DALHOUSIE
UNIVERSITY

Faculty of Computer Science

Summary Report on

"ENTVis: A Visual Analytic Tool for Entropy-Based Network Traffic Anomaly Detection"

CSCI 6708 – Advanced Topics in Network Security

Name: Arka Ghosh
Banner ID: B00911033

April 2022

Reference Paper:

F. Zhou, W. Huang, Y. Zhao, Y. Shi, X. Liang and X. Fan, "ENTVis: A Visual Analytic Tool for Entropy-Based Network Traffic Anomaly Detection," in *IEEE Computer Graphics and Applications*, vol. 35, no. 6, pp. 42-50, Nov.-Dec. 2015, doi: 10.1109/MCG.2015.97.

Introduction

Network anomalies refer to the disruption of the network's usual operation for a sudden and brief period of time [1]. Some anomalies are purposefully generated with malicious intents by intruders such as Distributed Denial-of-Service (DDoS) Attack in an IP network, whereas others anomalies might occur due to a faulty network device or overpass palling in a network path. In network communication, anomaly detection plays a significant part not only for the network administration but also for monitoring the network security issues [2] as it paves the way to discover any novel attacks, misconfiguration or failures in the network [3]. The fundamental idea behind network anomaly detection is specify the behavior of the network or system, which is then compared to the pre-defined behavior [4]. Based on the outcome, further investigation might take place by forwarding a trigger alarm to the network management system. In the network security field, network traffic analysis and anomaly detection have become a major topic within the research community and several approaches have been suggested to monitor traffic and identify network anomalies such as DDoS attacks, worm outbreaks, and device failure. All of these occurrences, however, are not always simple to identify from observed data, especially if the monitored network transports a huge quantity of traffic, drowning hostile activity [5]. Within this broad area of research, anomaly detection approaches based on entropy is intriguing because it provides more fine-grained insights than typical traffic volume analysis [6]. The ability to use different entropy-based metrics in combination with one another may improve the rapid detection of various vulnerabilities and attacks on a network [7]. In this regard, Fangfang et al. [7] have introduced a visual analytic tool called ENTVis that helps users comprehend entropy-based traffic metrics and detect traffic anomalies in a more accurate manner. A detailed summary of the problems being tackled with this tool and the proposed solution to the problems have been presented in this summary paper.

Problem Statement

Internet has grown at a breakneck pace for the past two decades, providing us with both convenient services and major network security concerns [8]. With the growing size and complexity of modern networking systems, as well as the change in cyber-attacks from a single user or application to the entire system, prompt detection of anomalies is required to launch a timely security check [9]. However, traditional security approaches such as firewalls or antivirus are, however, not always capable of detecting cyber security issues because of the sophistication of these attacks. Entropy based anomaly detection has been a hot topic in recent times. In information and network theory, entropy is a key notion that assesses the degree of ambiguity or impurity in a set of data points. The basic approach for entropy-based traffic anomaly detection is the time-series analysis [7], which considers the distribution of the studied traffic characteristic and detects an anomaly when the related entropy varies significantly [10]. Such kind of approach can be used to determine not just individual traffic distributions, but also changes in distribution over time. However, the application of entropy based approaches in anomaly detection has drawbacks, since short-term or distributed attacks are not readily identified as uncertainty is either minimal or distributed [11]. As entropy theory is an abstract mathematical matric of random variables, it becomes difficult for users to comprehend. Moreover, traffic analysis based on entropy are unable to offer precise information on traffic distribution, and if the timeline group has a large number of long time series, the visual congestion will make it difficult for viewers to see the underlying aberrant points. The traditional entropy based anomaly detection methods is quite incapable of providing intuitive and detailed information that may help the security analysts or network administrators to detect a legitimate attack on a network, severity of the attack or the measures that need to be taken to protect the network from similar attacks in the future. To aid in this process, Fangfang et al. has proposed a visual tool called ENTVis that can be used to help users perform entropy-based anomaly detection by visualizing network traffic entropy measures from numerous perspectives.

Proposed Solution

The interface of ENTVis (Appendix A) supports a cohesive visual examination from different perspective by providing three coordinated views and extensive interactivity: the timeline group view is used for general and drilldown scenario evaluations, the Radviz view is used to group comparable anomalies together across time, and the matrix view is used for deciphering traffic patterns and spotting irregularities in more depth. The proposed approach utilizes the IP entropy and port entropy where the IP entropy measures the randomization of the hosts engaging in the network activity and the port entropy detects the unpredictability of the ports participating in the network activity. If there are 100 different ports being visited within a specific time span and the visited amount is denoted by n_i, the probability for each port p_i is calculated as follows:

$$p_i = \frac{n_i}{\sum_1^{100} n_i} \text{ (i = 1…100)}$$

The destination port entropy is calculated at the same time using:

$$H(A) = \sum_{i=1}^{100} p_i log p_i$$

To compare various entropy-based traffic measures, H(A) is being standardized in the next step where Relative Uncertainty (RU) is used to scale the value of H(A) to the interval between 0 and 1. In the following equation, H_{max}(A) represents the destination port’s maximum entropy and log(100) is used to calculate the time span as 100 active destination ports are observed within this time span.

$$RU(A) = \frac{H(A)}{H_{\text{max}}(\text{A})} = \frac{H(A)}{\log(100)}$$

A detailed analysis of the three coordinated views of the ENTVis interface have been discussed below:

- **Timeline View:** The timeline view at the top of the interface combines a set of timelines where each timeline depicts a traffic feature’s temporal development tendency. There are two time span options in the timeline view: 5 minutes option for short-term analysis and 60 minutes for medium and long-term analysis. These two options allow the user to do a temporal drilldown where user can examine a selection of data by clicking on a ping in the timeline bar. The appropriated points are highlighted in the Radviz view once the pins have been selected in the timeline view.
- **Radviz View:** The Radviz view extends entropy-based traffic analysis from temporal to visual clustering space. Radviz is a basically radial visual clustering approach for mapping data from a multidimensional space to a planar image [12]. The data dimensions in Radviz are evenly allocated places on a circle's circumference called Dimension Anchors (DAs). As the time series of traffic in the timeline view are multidimensional data and each time period is a record with a variety of traffic characteristics, time spans with similar traffic

characteristics will congregate in the Radviz through the six spring tensions from the six DAs (Appendix B).

- **Matrix View:** The matrix view is intended to display all source/destination IPs/ports in a single interface, allowing the user to view the whole traffic distribution statistics for IPs and ports. As arranging all IPv4 addresses in a limited matrix space is difficult, a customized IP layout have been proposed by the authors according to the network architecture (Appendix C). Each port is represented as a grid in the port matrix view where the ports are divided into groups based on its port numbers. The detailed traffic distributions of any time span indicated in the timeline or Radviz view can be seen by the users in the Matrix view. If an IP address or a port is chosen, descriptive text will appear in the control panel's state box. In addition, the matrix view offers a traffic filter based on the specified IP or port.

The authors also evaluated the performance of the visualization tool based on three case studies: overall network analysis, DDoS attack detection and port scan analysis. In the first scenario, traffic patterns were observed over the course of a week for which the time limit was set to 60 minutes. ENTVis is used to analyze the patterns of a DDoS attack, locate the implicated hosts and estimate the damage in the second case while the third case analyzes the abnormal events in the cluster by zooming in on the time period, going from 60 minutes to 5 minutes intervals.

Comment and Discussion

The authors proposed and implemented a visualization tool in [7] named ENTVis to expand entropy-based traffic analysis from temporal to visual clustering space. As a result, it may easily identify linear time periods during which comparable network traffic attributes occurred. As entropy is a useful metric for quantifying a system's chaos, it is widely employed to aid data exploration and enhance visual mapping design to detect anomalies in a network system. Such kind of entropy based approach works really well for DDoS attacks and Port Scan to detect anomaly as it can provide a detailed visualization to the network analysts to figure out how random particular attributes in network packet headers are distributed. Though the proposed approach comes with some limitations mentioned in the paper, many entropy-based sectors of data analysis, such as identifying human movement aspects in urban computing and examining people's communication patterns on social media, can be benefitted from this analytical approach.

References

[1] T. Ahmed, B. Oreshkin and M. Coates, "Machine Learning Approaches to Network Anomaly Detection", *Usenix.org*. [Online]. Available: https://www.usenix.org/legacy/event/sysml07/tech/full_papers/ahmed/ahmed_html/sysml07CR_07.html. [Accessed: 05- Apr- 2022].

[2] Z. Tang, X. Zeng and Y. Sheng, “Entropy-based feature extraction algorithm for encrypted and non-encrypted compressed traffic classification”, *International Journal of ICIC*, *15*(3), 845-860.

[3] F. Iglesias and T. Zseby, "Analysis of network traffic features for anomaly detection", *Machine Learning*, vol. 101, no. 1-3, pp. 59-84, 2014. Available: 10.1007/s10994-014-5473-9.

[4] R. Abdulhammed, M. Faezipour, A. Abuzneid and A. AbuMallouh, "Deep and Machine Learning Approaches for Anomaly-Based Intrusion Detection of Imbalanced Network Traffic," in *IEEE Sensors Letters*, vol. 3, no. 1, pp. 1-4, Jan. 2019, Art no. 7101404, doi: 10.1109/LSENS.2018.2879990.

[5] Y. Ruo-Yu and Z. Qing-Hua, "Multi-scale Entropy Based Traffic Analysis and Anomaly Detection," *2008 Eighth International Conference on Intelligent Systems Design and Applications*, 2008, pp. 151-157, doi: 10.1109/ISDA.2008.167.

[6] G. Nychis, V. Sekar, D. G. Andersen, H. Kim, & H. Zhang, “An empirical evaluation of entropy-based traffic anomaly detection”. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement* (pp. 151-156).

[7] F. Zhou, W. Huang, Y. Zhao, Y. Shi, X. Liang and X. Fan, "ENTVis: A Visual Analytic Tool for Entropy-Based Network Traffic Anomaly Detection," in *IEEE Computer Graphics and Applications*, vol. 35, no. 6, pp. 42-50, Nov.-Dec. 2015, doi: 10.1109/MCG.2015.97.

[8] Y. B. Luo, B. S. Wang, Y. P. Sun, B. F. Zhang and X. M. Chen, "FL-LPVG: An approach for anomaly detection based on flow-level limited penetrable visibility graph," *2013 International Conference on Information and Network Security (ICINS 2013)*, 2013, pp. 1-7, doi: 10.1049/cp.2013.2470.

[9] G. Tian *et al.*, "CEFF: An efficient approach for traffic anomaly detection and classification," *2017 IEEE Symposium on Computers and Communications (ISCC)*, 2017, pp. 779-786, doi: 10.1109/ISCC.2017.8024622.

[10] C. Callegari, M. Pagano, S. Giordano and F. Berizzi, "CUSUM-based and entropy-based network anomaly detection: An experimental comparison," *2017 8th International Conference on the Network of the Future (NOF)*, 2017, pp. 132-134, doi: 10.1109/NOF.2017.8251234.

[11] D. Liu, C. -H. Lung, N. Seddigh and B. Nandy, "Entropy-based robust PCA for communication network anomaly detection," *2014 IEEE/CIC International Conference on Communications in China (ICCC)*, 2014, pp. 171-175, doi: 10.1109/ICCChina.2014.7008266.

[12] F. Zhou, Wei Huang, Juncai Li, Yezi Huang, Yang Shi and Ying Zhao, "Extending Dimensions in Radviz based on mean shift," *2015 IEEE Pacific Visualization Symposium (PacificVis)*, 2015, pp. 111-115, doi: 10.1109/PACIFICVIS.2015.7156365.

Appendix

Appendix A

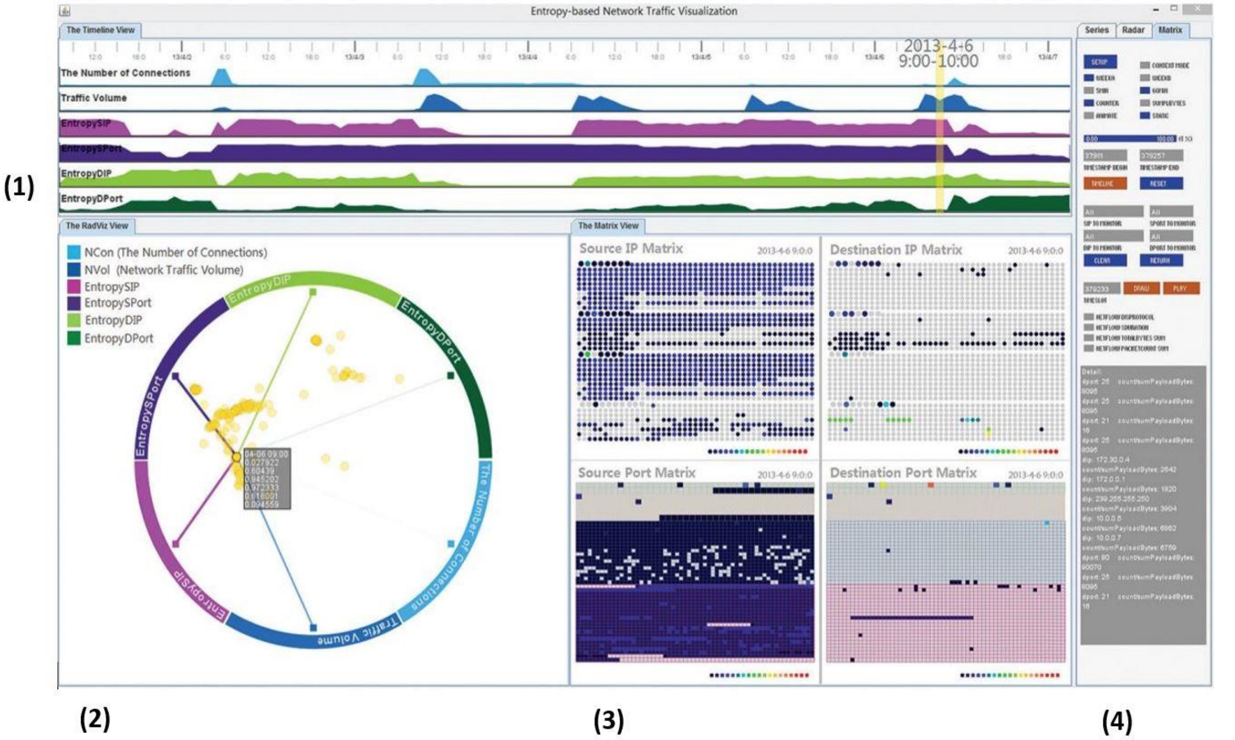

Figure 1: Interface Overview of ENTVis; (1) Timeline View, (2) Radviz View, (3) Matrix View and (4) Control Panel [7]

Appendix B

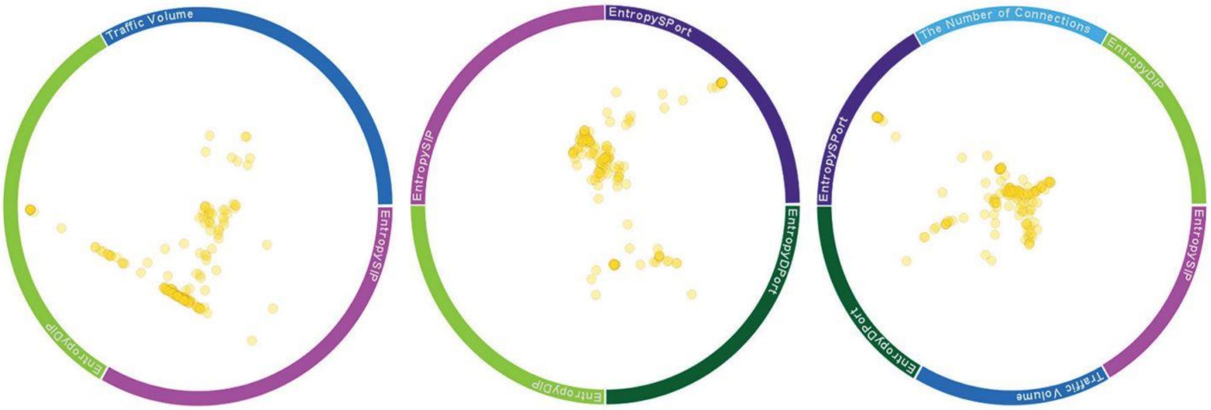

Figure 2: Visual Clustering Results using Radviz [7]

Appendix C

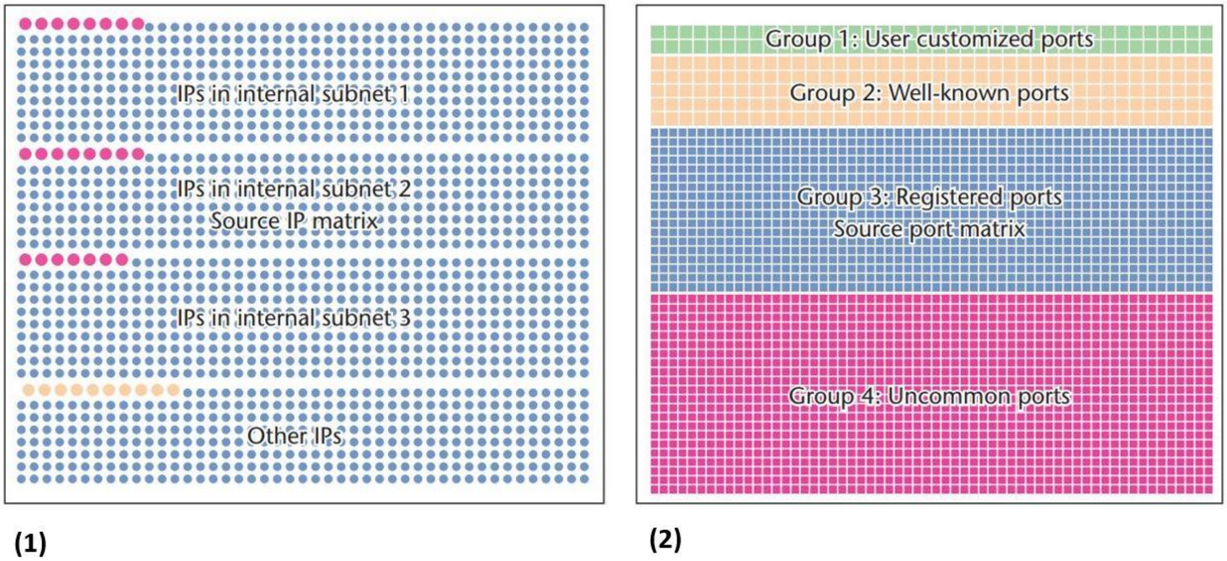

Figure 3: Matrix View with the Subnets. (1) The grouping and layout of an IP Matrix, (2) The grouping and layout of a Port Matrix [7]