, 9 2019, doi: 10.1049/iet-ifs.2018.5141.

[13] W. Zhang, B. Zhang, Y. Zhou, H. He and Z. Ding, "An IoT Honeynet Based on Multiport Honeypots for Capturing IoT Attacks", in IEEE Internet of Things Journal, Vol. 7, No. 5, pp. 3991-3999, May 2020, doi: 10.1109/JIOT.2019.2956173.