Information & Security

I. Stetsenko & M. Demydenko
vol. 47, no. 2 (2020): 221-231
https://doi.org/10.11610/isij.4715

Published since 1998

ISSN 0861-5160 (print), ISSN 1314-2119 (online)

Research Article

Signature-based Intrusion Detection Hardware-Software Complex

Inna V. Stetsenko (✉), *Maksym Demydenko*

Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine, https://kpi.ua/en

ABSTRACT:

Nowadays hackers are able to find many software vulnerabilities, which can be exploited for malicious purposes such as to destroy the operating system, to steal users' private data, to demand a ransom not to affect the data and retain their validity. The majority of attacks use an Internet connection; therefore, the efforts should be directed to the way in which data packets are transmitted. The hardware-software complex, which is the main subject of the presented research, serves as a firewall for the devices on one subnetwork with access to the Internet, simultaneously analysing and filtering both downstream and upstream traffic at packet level, resolving scumware and securing the perimeter of each device in the subnet. The concept and the architecture of the developed hardware-software complex are described. The implemented complex will not allow malicious traffic to pass through, providing protection of all endpoint devices in a subnetwork. The experimental results of malware detected are presented, and the security related metrics are evaluated.

ARTICLE INFO:

RECEIVED: 08 JUNE 2020
REVISED: 08 SEP 2020
ONLINE: 22 SEP 2020

KEYWORDS:

cybersecurity, vulnerability, web attack, SQL injection

Creative Commons BY-NC 4.0

Introduction

Nowadays the problem of cybersecurity is highly critical. Bad rabbit, Petya, Not Petya have been targeting many important data sectors worldwide. Many devices, which got affected by those viruses, were simply blocked and, in most cases, it turns out that it may be very difficult to go through backing up process after the sustained damage. Nowadays intruders are able to find many security vulnerabilities in software implementation both application and system

✉ Tel.: + 380688170742; E-mail: stiv.inna@gmail.com

software. Using security holes in software they destabilize the operating system, steal user's private data, demand a ransom for remaining the data unaffected and retaining their validity.

It is noteworthy that the main source of security vulnerability is Internet. Well-known web attacks are Cross-site request forgery (CSFR), Cross-site scripting (XSS), Denial of service (DoS), SQL injection. The actual list of HTTP/HTTPS threats are described in source.[1] Almost ten percent of URLs are infected by malicious malware according to the statistical report.[2] Simulating the malware propagation, it becomes clear how fast this process and how significant the impact of factors.[3]

Web attacks are reasonably divided onto attacks, being targeted server software, and attacks, being targeted web app. As a result of data gathering and bug-track analysis, it is revealed that the majority of security vulnerabilities are concentrated in web apps. According to the statistical investigation represented in source, 4 in 5 vulnerabilities were located in application code, and 1 in 5 vulnerabilities were of high severity in 2019.[4]

The development of the Internet of Things (IoT) aimed for the increase of human comfort by automation of daily routines. However, it also caused the increase of cyber threat, which the user is subjected to. Since the IoT device undertakes data collection, processing, dissemination and storage in real-time, the personal information, personal photo- or video content can be stealing as a result of web attack. Moreover, smart devices as a rule are connected to the same local network as a personal computer. That is why they can be used to get an access to the local network by malicious person.

According to the Symantec's Internet Security Threat Report, 57533 IoT attacks were registered in a global honeypot in 2018, and routers were the most infected devices accounted 75 percent of attacks (Internet Security Threat Report).

That is why the main goal of the research is the development of hardware-software complex serving as a firewall for the devices of one subnetwork with the access to the Internet, simultaneously analysing and filtering as downstream as upstream traffic at packet level, resolving scumware and securing the perimeter of each device in the subnet. The system developed will not let 'bad answer,' or scumware, in and 'sensitive' data out.

Section 1 introduce to the web attacks research. The second section describes related works. The next section describes the object under investigation. Section 4 describes the developed hardware-software complex. Experimental results are given in the next section. The last section summaries the article and gives the perspective of future research.

Related Works

An intrusion detection system (IDS) is a software tool being able to monitor and identify malicious activity directed to compromise information system security. A detailed survey of such systems is represented in work.[5] The author defined the following classes of detection: anomaly detection, signature detection, and

compound detectors. To detect an abnormal network traffic wavelet analysis can be used as it is described in work.[6] In contrast to anomaly detection, signature detection is able to identify intrusion when normal behaviour of system is undefined. Basing on the knowledge of the intrusive process, designers of the security system should determine patterns and clues which reveal unwilling events.

The malicious attacks that can be identified by means of deep inspection of packets, arriving at the computer system, are discussed in work.[7] Network, host, software, physical and human attacks have been described from the point of detection by monitoring network traffic. However, the main problem of anomaly-based detection system is a lack of real-life training datasets as the authors of work highlighted. Because of the time-consuming process of classification of unknown attack, a relevant dataset appears when many attempts to attack have been successful. The comparison of available datasets and discussion about their features is represented in work.[8]

Signature-based detection retains in modern IDS as a high accuracy method. It allows identifying known attacks. Nowadays special sources, as source,[9] gather information about signatures marked as malicious.

The adaptive anomaly-based IDS named AMODS was presented in work.[10] As the authors write, this system 'bridge the gap between adaptive detection and web attack detection'.

Both signature and anomaly detection are defenceless against different evasion techniques such as fragmentation, flooding, obfuscation, encryption, described in book.[11]

In addition, hybrid IDS were developed as a combination of anomaly and signature techniques and discussed in work.[12] Other hybrid systems combined host-based IDS and network-based IDS described in work.[13] All hybrids aim to reduce false positive rate and to rise the detection of unknown attacks.

Thus, researches constantly search for effective methods to identify cyberattacks before they start their malicious activity inside operating system (or personal computer). In this research, the complex of hardware and software is developed for the purpose of intrusion detection using signature detection and man in the middle (MITM) attack. This type of attack changes the original connection between a user and web application on the new connection providing by a perpetrator as described in source.[14] Instead of directing packet from application to the user it will be redirected to another computer resource, where it can be modified and after that, the message will be delivered to the user. The message passing in opposite direction performs in the same way. An ordinary MITM attack is used for malicious purpose however in this development it will be used for the defence purpose.

The Object under Consideration

We consider a subnet of independent devices connected with a crosspoint switch (router in this research). Devices generate requests which are processed by switch. The scheme of interaction is depicted in Fig. 1.

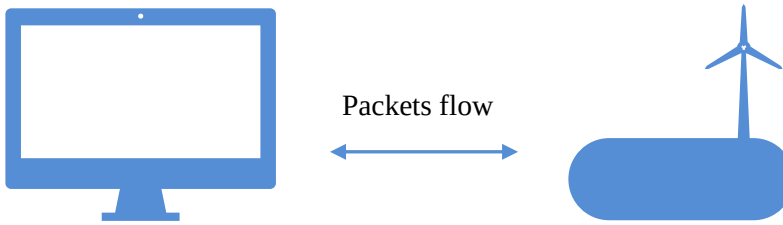

Figure 1: The interaction between the router and a particular device.

When a request is generated by any device the packet arrives to router. The router, in turn, transmits a request to the World Wide Web and waits for an answer. After getting the answer, the router sends it to the requesting device. The device gathers set of packets, interprets it into a computer-friendly format and after that displays it. The process of interpretation starts when the last packet of the answer has arrived.

If an attack was launched, certain packets were intentionally falsified or substituted in the transfer process. In an ideal situation firewall of the operating system would recognize the attack and would block it. If the firewall does not identify the threat, which is included in the server answer, the device will be infected. After that only antivirus software is available to identify the attack.

The intrusion Detection Hardware-Software Complex

The Concept of the Development

The solution is the developed hardware-software complex called Extendable Malware Detection System (EMDS). The term hardware is used because an algorithm has been implemented as a specific device, which should be connected to the subnet of multiple devices. The user is able to activate protection for one or all devices.

The key idea is to use man-in-the-middle attack for checking traffic. Because the man-in-the-middle is not seen as an attacker, it cannot be evaded.

During the protection activation, the attack man-in-the-middle occurs on the level of traffic routing. The whole traffic passes through connected device that means the interaction of a particular device can be represented by the scheme depicted in Fig. 2.

Now all devices, connected to the router, recognize EMDS as the router. Simultaneously the router perceives EMDS as a final device. In this case, EMDS pursue the policy of assuming responsibility to send upstream packets from the devices to the router and downstream packets from the router to the device. Thus, while traffic passes through EMDS each of the packets is being analysed. For the complex analysis, to store the packets and their content in the database is needed.

EMDS is implemented on a single-board computer Raspberry PI 4 with installed Ubuntu Core operating system on it. The following utilities provide the

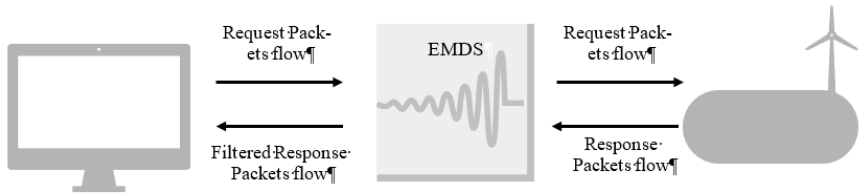

Figure 2: The interaction between the router and a particular device provided by EMDS.

proper system functioning: Nmap 7.6 used for scanning devices in the subnetwork, TShark 3.2 used for capturing live network data packets, arpspoof used for MITM attack performing, and MongoDB 3.2 used to store packets data.

Intrusion detection process

The analysis of the packets consists of three phases. In the first phase, the checking of signatures of already known attacks is performed. Evidently, the detection accuracy strongly depends on the set of using signatures that should be updated periodically to include new known web attacks. During this phase, EMDS detects intrusion in real time. It functions as an intrusion detection system, intrusion prevention system and network security monitoring system simultaneously. Network characteristics are analysed in real time and the alarm is generated when a strange event, which is potentially threatful, is detected. Such characteristics as traffic volume, bandwidth usage, and protocol usage are under control. Network traffic analysis is based on extensive rules and signature language. To identify complex threats Lua scripting language is used. Specific features of this language greatly increase the ability to detect malware scripts.[15] Implementation of the complicated rule of signature detection in EMDS software provides in case of a threat do not skip it to the end device.

The second phase is the applying of behavioural analysis. This method is based on an analysis of a sequence of different requests. Network behaviour analysis is able to identify normal everyday traffic. In other words, it recognizes disturbance if the traffic has exceeded the threshold determined in advance. By this type of analysis, a distributed denial-of-service (DDoS) attack can be recognized, which is the most well-known network security violations. This kind of attacks is a serious security threat, which can be harmful to Internet providers and large network infrastructures.

Take for instance a sequence of requests represented in Table 1. There is a SQL-injection attempt using manipulating the characteristics, adding the quotation mark and calling the ‘sleep’ function. Separately considered, these signs do not point to an obvious attack vector however, considering a set, it can be clearly suggested that a criminal person tries to probe the web application. The mathematical model summarizes signs of an attacker's behaviour over a period of time and the decision to block an attack, based on the results, can be made.

Actually, it allows not to miss the start of an attack without blocking normal requests.

Table 1. Attempts of SQL-injections.

Number	Request
1	`/api/products?userId=1`
2	`/api/products?userId=3-2`
3	`/api/products?userId=-1`
4	`/api/products?userId=1'`
5	`/api/products?userId=‘1`
6	`/api/products?userId=1 and sleep(5)`

The third phase is the heuristic scanning. This method aimed to catch unknown attacks (or zero-day attacks). Using rules and algorithms, it provides searching suspicious samples of code. Generic detection, which is the newest modification of heuristic analysis, provides smaller false positive ratio.[16]

The architecture of the complex is depicted in Fig.3. The devices EMDS use share signatures database and each of them includes a traffic database. The direction of package passing can be both from the user, as first device EDMS1 represented, and to the user, as the second device EDMS2 represented.

The current version of EMDS realizes the first and second phases however the third phase is provided in the hardware-software complex construction. Notice that according to the MetaDefender Cloud statistics, multi-scanning provides the greater chance to detect malware (Miao Y., 2015). Therefore, the developed complex exploits several techniques to detect malicious activity.

Discussion

Malicious software is always trying to cover their tracks and stay invisible as long as possible to accomplish their goals. To hide the intrusion, the malware tries to manipulate information about registered packages. To prevent this, the blockchain will be used in future as a database to guarantee the integrity and immunity of the data stored. This approach to store EMDS data could help avoid unauthorized modification of stored data retroactively.

The implementation of the developed hardware-software complex provides a solution of the following cases.

Case 1. Blocking a detected attack

A directed attack is being conducted on one of the network devices. In the stream of packets passed through EMDS, one of the three phases of intrusion detection revealed a threat. The packet, in which the attack has been detected, and the following response packets are blocked. Thus, the answer will not be fully interpreted on the end device and the attack will not be performed.

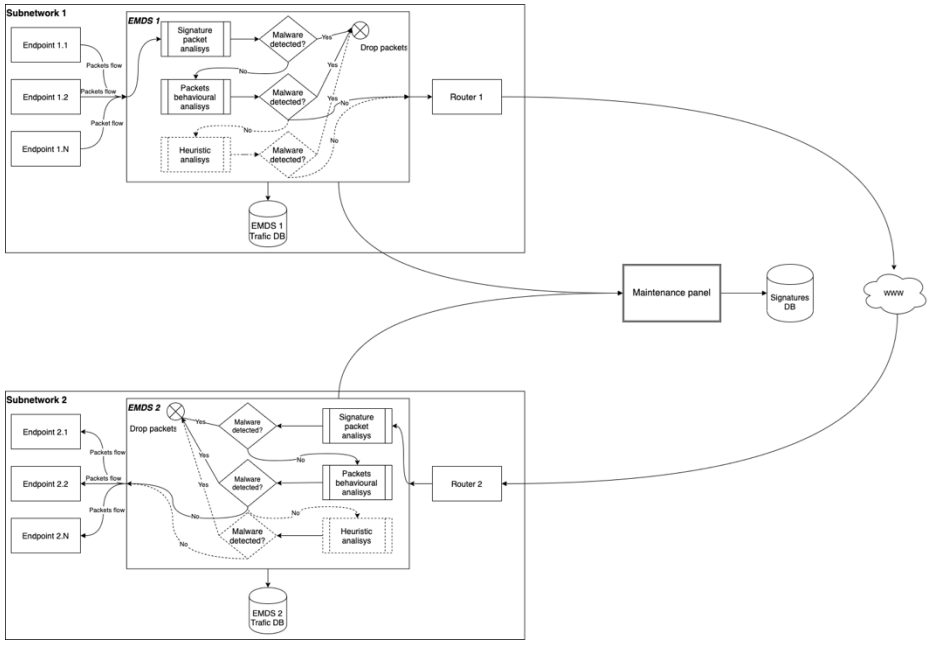

Figure 3: The architecture of Extendable Malware Detection System (EMDS).

Case 2. Blocking an undetected attack

A directed attack is conducted on one of the network devices. In the stream of packets, passed through the EMDS, none of three phases has been revealed the threat. The last packet of the response will attempt to interpret the packets into a clear answer and the attack will be replayed on the EMDS. Thus, the result of the verification service will not be performed and EMDS will not send the response to the end device.

Case 3. An attempt, performed by IoT devices, to 'merge' data onto the Internet

Outgoing packets from devices are filtered using established rules and will not skip files of certain sizes determined by the administrator of the subnetwork. It will help prevent personal data leakage such as audio, video, photos.

Case 4. Malware attempt to hide penetration during an attack

To hide the traces of an attack, malware can attempt to manipulate existing data on previous packages in the EMDS database. Using blockchain in EMDS data storage could help prevents frauds with data on packages.

Experimental results

Intrusion detection system performance is evaluated by metrics which can be divided into performance-related metrics and security-related metrics.[17] The performance-related metrics quantify non-functional properties of IDS such as resource-consumption and capacity. At the same time, the security-related metrics measure how accurate the detection of malware is. Therefore, this set of values is preferably used in IDS researches.

The security-related metrics are grounded on the matrix of false/true and negative/positive values, which corresponds to the results of detection according to the real situations. If an attack has been detected while it has not been in real life, that is the false positive case (FP). If it has not been detected in the same situation, it is the true positive case (TP). If an attack has been detected when it has been in real situation, that is true negative case (TN). If it has not been detected in the same situation, it is the false negative case (FN). Counting the number of cases occurred in series of experiment, the following ratio can be calculated: false positive ratio (FPR), true positive ratio (TPR), false negative ratio (FNR), classification ratio (CR). Corresponding formulas are (Khraisat A., 2019):

$$FPR = \frac{FP}{FP+TN}\ , TPR = \frac{TP}{TP+FN}\ , FNR = \frac{FN}{FN+TP}\ , CR = \frac{TP+TN}{TP+TN+FP+FN}\ . \quad (1)$$

An experiment has been conducted using 350 activities on web resources, 75 of which contained malicious traffic. Activities included visiting sites, video and audio traffic, loading files on the end device and sending files to remote servers. The test set contained DoS, DDoS, SQL-injection, XSS attacks. Total traffic bandwidth was more than 3.5 Gb, 23 percent of which was upstream traffic. The EMDS revealed 84 malicious activities, among them 26 were detected as a malicious wrong. 17 out of 75 malicious activities were not detected. The values of calculated ratios are represented in Table 2.

Table 2. Experimental results of EMDS intrusion detection research.

IDS metrics	Value
False positive ratio	9.5 percent
True positive ratio	77.3 percent
False negative ratio	23 percent
Classification ratio	87.7 percent

The value of false positive ratio measures the part of false security alarms and the value of true positive ratio measures the part of true detection of benign traffic. Obviously, the first should be as small as possible, and the second - as large as possible. However, the result of experiment strongly depends on a test set, which is used.

The growth of the number of detected malicious activity compared with actual number of malicious activities is represented in Fig. 4. The difference between detected and actual malicious activities is the number of false alarms.

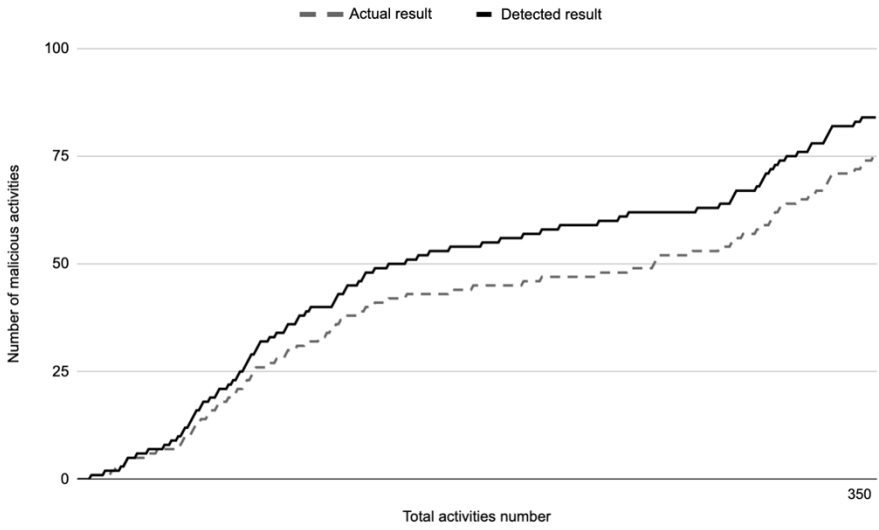

Figure 4: The growth of detected and actual malicious activity during the experiment.

Conclusions

The intrusion detection hardware-software complex Extendable Malware Detection System, that is able to isolate the subnet devices from identified threats, is developed. The novelty of the approach proposed is the usage of the man in the middle attack to catch and testify packages, hiding from malicious person a device that performs detection. Several phases of detection are provided to increase the percent of web attack detected and decrease a false positive number of alarms. The security-related metrics, evaluated experimentally, show the false positive ratio 9.5 percent and the true positive ratio 77.3 percent.

The complex can be improved by using blockchain technology to store data on packages. The most successful consensus mechanism of blockchain can be found by simulation proposed in work.[18] Furthermore, the heuristic analysis implementation will be investigated in the nearest future.

Acknowledgements

This work is funded by the NATO SPS Project CyRADARS (Cyber Rapid Analysis for Defense Awareness of Real-time Situation), Project SPS G5286.

References

1 "FortiWeb 5.7.1 Administration Guide: Solutions for specific web attacks," Fortinet, accessed June 5, 2020, https://help.fortinet.com/fweb/571/Content/FortiWeb/fortiweb-admin/solutions_for_specific.htm.

2 "Internet Security Threat Report. Vol. 24," *Symantec*, accessed June 4, 2020, https://docs.broadcom.com/doc/istr-24-2019-en.

3 Inna V. Stetsenko and Vitalii Lytvynov, "Computer Virus Propagation Petri-Object Simulation," In A. Palagin, A. Anisimov, A. Morozov, S. Shkarlet (eds.) *Advances in Intelligent Systems and Computing 1019* (Cham: Springer, 2020), 103-112, https://doi.org/10.1007/978-3-030-25741-5_11.

4 "Web Applications vulnerabilities and threats: statistics for 2019," *Positive Technologies*, accessed June 1, 2020, https://www.ptsecurity.com/ww-en/analytics/web-vulnerabilities-2020/.

5 Stefan Axelsson, "Intrusion Detection Systems: A Survey and Taxonomy," accessed September 8, 2020, https://www.semanticscholar.org/paper/Intrusion-Detection-Systems%3A-A-Survey-and-Taxonomy-Axelsson/550aec01bf61ff9fd271debc394a8c3dfa59657b.

6 Nicolai Stoianov, Vitalii Lytvynov, Igor Skiter, and Svitlana Lytvyn, "Traffic Abnormalities Identification Based on the Stationary Parameters Estimation and Wavelet Function Detailization," In *A. Palagin, A. Anisimov, A. Morozov, S. Shkarlet (eds.) Advances in Intelligent Systems and Computing 1019* (Cham: Springer, 2020), 83-95, https://doi.org/10.1007/978-3-030-25741-5_9.

7 Hanan Hindy, Elike Hodo, Ethan Bayne, Amar Seeam, Robert Atkinson, and Xavier Bellekens, "A Taxonomy of Malicious Traffic for Intrusion Detection Systems," *2018 IEEE International Conference on Cyber Situational Awareness, Data Analytics and Assessment (Cyber SA)*, Glasgow, UK., 2018, pp. 1-4, https://doi.org/10.1109/CyberSA.2018.8551386.

8 Ansam Khraisat, Iqbal Gondal, Peter Vamplew, and Joarder Kamruzzaman "Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges," *Cybersecurity* 2, (2019), https://doi.org/10.1186/s42400-019-0038-7.

9 "Attack Signatures," *Broadcom*, accessed June 1, 2020, https://www.broadcom.com/support/security-center/attacksignatures.

10 Ying Dong, Yuqing Zhang, Hua Ma, Qianru Wu, Qixu Liu, Kai Wang, and Wenjie Wang, "An Adaptive System for Detecting Malicious Queries in Web Attacks," *Science China Information Sciences* 61 (2018), https://doi.org/10.1007/s11432-017-9288-4.

11 "Evading IDS, Firewalls and Detecting Honeypots," In *Ethical hacking and countermeasures: Secure Network Operating Systems and Infrastructures,* Second edition, (Boston, USA: Cengage Learning, 2016), 77-134.

12 David J. Day, Denys A. Flores, and Harjinder Singh Lallie, "CONDOR: A Hybrid IDS to Offer Improved Intrusion Detection," *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2012)*, Liverpool, UK, 2012, pp. 931-936. https://doi.org/10.1109/TrustCom.2012.110.

[13] Megha Gupta, "Hybrid Intrusion Detection System: Technology and Development," *International Journal of Computer Applications* 115, no. 9 (2015): 5-8, https://doi.org/10.5120/20177-2384.

[14] "Man in the Middle (MITM) Attack," *Imperva*, accessed September 8, 2020, https://www.imperva.com/learn/application-security/man-in-the-middle-attack-mitm/.

[15] Bryant Smith, "Advanced Malware Detection with Suricata Lua Scripting," *Trustwave*, 2017, accessed September 8, 2020, https://www.trustwave.com/en-us/resources/blogs/spiderlabs-blog/advanced-malware-detection-with-suricata-lua-scripting/.

[16] Yiyi Miao, "Understanding Heuristic-based Scanning vs. Sandboxing," *Opswat*, 2015, accessed September, 8, 2020, https://www.opswat.com/blog/understanding-heuristic-based-scanning-vs-sandboxing.

[17] Aleksandr Milenkoski, Varco Vieira, Samuel Kounev, Alberto Avritzer, and Bryan D. Payne, "Evaluating Computer Intrusion Detection Systems: A Survey of Common Practices," *ACM Computing Surveys* 48, no. 1 (2015): 1-41, https://doi.org/10.1145/2808691.

[18] Ivan Burmaka, Nikolai Stoianov, Vitalii Lytvynov, Mariia Dorosh, and Svitlana Lytvyn, "Proof of Stake for Blockchain Based Distributed Intrusion Detecting System," In A. Palagin, A. Anisimov, A. Morozov, S. Shkarlet (eds.) *Advances in Intelligent Systems and Computing 1265* (Cham: Springer, 2021), 237-247, https://doi.org/10.1007/978-3-030-58124-4_23.

About the Authors

Inna V. **Stetsenko** , Dr. Sc., Professor, Department of Computer-Aided Management and Data Processing Systems, Igor Sikorsky Kyiv Polytechnic Institute. Research interests include simulation software, Petri nets, parallel computing, cyber-attack modelling and simulation.

Maksym **Demydenko** is PhD student in Software Engineering, Department of Computer-Aided Management and Data Processing Systems, Igor Sikorsky Kyiv Polytechnic Institute. Research interests include computer and network security, software security, distributed computing.