UDC 004.056.53
DOI: 10.15587/1729-4061.2021.225646

DEVISING A METHOD OF PROTECTION AGAINST ZERO-DAY ATTACKS BASED ON AN ANALYTICAL MODEL OF CHANGING THE STATE OF THE NETWORK SANDBOX

S. Buchyk
Doctor of Technical Sciences, Associate Professor*
E-mail: buchyk@knu.ua
O. Yudin
Doctor of Technical Sciences, Professor***
E-mail: yudin.ok8@gmail.com
R. Ziubina
PhD*
E-mail: ziubina@knu.ua
I. Bondarenko
Assistant***
E-mail: ab.imo.pectore@ukr.net
O. Suprun
Assistant
Department of Intelligent and Information Systems**
E-mail: oleh.o.suprun@gmail.com
*Department of Cyber Security and Information Protection**
**Taras Shevchenko National University of Kyiv
Volodymyrska str., 60, Kyiv, Ukraine, 01033
***Department of Information Systems and Technologies and
Protection of the State Interests in Information Sphere
National academy of the Security service of Ukraine
Mychaila Maksimovycha str., 22, Kyiv, Ukraine, 03022

This paper reports a method of protection against zero-day attacks using SandBox technology based on the developed analytical model with a probabilistic ranking of information system states. The model takes into consideration the conditions of a priori uncertainty regarding the parameters of the destructive flow on the system, accounting for the typical procedures of the network SandBox.

The proposed model of information system states makes it possible to analyze and track all possible states, as well as assess the level of security in these states, and the probability of transitions into them. Thus, it is possible to identify the most dangerous ones and track the activities that caused the corresponding changes. The fundamental difference between this model and standard approaches is the weight coefficients that characterize not the intensity of random events but the intensity of transitions between states.

Direct implementation and application of the proposed analytical model involved the technology of multilevel network "SandBoxes".

The difference from other popular anti-virus tools is the use of a priori mathematical threat assessment, which makes it possible to detect influences that are not considered threats by classical systems until the moment of harm to the system.

The combination with standard security tools makes it possible to separately analyze files that are too large in size, whether they enter the system not through a common gateway controlled by the network "SandBox" but from the external media of end-users.

The implementation of the developed analytical model has made it possible to improve the level of protection of the corporate network by 15 %, based on the number of detected threats. This difference is explained by the inability of classical software to detect new threats if they are not already listed in the database of the program, and their activity is not trivial

Keywords: zero-day attack, analytical model, state ranking, network SandBox, information protection

Received date 27.12.2020
Accepted date 05.02.2021
Published date 26.02.2021

1. Introduction

The current state of information society's development requires the protection of information resources and critical data at the highest level. Unprecedented requirements for software and hardware, technologies of modern infrastructure organization, etc. are characterized by a sharp increase in the demand of the state and public relations in the use of a system for providing reliable and high-quality IT services (information technology services) of different classes.

The increase in the number and level of complexity of various IT services, the increased technological level of cyber-attacks on state and public information resources form new tasks for the incident detection and prevention system.

The class of "zero-day" cyber-attacks is one of the most striking examples of the use of advanced information technologies in order to violate the properties of the information system and destroy relevant resources.

Antivirus SandBox is a solution for the protection of end devices, which makes it possible to prevent threats and attacks related to the use of critical data in integrated infrastructures, cloud environment, etc. under the condition of horizontal spread of the attack in information and communication networks [1].

The method of protection against zero-day attacks using SandBox technology in corporate networks is based on the methods of corporate (network) or antivirus SandBox. Antivirus SandBox makes it possible to emulate large files at end stations without loading the network shield and network SandBox.

Protection against targeted attacks of an indefinite class remains one of the most pressing issues in the field of information security. Over the past year, the number of cyberattacks in Ukraine increased tenfold. Almost everyone is familiar with the following expressions: "targeted attack", "zero-day vulnerability", "0-day", or even Advanced Persistent Threats (ATP). These topics can be safely called the main trend in the field of information security. Well-known encryption attacks are one of the subtypes of these threats. SandBoxes are the only means to combat the above threats [2].

Such means of protection conduct dynamic and statistical analysis of files in a virtual environment and block various attacks if necessary [3, 4].

That makes it possible to evaluate the behavior of suspicious files and the consequences of launching such files. At the same time, the main goal is not to detect malicious code using signatures but to evaluate the activities performed by the code, the security, and correctness in a given environment.

Now, the market for solutions for detecting and counteracting targeted attacks is only at the stage of formation. Manufacturers offer a wide range of protecting tools but, often, such products are made for marketing and do not reflect the real effectiveness of solutions. However, among the commercially-available protection products on the market, SandBoxes are one of the most effective solutions.

Zero-day attacks are a serious security threat to almost every organization. The traditional set of information security tools is not able to withstand undefined threat classes. SandBox technology is the most effective mechanism for detecting zero-day threats.

2. Literature review and problem statement

Ensuring the security of data in computer systems of various scales, and the proper level of functioning of such systems, in general, is one of the most important issues in the IT field. However, with the rapid growth of the industry, approaches and tools for obtaining unauthorized access are also changing; the main tools are shown and analyzed in work [5]. This requires continuous improvement of existing methods of protection and the creation of new approaches. One of the most striking examples of new methods is the creation of an antivirus system that mimics the immune system of living organisms. Such a system is developed and described in [6]; it makes it possible to detect non-trivial threats and address them but requires significant resources.

It should be noted that in the field of data protection and information systems in general, it is not enough just to respond to existing threats but it is necessary to prevent them. This is complicated by the fact that viruses often adapt to existing systems, so protection systems should detect new threats without human intervention. One of the first approaches to the creation of an adaptive system is demonstrated in [7]; however, the described approach is still not able to recognize specialized disguised attacks. Also, together with the rapid development of the information society and the increase in the amount of data that needs to be checked, the requirements for the hardware components of protection systems are increasing; the analysis of the main threats was carried out in [8]. One solution to the problem of lack of resources and time in ensuring security is to use artificial intelligence methods, such as neural networks, which were first proposed in the last century. Article [9] reports the developed method but points to such shortcomings as dependence on the quality of the network learning process and low efficiency using unprepared data.

At the same time, along with the increasing complexity of information systems, the concept of the intruder in this context also changes. That puts new requirements before existing systems. For example, the target of an attacker is increasingly not the destruction of the system or causing direct damage but access to personal or statistical data. This may be insignificant at first glance but can have significant consequences. Study [10] shows how the use of social media information can cause significant harm; a method of protection has been proposed. The disadvantage of that method is the slowness of its operation. Also relatively new is the concept of a "zero-day", but the problem described in [11] is one of the most relevant at the moment. The detection of such attacks requires the introduction of innovative methodologies, such as the trap system described in [12], or the technology of "honey pots". Paper [13] describes the operation of this method; it is quite effective in the presence of information about future attacks but irrelevant with complete uncertainty.

Also relatively new is the "SandBox" methods implemented in most well-known antivirus software, for example, Check Point for ESET and AvastSandbox for AVAST systems. This technology is a set of two key components: SandBlast and Threat Emulation – components that are a new kind of SandBox organization [14, 15] that make it possible to emulate probable attacks and predict system protection accordingly.

Attacks are detected at two levels of architecture: operating system levels (OS level) – as in traditional sandboxes, and at the CPU level [16]; SandBlast Threat Extraction is the component that makes it possible to analyze the files transmitted over the network, remove all dangerous content from them, reconstruct the files, and give these files to the user already clean.

For example, ESET Dynamic Threat Defense (EDTD) [17] provides another level of security, using ESET transitional technologies to detect new threats [14]. If the antivirus software recognizes malicious code, it prevents further threat activity, thereby keeping it in the quarantine zone. Other antivirus systems, such as Avast or NOD32, work on a similar principle.

It is because of the impossibility of introducing common standards and references on the large variety of existing information systems that it is necessary to create a general model that would adapt to the real state. This is achieved by implementing a mathematical threat detection basis. Based on the model, the protection method will be implemented.

3. The aim and objectives of the study

The purpose of this study is to create a method of protecting the information system from "zero-day" attacks, which will employ the probabilistic ranking of states of a given system under conditions of uncertainty.

To accomplish the aim, the following tasks have been set:

– to develop an analytical model of system states with the probabilistic ranking of transitions under conditions of *a priori* uncertainty to the parameters of the destructive influence flow, considering the dynamic changes in the system functionality over time;

– to develop a scheme for the implementation of the analytical model using a synthesized network and anti-virus SandBox;

– to test the proposed model implementation scheme using SandBoxes.

4. The study materials and methods

Taking into consideration the characteristic differences of "zero-day" attacks, as well as in order to synthesize a new method for detecting threats of this class against the background of a network SandBox, it is necessary to build a mathematical model of the method based on a multi-alternative approach to the number of possible types of attacks. This approach is characterized by the fact that the party that is the target of the attack does not have *a priori* data on the type, parameters, and time of the attack on the critical data of the owner of information resources.

In a given case, the process of determining the attack class should take place under conditions of *a priori* non-static uncertainty regarding the parameters and states of destructive influence on the information system, taking into consideration the typical procedures of the network SandBox. The analytical model should be formed in the context of the absence of preliminary probabilities about the type and state of the function of influence, as well as the *a priori* uncertainty of states regarding the system itself, against which destructive influences are directed.

Thus, the general approach to modeling threat on information resources [17] will be considered in the context of the built model of the analytical series. It should take into consideration discrete states and the continuous time of probabilistic ranking of input streams in order to calculate the necessary parameters and characteristics of the influence function (threat). In order to correct and simplify analytical representation, we shall use a functional series taking into consideration the dynamic sequence of random states (event flow) that occur in the system taking into consideration the vulnerabilities of information resources. Considering the peculiarities of the class of "zero-day" attacks, we shall choose an exponential distribution of the time of injection of the frequency of threats of the attack due to the vulnerabilities of the system [18].

Since a series of threats is simulated, in the formation of an analytical model it is advisable to take into consideration the sequence of vulnerabilities of the information system (or resources) used by the intruder. The difference in forming a model for "zero-day" attacks is that the correlation relationship between the threat and the corresponding vulnerability would not be taken into consideration in the analytical model of the series. These limitations are justified by the fact that the developed model would be further complicated by the set of relationships of critical data vulnerabilities to an undefined class of attacks. Thus, we established the difference from the typical criteria for describing threats – *a priori* unknown type and parameters of the attack, it is not known what vulnerability the threat would be directed at given the time dynamics. It should also be noted that *a priori* information system is considered to be protected at a certain level of guarantees and the risk of resource vulnerability is minimized.

We shall introduce a reasonable assumption that the incident is created by two classes of appropriate parameters of unauthorized influence on the information system. These parameters are as follows: in terms of the intensity of threats of different classes over time, as well as based on software implementation errors in identifying incidents and eliminating vulnerabilities. According to the defined approach, it is possible to form an analytical model of the system function for determining the procedure for identifying incidents of any complexity.

To directly implement the method of protection and implementation of the proposed model, the "SandBoxes" technology was used.

5. Analytical model of the information system states taking into consideration a probabilistic binary ranking of transitions

The state of the system, which is subject to unauthorized threats, is denoted through S_{ij}, where i and j are vulnerabilities of the i-th and j-th type. The flow of threats with unauthorized impact on the stationary state of the system enters the analytical model with the intensity of Q. We shall introduce a realistic assumption that the transitions between states in the analytical model are carried out instantly in time, which is typical of the "zero-day" attacks and modern data processing performance in information systems. Probabilistic binary ranking of the flow of unauthorized influences leads to the formation of changes in the state of the information system. That is, the dynamics of state change are taken into consideration depending on the intensity of the flow of influence per unit time. Given the possible transitions of the stationary state of the system, we shall introduce the binary ranking of probabilities P_{ij} of the system transition to each state S_{ij}. That is, P_{ij} is distributed between the states of the system S_{ij}, the event may occur at a random time when the system is in one of the possible states. Transitions between states in the analytical ranking model are carried out inert-free (instantly). The probabilistic ranking of the influence flow leads to the formation of a stream of events in the system and at the output. In this case, we shall introduce the ranking of the system model states, namely:

– S_{00} – the system is under a stationary mode of providing operational processes;

– S_{01} – a change in the stationary state with high intensity of external influence of the "zero-day" attack parameters on the vulnerability of the system over time (state – failure of the protection system based on an artificially created intense influence over time);

– S_{10} – a change in the stationary state with the external impact of "zero-day" attack on software vulnerability (state – failure of the protection system based on software vulnerability);

– S_{11} – a change in the stationary state with the external intense impact of a "zero-day" attack on the system and software vulnerability (state – failure of the protection system based on the mixed influence of two classes) [19].

When constructing an analytical model, the probability P_{ij} of the system entering any state in the initial model of probabilistic ranking is interpreted as a quantitative indicator of the correspondence of the system's stay in the corresponding state [20]. In this case, the set of states is considered discrete, and time is continuous.

State transitions have the weight coefficients g of the intensity of response to impact flows and state transitions in the system. The fundamental difference between this model and standard approaches is that the weight coefficients are characterized not by the intensity of random events in the system but by the intensity of transitions between states. That is, how significant the intensity factor g is for the transfer of the system to another state. The specified weight coefficients are determined by the level of ability of the software or security operator to respond to the flow of unauthorized influence in real time and the system's ability to restore processes. In order to ensure the correctness of this transformation, we shall take into consideration the weight coefficients in the construction of a model of the probabilistic binary ranking of input and output flows in the system.

Based on the description of the probabilistic binary ranking of system states and input and output flows, the intensity of the real threat of an attack in the system can be represented by an analytical series in the form:

$$Q = \sum_{S_i \in S_{(R+1)}} P_{S_i Q_{S_i s_R}}, \tag{1}$$

where $S_{(R)}$ is a set of system states characterized by the stationary nature of processes and the absence of a real threat of a "zero-day" attack in it.

The system can enter each state at probability $P_{S(R+1)}$, where $S_{(R+1)}$ is the state of the system that is at risk of a real attack. The transition to the state $S_{(R+1)}$ from $S_{(R)}$ in the system is carried out at intensity $P_{S(R+1),S(R)}$. For the binary ranking of the probability of transitions, the formula for determining the intensity of the state transition flow will be defined as:

$$Q_d = P_{10}Q_2 + P_{01}Q_1, \tag{2}$$

$$P_{S_{R+1}} = 1 - P_{0d},$$

where P_{0d} is the probability of the system's ability to remain stationary to ensure established operating processes in relation to the intensity of the input flow of class d of unauthorized influences.

Under a stationary mode of system functioning, taking into consideration the impact of the real threat at intensity $Q_{S(R+1),S(R)}$, the state of the system without a loss (the input flow does not change the state of the system) is a flow of events Q_d. Thus, it becomes possible to calculate the intensity of the elimination of real threats of attacks against the background of the introduction of the following weights:

$$q_d = \frac{Q_d}{1 - P_{0d}}. \tag{3}$$

For the simplest binary ranking of the probability of transitions, the formula for the weight q_a takes the form:

$$q_d = \frac{P_{10}Q_2 + P_{01}Q_1}{P_{11}}.$$

The probability of the information system being ready for safe (in relation to the threat of an attack) operation can be determined as follows:

$$P_{0d} = P_{00} + P_{10} + P_{01} = \frac{q_1 q_2 + Q\mu_2 + Q_2 q_1}{(Q_1 + q_1)(Q_2 + q_2)}. \tag{4}$$

It is clear that the number of possible random states of the system must be finite and can be determined according to the ordinal numbers. Such a random process is called a process with discrete states [14].

The adequacy of the presented analytical model is ensured by meeting the following restrictions related to the considered task of modeling the threat of a "zero-day" attack, namely:

– the model of system states that are discrete, with continuous time, is correct in general if each state of the system in case of a random process of unauthorized influences generates all N input flows of events at intensity Q_i, i=1,..., N;

– in a general case, to simulate the states of the system for threats of "zero-day" attacks one should use calculations based on the series model with an infinite number of discrete states and continuous time.

Within the framework of the above restrictions, such a model of states and transitions between them can be used to assess the reliability of information systems security tools and identify threats. The proposed model is characterized by the possibility of simultaneous occurrence in the system of two or more unauthorized external influences. The model of the probabilistic ranking of system states was used as the basis for the method of detecting "zero-day" attacks taking into consideration the intensity of the flows of influence [21, 22].

Since the modeling uses a sufficient number of possible sets of attacks or impact flows, it is possible to reasonably apply the normal law for the distribution of random events of influences on the system [23]. Taking into consideration the "zero-day" attacks' characteristics, it is necessary to account for the probability of several unauthorized events in the system at the same time. That is the simultaneous impact of multiple events on the vulnerabilities of the same type at a fixed time interval. For analytical modeling of such states, we shall introduce the load factor of the system depending on the intensity of the flow of influences. Using the load factor z=Q/g, it is possible to determine the required probability of simultaneous occurrence in the system of n events, taking into consideration the normal (Gauss law) distribution law $P_n(z)$. For each type of threat, taking into consideration the specified requirements for the accuracy of modeling, by calculating the probability $P_n(z)$ values, the number *max* is determined. The number of threats implemented due to vulnerabilities of the corresponding type is also determined. All states $S_{i>\max(ij)}$ and transitions between them are excluded from the model of a series of states of the random process system (zero probability is assigned to form an artificially finite sequence), as a result of which the desired final model with the ability to predict the probabilities of attacks is obtained.

Thus, the model makes it possible to track all possible states of the system S_{ij}, which are counted according to the corresponding limitation, and to assess the corresponding probabilities of being in these states $P_{S(R+1)}$. Accordingly, with the help of Q, the intensity of real threat in the system (1), it is possible to estimate P_{0d} – the probability of the system's readiness for safe operation (4). Using state emulation by applying a network "SandBox", it is possible to identify the most dangerous system states and factors that caused the transition to these states, and, accordingly, to track possible threats.

The proposed analytical model of system states (1) to (4) taking into consideration the peculiarities of "zero-day" attacks makes it possible to objectively assess the basic parameters, probability, and characteristics of the threat. To this end, we use statistics on the occurrence and elimination of vulnerabilities in the automated mode [24, 25].

6. Development of a scheme for the implementation of an analytical model using a synthesized network and antivirus SandBox

The method of protection against zero-day attacks is based on a combination of the developed model of the binary ranking of information system states and the methods of network and antivirus SandBoxes. The hardware solution, which includes a network SandBox, is installed along the perimeter of the network and acts as a gateway. The received traffic is decrypted, suspicious files are sent for analysis to the cloud. During the analysis, the file is converted to PDF and sent to the user with the option of obtaining the original.

Software is installed on end devices; it includes the antivirus SandBox technology based on full virtualization in the cloud service. A given solution makes it possible to prevent threats and attacks related to connecting to external networks. For example, connecting to Wi-Fi and cloud applications that cannot be decrypted by a gateway on the perimeter, connecting external media to a PC, and spreading a network attack horizontally. Antivirus SandBox makes it possible to emulate large files without creating a high load on the gateway (Fig. 1).

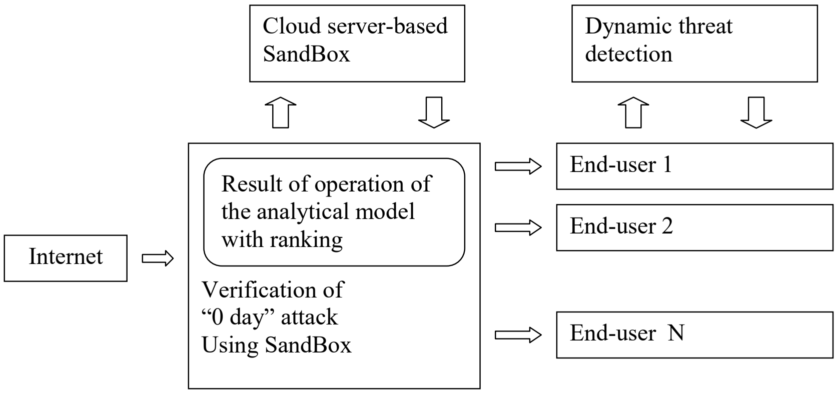
Cloud server-based SandBox
Dynamic threat detection
Internet
Result of operation of the analytical model with ranking
Verification of "0 day" attack Using SandBox
End-user 1
End-user 2
End-user N

Fig. 1. Implementation of the zero-day protection method using network and antivirus SandBox technologies in corporate networks

The network SandBox received small files for analysis. Files larger than 50 megabytes will be ignored for emulation in the antivirus SandBox. A given solution will reduce the load and speed up the emulation time.

Antivirus software will analyze suspicious files obtained from portable media and cloud applications that are not decrypted. This approach provides for the highest protection against malware (software) entering the network and endpoints [26, 27].

Thus, in general, the implementation scheme can be formally described as follows (Fig. 1):

1. A file comes from an external source to the system where it is decrypted and enters an external gateway represented by the network SandBox. Files larger than 50 megae bytes will not be ignored emulated in the antivirus SandBox.
2. File analysis is carried out using the analytical model with ranking. If a suspicious file is found, it is sent for further analysis to the cloud. A file converted to PDF format is sent to the user.
3. Verified files are sent to end-users.
4. Files larger than 50 MB and obtained from external sources, such as external drives and networks, are checked on end-user computers. To this end, use the algorithm of dynamic threat detection.

7. Testing the proposed model implementation scheme using SandBoxes

The proposed method and model were tested at a commercial enterprise with 500 end devices. Two antivirus software gateways assembled into the cluster to ensure fault tolerance were installed on the perimeter. A classic antivirus program is installed on each server and end device. Testing the built information security system based on the proposed method was carried out using a synthetic test and a real-traffic test.

The objectives of a given test were the assessment of the effectiveness and feasibility of using SandBoxes as part of an integrated information security system; the evaluation of the effectiveness of the built model of protection against zero-day attacks.

SandBox testing was carried out in two stages: testing using synthetic samples of malicious code; testing on real internet traffic of users [28].

Synthetic tests were carried out using temporary virtual machines. When delivering viruses to the test zone, methods that make it difficult to detect by traditional signature means of protection were used:

– file archiving using RAR, ZIP, 7-ZIP formats;

– mail messages with web links to a malicious file, including using URL shortening;

– encryption of malicious code (paya load) of a macro in Microsoft Word documents through macros.

It should be noted that all solutions were tested in a real network (in an isolated network environment). Therefore, before arriving at the analysis in the SandBox, files with malicious code were analyzed and blocked by available means of protection, using signature and reputation mechanisms. A given testing algorithm was employed to assess the effectiveness of existing means of protection as well.

As part of the tests, the main channels for obtaining malware were analyzed: files downloaded from web resources; E-mail attachments, files on external storage media.

As part of testing on the real Internet traffic, the network SandBox was set to a TAP mode and received a copy of internet traffic for analysis. Network traffic acquisition lasted 1 month.

The main channels for obtaining viruses from external networks were controlled: e-mail; interaction with web services on the Internet; interaction with cloud applications; interaction with external media. For synthetic testing of the built system, 55 malware samples were selected that were included in the signature database for 2018. The signature analysis method was disabled on the network and antivirus SandBox. Thus, the built system received malware without the possibility of verification in the database [29].

Of the 55 instances used in testing web traffic analysis, 32 files with malicious code, not detected by existing anti-virus protection tools, were analyzed in the SandBox. When testing on mail traffic, traditional protection means found only 1 out of 15 malware – 14 out of 15 files were sent for analysis in the SandBox. In the synthetic test, 55 different malicious files were used for web traffic. 23 out of 55 instances were blocked by existing security features (secure Internet access gateway). 32 out of 55 malicious files were sent to the SandBox for testing.

The results from a synthetic "network SandBox" test for web traffic, a synthetic test for postal traffic, as well as the malware, detected using SandBoxes, are given in Tables 1–3.

Table 1

Synthetic "network SandBox" test results for web traffic

	The best result of the built information security system	Existing Information Security Systems (WatchGuard)
Detected malware	29/32	0/32

Table 2

Synthetic test results for mail traffic

	The best result of the built information security system	Existing Information Security Systems (WatchGuard)
Detected malware	9/14	0/14

Table 3

Malware detected using SandBoxes

Type of malware	Detected (quantity)
Trojan	34
Worm	2
Backdoor	5
Trojan.Downloader	18
Ransomware	2
Spyware	2
Riskware/Adware	7
Using a vulnerability in the Web browser Web.Exploit	1
Use a vulnerability in a Web browser Mal/FakeAV-SE	1
Attempts to communicate with an external botnet management server (callbacks)	25
Total (disregarding callbacks)	72

In total, in one month, the system of protection against targeted attacks recorded 72 threats that are not blocked and are not recorded by existing signature means of protection.

The number of detected threats by the built information security system, in comparison with the existing one, over a month of monitoring at an enterprise is also shown in Fig. 2.

Along with the classic antivirus system and the use of the proposed method, two more approaches were also tested, namely Sensory Traps [12] and the use of Honey Pots [13].

Certain malware detected as part of the testing is not detected by some signature antiviruses even a few months after the end of testing.

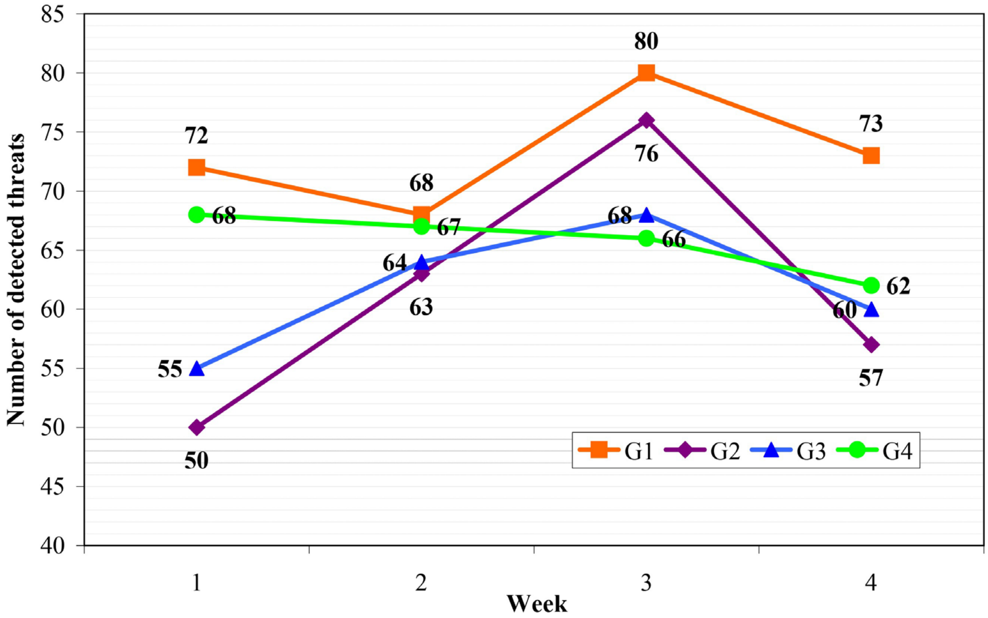

Fig. 2. The number of detected threats by the built by information security system compared to the existing one, over a month of monitoring at an enterprise. Charts: G1 — the developed model of protection; G2 — classic antivirus system; G3 — Sensory Traps [12]; G4 — Honey Pots [13]

8. Discussion of results of testing the developed method of protection and the corresponding model

The proposed method of detecting threats using "network SandBoxes" and the constructed mathematical model make it possible not only to detect the threat after its activity but to prevent penetration. The effectiveness of the method has been proven during testing; the results are given in Tables 1–3. Such significant efficiency can be explained by *a priori* use of a mathematical model that tracks changes in the state of the system, and not only collects statistics post-factum, like most antivirus systems.

The system of adaptive analysis of states can be compared to artificial intelligence algorithms, for example, neural networks [9, 31]. However, neural networks require a long learning process on reliable examples, which makes it impossible to identify new types of threats. Classic detection methods are not able to track an attack in the first stages and detect a threat only after a mass infection of the system.

Comparing the proposed method with the methods of Sensory Traps and Honey Pots, the following conclusions can be drawn: both above methods lose their relevance over time since they provide protection against well-known attacks. Therefore, such approaches are relevant if one knows in advance about potential weaknesses in the system or data that are most interesting for the attacker.

In addition, the proposed method is quite flexible due to the simplicity of the mathematical model, which allows it to be adapted to each specific task and system. This is both an advantage and a disadvantage since it requires the attention of an experienced specialist who will be able to properly adjust the system coefficients. But such flexibility will make it possible to identify events that have different interpretations in terms of danger to different systems.

As one of the drawbacks, we should note the probability of false triggering of the system, that is, the detection and recognition of secure messages as an attack. Such a case is possible, for example, with the constant mass mailing of files to all computers on the network.

Despite the significant advantage of the method, illustrated in Tables 1–3, it is currently in the early stages of development. The analytical model needs to be improved, for example, using weight coefficients and the ability to change parameters directly in the course of algorithm operation. The possibility of implementing existing cyber defense algorithms should also be investigated.

9. Conclusions

1. The creation of an analytical model of the states of the information system makes it possible to assess the impact of various factors in real time, and track possible destructive actions without endangering the system itself. Owing to this, the proposed model is universal but, at the same time, allows for fairly in-depth analysis and tracing destructive external influences. The introduction of the probabilistic binary ranking of transitions between system states makes it possible to automatically track potentially dangerous changes in the system. This is not possible manually due to the large number of requests coming in at each point in time. To identify and assess a potential threat in any state, the weight coefficients are used, which are characterized not by the intensity of random events in the system but by the intensity of transitions between states. The analytical series was also used taking into consideration the dynamic sequence of random states to ensure the correctness of the analytical representation.

Such a system makes it possible to detect dangerous activity even before harming the system, which is different from most classical methods of protection that already operate post factum, that is, after the attack. The introduction of weight coefficients makes it possible to individually configure the parameters of threat detection, assessing the probabilities of the transition of the system to dangerous states. That is, an expert could indicate what intensity of requests should be considered dangerous.

2. The scheme of implementation of the developed model with the help of network "SandBoxes" and integration with existing methods and programs of anti-virus protection was proposed and tested, which makes it possible to utilize the strengths of various approaches. As an example, an antivirus system was used, which gave significant positive results. The application of methods of a multilevel network SandBox makes it possible to more objectively assess the situation in the information system and detect hidden activity. In addition, the use of an external antivirus system also makes it possible to check large files obtained from external media and reduce the load on the network SandBox.

3. Based on the test results, the proposed combination of the network SandBox method based on the developed model and anti-virus systems shows its effectiveness in comparison with classical methods. Some of the attacks were not diagnosed by classical systems for months. At the same time, it is important to remember the possibility of erroneous detection, which requires further work on the algorithm. The constructed method of protection against zero-day attacks using SandBox technology was used in combination with the methods of a multilevel network SandBox and an antivirus SandBox based on full virtualization. The multilevel network SandBox is based on the analytical model of the system, taking into consideration the ranking of states. This approach has made it possible to improve the protection of the corporate network against unspecified

References

1. Moussouris, K., Siegel, M. (2015). The Wolves of Vuln Street: The 1st System Dynamics Model of the 0day Market. RSA Conference 2015. San Francisco. Available at: https://ic3-2017.mit.edu/sites/default/files/documents/MichaelSiegelKatieMoussouris_VulnMarketsRSAC2015Speaker.pdf
2. Schwartz, A., Knake, R. (2016). Government's Role in Vulnerability Disclosure: Creating a Permanent and Accountable Vulnerability Equities Process. Discussion Paper 2016-04. Harvard Kennedy School. Available at: https://www.belfercenter.org/sites/default/files/files/publication/Vulnerability%20Disclosure%20Web-Final4.pdf

3. Yudin, O., Ziubina, R., Buchyk, S., Bohuslavska, O., Teliushchenko, V. (2019). Speaker's Voice Recognition Methods in High-Level Interference Conditions. 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON). doi: https://doi.org/10.1109/ukrcon.2019.8879937
4. Gurzhiy, P., Gorodetsky, B., Yudin, O., Ryabukha, Y. (2019). The Method of Adaptive Counteraction to Viral Attacks, Taking Into Account Their Masking in Infocommunication Systems. 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT). doi: https://doi.org/10.1109/aiact.2019.8847893
5. Edwards, J. (2001). Next-generation viruses present new challenges. Computer, 34 (5), 16–18. doi: https://doi.org/10.1109/2.920606
6. Hedberg, S. (1996). Combating computer viruses: IBM's new computer immune system. IEEE Parallel & Distributed Technology: Systems & Applications, 4 (2), 9–11. doi: https://doi.org/10.1109/88.494599
7. Zhao, F., Li, Q., Jin, L. (2006). An Intrusion-Tolerant Intrusion Detection Method Based on Real-Time Sequence Analysis. 2006 International Conference on Machine Learning and Cybernetics. doi: https://doi.org/10.1109/icmlc.2006.258927
8. Jensen, M. (2013). Challenges of Privacy Protection in Big Data Analytics. 2013 IEEE International Congress on Big Data. doi: https://doi.org/10.1109/bigdata.congress.2013.39
9. Tesauro, G. J., Kephart, J. O., Sorkin, G. B. (1996). Neural networks for computer virus recognition. IEEE Expert, 11 (4), 5–6. doi: https://doi.org/10.1109/64.511768
10. Bonneau, J., Anderson, J., Danezis, G. (2009). Prying Data out of a Social Network. 2009 International Conference on Advances in Social Network Analysis and Mining. doi: https://doi.org/10.1109/asonam.2009.45
11. Azzedin, F., Suwad, H., Alyafeai, Z. (2017). Countermeasureing Zero Day Attacks: Asset-Based Approach. 2017 International Conference on High Performance Computing & Simulation (HPCS). doi: https://doi.org/10.1109/hpcs.2017.129
12. Popereshnyak, S., Suprun, O., Suprun, O., Wieckowski, T. (2018). Intrusion detection method based on the sensory traps system. 2018 XIV-Th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH). doi: https://doi.org/10.1109/memstech.2018.8365716
13. Tian, Z.-H., Fang, B.-X., Yun, X.-C. (2003). An architecture for intrusion detection using honey pot. Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693). doi: https://doi.org/10.1109/icmlc.2003.1259851
14. Yudin, O., Boiko, Y., Ziubina, R., Buchyk, S., Tverdokhleb, V., Beresina, S. (2019). Data Compression Based on Coding Methods With a Controlled Level of Quality Loss. 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT). doi: https://doi.org/10.1109/atit49449.2019.9030431
15. How to Choose your Next Sandboxing Solution. Featuring insight from gartner's market guide for network Sandboxing (2016). Check Point Software Technologies Ltd. Available at: https://www.checkpoint.com/downloads/products/check-point-gartner-how-to-choose-sandboxing-solution-whitepaper.pdf
16. Burnap, P., French, R., Turner, F., Jones, K. (2018). Malware classification using self organising feature maps and machine activity data. Computers & Security, 73, 399–410. doi: https://doi.org/10.1016/j.cose.2017.11.016
17. ESET Dynamic Threat Defense. Available at: https://www.eset.com/int/business/dynamic-threat-defense/
18. Lakhno, V., Kasatkin, D., Kozlovskyi, V., Petrovska, S., Boiko, Y., Kravchuk, P., Lishchynovska, N. (2019). A model and algorithm for detecting spyware in medical information systems. International Journal of Mechanical Engineering and Technology, 10 (1), 287–295.
19. The Problem with Traditional Sandboxing. Available at: https://blog.checkpoint.com/2015/09/14/the-problem-with-traditional-sandboxing/
20. Villalba, L. J. G., Orozco, A. L. S., Vidal, J. M. (2015). Malware Detection System by Payload Analysis of Network Traffic. IEEE Latin America Transactions, 13 (3), 850–855. doi: https://doi.org/10.1109/tla.2015.7069114
21. Yudin, O., Ziubina, R., Buchyk, S., Matviichuk-Yudina, O., Suprun, O., Ivannikova, V. (2020). Development of methods for identification of informationcontrolling signals of unmanned aircraft complex operator. Eastern-European Journal of Enterprise Technologies, 2 (9 (104)), 56–64. doi: https://doi.org/10.15587/1729-4061.2020.195510
22. Yudin, O., Symonychenko, Y., Symonychenko, A. (2019). The Method of Detection of Hidden Information in a Digital Image Using Steganographic Methods of Analysis. 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT). doi: https://doi.org/10.1109/atit49449.2019.9030479
23. D'Hoinne, J., Orans, L. (2015). Market Guide for Network Sandboxing. Gartner. Available at: https://www.gartner.com/en/documents/2995621
24. Cooke, E., Jahanian, F., McPherson, D. (2005). The zombie roundup: Understanding, detecting, and disrupting botnets. SRUTI '05: Steps to Reducing Unwanted Traffic on the Internet Workshop, 39–44.
25. Koller, D., Friedman, N. (2009). Probabilistic Graphical Models. Principles and Techniques. MIT Press.
26. National Vulnerability Database. Statistics. NIST. Available at: https://nvd.nist.gov/vuln/search?adv_search=true&cves=on&pub_date_start_month=0&pub_date_start_year=2010&pub_date_end_month=9&pub_date_end_year=2016&cvss_version=3
27. CVSS Severity Distribution Over Time. NIST. Available at: https://nvd.nist.gov/general/visualizations/vulnerability-visualizations/cvss-severity-distribution-over-time
28. Ablon, L., Libicki, M. C., Abler, A. M. (2017). Markets for Cybercrime Tools and Stolen Data: Hackers' Bazaar. RAND Corporation. Available at: https://www.rand.org/pubs/research_reports/RR610.html
29. Allodi, L., Massacci, F. (2014). Comparing Vulnerability Severity and Exploits Using Case-Control Studies. ACM Transactions on Information and System Security, 17 (1), 1–20. doi: https://doi.org/10.1145/2630069
30. Chandrasekaran, M., Baig, M., Upadhyaya, S. (2006). AVARE: Aggregated Vulnerability Assessment and Response against Zero-day Exploits. 2006 IEEE International Performance Computing and Communications Conference. doi: https://doi.org/10.1109/.2006.1629458