International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 08 | Aug-2016 www.irjet.net p-ISSN: 2395-0072

Network Traffic Visualization Framework for Threat Prediction and Detection

Athira Rajeev[1], Sheema Madhusudhanan[2]

[1]MTech Cyber Security, Dept. of CSE, SNGCE, Kadayiruppu, Kerala, India.
[2]Asst. Prof. MTech Cyber Security, Dept. of CSE, SNGCE, Kadayiruppu, Kerala, India

---

Abstract- *Computer network plays a very important role in information and which also suffers from all kinds of illegal access and attacks. So network forensic analysts may need to check the entire network traffic data, which takes long time to complete the evaluation. Visualization techniques can convert abstract data into visual sensitive graphics and here considering a matrix-based visualization system for visualized forensic analysis and prediction and detection of attacks on unintelligible traffic datasets. We describe the forensic process of this system, on the datasets in VAST Challenge 2013*.

Keywords - Cyber security, information visualization, traffic forensics, VAST Challenge, prevention, detection

1. INTRODUCTION

We are in the midst of a networking revolution. Increasing network sizes has been a concomitant increase in the collection of network measurement data. Understanding this data is of crucial importance as we move to a modern, information-rich society. Fortunately, there were different tools available for analysing network data but one problem is that the data volumes. So providing network security is crucial step in network communication. Traditional network analysis software and graphs cannot cope with the size of today's networks and their data collection capabilities.

Network forensics is not another term for network security and which is an extended phase of network security for forensic analysis. For this security analysis data's are collected from security products like firewalls and intrusion detection systems. Finally these results are used to detect the attacks. However, there may be certain crimes which do not breach network security policies but may be legally prosecutable. These crimes can be handled only by network forensics. Information visualization is a good technique to investigate the crimes in network traffic and helpful for network forensics. Network forensics requires analysts to efficiently reason about various attack phenomena from massive data. Visualization techniques can convert abstract data into visual sensitive graphics; thus, forensic officers can extract useful information quickly. Shi [1] proposed a matrix-based visualization system for visualized forensic analysis on unintelligible traffic datasets.

Information visualization has become an integral part of network security tools, as it is well said that, "a picture depicts thousands words" [2]. Human brain responds rapidly to a picture than text because the eye and the visual cortex of the brain form a massively parallel processor that provides the highest bandwidth channel into human cognitive centres'. Numerous commercial, private, public and social web applications generate voluminous network traffic over Internet. The Internet users are observing bitter experience as computer machines; websites and network resources are in trap under different network attacks.

There were different network visualization techniques are available for mapping the IPs and ports into pixilated format and applying color coding, and are PortVis, IDS RainStorm and IPMatrix.

Here we were presenting a matrix view in addition with timeline view. When combined together for traffic forensics, the entropy and the matrix can obtain mutual help from each other. Finally we predict and detect the attack in the network traffic by analyzing the matrix view.

The scheme of this paper is as follows: Section 2 is about the background and related work. System design is explained in next Section 3. Section 4 is about the Results and Discussion of the work done and finally conclusion and future work is in section 5.

2. BACKGROUND AND RELATED WORKS

In this section, we are discussing data preprocessing and the work of many key researchers in different domains of network visualization and security.

2.1 Data Preprocessing

Implementation details will be influenced by available data sources and the actual network architecture. Here the data used are NetFlow logs offered by MC3 (VAST Challenge 2013 Mini Challenge 3) [3]. The dataset recorded behaviors of the Big Marketing enterprise, which consists of three different branches within each of them possessing around 400 workstations and many servers.

Network flow data captures, to the extent feasible, the traffic moving across the network. Big Marketing captures network flow at the firewall, so transactions that go from Big Marketing to the internet, or come from the internet into Big Marketing, are captured.

A network flow is an abstraction of a sequence of packets between two terminals. A typical NetFlow is identified by the following unique keys: the timestamp, the source IP address, the destination IP address, the source port, the destination port, the protocol type, the type of service, and the amount of traffic, etc. Characteristics of network activities are often recorded in these fields. The data after integration and processing are neat and orderly, which is in favour of the system performance.

2.2 Network Traffic Forensics

Network forensics can be performed as a standalone investigation or alongside a computer forensics analysis (where it is often used to reveal links between digital devices or reconstruct how a crime was committed). Network forensics systems can be one of two kinds:

- "Catch-it-as-you-can" systems, in which all packets passing through certain traffic point are captured and written to storage with analysis being done subsequently in batch mode. This approach requires large amounts of storage.
- "Stop, look and listen" systems, in which each packet is analyzed in a rudimentary way in memory and only certain information saved for future analysis. This approach requires less storage but may require a faster processor to keep up with incoming traffic.

2.3 Network Traffic Visualization

A primary goal of data visualization is to communicate information clearly and efficiently. There for normally use statistical graphics, plots, information graphic charts etc are selected for visualization and numerical data may be encoded using dots, lines, or bars. Visualization technique may visualize the network traffic data into graphical one, which makes the network security analysts to easily analyze the data.

Information visualization [1] involves the use of computer supported visual representations for abstract data to amplify cognition by taking advantage of human perceptual capabilities.

Card et al. [4] propose six major ways in which visualizations can amplify cognition.

- By increasing the memory and processing resources available to the users.
- By reducing the search for information,
- By using visual representations to enhance the detection of patterns,
- By enabling perceptual inference operations,
- By using perceptual attention mechanisms for monitoring,
- By encoding information in a manipulable medium.

Other different methods used for network security analysis is node connection diagram, parallel coordinates because of its advantage in handling high-dimensional data [5]. Three-dimensional effect has been also introduced to display network activities [6]. Inoue at el [7] proposed novel real time 3D visualization for Darknet monitoring-based alert system. This also shows IPs as 3D visualization.

Lakkaraju et al. [8] developed 'NVisionIP', a visualizing tool for entire Class B network node visualization on a single screen. It facilitates drill down and collects information in more details about a particular host in a network.

Abdullah et al. [9] proposed IDS Rainstorm to overcome the problem of text-based evaluation of log files. IDS Rainstorm works offline and displays scattered plot and parallel coordinate plot visualization.

Allen and McLachlan [10] developed 'NAV' Network Analysis Visualization in JAVA to visualize the high-level network events, to investigate the malicious activities for

large network, while operating on network and transport layer visualizations.

3. SYSTEM DESIGN

Here, we will present the system design of matrix based visualization and attack prevention and detection. The visualization system presents matrix as well as timeline view and predict and detect the attack in the network traffic by using how active it is and using set theory respectively.

3.1 Matrix View

Matrix view displays a comparative interface of IPs and Ports. Analyzing both IPs and ports are helpful in network traffic forensic analysis. There are four matrices designed for source IPs, destination IPs, source ports, and destination ports, respectively. In IP matrices, thousands of dots on behalf of hosts and servers are arranged as four groups.

All IP addresses exist are automatically grouped according to IP fields with artificial settings. Where Shi et al [1] proposed a matrix-based visualization system for visualized forensic analysis on unintelligible traffic datasets. The advantage is that the network architecture will display intuitively. That is all internal IPs are grouped into different parts according to the last field, and IPs in the same group are arranged in order. When the ports are visualized as 256*256 matrix of 65536 ports.

IP:

- Internal subnets: Each of them is composed of hundreds of hosts and several subnet servers located in the same internal network.
- Others: Here is external IPs that appeared frequently and are included in the registered list of the company. It is worthy of serious analysis on their communication records with each other and with internal network.

Port:

- The user-customized ports: They are the most popular used
- The well-known ports: The port numbers from 0 to 1023 are the well known ports, and multiple applications need to use them.
- The registered ports: The range of port numbers is from 1024 to 49151.
- The uncommon ports: They are all the other dynamic or private ports that are not declared in previous groups.

In the matrix view the IPs and ports are mapped into dots. Then this pixel is filled with colors from warm tones to cool tones. Colors represented in red-green-yellow. The advantages of using dots representation avoids overlapping of data, because we considering pixilated format.

3.2 Timeline View

The timeline view provides clearest and easiest method for time-series analysis. It represents several statistical time series, such as network traffic and distributions of IPs over time.

As every timeline consist of two dimensions: and are represented in X-axis and Y-axis. The X-axis represents the time attribute as When, and the Y -axis represents one of the context attributes as how many or how serious of the events or activities, i.e which simply denotes the entropy value of ip address and ports of corresponding time interval.

In our interface, four attributes of network status data are grouped in the Timeline view. Entropies of four metrics present the disorder of active individuals in NetFlow logs. They are calculated on categorical datasets, including

- source IPs
- destination IPs
- source ports
- destination ports

The number of different IPs and ports appeared in network activities are essential to detecting anomalies.

3.2.1 Entropy Calculation

In the computer security as well as cyber security field, security experts or investigators or scholars made different attempts to found or measure the network activities having any disorders and discovering new types of attacks by calculating diverse forms of entropies such as Shannon entropy [11], conditional entropy and joint entropy.

Here we were using Shannon entropy, the classic information entropy, to calculate on source IPs, destination IPs, source ports, and destination ports separately.

Take the source ip data as an example. Let the source ips in the NetFlow to be a discrete random variable, known as X

with variables $x_1 \ldots x_n$ representing the possible realizations. In the time span t_j, j = 1 to ∞, the amount of each source ip occurred is n_i, i = 1 to n, with probability of p_i, i = 1 to n. Thus, each source ip is calculated by the following equation:

$$p_i = \frac{n_i}{\sum_1^n n_i}$$

The entropy value is calculated by using following equation:

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i$$

In order to make the result of the entropy analysis easier to interpret, we scaled the entropy to the interval [0, 1], with the standardized entropy that is defined as relative uncertainty (RU):

$$RU = \frac{H(X)}{H_{max}(x)}$$

3.3 Attack Prediction and Detection

Presents an effective threat prediction and detection system, which detects the suspected list of ip address. The prediction and detection done via based on how active the ip address in the network traffic and another one is based on a set theory algorithm.

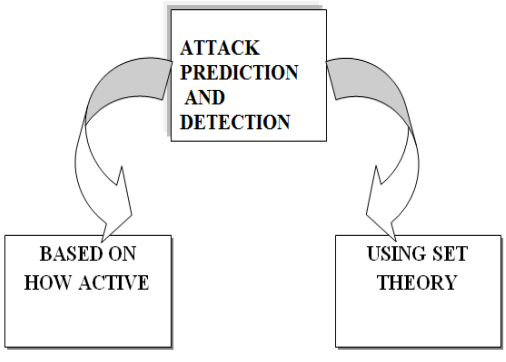

Fig -1: Overview of attack prediction and detection

- Based on how active it is

Here which explains the prediction is done based on how active it is and its working. Basically visualization of abstract data into graphical representation is consists of six different ways and are explained earlier this work is based on "by reducing the search of information" and which is basically depends on how active the ip address in the traffic.

The prediction via how active is depending on a threshold value set by the network administrator. The network administrator, who knows the network traffic. So they actually set a threshold data usage in the network for the internet connection. If the data usage is more than the threshold value set by the admin, then those users ip address automatically detect in the suspected list.

Considering the prediction in the case of source ips, single user ip is detected. But case of destination ips entire network is added in the suspected list. Because if we block a single ip those exceed the threshold limit get blocked. May be it is an attacker and they found that they were blocked, and then there is chance to attack via another ip address from the same network. There for in the case of entire network is blocked.

- Using Set theory

Presents another way of prediction is using set theory. Using these method blocks the unauthorized access to network.

Algorithm

1. Begin
2. A representing the set of ip address of legitimate users from inside of the network
3. B representing the set of authorized users from extranet
4. N representing Netflow
5. Malicious traffic = (A∪B)∩N
6. End

Here the algorithm which effectively detect the malicious user who tries to intrude the network. The above algorithm proposed by amit et al [2]. Here we consider a set of ip address of legitimate users those inside the particular network and a set of authorized users from extranet. Another data taken is the actual netflow of the network. Using this data apply the set theory. First taken union of first two sets and then taken the intersection with netflow.

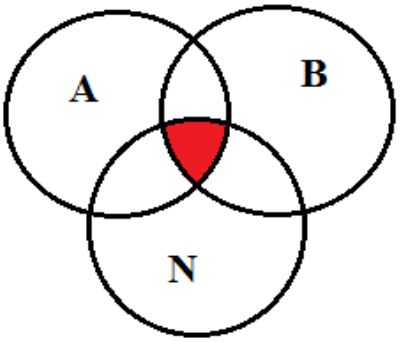

Fig -2: Venn diagram for detection of attack

Figure 2 shows the Venn diagram for detection of attacks in network. The darken portion denote the malicious users based on the equation of (A∪B) ∩N.

4. RESULTS AND DISCUSSION

4.1 Experimental Setup

It is often the case that a network analyst is not interested so much in what occurred during a particular time unit but rather what changed across a range of time units. Therefore, ipmatrix offers a feature that allows analysts to select any arbitrary set of time units and see on the main visualization not a depiction of the actual values at each ip address and port but rather a depiction of the variance of the values at each ip address and port.

The experiment was conducted by using a dataset of mini vast challenge3 2013. During the visualization of ip address and port into matrix view is reshaped into small size for the big data visualization.

4.2 Experimental Results

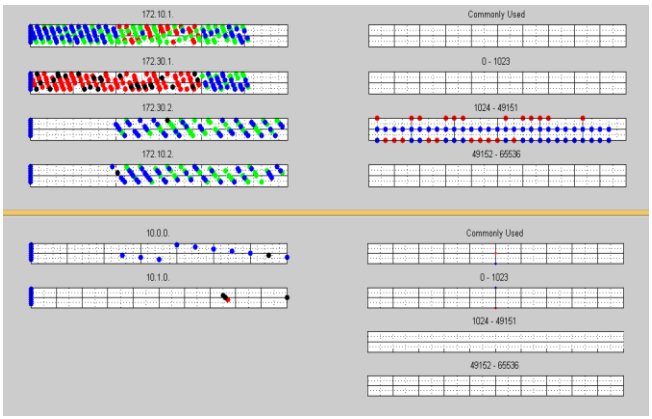

Fig -3: Matrix view of IPs and Port

Figure 3 shows the matrix view of source and destination ip address, source and destination port. IP address and port are mapped into dots and filled with colors. The color coding is applied from warm tones to cold tones. The ip addresses are visualized into four different sections with grouped ip address. Where ports are visualized based on how they active in the traffic and which is also filled with colors.

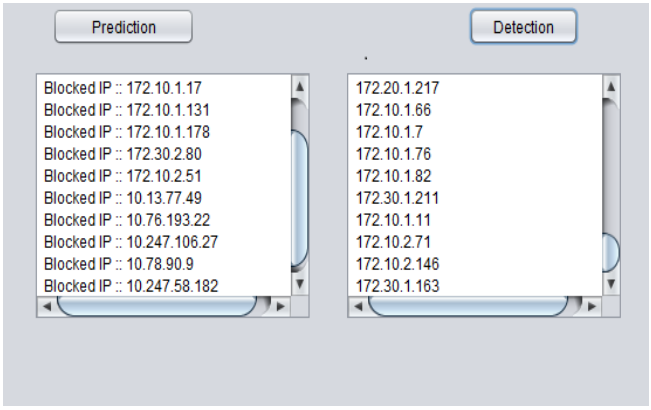

Fig -:4 Attack prediction and detection

Figure 4 shows the attack prediction and detection in the network traffic netflow. Attack is predicted based on how active in the network and detection is based on using set theory. ie considering a set of ip address of legitimate users those inside the particular network and a set of authorized users from extranet. Another data taken is the actual netflow of the network. Using this data apply the set theory. First taken union of first two sets and then taken the intersection with netflow.

4.2 Problems of the system

- Usability: Usability refers to the ease with which users can execute the desired tasks in our application. One usability issue about our system is that ordinary users may don't know how to get start, because it is only suitable for expert users who can acclimate to the variety of settings and interactions.
- Robustness: The work is based on time series and is the ability of the system to cope with errors during execution and cope with erroneous input

5. CONCLUSION AND FUTURE WORK

In conclusion, we present a matrix based visualization scheme. This method consists of four matrices of source and destination IPs and ports. The timelines of four kinds of entropies are also visualizing, ie both views are combined together to obtain mutual help from each other. Also done prediction and detection of attacks in the network, thus enable secure networks. When a threshold limit is setting for communication between users in the network, and which exceeds, those users are added to the suspected list. Using set theory algorithm detects the unauthorized users in the network. The process of solving the work in Mini Challenge 3 of VAST Challenge. Thus as a part of future work we focus on capture the packet in real time and improve usability and robustness of the system.

REFERENCES

1. Ronghua Shi, Mengjie Yang, Ying Zhao, Fangfang Zhou, Wei Huang, and Sheng Zhang "A Matrix-Based Visualization System for Network Traffic Forensics," IEEE Systems Journal, 2015, PP. 1-11
2. Amit Kumar Bhardwaj and Maninder Singh,"Data mining-based integrated network traffic visualization framework for threat detection," Neural Comput. Applic. Springer, 2015, pp. 117-130
3. VAST Challenge 2013. [Online]. Available:http://www.vacommunity.org/VAST+Chall-enge+2013
4. S. K. Card, J. D. Mackinlay, and B. Shneiderman, Eds., "Readings in Information Visualization: Using Vision to Think", San Mateo, CA, USA: Morgan Kaufmann, 1999.
5. H. Choi, H. Lee, and H. Kim, "Fast detection and visualization of network attacks on parallel coordinates," *Comput. Security*, vol. 28, no. 5, pp. 276–288, Jul. 2009.
6. T. Nunnally *et al.*, "P3D: A parallel 3D coordinate visualization for advanced network scans," in *Proc. IEEE ICC*, 2013, pp. 2052–2057.
7. D. Inoue, M. Eto, K. Suzuki, M. Suzuki, and K. Nakao, "DAEDALUSVIZ: Novel real-time 3D visualization for darknet monitoring-based alert system," in *Proc. 9th Int. Symp. Vis. Cyber Security*, 2012, pp. 72–79.
8. K. Lakkaraju, W. Yurcik, and A. J. Lee, "NVisionIP: NetFlow visualizations of system state for security situational awareness," in *Proc. ACM Workshop Vis. Data Mining Comput. Security*, 2004, pp. 65–72.
9. K. Abdullah, C. P. Lee, G. Contin, J. A. Copeland, and J. Stasko, "IDS RainStorm: Visualizing IDS Alarms," in Proc. VizSEC, 2005, pp. 1–10.
10. Allen, Meghan, and Peter McLachlan. "NAV network analysis visualization."University of British Columbia,[Online, 29 May 2009] (2009).
11. G. Nychis, V. Sekar, D. G. Andersen, H. Kim, and H. Zhang, "An empirical evaluation of entropy-based traffic anomaly detection," in *Proc. 8th ACM SIGCOMM Conf. Internet Meas.*, 2008, pp. 151–156