DEVELOPMENT OF A HEURISTIC SCANNER FOR AN ANTIVIRUS PROGRAM ON THE BASIS OF THE MAMDANI FUZZY LOGIC METHOD

S. Y. Gavrylenko, V. V. Chelak, A. A. Gornostal

Abstract: The article considers the means of antivirus protection of information, their advantages and disadvantages. An analysis of modern decision-making systems is carried out. The system of fuzzy logic is chosen. A module based on the Mamdani fuzzy logic method was developed, and the developed system was tested. The obtained results of the research showed the possibility of using the developed module in heuristic analyzers of intrusion detection systems.

Keywords: anti-virus protection of information, malicious software, intrusion detection systems, heuristic analyzer, Mamdani fuzzy logic.

1. Introduction

New information technologies are successfully introduced into all spheres of human activity and are the greatest value of modern society. With the growth of the value of information, the demand for it has grown, and with it - the number of people wishing to obtain unauthorized access to it by using computer viruses. This problem is exacerbated by the spread of Internet technologies, the transition to cloud technologies and the dynamic growth in the number of mobile devices. All these leads to an increase in the number of malicious software (VPO). Despite the laws adopted in many countries to combat computer crimes and the development of special software tools for virus protection, the number of new software viruses is constantly growing. Therefore, an actual topic is the development of effective methods and technologies to counter computer viruses.

Analysis of the literature showed many different antivirus programs, the basis of which is most often formed on signature and heuristic analysis technologies [1-3]. Threat data is collected from a variety of sources, including cloud infrastructure, botnet monitoring service, search robots, spam traps. New cyber threats are determined based on checking IP addresses, URLs, checksums of files, domains, timestamps, file names, DNS data, and other characteristics inherent in programs. The received information is carefully checked, systematized, cleaned and analyzed by the forces of company analysts developing antivirus software and using heuristic analysis tools [4].

Heuristic analysis is based on intellectual subsystems and on the theory of artificial intelligence, fuzzy logic methods, neural network theory, cluster analysis, genetic algorithms, etc. The main drawback of the heuristic method is the high frequency of false triggering. Eliminate this shortcoming is possible due to a reasoned choice of criteria for assessing the abnormal behavior of computer systems, improving existing and developing new intelligent anti-virus systems.

2. Development of a heuristic intrusion detection scanner into computer systems based on the Mamdani fuzzy logic method

Under the impact on the computer system of HPE: the system performance indicators change (CPU load, RAM, traffic, etc.). The conducted studies showed the effectiveness of using various statistical methods of processing these data for assessing the state of the computer system [5-10].

In this paper, some of the criteria considered above were used as inputs to a developed heuristic scanner: the value of the BDS test in statics and dynamics [5, 6], Hurst index [7], Shewhart control charts, CUSUM, EMWA [8-9], Pareto, as well as qualitative metrics obtained through analysis of the PE file structure and actions performed by malicious software [10].

The conducted researches have shown that one of the perspective directions of heuristic analysis of computer viruses is the use of fuzzy logic [11,12].

The process of fuzzy inference is a procedure, or an algorithm for obtaining fuzzy inferences based on fuzzy conditions or prerequisites. This process combines all the basic concepts of fuzzy sets theory: membership functions, linguistic variables, fuzzy logical operations, methods of fuzzy implication and fuzzy composition.

In this paper, we propose a heuristic scanner for

an antivirus program based on the Mamdani fuzzy logic method. The structural scheme of the scanner is shown in Fig. 1.

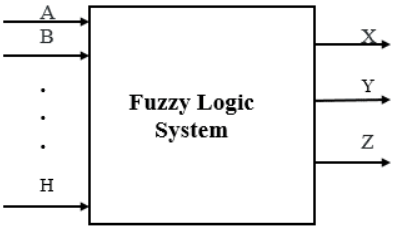

Figure 1 - Block diagram of the heuristic scanner

where A – BDS test value, B – value Shewhart control charts, C – the value of CUSUM cards, D – EWMA card value, E – Hurtst value, F – the value of the qualitative metric (QM), G – the BDS value of the virtual machine, H – the value of Pareto maps, X – activation of the subprogram "Treatment", Y – activation of the "Delete" subprogram, Z – activation of the "Quarantine" subprogram.

Input linguistic variables for the system of fuzzy inference by the Mamdani method were described by such a tuple:

$$< \alpha, T, X, G_r, M>, \qquad (1)$$

where: α - name of the linguistic variable (A, B, C, D, E, F, G, H), T - set of values (terms) of the input linguistic variable {"Dangerous", "Perhaps dangerous", "Uncertain," "Safe," "Somewhat secure"};

X – the set of calibrated values of the input variable.

G_r – the procedure of conditions' aggregation (new terms)

M – the function of forming a fuzzy set of values for each term of a given linguistic variable.

For each of the 8 input values, a membership area is defined. In Fig. 2 shows the range of belonging for the input variable A (BDS test value) in Fig. 3 - for variable B (Shewhart control charts).

The base of rules is formed on the basis of the prevailing conditions and conclusions, that is, input and output linguistic variables. Figure 4 shows a fragment of the rule base for a given heuristic analyzer based on the Mamdani fuzzy logic method. The rules database of this analyzer contains 517 different rules.

In the future, to solve the problem, the system of fuzzy inference makes a decision based on the weight coefficients of each of the rules subwords. It relies on the established rules base gives a certain

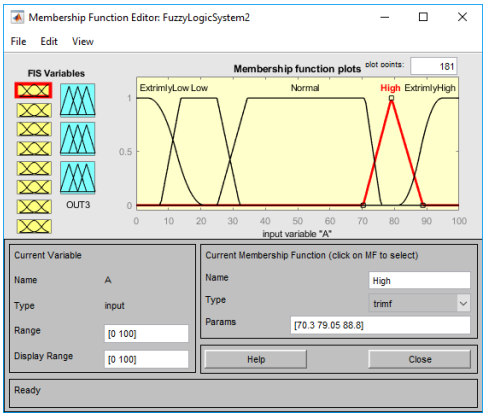

Figure 2. The graph of the membership function of the input linguistic variable A

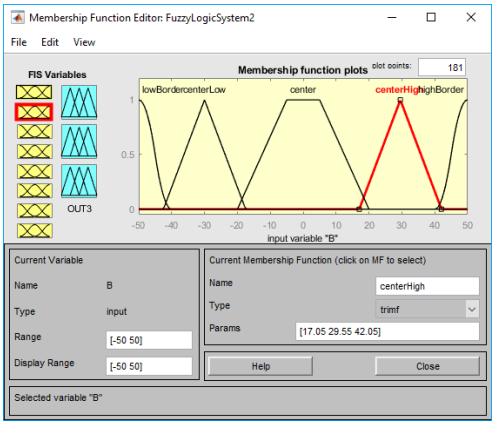

Figure 3. Graph of the membership function of the input linguistic variable B

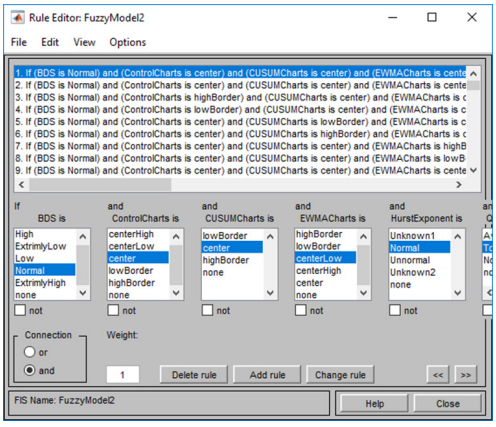

Figure 4. The Fragment of the base of rules of the fuzzy system inference

fuzzy conclusion. The subjective initial estimate goes through the process of defuzzification, that is, the process of transition from the membership function of the output linguistic variable to its clear numerical value [11]. The defuzzification process can be carried out by various methods, for example, by the center of gravity method:

$$\Delta = \frac{\int_{min}^{max} xMF(x)dx}{\int_{min}^{max} MF(x)dx} \qquad (2)$$

The results of modeling and visualization are shown in Fig. 5-7.

As can be seen from the results, the heuristic scanner of the anti-virus program is designed to build multi-level fuzzy product models, and the fuzzy output mechanism based on the Mamdani algorithm allows to obtain a numerical value of the risk of malware detection, a linguistic description of the level of risk, in the occurrence of such an event. The received information will allow to decide and develop measures to prevent the infection of the computer system.

3. Conclusion

1.The results of the research showed that when the computer system is influenced by virus software:

– the statistical indicators of the functioning of the system change the tendency of the Hurst index to an average value of 0.5 (which indicates the randomness of the process), the output of the system beyond the templates of the normal state of the system based on the value of BDS statistics, Shewhart control cards, CUSUM, EWMA, Pareto.

– Indirect signs of infection of the computer are fixed: removal of anti-virus programs and monitoring programs; disabling the Task Manager; Deletion of anti-virus database files; disable the boot of the computer in protected mode; closing programs that contain headers that indicate that they belong to the antivirus software; disabling User Account Control (Microsoft Windows security components and technologies); Disable Windows Firewall; ban editing of the registry; registration and launch of the service to block access to the sites of antivirus companies;

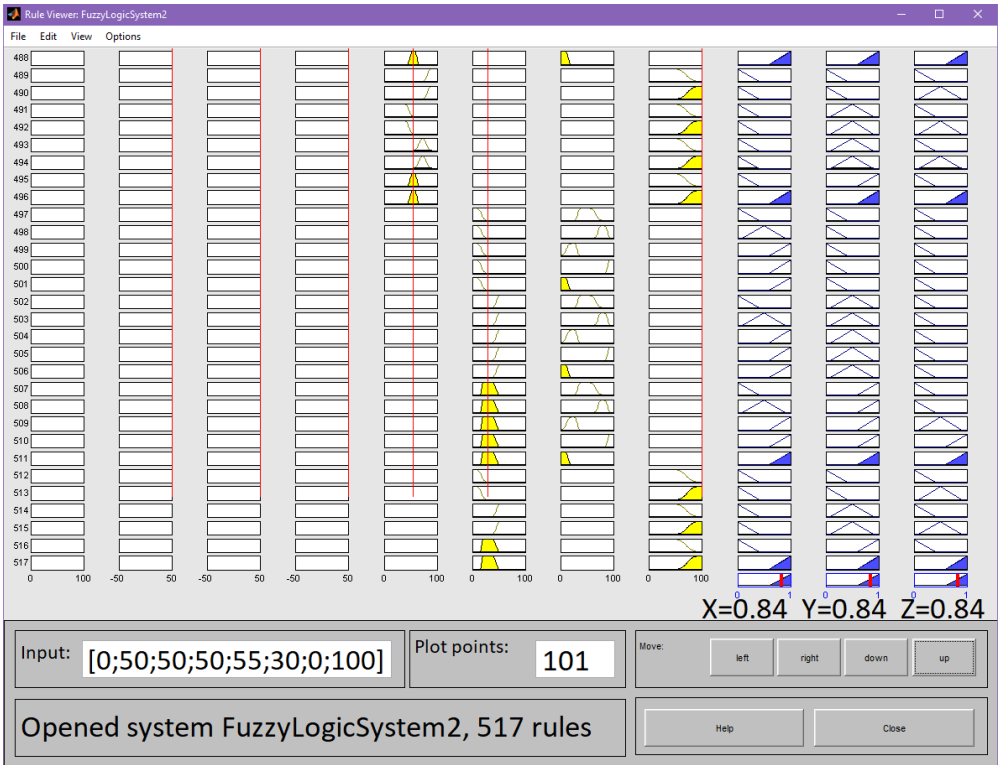

Figure 5 - The result of a system simulation for a malicious file

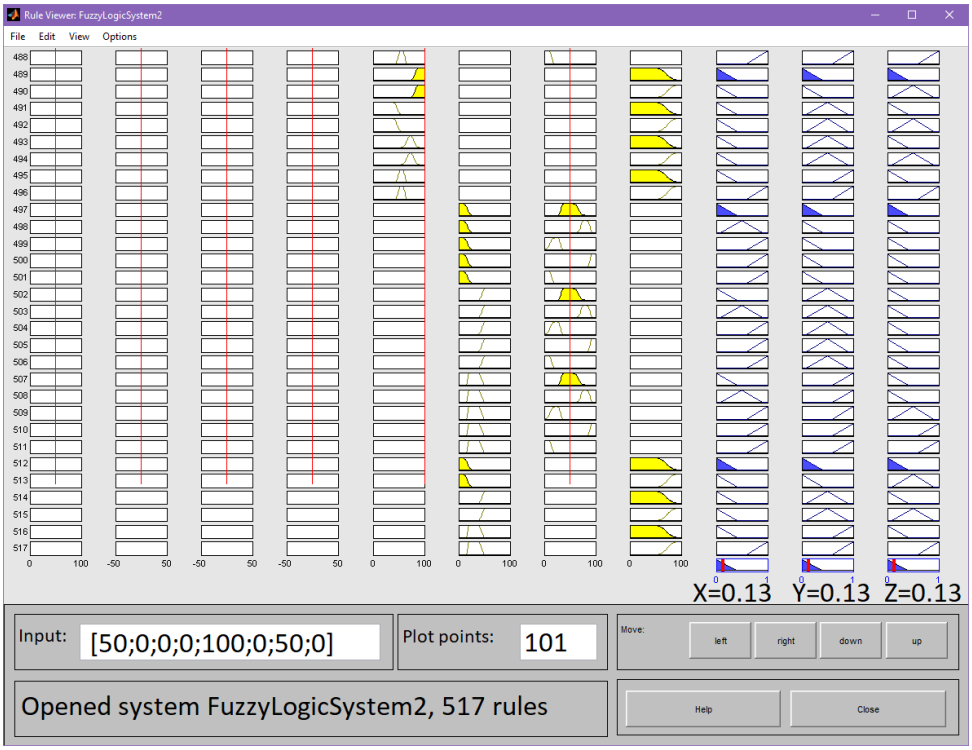

Figure 6 - The result of the system simulation for a secure file

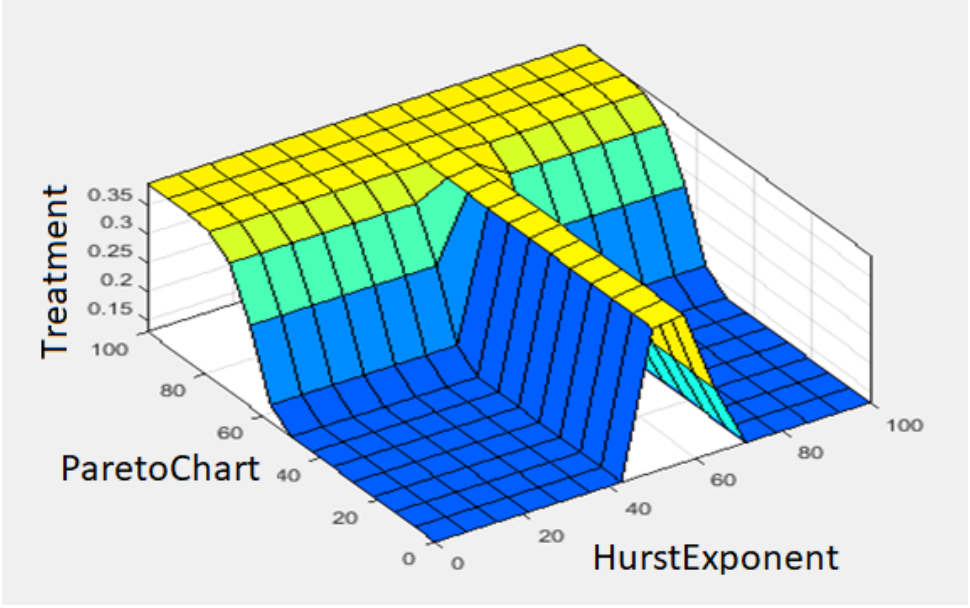

Figure 7. Visualization of the dependence between input variables: Hurst exponent and Pareto maps.

stopping the service and unloading processes related to antivirus programs and Firewall, etc.

2. A model of a heuristic scanner based on the Mamdani fuzzy logic method was introduced, with input data for it: the BDS value in static and dynamic, the Hurst index, Shewhart's control charts, CUSUM, EMWA, Pareto, and qualitative metrics obtained by analyzing the PE structure file and fixing the above listed indirect signs of infection of the computer. Assignment functions are assigned for each of the variables. A database of 517 rules based on input and output linguistic variables is formed. The developed system was tested.

2.The conducted experimental studies confirm the possibility of using the heuristic scanner of the anti-virus program based on the Mamdani fuzzy logic method for detecting virus attacks in the general malware detection system

References:

[1] **Semenov. S., Davydov V. and Gavrilenko S.** (2014), *Data Protection in Computer-Aided Control Systems,* "LAP LAMBERT ACADEMIC PUBLISHING", Germany, 236 p.

[2] **Shelukhin, O. Sakalema Zh. and Filinov A.** (2013) Intrusion Detection into Computer Networks, Hot line-Telecom, Moscow, 220 p.

[3] **Lukatsky, A.** (2001), Attack Detection, VHV-Petersburg, St. Petersburg, 624 p.

[4] **K. Kaspersky.** Play as Kaspersky Lab: the company opens access to its knowledge base about cyber threats in the framework of a new business service – [Electronic resource]. – Access mode: *https://www.kaspersky. ru/about/press-releases/2017_kompaniya-otkryvayet-dostup-k-svoyey-baze-znaniy-o-kiberugrozakh-v-ramkakh-novogo-biznes-servisa*

[5] **Semenov S., Gavrylenko S. and Chelak V.** (2016), Development of templates for the identification of the state of computer systems based on BDS-testing. *Herald of the National Technical University "KhPI".* Subject issue: Information Science and Modelling, Vol. 21, pp.118-125.

[6] **Semenov S., Gavrylenko S. and Chelak V.** (2016,) Developing parametrical criterion for registering abnormal behavior in computer and telecommunication systems on the basis of economic test. *Actual problems of economics.* Kiev, Vol 4(178), pp. 451-459.

[7] **Gavrylenko S., Chelak V, Bilogorskiy N.** (2017) Investigation of intrusion in computer systems based on the hurst exponent. *Advanced Information System. Quarterly scientific and technical journal.* Kharkiv: KhPI, Volume 1, No2, pp.58-61.

[8] **Gavrylenko S., Chelak V., O. Hornostal** (2016). Intrusion detection in computer systems. *Proceedings of the symposium "Metrology and metrology assurance"*. Sozopol, Bulgaria, pp. 342-347.

[9] **Gavrylenko S., Semenov S. and O. Hornostal** (2016). Development of adaptive patterns for detecting the abnormal behavior of a computer system. *Systems of information processing.* Kharkov, №. 3(140), pp.11-14..

[10] **Gavrylenko S., Saenko D.** (2017). Development of the method and program model of the static analyzer of harmful files. *Advanced Information System. Quarterly scientific and technical journal.* Kharkiv: KhPI, Volume 1, No1, pp.44-48.

[11] **Hurst H. E.,** (1951). Long-term Storage of Reservoirs. Transactions of the American Society of Civil Engineers, 116 p.

[12] **Gavrylenko S., Melnyk M.and Chelak** (2017). Development of a heuristic antivirus scanner based on the file's PE-structure analysis, *Information Technology and Computer Engineering. International scientific and technical journal.* Vinnitsa: VNTU, – №3 (40), pp. 23-29.

[13] **Zaichenko Yu. P.** (2008). Fuzzy models and methods in intelligent systems. K.: Slovo, 344 p.

Information about the Authors:

Gavrylenko Svetlana Yuryevna, Kharkiv Polytechnic Institute (1986), candidate of technical sciences (2001), associate professor (2005). Workplace – Professor of the Computer and Information Technologies Faculty of the National Technical University "KhPI". Scientific interests: safety in technical systems, digital automata. 61002, Ukraine, Kharkiv, Kirpicheva Str., 21, *gavrilenko08@gmail.com*

Chelak Viktor Vladimirovich, student of the Computer and Information Technologies Faculty of the National Technical University "KhPI". Scientific interests: protection of information in systems and networks. 61002, Ukraine, Kharkiv, Kirpicheva Str., 21, *victor.chelak@gmail.com.*

Hornostal Oleksii Andriiovich, student of the Computer and Information Technologies Faculty of the National Technical University "KhPI". Scientific interests: information protection, investigation of the efficiency of algorithms for solving applied problems. Ukraine, Kharkiv, Kirpicheva Str, 21, *agornostal@toidev.com.*