Data Mining Techniques in Intrusion Detection: Tightening Network Security.

David Ndumiyana[1]*, Richard Gotora[2] and Hilton Chikwiriro[3]

[1,2,3]Bindura University of Science Education

Computer Science Department

P. Bag 1020

Bindura, Zimbabwe

Abstract

In any part of business environment, providing adequate protection of company networked resources is extremely important. Intrusion detection systems which should provide the pillar of strength for infrastructural defence against any form of malicious activity had been found wanting in that mandate: alerting the administrators and network managers in case of severe violation of the company security policy. The reason emanates from the increase in the number of false alarms which makes the system to operate at weaker levels. The biggest challenge facing intrusion detection systems today is dealing with both an attempted attack and a successfully launched attack.

This paper develops data mining - based model of intrusion detection system on both Network Intrusion Detection to monitor all network traffic passing on segment, where a detector is installed to alert the administrator of any signature based activity or suspicious anomaly, and Host Intrusion Detection to monitor inbound and outbound packets from a network device,and will alert the user or network administrator of suspicious behaviour detected. The model designed addresses negative effects of its weaknesses so as to enhance operational effectiveness. The importance of intrusion detection systems and the old techniques, type, characteristics and limitations would be given special attention in this research.

Keyword: Data mining techniques, intrusion detection system, false alarms, false negatives, network security.

Introduction

The mandate of an intrusion detection system is to monitor and analyse events happening in the computer network in order to detect signs of security breaches. According [1] "Data mining offers firms in many industries the ability to discover many patterns in their data patterns that can help them understand current market trends”. Intrusion detection (ID) is a type of security management system for computers and networks. An ID system gathers and analyses information from various areas within a computer or a network to identify potential security violations [12]. The main functions of intrusion detection systems include:

- Monitoring and analysing both user and computer activities.
- Analysing system configurations and vulnerabilities.
- Assessing system and file integrity.
- Ability to recognize patterns typical of attacks.
- Analysis of abnormal activity patterns.
- Tracking user policy violations. [12]

The number of attacks on networks and other sites is on the increase prompting development of ID systems so as to reduce the negative impact of these malicious attacks. Network security is fast becoming vulnerable because the authors of possible technologies of attack are becoming complex.

Secure Computing Infrastructure: Objectives

Most businesses depend entirely on a functional security tightened computing infrastructure to the extent that any lapse could be critical to their doing business [18]. The increase in the amount of interactive internet applications in the field of e-business and e-commerce has facilitated a rise in risks and possibilities of misuse, hence the following core objectives have to be met in order to protect the network of misuse [1,2,3]:

- Confidentiality – means secret information remains undisclosed.
  Possible countermeasures: encryption
- Integrity –means the information is protected against unexpected modification.
  Possible countermeasure: Message Authentication Code
- Availability – the objective guarantees that a resource is available when needed.
  Possible countermeasure: firewall blocking denial of service attacks
- Authentication – means the identity of a party is confirmed.
  Possible countermeasure: digital certificate
- Non-repudiation – means the denial of integrity and authenticity of information is not possible.

Possible countermeasure: digital signatures

Therefore to meet all the objectives described, it is crucial to protect the system as much as possible against intruders. The job of IDS is to at least detect an anomaly should intrusion occurs despite of defence.

Types ofIntrusion Detection Systems

Intrusion detection systems are designed principally to monitor network traffic and monitor for suspicious activities and alert the system administrator. Intrusion detection technology can be divided into Misuse Detection and Anomaly Detection.

Anomaly Detection Technique

Anomaly Detection defines distinction between detection and unacceptable behaviour according to security programming done by the network administrator. It clearly defines each acceptable activity and unacceptable behaviour is labelled as intrusion [14, 12, 4].The technique gives a detailed summary of characteristic which normal operation should have in the event that user activity has significantly deviated away from the normal behaviour, undoubtedly considered intrusion. The rate of missing report in this kind of detection is low and even though it is able to effectively detect unknown intrusion, the rate of false alarm is high [11]**.**

Misuse Detection Technique

Misuse Detection describes the level of matching between detection and unacceptable behaviour according security programming done by the network security expert. The behaviour characteristic of unusual operation is collected by security personnel andbuilds a repository of unusual operation characteristics. If the monitored user or system behaviour matches record of repository, then this behaviour is regarded by the system as intrusion [6]. It has been observed that this kind of detection is characterised by a low rate of false alarm and a high rate of missing report.

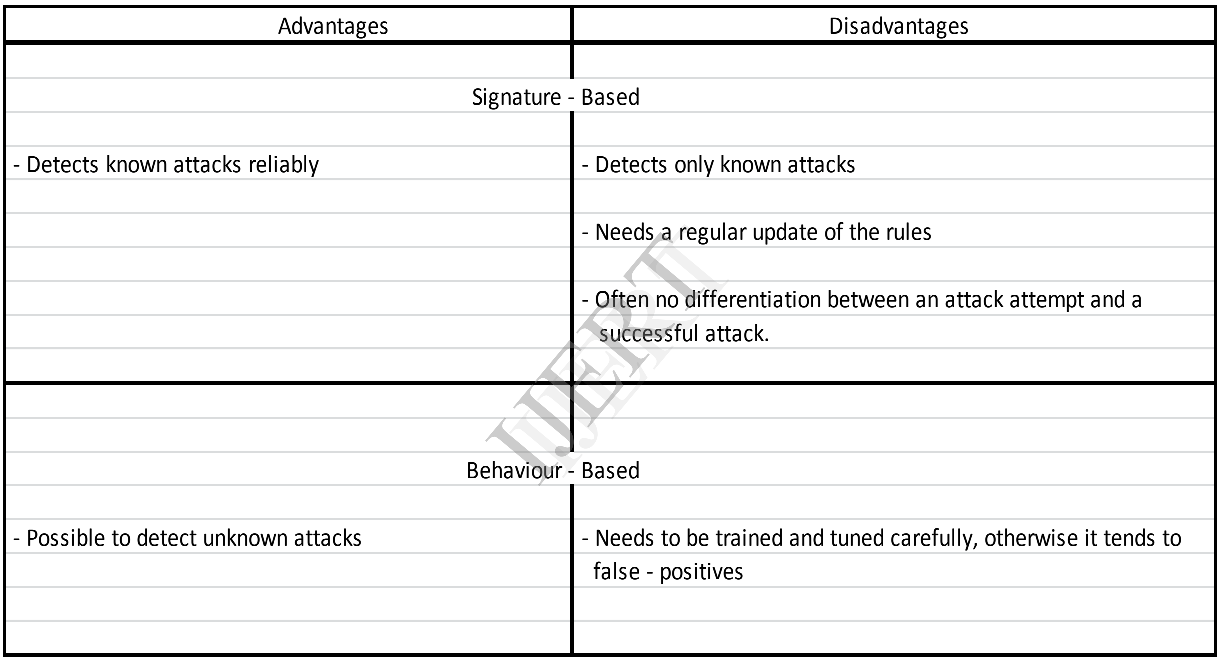
For the known attack, it may report a detailed and accurate report; for the unknown attack, its function is limited, addition, the characteristic repository is renewed continually [11].

Signature - Based versus Behaviour – Based IDS

A signature based (misuse detection) IDSis used to monitor packets on the network and compares the packets against a library of signature or characteristics from known malicious threats [6, 16]. In short it mimics the way most antivirus software packages detect malware (viruses, worms, Trojan viruses).

Behaviour – based IDS (anomaly detection system) analyses packets of data in the network at the very first instant. The main objective is to distinguish between normal and abnormal traffic examining the fundamental behaviour of a system [2, 6, 18]**.**

ffp



Advantages	Disadvantages
Signature - Based	
- Detects known attacks reliably	- Detects only known attacks
	- Needs a regular update of the rules
	- Often no differentiation between an attack attempt and a successful attack.
Behaviour - Based	
- Possible to detect unknown attacks	- Needs to be trained and tuned carefully, otherwise it tends to false - positives

Table 1.1: Signature – Based versus Behaviour – Based [18]

Network Intrusion Detection System

Intrusion detection systems can further be divided into Network Intrusion Detection Systems (NIDS) and Host Intrusion Detection System (HIDS), the fundamental goal is to detect suspicious traffic using a variety of techniques as each method brings its own weaknesses and strengths.

NIDS – These systems are placed at a strategic point or points within the network to monitor traffic to and from all devices on the network[15]. The system normally scans both inbound and outbound traffic but has the disadvantage of reducing the overall speed of the network [12].

Host Intrusion Detection System

HIDS – The systems are run on individual hosts or devices on the network[15] .Basically a HIDS monitors both inbound and outbound packets from the device only and will alert the user or network administrator of any suspicious activity detected [12].

Network – based versus Host – based IDS

A number of IDS integrate parts from the two approaches to enhance the effectiveness and efficiency of operations. This is important because each technique has its own weaknesses and strengths.

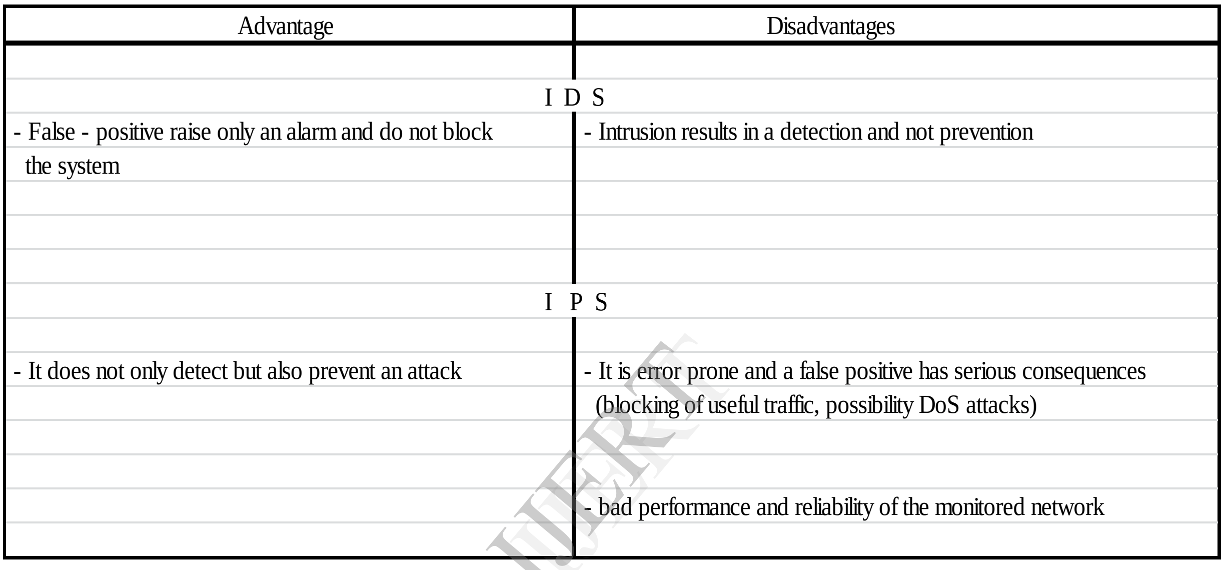
Advantages	Disadvantages
Network - Based	
- No impact on end systen	- High requirements on computing performance to scan every packet
- Detection of distributed attacks e.g denial of service (DoS)	- Detection of attacks that manifest themselves in the network
- Invisible configuration (stealth mode)	- Cannot be used if encrypted communication is allowed (unless a workaround like an SSL proxy is available)
	-
Host - Based	
- Monitors the actual reaction of the host	- installation on every single host
- Access to host - specific information e.g integrity checker, process or system call monitoring	- Adaptation to the different platforms and operating systems
- Monitoring on all protocol layers	- Performance requirements on every supervised host
- Encryption is no hindrance (except on the application layer	- No detection of distributed attacks on multiple targets

Table 1.2: Network – Based versus Host – Based IDS **[7]**

Intrusion Prevention System (IPS) versus Intrusion Detection System (IDS)

Theactive IPS monitors and analysespackets on a continuous basis reporting any security violation in real time. IPS detects, blocks and prevents an attack independently as human intervention is of no or little significance.

In addition to inspecting packet headers such as ports and IPs, other attributes such as packet flow or content of a packet are also subjected to scrutiny. The passive IPS on the other hand depends on human input to sound an alarm in case of a supposed attack [1, 7]**.**



Advantage	Disadvantages
I D S	
- False - positive raise only an alarm and do not block the system	- Intrusion results in a detection and not prevention
I P S	
- It does not only detect but also prevent an attack	- It is error prone and a false positive has serious consequences (blocking of useful traffic, possibility DoS attacks) - bad performance and reliability of the monitored network

Table 1.3: IDS versus IPS showing advantages and disadvantages [18].

IDS versus Extrusion Detection System

Most currently used IDS concentrate on intrusion from outside of the network into the monitored network. This detection of attacks yields a good number of false alarms.Extrusion detection is a new technique which focuses on traffic whose source address is found inside of the monitored network**.**

Advantages	Disadvantages
IDS	
- Detection during infection	- Detection of unsuccessful attacks without a real intrusion
EDS	
- Detection denotes a successful intrusion and not only an attempted attack.	- Not detectable before the host is infected

Table 1.4: IDS versus EDS showing advantages and disadvantages [18]

Data Mining Technology

Data mining is defined as the activity of discovering interesting patterns from a large amount of data, where the data can be stored in a database, data warehouses or other information repositories [5, 10]. Data mining involves integration of techniques from multiple disciplines such as databases and data warehouse technology, statistics, machine learning, high – performance computing, pattern recognition, neural networks, data visualisation, information retrieval, image and signal processing, and spatial or temporary data analysis [8].

The process of data mining is very useful in the extraction and in understanding patterns from a lot of incomplete, noise, non – stable, vague and random data. It is the science that allows systems to extract useful information from a large data set or databases [13]. Data mining techniques therefore extract implicit, previously unknown and potentially useful information from data [4, 14, 15]

Intrusion detection systems generate important information from a variety of sources such as the host log, the network data package, the system's data against applications and alarm messages among others. Data mining technology has a huge advantage in data extracting characteristic and the rule, hence it is of great importance to use data mining technology in the intrusion detection[11, 3]. The main idea behind data mining tools: Through analysis mining of the network data and the host call data discover misusing detection rule or exception detection model [2]. The importanceof applying data mining technology in intrusion detection is that it can audit data to obtain the model, thereby facilitating urgent catching of actual invasion and the normal behaviour pattern accurately [11].

Therefore the main benefit of using data mining technology is that the same data mining tool can be applied to many data streams, hence it is an advantage to build strong intrusion detection systems.

However, distinguishing between the normal and abnormal behaviour from a huge pool of raw data attributes and how to effectively generate automatic intrusion rule soon after collecting raw network data remain the biggest challenge affecting intrusion detection systems in general.The negative repercussion of this problem could be reduced by correlation analysis algorithm to discover relationship of attributes in network connection record, sequence analysis algorithms can discover the timing relationship of network connection records, both correlation analysis and sequence analysis algorithms can be used to obtain the normal behaviour pattern which is used in anomaly intrusion detection. In addition, classification of data mining algorithm can be used to generate mining rule from trained datasets, which can identify normal behaviour and intrusion [11].

System Model Design

Adata mining technology network intrusion detection system model was designed by fusing the general process of intrusion detection and the features of data mining **(**as shown in Diagram 1).

Architecture for Intrusion Detection System

Data Capture

Captures data which will be used for analysis on the network influencing internet optionsattract a variety of patterns and intercepting every packet on the network for further mining.

Data Preprocess

Data packet is transformed into appropriate forms using many approaches such as data cleaning, data integration, data reduction techniques and standardized data analysis [10]. The quality of extracted user features and accuracy of derived rules are all enhanced in the Data Preprocess mining activity. Therefore Data Preprocessing is done to enhance the accuracy and effectiveness of data mining algorithms [7, 17]. Data preprocessing is important because the real world data to be analysed by data mining techniques are:

- Incomplete: lacking certain attributes of interest.
- Noisy: containing errors
- Inconsistent: containing discrepancies between different data items.
- Enhancing mining process: by reducing number of data sets to enhance performance
- Improve data quality: Data preprocessing techniques can improve the quality of data, thereby helping to improve the accuracy and efficiency of the subsequent mining process.Data preprocessing is an important step in knowledge discovery process because quality decision must be based on quality data. Detection of data anomalies, rectifying them early and reducing the data to be analysed can lead to quality decision making [8]**.**

Data cleaning is a technique of data preprocessing which deals with the removal of imperfect, such as noise for continuous data attributes [8]

Data Mining Processor

Data warehouse is responsible for keeping data obtained from the Data Preprocess, and mines the training data which the control module produces. Data mining algorithm library module is a collection of a variety of mining algorithms (e.g connection rule algorithm, sequence pattern analysis algorithm). The efficiency of each method in the library of algorithms is improved by making sure that each technique has good time complexity, extensible, in order to initiate search process of data mining, and be used for forecast next time. Data mining process control is responsible for choosing a suitable mining algorithm from the data mining algorithm library. The traditional data mining intrusion detection system labels the data as normal data or intrusion behaviour, meaning it needs well labelled data to obtain training data sets. During the process of marking training data sets, the data mining control module may extract the features and detection characteristics from the library of algorithms. The mining control module can also be used when system starts running before training data sets is available, by calling data from a data warehouse and using clustering algorithm, the data is labelled as normal or abnormal data and returns data warehouse as a training data.

Finally, results obtained from the mining process are carried to intrusion detection module to produce appropriate communication to the security officer or for the system to take a suitable response.

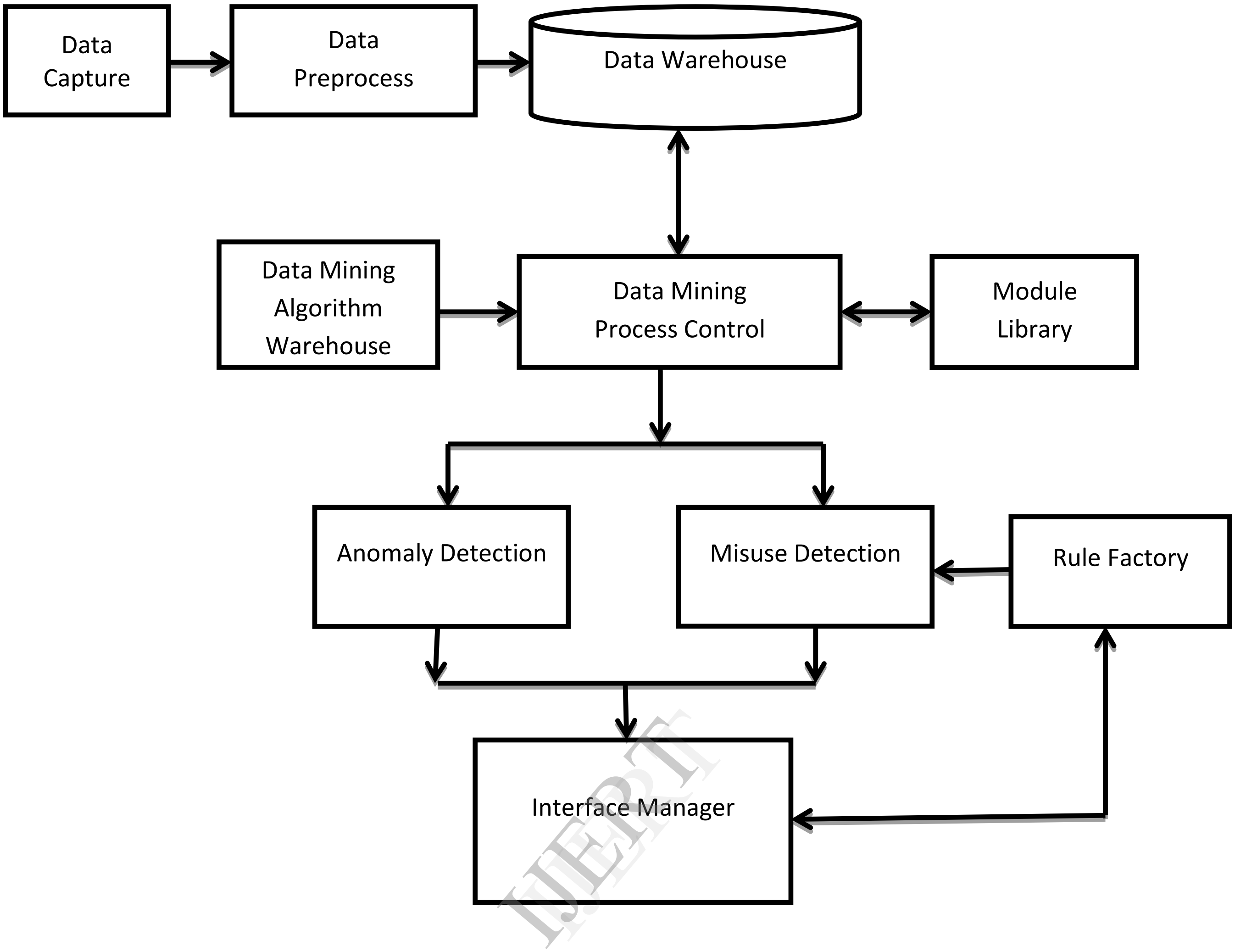


Diagram 1: Architecture for Intrusion Detection System [11]

Intrusion Detection Module

Rules Factory: Keeps the rules that intrusion detection system needs for matching with data mining output modules. Anomaly Detection Module: This module is responsible for generating a summary of characteristic which the normal behaviour should have, comparing the rules from current network data streams with the rules of Rules Factory, if detected data is beyond the threshold, that is intrusion, else normal behaviour.

Misuse detection module: The system believes an intrusion has occurred if detected system behaviour matches records of Rules Factory.

Interface Manager

This module is responsible for making quality decision on normal or unknown pattern. If the decision reached is normal pattern, it adds it to close normal pattern in the regular library, but if the decision reached indicates abnormal pattern, adds it to close abnormal pattern in the regular library resulting in the execution of important processing measure instantly.

Conclusion

Data mining technology can help enhance intrusion detection by tightening IDS through paying attention at anomaly detection. We have shown the methods in which data mining is helpful in the process of Intrusion Detection and the various ways of applying these methods.

Finally, we proposed a data mining model that we are confident can contribute significantlyin our quest to create better and more effective Intrusion Detection Systems.

In future, we recommend the use of Comprehensive Defense which deals with all network security concerns infusing new ideas and methods of safety engineering risk management and treating network security as an investment.

References

[1] Julisch K. Data mining for intrusion detection: A critical review. IBM Research. Zurich Research Laboratory.

[2] “Data Mining Approaches for Intrusion Detection” Lee, Wenke and Stolfo, Salvatore.

[3] Rothleder, Neal. “Data Mining for Intrusion Detection”. The Edge Newsletter

[4] Fayyad, Usama; Gregory Piatetsky – Shapiro, and Padhraic Smyth (1996). “From Data Mining to Knowledge Discovery in Databases”.

[5] The architecture of a network level intrusion detection system. Technical Report, R. Heady, G. Luger, A. Maccabe, and M. Servilla. Computer Science Department, University of New Mexico. August 1990

[6] Geer, D. (2006), Behaviour-Based Network Security Goes Mainstream.

[7] Pyle, Dorian. Data Preparation for Data Mining. San Francisco, CA: Morgan Kaufmann, 1999. Print

[8] http://loremate.com/

[9] http://ptucse.loremate.com/mp/

[10] Kovac, S. (2012) Suitability analysis of data mining tools and methods. Bachelor's Thesis, Faculty of Informatics, Masaryk University.

[11] LI Min, Application of Data Mining Techniques in Intrusion Detection, An Yang Institute of Technology.

[12] Jadhav, R.J. & Pawar, U.T. (2011), "Data Mining for Intrusion Detection", International Journal of Power Control Signal and Computation (IJPCCSC), 1 (4), pp. 45-48.

[13] Hand, D., Smyth, P., Mannila, H. (2001) Principles of Data Mining, MIT Press.

[14] Frawley, W., Piatetsky-Shapiro, G., Matheus, C. (1992) "Knowledge Discovery in Databases: An Overview", AI Magazine, pp. 213-228.

[15] Siddiqui, M.A. (2008), PhD Thesis, "Data Mining methods for Malware Detection", University of Central Florida, Orlando, Florida.

[16] Shabtai, A., Moskovitch, R., Feher, C., Dolev, S. & Elovici, Y. (2012), "Detecting unknown malicious code by applying classification techniques on Opcode patterns", Security Informatics, a SpringerOpen Journal, 1(1).

[17] Azam, N. (2002) Comparative Study of Features Space Reduction Techniques for Spam Detection,

MSc Thesis, Department of Computer Engineering, College of Electrical and Mechanical Engineering, National University of Science Technology.

[18] Schweikert, D. (2008). Master Thesis, "Signature-based Extrusion Detection", Swiss Federal Institute of Technology, Zurich.