DISCLOSING MALICIOUS TRAFFIC FOR NETWORK SECURITY

Kamal Shah and Tanvi Kapdi
Thakur College of Engineering and Technology,
Mumbai University, Mumbai-400101, India

ABSTRACT
Network anomaly detection is a broad area of research. The use of entropy and distributions of traffic features has received a lot of attention in the research community. While previous work has demonstrated the benefits of using the entropy of different traffic distributions in isolation to detect generalized anomalies, there has been little effort in unconditionally understanding the detection power provided by entropy-based analysis of multiple traffic distribution used in affiliation with each other. In support to the previous work we are disclosing malicious traffic for network security using entropy based approach. To calculate entropy features like source and destination IP address, port numbers, packet size, connection time and the total number of packets flowing are considered. A framework has been presented for disclosing malicious traffic for network security by providing artificial traffic sets. The results suggests a metric for choosing traffic features for entropy based anomaly detection that are inherently complementary to one another i.e. selection of traffic distribution should be made judiciously and in particular should look beyond simple port and address based distributions. The major strength of the new scheme is that it can detect attacks with altered packet size but there are some drawbacks as well i.e. that the data set taken is relatively very small and hence does not cover all major attacks in the world. Thus in future more research can be done by inculcating as many attacks possible with a relatively large data set by keeping an eye on the performance of the system.

KEYWORDS: *Network anomaly detection, Intrusion detection, Network entropy, Relative network entropy.*

I. INTRODUCTION

Network systems and internet are so important in today's business. They also invite accidental or malicious attack on company's important data which can costs losses in terms of working hours, customers' trust and actual revenue. Network security [1] consists of the provisions and policies adopted by a network administrator to prevent and monitor unauthorized access, misuse, modification, or denial of a computer network and network-accessible resources. Network Security helps to understand that no single solution protects you from a variety of threats. You need multiple layers of security. If one fails, others still stand [2].Network security is accomplished through hardware and software. The software must be constantly updated and managed to protect you from emerging threats. A network security system usually consists of many components. Ideally, all components work together, which minimizes maintenance and improves security.
Network security components often include:

- Anti-virus and anti-spyware
- Firewall, to block unauthorized access to your network
- Intrusion prevention systems (IPS), to identify fast-spreading threats, such as zero-day or zero-hour attacks
- Virtual Private Networks (VPNs), to provide secure remote access

1.1 Objective

The development of computer networks has resulted in an important class of computers: network servers. The primary purpose of these machines is to provide services, including both computational and data services, to other computers on the network. Because of their service role, it is common for

servers to store many of an organization's most valuable and confidential information resources. They also are often deployed to provide a centralized capability for an entire organization, such as communication (electronic mail) or user authentication. Security breaches on a network server can result in the disclosure of critical information or the loss of a capability that can affect the entire organization. Therefore, securing network servers should be a significant part of your network and information security strategy. Many security problems can be avoided if servers and networks are appropriately configured. Default hardware and software configurations are typically set by vendors to emphasize features and functions more than security. Since vendors are not aware of your security needs, you must configure new servers to reflect your security requirements and reconfigure them as your requirements change.

Disclosing malicious traffic in the network through the concept of entropy of packet dynamics as in the proposed system will help to achieve effective threshold, compare the flow rate of other samples and help in optimizing the utilization of filtering and categorizing mechanism thereby reducing contradiction between improving detection rate and redundancy false alarm rate.

1.3 Motivation

Security is a very complex topic as everyone has a different aspect of the risk involved in their enterprise. The master key to build a secure network is to find out the key aspects of the enterprise which are critical and security sensitive and based on the policy enforced by the organization measures should be taken to make the network secure. Also the system can be broken down into smaller components which will make the task simpler to decide whether what is implied will conflict with your security policies or not.

1.4 Scope

The scope of the system is very vast as it can be implemented in any organization where security is a major concern. The techniques and the concepts used in providing a base to the system rely on well known theoretical assumptions and to proof their soundness we provide experimental results based on the simulations to illustrate their efficiency.

1.4Organization of the paper

The organization of the report is as follows:

Section 2 describes the problem definition of the existing system and how the proposed algorithm overcomes the existing algorithm.

Section 3 describes the design of the system which includes the proposed algorithm as well as hardware and software specification.

Section 4 describes the results and performance analysis

Section 5 describes the conclusion and future work of the system

II. PROBLEM DEFINITION

With the advent of new technology network security has become a major concern for commercial bodies as well as academic research. Network monitoring is a vital part in achieving a secure network. In network monitoring, a service provider is often attentive in seizing network attributes as heavy flows that use a link with a given capacity, flow size distributions, and the number of distinct flows. In network security, the interest lays in detecting known or unknown patterns of an attack.

A prevailing definition of a network anomaly reports an occurrence that diverges from the normal network behavior. However since there are no known models available for normal network behavior, it is strenuous to develop an anomaly detector in the toughest sense. Researchers have recently proposed the use of entropy based metrics for traffic analysis [6] and anomaly detection [7]. Entropy based anomaly detection techniques captures more fine grained traffic patterns as compared to normal volume based metrics. Many traffic features such as IP address, port number, flow size etc are considered as attributes is calculating entropy. However a very little work has been carried out in understanding the detection capabilities provided by a set of entropy metrics.

The goal of this dissertation is to provide a better understanding of entropy based traffic anomaly detection using different traffic features. We have considered two classes of traffic feature distribution i.e. flow header features and behavioral features. The flow header features include attributes like IP

addresses of the source and destination, port number, packet size, packet rate and connection time [6,8] . Whereas the behavioral features captures the structure of the end-host communication patterns. Our key analysis is

- The port and address distribution are very correlated, the anomalies detected by port and address distributions overlap significantly.
- To complement this study we evaluate several artificial anomaly scenarios using the collected flow data as background traffic.

While ports and addresses have been commonly suggested [9] as good candidates for entropy based anomaly detection, our results give us reason to question this rationale. Our results also suggest a natural metric for choosing traffic features for entropy based anomaly detection: select traffic distributions that are inherently complementary to one another and thus provide different views into the underlying traffic structure. Our results thus suggest that the selection of traffic distributions in entropy-based anomaly detection should be made more judiciously, and in particular we should look beyond simple port and address based distributions.

III. DESIGN AND METHODOLOGY

We propose a method for identifying abnormal traffic behaviour based on entropy. Main challenge is to distinguish between normal traffic and attack traffic since there is no major difference between normal and attack traffic. Our objective is to extract network features and make a model to identify the attack traffic. Entropies of network parameters are extracted from the traffic coming in the network. The entropy of network traffic is calculated in certain duration, and then sends its outputs directly for analysis. Experiments are performed on set a sample data where in the first sample case an attack free data is collected and then in the second sample case the data with attacks i.e. DOS , IP spoofing and half open connection. Experiment result demonstrates that our method works well with high detection rate of attack traffic and very less false alarm rate.

Entropy-based analysis of traffic feature distributions is used for anomaly detection. Entropy-based metrics are appealing since they provide more fine-grained insights into traffic structure than traditional traffic volume analysis. Features extracted for the detection of anomaly based attack are as follows.

• Entropy of source IP address and port number.
• Entropy of destination IP address and port number.
• Entropy of packet type.
• Occurrence rate of packet type.
• Number of packets per unit time.
• Entropy of packet size.

3.1 Proposed Algorithm

Step1- First of all capture the packets and add packets in the current queue L.
Step2- Now compute the current queue length.
Step3-Select the desired features required for calculations i.e. ip address of source and destination, port number of source and destination, packet size, packet rate and connection time.
Step4-Calculate the entropy.

$H(X) = \sum P(xi) \log P(xi)$ Here,

X for a fixed time window w is, $P(xi) = mi/m$,

Where mi is the frequency or number of times we observe X taking the value xi as

$m = mi$.

$H(X) = -(mi/m) \log (mi/m)$.

H (X) = Entropy

If we want to calculate probability of any source (destination) address then,

mi = number of packets with xi as source (Destination) address

M = total number of packets

P(xi) =Number of packets with xi as source/destination address/M

Here total number of packets is the number of packets seen for a time window T.

Similarly we can calculate probability for each source (destination) port as

P(xi)=Number of packets with xias source (destination) address/M
Normalized entropy calculates the over all probability distribution in the captured flow for the time window T.
Normalized entropy = (H/log(n)) Where n is the number of distinct xivalues in the given time window.
Step 5- Determine the threshold on the basis of the maximum and minimum deviations calculated for a number of times.
Step 6- If the result exceeds the threshold an attack is found.

3.2 System specification

The proposed algorithm is being implemented on the following system specification.
Hardware

- ✓ Processor - AMD A8 4500M APU with Radeon (tm) HD Graphics 1 GHz
- ✓ RAM – 4.00 GB
- ✓ Harddisk – 1 TB

Software

- ✓ Operating system - Windows 8 Single language
- ✓ Programming language – C#
- ✓ Development kit – Visual Studio 2010

IV. RESULTS

The result of the proposed scheme hereby states the variation in the flow of traffic and thereby detection of the attack with reduced false alarm rate. In the below figure we have executed the system for a time interval of 10 minutes to collect a sample data set. Hence for that first of all we will start the router and internet and allow the incoming flow of packets into our network. Once the allotted period of time i.e 10 minutes is completed then we have our sample data set ready from where we can insert the values of start time and end time i.e. T1 and T2 respectively.
Here X-axis shows time in seconds and Y-axis shows the value of entropy at a particular time. The below graph is divided into four sections which shows the traffic rate, packet size, IP address and connection time entropy which can be regarded as our sample data set.

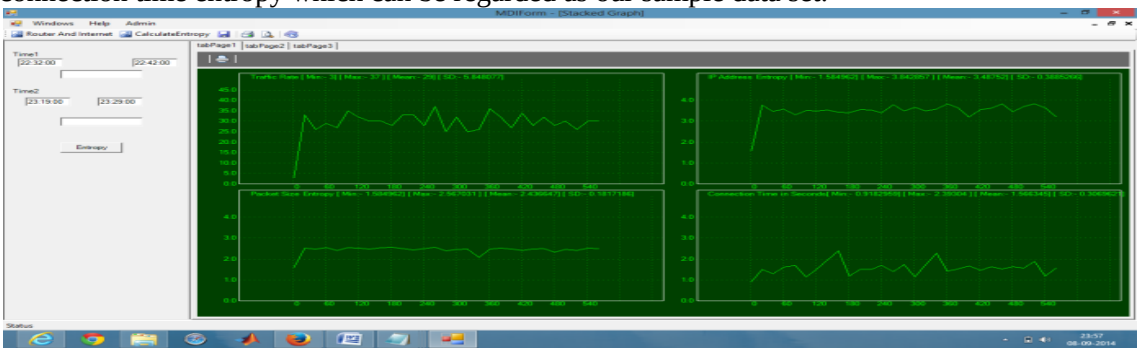

Figure 1.1: Resulting snapshot in proposed scheme with normal connection

Now in order to detect an attack again we have to execute the system for 10 minutes and calculate the entropy which will be our base class and meanwhile introduce the attacks and after the completion of 10 minutes again calculate the entropy and then compare both the values as shown in below figure.

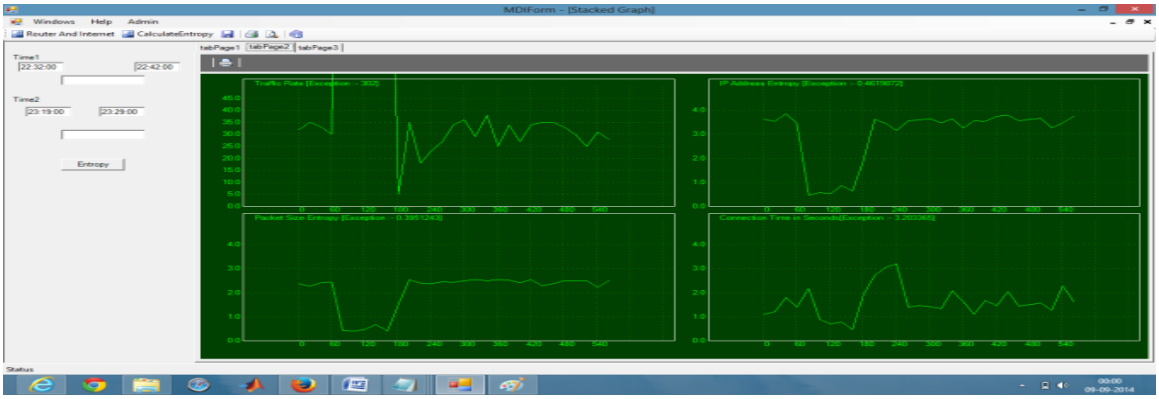

Figure 1.2: Resulting snapshot of proposed scheme after attack

Comparing the graphs of normal connection V/S attack we found the value of entropy fluctuates and goes beyond the threshold and hence anomaly is detected.

V. CONCLUSION AND FUTURE WORK

The prevailing definition of network anomaly reports an occurrence that diverges from the normal behavior. However there are no known models available for normal network behavior. The proposed scheme introduces an alternative technique by using features selection like IP address of source and destination, port number of source and destination, packet size, packet rate and connection time to detect an attack. In addition it also reduces the false alarm rate thereby increasing the true positive rate.

The major strength of the new scheme is that it can detect attacks that come with altered packet size. We are not claiming that our method is superior to all other methods. There are some drawbacks as well in our approach. Experimental sample data set which we have taken is relatively small and hence this data-set won't cover all the attacks in the world. There are a lot of new attacking methodologies introduced by the attackers nowadays. All the methods are not available to the public due to security reasons, so it is difficult to study about the attack schemes and prevention mechanisms. Still a lot amount of data is available for academic purposes.

After executing the proposed system for number of times the experimental results shows that:

1. Memory utilization of the proposed scheme is less as compared to the normal memory utilization of the machine.
2. False positives i.e. the number of records that were incorrectly identified as attack is reduced.
3. True positives i.e the number of malicious records are correctly identified.
4. Relatively small data set is used to detect an anomaly in the network.
5. CPU utilization is more which makes the machine work slow.

5.1 Future Work

In future this system can be extended by using large data sample set and by incorporating as many major attacks possible also more accurate threshold value can be found which can help in making the system more precise. Research can be carried out in order to extend on cloud platform thereby making the system as a generic service provider.

REFERENCES

[1] Simmonds, A; Sandilands, P; van Ekert, L (2004). "An Ontology for Network Security Attacks". *Lecture Notes in Computer **Science***. Lecture Notes in Computer Science **3285**: 317–323. doi:10.1007/978-3-540-30176-9_41. ISBN 978-3-540-23659-7.

[2]http://www.cisco.com/cisco/web/solutions/small_business/resource_center/articles/secure_my_business/what_is_network_security/index.html

[3] A Role-Based Trusted Network Provides Pervasive Security and Compliance - interview with Jayshree Ullal, senior VP of Cisco

[4] Dave Dittrich, *Network monitoring/Intrusion Detection Systems (IDS)*, University of Washington.
[5] C. Cranor, T. Johnson, O. Spatscheck, and V. Shkapenyuk. Gigascope: A stream database for network applications. In *Proc. of the 2003 SIGMOD Conf.*, June 2003.
[6] K. Xu, Z. Zhang, and S. Bhattacharyya. Profiling internet backbone traffic: Behavior models and applications. In Proc. of ACM SIGCOMM, 2005.
[7] D. Brauckhoff, B. Tellenbach, A. Wagner, A. Lakhina, and M. May. Impact of traffic sampling on anomaly detection metrics. In Proc. of ACM/USENIX IMC, 2006.
[8] L. Feinstein, D. Schnackenberg, R. Balupari, and D. Kindred. Statistical Approaches to DDoS Attack Detection and Response. In Proc. of DARPA Information Survivability Conference and Exposition, 2003
[9] International Journal of Computer Applications (0975 –8887) Volume 62–No.15, January 2013 Entropy based Anomaly Detection System to Prevent DDoS Attacks in Cloud A.S.Syed Navaz, V.Sangeetha C.Prabhadevi
[10] . Shannon. A mathematical theory of communication.Bell System Technical Journal, 27:379–423,July 1948.
[11] Dorothy E. Denning. An Intrusion-Detection Model. IEEE Transactions on Software Engineering, 13:222-232, 1987.
[12] N. N. Wu, "Audit data analysis and mining", Ph.D.Thesis, George Mason University, USA, 2001.

AUTHOR

Kamal Shah working as Dean, research and development in Thakur College of Engineering and Technology and having an experience of more than 15 years in field of teaching. Her area of specialization is network security and image processing.

Tanvi Kapdi obtained her degree of Bachelor in engineering from Sardar Patel University in 2011. She is currently pursuing her Masters in engineering in Information Technology from Mumbai University under the guidance of Dr. Kamal Shah, dean R&D, Thakur college of engineering and Technology. Her research is centred on network security and malicious traffic.