A SURVEY ON ANOMALY AND SIGNATURE BASED INTRUSION DETECTION SYSTEM(IDS)

Roshni Dubey*

Pradeep Nandan Pathak**

ABSTRACT

Denial of Service (DoS) and Distributed DoS (DDoS) attacks are evolving continuously. These attacks make network resources unavailable for legitimate users which results in massive loss of data, resources and money. Combination of Intrusion detection System and Firewall is used by Business Organizations to detect and prevent Organizations" network from these attacks. But this combination cannot prevent network from novel attacks as Signatures to detect them are not available. This paper presents a light-Weight mechanism to detect novel DoS/DDoS (Resource Consumption) attacks and automatic Signature generation process to represent them in real time. Experimental results are provided to support the proposed mechanism.

Keywords

Novel DoS attack detection, automatic Signature generation, Main Memory Database Management System

* Assistant Professor (I.T.), Shri Ram Institute of Technology, Jabalpur (M.P)

** Shri Ram Institute of Technology, Jabalpur (M.P)

1. INTRODUCTION

DoS/DDoS attack is attempt by attacker to prevent Internet site or Server from functioning efficiently or properly. There are several ways of launching DoS/DDoS attacks against a server. Every attack uses any one of the following technique:

i. Consume Server resources

ii. Consume network bandwidth

iii. Crash the server using vulnerability present in the server

iv. Spoofing packets

Even though there are different ways to launch attack but every attack makes server either nonresponsive or extremely slow. Firewall and Intrusion Prevention System (IPS) can prevent Server from known DoS/DDoS attacks and sometimes from their variations; as their working mechanism is known in advance. But no one can build a prevention system which will prevent Server from every novel DoS/DDoS attack. One possible solution is to detect a novel attack in real time and automatically generate a signature to represent it. Once the signature of attack is available; defense mechanism against that attack can be developed.

Time required for microprocessor to access data stored in main memory (RAM) is very less compared to data stored in secondary memory (disks). With continuous increase in density of semiconductor chips and decrease in their cost, it is economically feasible to store and process huge amount of data in main memory. Main Memory Database Management System (MMDBMS) uses main memory as primary storage for data and provides high speed access to it. This makes it suitable option for implementation of real-time network security systems. Proposed approach uses MMDBMS to store network traffic information in order to increase the speed of novel attack signature generation module.

Section 2 gives overview the related work and section 3 gives overview of KDD 99 dataset used for training and testing IDS. Section 4 describes advantages of attribute selection process in design of IDS and Section 5 describes requirement of automating the Signature Generation process. Section 6 describes proposed mechanism and section 7 presents the Experimental results. Section 8 compares the proposed mechanism with existing solutions and section 9 concludes the paper.

2. RELATED WORK

Gang Xiong and Minxia Zhang [1] proposed a clustering based outlier detection method to detect unknown (novel) intrusive activities. They considered intrusive activities as outliers and used DOExMi Cluster (proposed by them) to detect outliers of unknown type. The micro-cluster concept, data normalization and k-mean measure are used interactively to create sub micro-clusters of normal activities till two micro-clusters can be merged to create new micro-cluster. After this network activity instances which cannot fall into any micro-clusters are recognized as outliers.

Jie Yang et al [3] proposed Hybrid (Anomaly-Misuse) Intrusion Detection System (IDS) using network protocol analysis. It is consist of four modules: Data Preprocessing, Misuse detection, Anomaly Detection and Decision making module. Both Misuse and Anomaly detection modules are built using Decision tree. Decision making module classifies any network activity as intrusion if both (Misuse and Anomaly) detection modules classify it as intrusion.

Bharanidharan Shanmugam and Norbik Bashah Idris [4] proposed Fuzzy logic based Hybrid (Anomaly-Misuse) (IDS). It has three main components: Data Analyzer, Fuzzy Data Miner and Fuzzy Inference Engine. Data Analyzer module analyzes network packets and performs packet grouping to get aggregate information. This aggregated information is then used by fuzzy data miner to generate the rules for Intrusion detection. These rules and network traffic data is given as an input to the fuzzy inference engine which determines whether there is any

intrusive activity or not. It cannot be used for real time detection due to its high computational complexity and high false positive rate.

Imen Brahmi et al [5] proposed Hybrid (Misuse-Anomaly) IDS using data mining and mobile agent technology to detect known and novel attacks. It uses mobile agents to collect and analyze network traffic. Multiple copies of sniffing agents are created and distributed in the network to collect network traffic data in a file. Data present in this file is processed by filter agent to transform it into a form suitable for Misuse and Anomaly detection. Distribute clustering technique is used by Anomaly detection agent to process huge amount of information in efficient manner and early detection of novel attacks. Rule mining agent uses Apriori algorithm to generate signatures of novel attacks detected by Anomaly detection agent.

Yu-Xin Ding et al [8] proposed a Snort-Based Hybrid (Misuse-Anomaly) IDS. It is divided into three modules: misuse detection, anomaly detection and signature generation module. Snort is used as misuse detection module to detect known attacks. Anomaly detection module used uses Frequent Episode Rule mining algorithm with a sliding window to generate rules for Anomaly detection. Signatures of newly detected attacks by Anomaly detection module are generated by using Signature generation module. It uses Apriori algorithm for signature generation. IT provides good performance in offline detection, but cannot be used for real- time detection.

3. KDD 99 DATA SET

KDD Cup 99 data set [15] is widely used and authoritative data set to train, test newly developed IDS. It contains 4,94,021 connection records in training data set which belongs to any of 22 different type of attack or legitimate network connections. Each record is consist of 41 fields (features) and is labeled as either normal or intrusive connection (attack). First 9 features of connection record are basic features (e.g. duration, protocol type, etc…). Next 13 features are obtained by using domain knowledge (e.g. number of failed login attempts, number of root accesses, etc...). Last 19 features are obtained using two second time windows (number of connections to the same host as the current connection, the number

of connections to the same service, etc…). A complete listing of features, attacks in data set and details description can be found in [15].

4. ATTRIBUTES SELECTION

Before using collected raw network traffic data for Intrusion detection, it is transformed into records format which contains number of network traffic features such as service, duration, protocol, source bytes, destination bytes, etc… Selection of network traffic features for intrusion detection is important step and decides the success, effectiveness of IDS. Intelligent selection of features results in simplified, faster and more accurate IDS.

Adetunmbi A.Olusola et al [9] performed experiments using KDD 99 dataset. Their experimental results show that, two network traffic features num_outbound_cmds and is_hot_login have no relevance in Intrusion Detection. Their results also show that, derived features namely num_compromised, su_attempted, num_file_creations, is_guest_login, and dst_host_rerror_rate are of little significance for Intrusion Detection process. So if these features are not used, it will increase speed of IDS and reduce the resource requirements without affecting the accuracy.

Neveen I. Ghali [10] performed experiments using KDD 99 data set and experimental results show that, only 7 features are enough to detect DoS attack with high accuracy. This reduction in number of attributes for detection process reduces

i. Amount of data to be processed by 83%

ii. Mean square error in detection of novel attack by approximately 90%

iii. Memory and CPU time required to detect attacks

H. Güneş Kayacık et al [11] performed experiments using KDD 99 data set and used information gain to express Feature relevance with different attacks. Their experimental results show that, normal network traffic, neptune and smurf attacks are highly related to certain network traffic features compared to others. If only these are used for Intrusion detection, attack detection task becomes much easier and provides good results.

Wei Wang et al [12] performed experiments using KDD 99 dataset and their experimental results show that, network traffic record containing only 10 relevant features with highest information gain can be used for Intrusion detection with same or improved detection rate.

5. AUTOMATIC SIGNATURE GENERATION (ASG)

Every activity (legitimate and intrusive) over network has a unique pattern. These patterns can be used to detect which activities are going on the network. So these patterns are also called as Signatures. Signatures of known intrusive activities
are defined and used to detect their existence. But there are two major problems in this approach; both problems are related with the manual process usually carried out to create Signature of attack.

First, a detailed and precise knowledge about attacks process is required to define its Signature. If defined Signature is too simplified, it will generate high false positive rate. On other hand if it is too specific, it will result in high false negative rate. Second, some time is required to gain detailed and precise knowledge about attack. This introduces delay between the first time attack is reported and generation of signatures to detect it. Thus zero-day or novel attacks are serious threat for computer systems. M. Soleimani et al [13] proposed some approaches to make this process easier by correlating and thus reducing the number of alerts to analyze, but major problems are still unsolved.

According to I. Qualys [14] approximately twenty to forty new vulnerabilities in commonly used networking and computer products are discovered and published every month by users or attackers. Such wide-spread availability of known vulnerabilities leads to

launch of novel („Zero-day") attacks. Since Firewall and IDS cannot protect network against novel attack it leads to massive loss of data, resources and money. Thus, both the activities are very crucial

i. Detect novel attacks in real time without human involvement

ii. Generate signatures of novel attack in real time without human involvement

Once the signature of attack is known, security expert can design a prevention mechanism against that attack. Thus Automatic Signature Generation for novel attacks using selected features of network traffic is one of the hot research areas in network security domain.

6. PROPOSED MECHANISM

Figure 1 shows the proposed mechanism. Signature-based IDPS is used to detect and prevent server from known DoS/DDoS attacks. Anomaly Detection based Filter (ADF) and Signature Generator (SG) are used to generate signatures which can represent Novel attack. Known Attack Signature DB (KAS DB) contains signatures of known attacks and used by Signature based IDPS to detect them. LogDB contains all the connection records which do not match with known

attacks. These records are used to generate signatures of novel attack after filtering them by using ADF.

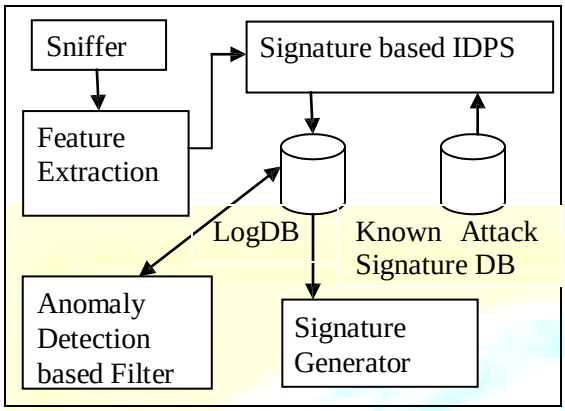


Fig 1: Proposed Mechanism

6.1 Signature-based IDPS Module

KDD Cup 99 dataset and One-pass Incremental Apriori algorithm [16] are used to generate signatures of known attacks. These signatures are used by IDPS for attacks detection. After detection it performs preventive steps according to type of attack detected. For example, when it detects „Land", „Teardrop" attack; packet containing that attack is dropped.

Working:

Step 1: Receive network traffic feature record

Step 2: Search for network traffic feature record in KAS DB Step 3: If (Attack detected) Then

a) Take preventive action according to type of attack detected

b) Go to step 1

Else

a) Insert traffic feature record in LogDB

b) Go to step 1

End If

Step 4: Go to Step 1

6.2 ADF

Attack free traffic to the server is collected during training period and used to train Anomaly Detection System (ADS). ADS is then used by ADF to filter out legitimate connection records from LogDB. This filtering reduces the load of SG module.

Working:

Step 1: Wait for 2 minutes

Step 2: For every record in LogDB

a) If (Record matches with signature of ADS) Then

a. Remove record from LogDB Step2: Go to Step 1

6.3 SG

One pass Incremental Apriori algorithm [16] is proposed by us to generate signature of DoS/DDoS attack. SG uses this algorithm to identify large item sets of size 9 and these item sets are considered as Novel attack Signatures. These 9 features are: Duration, dst_bytes, src_bytes, service, flag, Count, srv_count, num_compromised, Wrong_fragment [15].

6.4 Novel DoS/DDoS attack detection Strategy

As stated in Section-I every DoS/DDoS attack either makes Server nonresponsive or extremely slow. IDPS module periodically checks whether the Server is responding within predefined time duration or not. If it is not responding then Sever is considered to be under novel DoS/DDoS attack and IDPS executes ASG procedure.

Working of Novel attack Detection process:

Step 1: Wait for 2 Minutes

Step 2: Send request to server to detect its responsiveness

Step 3: If (Response is received with predefine interval) Then a) Remove all connection record from LogDB which are older than 30 minutes

Else

a) Execute ASG Procedure

End If

Step 4: Go to step 2

It is almost impossible to define exact behavior of legitimate users, as it can change over a

time period and according to situation. Thus Network traffic that does not match with signatures present at ADS need not be a malicious traffic. If all traffic which does not match with ADS signatures is used to generate signatures of novel attacks:

i. It will increase load of SG module

ii. Decrease the accuracy of generated signatures.

So if signs of DoS attacks are not present for long time then older network connection record which does not match with Anomaly detection signatures can be treated as non malicious (Deviation in normal user behavior) and discarded from SG process.

6.5 ASG Procedure

Working:

Step 1: Trigger ADF to remove legitimate connection records from LogDB

a) If (LogDB is empty after ADF process) Then

a. Report that, "Server is facing Heavy legitimate load"

b. Go to Step 4

Step 2: Trigger SG to generate signatures using data present in LogDB

Step 3: Submit signatures for evaluation

Step 4: Stop

Some time there exist sudden rise in legitimate traffic to Server, which creates impact

similar to DoS attack. E.g. After the deaths of Michael Jackson in 2009 websites like Google or

Twitter were slowed down for a long time due to tremendous inquires of users related to him.

7. EXPERIMENTAL RESULTS

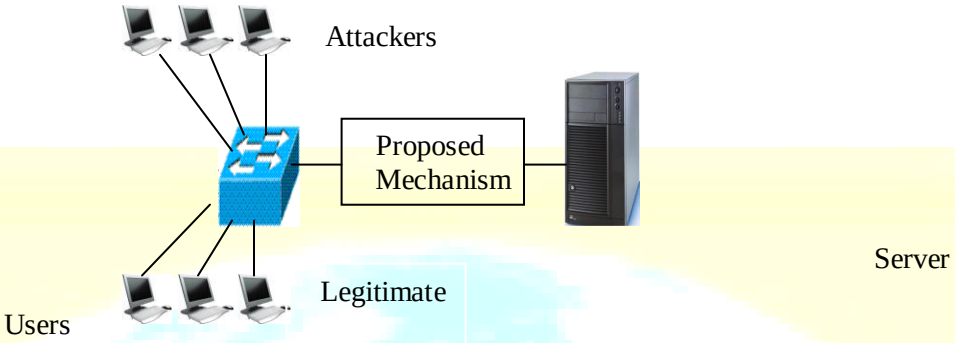


Fig 2: Experimental Setup

Experimental setup is shown in Figure 2. Proposed mechanism is implemented using Java, JPCAP, WinPcap and PC (Intel Core 2 Duo, 1.60GHz, 1 GB Ram) with 2 network cards. First network card is used to receive data packets from external world (Clients). Packets arrived on this network card are analyzed by IDPS to detect known intrusions. It discards intrusive packets and forwards legitimate packets to Server using second network card.

Client, Attacker and Server programs are implemented using Java. Client connects to server and sends a series of complex mathematical equations. Server solves these equations and sends reply back to Client. Server can accept and handle maximum 100 client connections. Signature-based IDPS is designed to detect and prevent server from Land, Pod, smurf, teardrop attacks [17] i.e. these attacks are considered as known DoS attacks. Neptune (SYN Flood) [17] and two resource consumption attacks are used as Novel attacks.

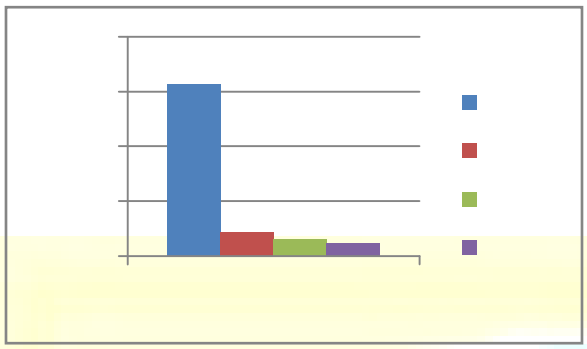
Originally Duration, Src_bytes and Dst_Bytes fields contain continuous values, but these values are not suitable for Frequent Item set identification. So these fields are converted into discrete by dividing them by 10.

7.1 Test Case1: Neptune (SYN Flood) attack

SYN Flood attack is launched using 5 attacking machines. Signatures generated by SG after SYN flood attack are:

a) Duration-> 0, dst_bytes-> 0, src_bytes-> 0, service-

> private, flag-> REJ, Count-> 38, srv_count-> 38, num_compromised,->0, Wrong_fragment-> 0



NRNM: No of Records which do Not Match with Anomaly and Signature based IDS

NRSG: No of Records used for Signature Generation

NAIR: No of Actual Intrusive Records

Here, 15,640 records were analyzed by IDPS. Out of these 2217 records were not matching with Anomaly Filter as well as Signature based IDS. Still only 1548 records were used for Signature generation. This reduces the time and processing power required to generate Novel attack Signature. 1200 out of 1548 records were actual intrusive records; this increases the accuracy of generated Signature.

7.2 Test Case 2: Resource Consumption

Attack 1(RCA 1)

Attackers establishes many connections with Server and remains ideal (no data transfer) forever. Thus after some time connection acceptance resources are drained and server becomes nonresponsive. RCA 1 attack is launched using 5 attacking machines. Signatures generated by SG after RCA 1 flood attack are:

a) Duration-> 18, dst_bytes-> 0, src_bytes-> 0, service-> private, flag-> S1, Count-> 100, srv_count->100, num_compromised,->0, Wrong_fragment-> 0

b) Duration-> 14, dst_bytes-> 0, src_bytes-> 0, service-> private, flag-> S1, Count-> 100, srv_count->100, num_compromised,->0, Wrong_fragment-> 0

c) Duration-> 10, dst_bytes-> 0, src_bytes-> 0, service-> private, flag-> S1, Count-> 100, srv_count->100, num_compromised,->0, Wrong_fragment-> 0

These generated signatures perfectly represent the attack, i.e. there exist many connections with Server but data transfer does not take place.

NRSG

NAIR

Where;

Neptune

Fig 3: Neptune attack Data Analysis

Here, 6,543 records were analyzed by IDPS. Out of these 752 records were not matching with Anomaly Filter as well as Signature based IDS. Still only 178 records were used for Signature generation. 100 out of 178 records were actual intrusive records.

NCR: No of Records inspected by IDPS

7.3 Test Case 3: Resource Consumption

Attack 2(RCA 2)

Attackers establish many connections with Server and send requests (mathematical equations) at extremely slow rate (2/Minutes) and do not terminate the connection. Thus after some time connection acceptance resources are drained and server becomes nonresponsive. RCA 2 attack is launched using 5 attacking machines. Signatures generated by SG after RCA 2

flood attack are:

a) Duration-> 18, dst_bytes-> 6, src_bytes-> 30, service-> private, flag-> S1, Count-> 100, srv_count-> 100, num_compromised,->0, Wrong_fragment-> 0

b) Duration-> 14, dst_bytes-> 5, src_bytes-> 25, service-> private, flag-> S1, Count-> 100, srv_count-> 100, num_compromised,->0, Wrong_fragment-> 0

c) Duration-> 10, dst_bytes-> 3, src_bytes-> 15, service-> private, flag-> S1, Count-> 100, srv_count-> 100, num_compromised,->0, Wrong_fragment-> 0

These generated signatures perfectly represent the attack, i.e. there exist many connections with Server but the amount of data transfer is very less.

Here, 5,943 records were analyzed by IDPS. Out of these 791 records were not matching with Anomaly Filter as well as Signature based IDS. Still only 162 records were used for Signature generation. 100 out of 162 records were actual intrusive records.

7.4 Test case 4: Flood of normal traffic Flood of normal requests is generated using 10 computers. All records in LogDB are deleted by ADF filter process initiated by ASG. Thus ASG generates an alarm stating that, "There is a flood of legitimate traffic".

8. COMPARISON WITH EXISTING APPROACHES

Clustering based IDS methods are based on two assumptions. The first assumption is; number of normal instances in dataset is much higher than number of intrusive instances and second is; enough difference exists between the intrusive and the normal instances. But these assumptions are not true in every situation. Proposed method does not rely on any assumption. It performs incremental analysis while remaining approaches does not. This difference makes it real time solution for detection of Novel attacks. Every other mechanism treat every network traffic record which

does not match with Signature and Anomaly based IDS as possible attack which increases the computational complexity; this is not done by proposed mechanism. This reduces the processing time and increases the accuracy of Generated Signature. Comparison of proposed mechanism with existing solutions is shown in table 1.4000

Table 1. Comparison of proposed mechanism with existing solutions

Method Name	Is Incremen	Assumptions Used	Data Processed for new attack	Relative Complexity	Real time
Outliers Detection Based on Clustering	No	Yes	Every data instance which does not match with	Moderate	No
DecisionTree based Hybrid IDS	No	No	Every data instance which does not match with	Moderate	No
Fuzzy Logic based Hybrid IDS	No	No	Every data instance which does not match with	High	No
MAD-IDS	No	Yes	Every data instance which does not match with Signature and Anomaly	High	Yes
Anomaly Detection by Clustering in the	No	Yes	Every data instance which does not match with	High	No
K-Means Clustering and Naïve Baye	No	Yes	Every data instance which does not match	High	No
Snort-Based Hybrid IDS	No	No	Every data instance which does not match with	High	No
Proposed Mechanism	Yes	No	Much lesser compared to	Low	Yes

IJMRA

9. CONCLUSION AND FUTURE WORK

This paper presents a Light weight mechanism for novel DoS/DDoS attack detection and signature generation to represent those using MMDBMS. Condition based network connection records omission used for Novel attack Signature Generation increases the speed and accurate. This paper only concentrates on Resource consumption based attacks; further work should be done to detect novel Bandwidth consumption, Vulnerability and Packet spoofing based DoS/DDoS attacks.

10. REFERENCES

[1] Gang Xiong, Minxia Zhang, "A Novel Method of Outliers within Data Streams Based on Clustering Evolving Model for Detecting Intrusion Attacks of Unknown Type", 2010 International Conference on Multimedia Information Networking and Security

[2] Pedro García Teodoro, Pablo Muñoz Feldstedt, David Ruete Zúñiga, "Automatic Signature Generation for Network Services Through Selective Extraction of Anomalous Contents", 2010 Sixth Advanced International Conference on Telecommunications

[3] Jie Yang, Xin Chen, Xudong Xiang, Jianxiong Wan, "HIDS-DT: An Effective Hybrid Intrusion Detection System Based on Decision Tree", 2010 International Conference on Communications and Mobile Computing

[4] Bharanidharan Shanmugam, Norbik Bashah Idris, "Improved Intrusion Detection System using Fuzzy Logic for Detecting Anamoly and Misuse type of Attacks", 2009 International Conference of Soft Computing and Pattern Recognition

[5] Imen Brahmi, Sadok Ben Yahia, Pascal Poncelet, "MAD-IDS: Novel Intrusion Detection System Using Mobile Agents and Data Mining Approaches", Lecture Notes in Computer Science, 2010, Volume

6122/2010, 73-76, DOI: 10.1007/978-3-642-13601-6_9

[6] Feng Guo, Yingzhen Yang , Lian duan , "Anomaly Detection by Clustering in the Network", International Conference on Computational Intelligence and Software Engineering, 2009, ISBN: 978-1-4244-4507-323

[7] Z. Muda, W. Yassin, M.N. Sulaiman, N.I. Udzir, "Intrusion Detection based on K-Means Clustering and Naïve Bayes Classification", International Conference on Information Technology in Asia (CITA 11), IEEE 2011, ISBN: 978-1-61284-128-1

[8] Yu-Xin Ding, Min Xiao, Ai-Wu Liu, "Research And Implementation On Snort-Based Hybrid Intrusion Detection System", Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July IEEE 2009, DOI: 10.1109/ICMLC.2009.5212282

[9] Adetunmbi A. Olusola, Adeola S. Oladele, Daramola O.Abosede, "Analysis of KDD ‟ 99 Intrusion Detection Dataset for Selection of Relevance Features", Proceedings of the World Congress on Engineering and Computer Science 2010 Volume I, IEEE 2010

[10] Neveen I. Ghali, "Feature Selection for Effective Anomaly-Based Intrusion Detection", International zJournal of Computer Science and Network Security, VOL.9 No.3, March 2009, pp. 285-289

[11] H. Güneş Kayacık, A. Nur Zincir-Heywood, Malcolm I. Heywood, "Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets", Proceedings of the Third Annual Conference on Privacy, Security and Trust, October
2005, St. Andrews, Canada

[12] Wei Wang, Sylvain Gombault, Thomas Guyet, "Towards fast detecting intrusions: using key attributes of network traffic", The Third International Conference on Internet Monitoring

and Protection, 978-0-7695-3189-2/08, 2008 IEEE, pp. 86 – 91

[13] M. Soleimani, E. Khosrowshahi, M. Doroud, M. Damanafshan, A. Behzadi, M. Abbaspour, "RAAS: A Reliable Analyzer and Archiver for Snort Intrusion Detection System," ACM SAC, 2007

[14] I. Qualys, "The laws of vulnerabilities: Six axioms for understanding risk" http://www.qualys.com/docs/Laws- Report.pdf

[15] KDD99CUP Dataset, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[16] Vijay Katkar, Rejo Mathew, "One Pass Incremental Association Rule Detection Algorithm For Network Intrusion Detection System", International Journal of Engineering Science and Technology (IJEST), ISSN : 0975-5462 Vol. 3 No. 4 Apr 2011

[17] Kristopher Kendall, "A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems", http://www.ll.mit.edu/mission/communications/ist/files/k kendall_thesis.pdf