UDC 004.8

Simon C.K., Sochenkov I.V.

Peoples' Friendship University of Russia, Moscow, Russia

EVALUATING HOST-BASED INTRUSION DETECTION ON THE ADFA-WD AND ADFA-WD: SAA DATASETS*

Abstract

With the growth of the internet and the development of new technologies also originates advancements in methods of cyber-attacks such as zero-day and stealth attacks, a more effective method of network safety is essential for network stability for both personal use and businesses. This research paper will assess anomalous patterns of Normal Pattern and Abnormal Pattern comprised of system calls based on the Dynamic-Link Library. The two datasets assessed are designed on the Windows Operating System on a Host-based Intrusion Detection System; comprised of the Australian Defence force Windows Dataset (ADFA-WD) and Australian Defence Force Academy Windows Dataset: Stealth Attacks Addendum (ADFA-WD:SAA). The development of a binary feature space is developed based on the common vulnerabilities and exposures at the time of the creation of the dataset. The data mining techniques implemented are Support Vector Machine classifier with sigmoid and RBF kernels is compared to the Random Forest classifier.

Keywords

Host-based Intrusion Detection; machine learning; random forest; SVM, RBF; Sigmoid kernel.

Симон К.К., Соченков И.В.

Российский университет дружбы народов, г. Москва, Россия

ОЦЕНКА ХОСТОВОЙ СИСТЕМЫ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ В НАБОРАХ ДАННЫХ ADFA-WD И ADFA-WD: SAA

Аннотация

В связи с повышением значения Интернета и развитием новых цифровых технологий в современном мире происходит улучшение методов кибератак таких, в частности, как атаки нулевого дня и стелс атаки. Данные факторы обусловливают необходимость в разработке более эффективных методов сетевой безопасности для обеспечения стабильной работы в сети как для личного использования, так и для бизнеса. В данной исследовательской работе будут оцениваться аномальные паттерны, проявляющиеся в работе Нормального шаблона и Аномального шаблона, состоящие из системных вызовов на базе динамически подключаемой библиотеки. Анализируемыми критериямив данной статье выступают критерии скорости, точности и возможных ошибок. Два рассматриваемых набора данных разработаны в операционной системе Windows и предназначены для системы обнаружения вторжений на базе ОС Windows ADFA-WD и ADFA-WD: SAA. В статье обсуждается развитие бинарного пространства на основе общих уязвимостей и воздействий на момент создания набора данных. Используемые методы интеллектуального анализа данных включают в себя классификацию по методу опорных векторов, который сравнивается с классификацией по методу случайного леса.

Ключевые слова

Хостовая система обнаружения вторжений; машинное обучение; метод случайного леса; метод опорных векторов, радиальная базисная функция; сигмоидное ядро.

[*] **Труды II Международной научной конференции «Конвергентные когнитивно-информационные технологии» (Convergent'2017), Москва, 24-26 ноября, 2017**

Proceedings of the II International scientific conference "Convergent cognitive information technologies" (Convergent'2017), Moscow, Russia, November 24-26, 2017

Introduction

Currently, the Web is actively developing in its use, speed and the amount of that can be stored on it. In regard to the growth of the network, the importance of network security increases, since effective information protection is becoming one of the main tasks, both for business entities and individuals. With increased network protection, we reduce the risk of threats to data protection, in particular [1]:

1) Violations of the confidentiality: In spite of the fact that currently, there is a "removal of the corporate veil" in regard to responsibilities of companies, which includes the disclosure of information to shareholders and transparency of certain data that must be published in open sources. There are such information that should be inaccessible to competitors (commercial secret) and to some employees (state secret), and personal data as a whole.

2) Data Manipulation: Even in a brief moment of a intrusion in a network, data can be manipulated, the victim or company issues that could be insuperable for the Information System Staff to return to its original state. Documents that were manipulated due to a hacker who attacked the system can cause mass corruption in data which can cause an uproar in the inner working of the business be it immediately or years from now.

3) Data destruction: Data is a priceless commodity for normal users and companies alike hence why the importance of backup technology has been so widely used. What happens when this important data is destroyed by a malicious act be it financial data, contracts, raw data, company secrets and the like. Destruction of data can severely cripples the victim or company involved.

Based on threats mentioned 1) Violations of the confidentiality, 2) Data Manipulation, 3) Data destruction and the fact that Windows havs the highest market share, it is safe to state that Windows is the most dominant Operating System (OS) on the market at present making Windows OS an optimal OS to do a synopsis on vulnerability to cyberattacks. There is a need to create additional tools to ensure network security. Intrusion detection system (IDS) is traditionally used in one of three forms: 1) Host-based Intrusion Detection System (HIDS), Network based Intrusion Detection (NIDS), and a hybrid system that is a combination of HIDS and NIDS. In this research paper, the system calls based on Windows Dynamic-Link Layer (DLL) to investigate in regard of HIDS. In the present research work, the ability for the system to detect violations of rules established by the IDS will be analysed due to patterns of system attributes to normal system actions (Normal Pattern) and vulnerable attacks (Abnormal Pattern) in regard to [2]:

1) Australian Defence force Windows Dataset (ADFA-WD),
2) Australian Defence Force Academy Windows Dataset: Stealth Attacks Addendum (ADFA-WD:SAA).

ADFA-WD and ADFA-WD:SAA are both datasets that are based on Windows OS HIDS they represent a new milestone and standard in HIDS in regard to targeting zero-day and stealth attacks on Windows OS.

The goal of this research is to solve the following problems:

1) To identify the accuracy that the HIDS can achieve with the help various machine learning algorithms.
2) The measurement of accuracy that is used is False Negative Rate (FNR), False Alarm Rate (FAR), Detection Rate (DR), False Positive Rate (FPR) [2-6].
3) Getting the highest DR possible while maintaining the lowest FAR possible.
4) Acquiring a lowest possible processing time for each algorithm.

2 Related Researches

2.1 Development of the datasets from Australian Defence Academy

In ADDA-WD, The 12 "zero-day" and stealth attacks vulnerabilities used in respect to the dataset are CVE: 2006-2961, CVE: 2004-1561, CVE: 2009-3843, CVE: 2008-4250, CVE: 2010-2729, CVE: 2011-4453, CVE: 2012-0003, CVE: 2010-2883, CVE: 2010-0806, EDB-ID: 18367, a virus based attack and Background usage (Normal). These attacks are used because of the trends identified at the time against threats on Windows [7]. The focus of these attacks is given to the TCP port, web applications, browsers and malicious applications.

ADDA-WD:SAA contains four stealth attack theories: 1) Doppelganger , 2) Chimera Attack , 3) Chameleon Attack (Network) 4) Chameleon Attack (Malware). All three stealth attack theories provide full interactivity with the target attacks, and was based on replacing generic, non-stealth shellcode in an existing exploit skeleton with the various stealth. The focus of these attacks is based on TCP port and on two targeted server programs were Icecast V2.0 and CesarFTPV0.99g.

The Dataset was collected on the Windows XP SP2 host. It had configured as FTP server, web server, the Hotspot, wireless network or Ethernet network. An array of compounds and protocols is the standard working network, which can become a victim of a cyber- attack.

The purpose of the designed dataset is to provide a contemporary look at modern IDS, when compared with earlier methods used in the IDS such as KDD99 [10-12] which are now being used less, even despite the fact that they are effective.

The idea of creating a standard for Windows IDS was due to the lack of credible modern methods of intrusion detection and availability of a dataset for OS Windows.

The choice of audit data: analyzes an array of DLL system calls, as these system calls can reflect the state in which the HIDS is currently in. The system calls which are used DLL – Kernel32, ntdll, user32, comctl32, ws2_32, mswsock, Msvcrt, msvcpp,ntoskrnl.

2.2 Types of machine learning algorithms, dedicated to the works of Creech, Borisanya and Patel, V. Hadera for Windows Australian Defense Academy

The fundamental work and the design of the datasets were done in the dissertation of G. Creech. The ADFA-WD and ADFA-WD:SAA who brought to science a new understanding of the IDS on Windows OS [13]. In that paper he considered algorithms such as a hidden Markov model (HMM), Extreme Learning Machine (ELM) , support vector machine (SVM).

The joint study by Borisanya and Patel [14], also devoted to ADFA-WD, considered such algorithms as an algorithm Naïve Bayes algorithm sequential minimal optimization (SMO), LIBSVM, algorithm instant training (IBK), as well as algorithms, KMeans, ZeroR, ONeR, JRIP, J48.

In a joint paper by Haider J. Creech, G. and J. Xu Hu [2], which is dedicated to algorithms for data ADFA-WD, the report focuses on algorithms such as SVM, K Nearest Neighbor (KNN) method, the method of Artificial Neural Network (ANN), and a method of extreme learning machine algorithm Naïve Bayes.

3 Methodology

KDD99 [15] is one of the classic Linux datasets on IDS [2], which has attack types that have become obsolete in terms of the approach to the attack type and do not represent the modern day approaches used [16]. Most of today's work force and personal computers run Windows OS, which leads to the need for modern IDS dataset for Windows, such as ADFA-WD (Table 1). The available dataset is in “.ghc” format.

Diagram 1. *Data Process:*

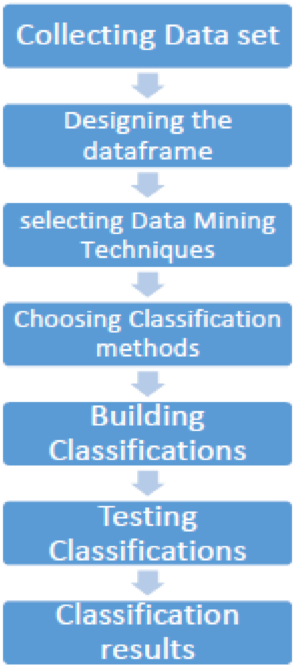


Table 1-3 *consists of Windows OS vulnerabilities used in the ADFA-WD:*

Table 1. *ADFA-WD Attack Data*

VID	Vulnerability	Program	Trace Count
V1	CVE: 2006-2961	CeasarFTp 0.99g	454
V2	EDB-ID: 18367	XAMPP Lite v1.7.3	470
V3	CVE: 2004-1561	Icecastv2.0	382
V4	CVE: 2009-3843	Tomcast v6.0.20	418
V5	CVE: 2008-4250	OS SMB	355

V6	CVE: 2010-2729	OS Print Spool	454
V7	CVE: 2011-4453	pMWiki v2.2.30	430
V8	CVE: 2012-0003	Wireless Karma	487
V9	CVE: 2010-2883	Adobe Reader 9.3.0	440
V10		Backdoor executable	536
V11	CVE: 2010-0806	IE v 6.0.2900.2180	495
V12		Infectious Media	621

Table 2. *ADFA-WD Validation Data*

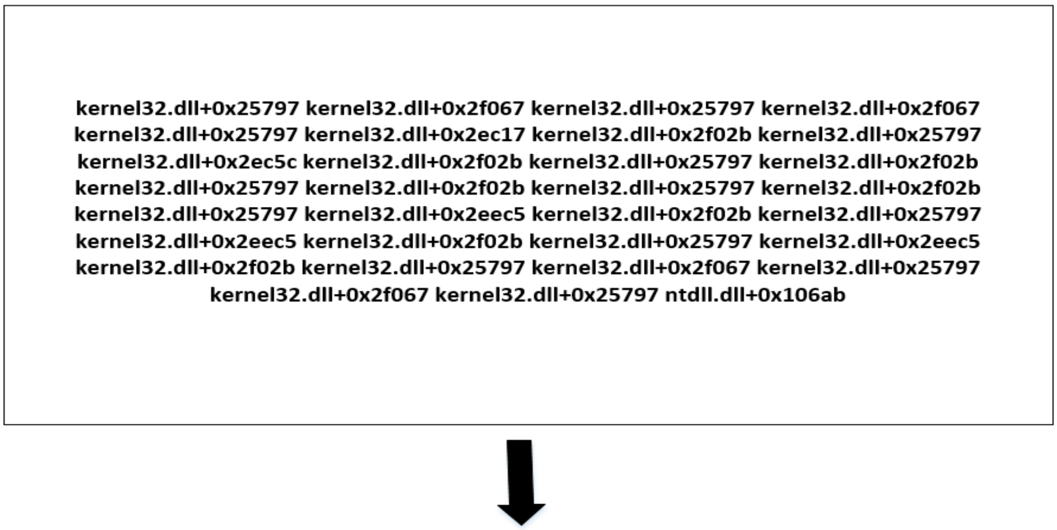
VID	**Vulnerability**	**Program**	**Trace Count**
V1	CVE: 2006-2961	CeasarFTp 0.99g	17
V2	EDB-ID: 18367	XAMPP Lite v1.7.3	105
V3	CVE: 2004-1561	Icecastv2.0	24
V4	CVE: 2009-3843	Tomcast v6.0.20	51
V5	CVE: 2008-4250	OS SMB	17
V6	CVE: 2010-2729	OS Print Spool	115
V7	CVE: 2011-4453	pMWiki v2.2.30	18
V8	CVE: 2012-0003	Wireless Karma	320
V9	CVE: 2010-2883	Adobe Reader 9.3.0	103
V10		Backdoor executable	127
V11	CVE: 2010-0806	IE v 6.0.2900.2180	242
V12		Infectious Media	610
V13	Normal	Background	17

Table 3. *ADFA-WD Training Data*

VID	**Vulnerability**	**Program**	**Trace Count**
V1	CVE: 2006-2961	CeasarFTp 0.99g	22
V2	EDB-ID: 18367	XAMPP Lite v1.7.3	23
V3	CVE: 2004-1561	Icecastv2.0	19
V4	CVE: 2009-3843	Tomcast v6.0.20	20
V5	CVE: 2008-4250	OS SMB	12
V6	CVE: 2010-2729	OS Print Spool	29
V7	CVE: 2011-4453	pMWiki v2.2.30	21
V8	CVE: 2012-0003	Wireless Karma	28
V9	CVE: 2010-2883	Adobe Reader 9.3.0	29
V10		Backdoor executable	23
V11	CVE: 2010-0806	IE v 6.0.2900.2180	26
V12		Infectious Media	90
V13	Normal	Background	13

Data Design: Dataframes were designed and named “ADFA-WD-TRAIN” where all data training data was placed, “ADFA-WD-VALIDATION” where all validation data was placed, ADFA-WD-ATTACK, where all attack data dataset gathered [17]. In the ADFA-WD dataset it contains 9 attributes based on the Distinct Dynamic Link Count (DDLLC) and the primary key. The before mentioned dataset which 9 attributes of were provided by distinct DLL system calls; Kernel32, ntdll, user32, comctl32, ws2_32, mswsock, msvcrt, msvcpp, ntoskrnl, and then placed in a table (Figure 1).

Under this conditions the training and testing data contains 12 types of vulnerability attacks, with a binary classification as 0, and normal activities that are classified as 1. The binary approach is used because of the similarities between vulnerability attacks due to the attack being too precise to make a distinction in the class of attacks. All of the 12 vulnerabilities are classified as attacks (anomalies). Any deviation from the normal class type will be considered an attack. Testing was conducted using data ADFA-WD-VALIDATION, which are then followed for classification results obtained from the data ADFA-WD-ATTACK.



```
kernel32.dll+0x25797 kernel32.dll+0x2f067 kernel32.dll+0x25797 kernel32.dll+0x2f067
kernel32.dll+0x25797 kernel32.dll+0x2ec17 kernel32.dll+0x2f02b kernel32.dll+0x25797
kernel32.dll+0x2ec5c kernel32.dll+0x2f02b kernel32.dll+0x25797 kernel32.dll+0x2f02b
kernel32.dll+0x25797 kernel32.dll+0x2f02b kernel32.dll+0x25797 kernel32.dll+0x2f02b
kernel32.dll+0x25797 kernel32.dll+0x2eec5 kernel32.dll+0x2f02b kernel32.dll+0x25797
kernel32.dll+0x2eec5 kernel32.dll+0x2f02b kernel32.dll+0x25797 kernel32.dll+0x2eec5
kernel32.dll+0x2f02b kernel32.dll+0x25797 kernel32.dll+0x2f067 kernel32.dll+0x25797
kernel32.dll+0x2f067 kernel32.dll+0x25797 ntdll.dll+0x106ab
```

Key	Kernel32	Ntdll	User2_32	Comctl32	Ws2_32	Mswsock	Msvcrt	Msvcpp	Ntoskrnl
1	30	1	0	0	0	0	0	0	0

Figure 1. Gathering the attributes provided by distinct DLL system calls

The choice of classification methods: Classification Algorithm of Support Vector Machines – is a two rudimentary variation of feature space, aimed at the solution of the problem of binary classification. We decided to test the ability of two kernel functions to separate the attack and normal classes using the binary features: Sigmoid and Radial basis function (RBF).

The other machine learning method tested is the Random forest [18] – a classification algorithm, under which there is the construction of a plurality of decision trees during a training class and excretion, which is a mode of individual classes or regression trees.

Construction of Classifiers: The decision parameters were selected for the algorithms. The classification has been processed using Jupyter Notebook based on the desired parameters for classification.

Test classification: using ADFA-WD-VALIDATION we can carry out an effective process of comparison of the classification, which announced the results of predication in comparison with the level of accuracy of predicate data.

4 Evaluation and Discussion

4.1 Problems of acquiring an effective dataset

Optimization of the algorithm parameters: All weight classes were "balanced" to create a more accurate representation of the classes in which there would be less samples compared with bulkier class.

Cross-validation and Grid Search optimizes the parameters in order to create a better model for the algorithms used. Cross-validation k-FOLD = 5 is used for all algorithms scoring parameter "Accuracy". Search parameters of Grid Search vector: 'C': [1,10,100, 1000], 'gamma': [0.14], 'kernel': ['rbf'], 'decision_function_shape':['ovr'], 'class_weight':['balanced'] and setting method for random forest 'n_estimators': [5,10,15,20,25,30,35,40,45,50], 'max_depth':[5,7, 9,11,13,15,17,19],'min_samples_leaf': [1,2,3,4,5,6,7,8, 9,10], 'criterion': ['entropy', 'gini '],' class_weight ':['balanced '].

Table 2. *Results Data on Windows ADFA-WD*

Algorithm	**Detection Rate (DR)**	**False-positive Rate (FPR)**	**False-negative Rate (FNR)**	**False alarm Rate (FAR)**	**Processing Time (Seconds)**
SVM (RBF)	68%	71%	**1%**	36%	0.59
SVM (Sigmoid)	71%	**65%**	**1%**	**33%**	0.63
Random forest	**82%**	82%	10%	46%	**.019**

Table 3. *Results Data on Windows ADFA-WD:SAA*

Algorithm	**Detection Rate**	**Processing Time**
SVM (RBF)	68%	0.59
SVM (Sigmoid)	71%	0.63
Random forest	**82%**	**.019**

DR – is a representation of the accuracy of the attack data, calculated from the total amount of exactly predicted data of the attack, divided by the total number of data in said dataset. FPR represents an estimate of the total number of normal activities predicted to be an attack, divided by the total number of hectares of normal activities in this dataset. False Negative Rate is an estimate of the total number of attacks predicted as a normal action, divided by the total number of attacks in the dataset. FAR is (FPR + FNR)/2

With supervised learning, we were able to classify the attacks, used on the ADFA-WD and ADFA-WD:SAA. A binary method implementation is due to similarities in the approach of the attack types. The DR of SVM RBF was 68%, Sigmoid was 71%, Random Forest was 82%, but the FAR was fixed at 33%, 36% and 46%, respectively, as shown in Table 2 and Table 3, was built through the use of the confusion matrix (Table 4).

Table 4. Confusion matrix

Actual Classification	Predicated Normal	Predicted Attack
Normal	True Negative	False Positive
Attack	False Negative	True Positive

In regard to the ADFA-WD:SAA, The classification was done based on original training data of ADFA-WD , with the same DR as ADFA-WD.

Conclusion

In this paper, we were able to evaluate the system calls made on the DLL of the ADFA-WD. To evaluate the data a binary classification is implemented due to similarities in attack types making multiclass insufficient for the evaluation. SVM Algorithms (Sigmoid and RBF) though having a lower DR than Random Forest, it did achieve a better FAR, balancing the class weight played a key difference in getting an optimal DR and FAR as when looking at the 12 vulnerabilities and Normal Pattern.

Исследование частично поддержано Российским фондом фундаментальных исследований, проект 15-29-06031.

The research was partially supported by Russian Foundation for Basic Research, project 15-29-06031.

References

1. Sundaram, K. Why is Network Security Important? // Stonecypher, L. (ed.) — 2010 [Jelektronnyj resurs] // URL: http://www.brighthub.com/computing/enterprise-security/articles/69275.aspx (data obrashcheniya 15.10.2017).
2. Haider et al. Windows Based Datasets for Evaluation of Robustness of Host Based Intrusion Detection Systems (IDS) to Zero-Day and Stealth Attacks // Future internet 20168, 29. (2016) — C 1-8.
3. Burke et al. Measurement of the False Positive Rate in a Screening Program for Human Immunodeficiency Virus Infections // The New England Journal of Medicine.319 — 1988 — C. 961-964. doi: 10.1056/NEJM198810133191501.
4. Wu S. and Banzhaf W. The use of computational intelligence in intrusion detection systems: A Review // Applied Soft Computing 10 — 2010 — C.1-35. doi: 10.1016/j.asoc.2009.06.019.
5. Wu, S. and Yen, E., Data mining-based intrusion detectors // Expert Systems with Applications 36, Elsevier Ltd — 2009 — C. 5605-5612.
6. Manning et al. Introduction to Information Retrieval // Cambridge University Press, Cambridge — 2008 —
7. Common Vulnerabilities and Exposures [Jelektronnyj resurs] // URL: http://cve.mitre.org/data/refs/refmap/ (data obrashcheniya 15.10.2017).
8. Limin, L. Launching Return-Oriented Programming Attacks against Randomized Relocatable Executables // In Trust, Security and Privacy in Computing and Communications (Trust-Com) // 2011 IEEE 10th International Conference on 2011 — 2011 — C. 37-44.
9. Prandini M. and Ramilli M.Return-Oriented Programming /. Security Privacy, IEEE, 10 (6). – 11.12.2012 — C. 84-87.
10. Mahoney M and Chan P. An Analysis of the 1999 DARPA // Lincoln Laboratory Evaluation Data for Network Anomaly Detection / In Recent Advance in Intrusion Detection volume 2820 of Lecture Notes in Computer Science. Springer Berlin — Heidelberg. — 2003 — C. 220-237.
11. McHugh J.Testing Intrusion Detection Systems: a critique of the 1998 and 1999 DARPA Intrusion Detection System evaluations as performed by Lincoln Laboratory // ACM Trans. Inf. Syst. Secur., 3 (4) — 11.2000 — C. 262-294.
12. Owezarski P.A Database of Anomalous Traffic for Assessing ProfileBased IDS // In Traffic Monitoring and Analysis, volume 6003 of Lecture Notes in Computer Science. Springer Berlin // Heidelberg — 2010 — C. 59-72.
13. Creech G. Developing a high-accuracy cross platform Host-Based Intrusion Detection System capable of reliably detecting zero-day attacks // Ph.D. Dissertation, University of New South Wales, Sydney — 2014

14. Borisaniya B. and Patel, D.Evaluation of Modified Vector Space Representation Using ADFA-LD and ADFA-WD Datasets // Journal of Information Security, Vol. 6 — 07.2015 — C. 250-264,
15. KDD Cup 1999 Data. (1999) [Jelektronnyj resurs] // URL: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (data obrashcheniya 15.10.2017)
16. Creech G. and Hu J A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns // IEEE Trans on Computers — 2014 — C. 807-819.
17. Creech G. and Hu J. Generation of a New IDS Test Dataset: Time to Retire the KDD Collection // Wireless Communications and Networking Conference (WCNC 2013), Shanghai, 7-10 th April 2013 — C. 4487-4492.
18. Breiman, L. Random Forests. Machine Learning 2001. 45 (1) — C. 5–32. doi: 10.1023/A:1010933404324

Литература

1. Sundaram, K. Why is Network Security Important? // Stonecypher, L. (ed.) — 2010 [электронный ресурс] // URL: http://www.brighthub.com/computing/enterprise-security/articles/69275.aspx (дата обращения 15.10.2017).
2. Haider et al. Windows Based Datasets for Evaluation of Robustness of Host Based Intrusion Detection Systems (IDS) to Zero-Day and Stealth Attacks // Future internet 20168, 29. (2016) — C 1-8.
3. Burke et al. Measurement of the False Positive Rate in a Screening Program for Human Immunodeficiency Virus Infections // The New England Journal of Medicine.319 — 1988 — C. 961-964. doi: 10.1056/NEJM198810133191501.
4. Wu S. and Banzhaf W. The use of computational intelligence in intrusion detection systems: A Review // Applied Soft Computing 10 — 2010 — C.1-35. doi: 10.1016/j.asoc.2009.06.019.
5. Wu, S. and Yen, E., Data mining-based intrusion detectors // Expert Systems with Applications 36, Elsevier Ltd — 2009 — C. 5605-5612.
6. Manning et al. Introduction to Information Retrieval // Cambridge University Press, Cambridge — 2008 —
7. Common Vulnerabilities and Exposures [электронный ресурс] // URL: http://cve.mitre.org/data/refs/refmap/ (дата обращения 15.10.2017).
8. Limin, L. Launching Return-Oriented Programming Attacks against Randomized Relocatable Executables // In Trust, Security and Privacy in Computing and Communications (Trust-Com) // 2011 IEEE 10th International Conference on 2011 — 2011 — C. 37-44.
9. Prandini M. and Ramilli M.Return-Oriented Programming /. Security Privacy, IEEE, 10 (6). – 11.12.2012 — C. 84-87.
10. Mahoney M and Chan P. An Analysis of the 1999 DARPA // Lincoln Laboratory Evaluation Data for Network Anomaly Detection / In Recent Advance in Intrusion Detection volume 2820 of Lecture Notes in Computer Science. Springer Berlin — Heidelberg. — 2003 — C. 220-237.
11. McHugh J.Testing Intrusion Detection Systems: a critique of the 1998 and 1999 DARPA Intrusion Detection System evaluations as performed by Lincoln Laboratory // ACM Trans. Inf. Syst. Secur., 3 (4) — 11.2000 — C. 262-294.
12. Owezarski P.A Database of Anomalous Traffic for Assessing ProfileBased IDS // In Traffic Monitoring and Analysis, volume 6003 of Lecture Notes in Computer Science. Springer Berlin // Heidelberg — 2010 — C. 59-72.
13. Creech G. Developing a high-accuracy cross platform Host-Based Intrusion Detection System capable of reliably detecting zero-day attacks // Ph.D. Dissertation, University of New South Wales, Sydney — 2014
14. Borisaniya B. and Patel, D.Evaluation of Modified Vector Space Representation Using ADFA-LD and ADFA-WD Datasets // Journal of Information Security, Vol. 6 — 07.2015 — C. 250-264,
15. KDD Cup 1999 Data. (1999) [электронный ресурс] // URL: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (дата обращения 15.10.2017)
16. Creech G. and Hu J A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns // IEEE Trans on Computers — 2014 — C. 807-819.
17. Creech G. and Hu J. Generation of a New IDS Test Dataset: Time to Retire the KDD Collection // Wireless Communications and Networking Conference (WCNC 2013), Shanghai, 7-10 th April 2013 — C. 4487-4492.
18. Breiman, L. Random Forests. Machine Learning 2001. 45 (1) — C. 5–32. doi: 10.1023/A:1010933404324

Note on the authors:

Simon Conrad Kenyon, 5th year student of the "Fundamental Informatics and Information Technologies, Russian Peoples' Friendship University, Saint Vincent and Grenadini, conradsimon@hotmail.com

Sochenkov Ilya V., Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of Information Technologies, Peoples' Friendship University of Russia, sochenkov_iv@rudn.university

Об авторах:

Симон Конрад Кеньон, студент 5 курса направления «Фундаментальная информатика и информационных технологий, Российский университет дружбы народов, Сент-Винсент и Гренадиры, conradsimon@hotmail.com

Соченков Илья Владимирович, кандидат физико-математических наук, доцент кафедры информационных технологий, Российский университет дружбы народов, sochenkov_iv@rudn.university