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

Fangfang Zhou, Wei Huang, Ying Zhao, and Yang Shi ▪ *Central South University, China*

Xing Liang ▪ *Arizona State University*

Xiaoping Fan ▪ *Central South University and Hunan University of Finance and Economics, China*

Entropy-based traffic metrics have received much attention in network traffic anomaly detection, but practical issues still hinder widespread adoption. The visual analytic tool ENTVis provides coherent visual analysis that makes entropy-based traffic features more intuitive and helps users interpret network data and more quickly identify traffic anomalies.

Network anomalies can arise from internal or external factors, such as network device failures, worms, and distributed denial of service (DDoS) attacks. Most of these causes will lead to traffic patterns that differ from benign traffic. Entropy-based traffic anomaly detection has received considerable attention because entropy can provide more fine-grained metrics about traffic distribution patterns than traditional traffic volume analysis.[1] Several flow-header attributes (such as source and destination addresses and ports) and flow-behavior features (such as in and out degrees) have been suggested as candidates for entropy-based traffic anomaly detection.[1,2] The power of using multiple entropy-based metrics in conjunction with each other may improve the rapid detection of various attacks.

Time-series analysis is the main method for entropy-based traffic anomaly detection. The method can help identify not only single traffic distributions but also the distributional changes over time. The timeline group is the most common representation of multiple entropy-based time series, and it can be analyzed by intelligent or manual methods to detect potential traffic anomalies.

However, the application of entropy-based methods to detect traffic anomalies is hindered by several problems. First, entropy theory is not intuitive enough for users to understand because such a theory is an abstract mathematic metric of random variables. Second, entropy-based traffic metrics cannot provide detailed traffic distribution information. Third, when the timeline group includes many long time series, the visual clutter will prevent users from recognizing the underlying abnormal points.

Consider this scenario: a security analyst receives an alert through the automatic or manual analysis of entropy-based traffic features. If the alert is about a potential malicious scan for vulnerable ports, the security analyst will want to know which hosts and ports have been scanned, who the attackers are, and if the timeline contains similar scans. A comprehensive analysis of this information will help the analyst make informed decisions, including whether the scan is a true attack or a false alert, the severity of the network attack, and what actions must be taken to safeguard the network from further attacks. The traditional entropy-based anomaly detection method is weak in that

Published by the IEEE Computer Society

0272-1716/15/$31.00 © 2015 IEEE

it does not provide intuitive information and thus does not effectively support the analysis process.

The ENTVis visual tool visualizes the entropy measures of network traffic from multiple perspectives to help users achieve entropy-based anomaly detection. This tool provides three coordinated views and rich interactions to support a coherent visual analysis from multiple perspectives: the timeline group view for overall and drill-down situation analyses, the Radviz view for clustering similar anomalies during a given period, and the matrix view for understanding traffic distributions and diagnosing anomalies in detail. We performed several case studies and an expert review to verify the usability and effectiveness of our method.

Related Work in Entropy-Based Detection

Entropy is an important concept in information and communication theory. This concept measures the uncertainty or impurity of a collection of data items. In the cybersecurity field, Wenke Lee and Dong Xiang proposed using several information-theoretic measures, such as entropy and conditional entropy, to describe the regularity features of audit data for novel anomaly detection.[1] Since then, entropy-based traffic analysis has received considerable attention. For example, the port and address distributions have been commonly suggested as favorable candidates for entropy-based anomaly detection[2] because they lead to a highly sensitive detection of a range of anomalies and to the automatic classification of anomalies via unsupervised learning. Furthermore, George Nychis and his colleagues found that behavioral and flow-size distributions can provide distinct anomaly detection abilities to complement the port and address distributions.[3]

In the visualization community, entropy is a powerful measure for quantizing the chaos of a system or the saliency among multiple variables. Therefore, entropy is frequently used to guide efficient data exploration and improve visual mapping design. For example, Chaoli Wang and Han-Wei Shen discussed how information theory principles can be applied to scientific visualization.[4] Jamal Alsakran and his colleagues used entropy and joint entropy in reordering dimensions within the parallel set visualization to reduce visual clutter.[5]

Visualization for cybersecurity is a novel interdisciplinary field that offers humans visual tools to help solve cybersecurity problems.[6] In this area, entropy is generally used for data preparation or is directly visualized into the interface for high-level situational awareness. For example, in TVi, a visual tool for traffic traces,[7] the port and address distributions are taken as inputs to its core PCA (principal component analysis) based anomaly detector. In IDSRadar, which is a visual tool for intrusion detection system alerts,[8] five entropy timelines are mapped to the five tracks around a radial graph to distinguish false alarms from real risks. An online situational analysis tool, OCEANS,[9] integrates four entropy timelines that correspond to source/destination IPs/ports into the main visualization view to help users comprehend situations and detect high-risk periods.

References

1. W. Lee and D. Xiang, "Information-Theoretic Measures for Anomaly Detection," *Proc. IEEE Symp. Security and Privacy* (S&P), 2001, pp. 130–143.
2. A. Lakhina, M. Crovella, and C. Diot, "Mining Anomalies Using Traffic Feature Distributions," *ACM Sigcomm Computer Comm. Rev.*, vol. 35, no. 4, 2005, pp. 217–228.
3. G. Nychis et al., "An Empirical Evaluation of Entropy-Based Traffic Anomaly Detection," *Proc. 8th ACM Sigcomm Conf. Internet Measurement*, 2008, pp. 151–156.
4. C. Wang and H.W. Shen, "Information Theory in Scientific Visualization," *Entropy*, vol. 13, no. 1, 2011, pp. 254–273.
5. J. Alsakran et al., "Using Entropy-Related Measures in Categorical Data Visualization," *Proc. IEEE Pacific Visualization Symp.* (PacificVis), 2014, pp. 81–88.
6. H. Shiravi, A. Shiravi, and A.A. Ghorbani, "A Survey of Visualization Systems for Network Security," *IEEE Trans. Visualization and Computer Graphics*, vol. 18, no.8, 2012, pp. 1313–1329.
7. A. Boschetti et al., "Tvi: A Visual Querying System for Network Monitoring and Anomaly Detection," *Proc. ACM 8th Int'l Symp. Visualization for Cyber Security*, 2011, pp. 1–10.
8. Y. Zhao et al., "IDSRadar: A Real-Time Visualization Framework for IDS Alerts," *Science China Information Sciences*, vol. 56, no. 8, 2013, pp. 1–12.
9. S. Chen et al., "OCEANS: Online Collaborative Explorative Analysis on Network Security," *Proc. ACM 11th Workshop on Visualization for Cyber Security*, 2014, pp. 1–8.

Calculating Entropy-Based Traffic Metrics

Entropy is a measure of the uncertainty of a random variable in information theory. Let X be a discrete random variable with r states, x_i $i = 1 \ldots r$, and probability function $p_i = P\{X = x_i\}$, $x \in X$, $\Sigma p_i = 1$, $0 \leq p_i \leq 1$. The classic Shannon entropy $H(X)$ of the discrete random variable X is defined as follows:

$$H(X) = \sum\nolimits_{i=1}^{n} p_i \log p_i \ .$$

Entropy can be easily adopted to analyze network traffic distributions because many traffic attributes (such as source IP, destination IP, source port, destination port, protocol type, and flow size) can be considered discrete random variables.

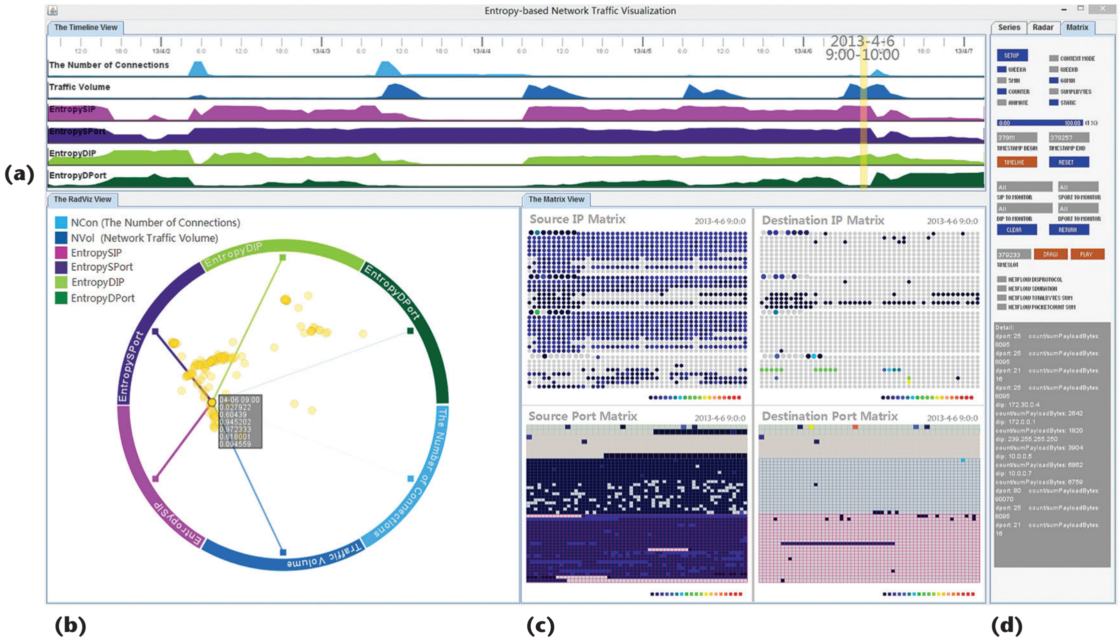

Figure 1. Interface overview. (a) Timeline view, (b) Radviz view, (c) matrix view, and (d) control panel.

Our approach uses the IP entropy and port entropy. The IP entropy reflects the randomness of the hosts involved in network activities. For example, if an attacker scans for vulnerable hosts, the source IPs of the scanning hosts will be seen in many traffic flows, and because they are relatively few, the source IP fields will contain less entropy than normal traffic. On the other hand, the destination IPs seen in traffic flows will be much more random than in normal traffic, while the destination IP fields will contain more entropy.

A similar phenomenon happens on the port level. The port entropy depicts the randomness of the ports involved in network activities. A larger port entropy corresponds to a higher number of ports that are visited with similar probability. By contrast, a smaller port entropy means fewer ports are visited. For example, if malware calls for a scan in all ports of some hosts with similar probability, the destination port entropy will become large. If a DDoS attack is aimed at a special destination port to paralyze the service on the host, the probability of the attacked port will increase rapidly, and the destination port entropy will suddenly decrease.

The IP entropy and port entropy calculations are similar. In the following example, we demonstrate how to calculate the destination port entropy. Let the destination port in a traffic log be a discrete random variable A. If 100 different ports (a_i i = 1–100) are visited within a specific time span, and the visited amount is n_i i = 1–100, the probability p_i i = 1–100 of each port is calculated as follows:

$$p_i = \frac{n_i}{\sum_{1}^{100} n_i} \left(i = 1 \ldots 100\right) .$$

The destination port entropy at this time is computed as follows:

$$H(A) = \sum\nolimits_{i=1}^{100} p_i \log p_i .$$

The next step is to standardize the value of $H(A)$ so we can compare different entropy-based traffic metrics. We scale the value of $H(A)$ to the interval [0, 1] with relative uncertainty (RU):

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

where $H_{\max}(A)$ denotes the maximum entropy of the destination port. For the previous example, 100 active destination ports are observed within a specific time span. Thus, the maximum entropy in this time span can be calculated by log(100). That means that 100 destination ports are visited with the same times.

Visualization

Figure 1 provides an overview of our interface, which consists of three coordinated views: the timeline, Radviz, and matrix.

Timeline View

The timeline view integrates a group of timelines at the top of the interface. Each timeline shows the temporal developing trend of a traffic feature. The group of timelines consists of two different types. The first type shows four entropy timelines that are widely used in network traffic anomaly detection, including the source IP entropy (EntroSIP), source port entropy (EntroSPort), destination IP

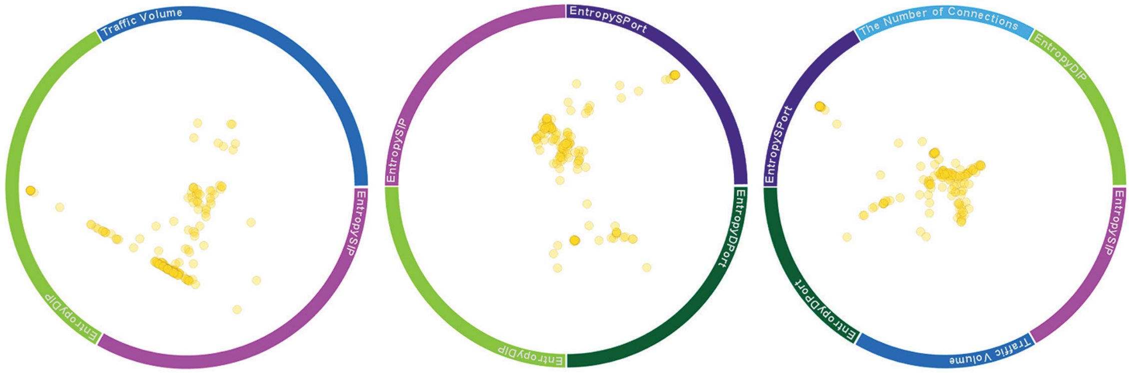

Figure 2. Visual clustering results when using Radviz. Users can freely explore the data by selecting single or multiple records, choices that change the number of DAs and reorder the DAs in Radviz.

entropy (EntroDIP), and destination port entropy (EntroDPort). The second type is about network traffic states, including the number of network connections (NCon) and traffic volume (NVol) across the whole monitored network. The NCon and NVol not only show the overall state of network activities, but they also supplement entropy-based traffic distributions.

The timeline view offers two time span choices—namely, 5 minutes and 60 minutes. Five minutes is suitable for short-term analysis in hours, whereas 60 minutes is suitable for medium- and long-term analysis in days. These two choices support temporal drilldown analyses. Using the timeline view, users can click on a pin in the timeline bar to view a selection of data. When the pins in the timeline view have been selected, the relevant points are highlighted in the Radviz view. Figure 1a shows six traffic features of a network for one week, 1–7 April 2013. The selected time span option is one hour. Each timeline includes 140 time spans. Figure 1 highlights the time span from 9:00 a.m. to 10:00 a.m. on 6 April in the timeline and Radviz views.

Radviz View

The timeline group provides the changing trends in the network traffic from a variety of metrics. However, it is difficult for users to quickly identify the time spans when the network has similar traffic features over a long period. In our approach, the Radviz view is designed to extend the entropy-based traffic analysis from the temporal space to the visual clustering space. In the visual clustering space, each data point corresponds to a time span, and the locations of data points with similar traffic features will be close together. Thus, users can quickly recognize similar time spans.

Radviz is a radial visual clustering technique that maps data from a multidimensional space onto a planar picture.[3] The data dimensions in Radviz are uniformly assigned positions on the circumference of a circle are called dimension anchors (DAs). The data records are mapped to the points within the circle. The position of each point in the circle is determined by the equilibrium of the attractions from the DAs. A significant feature of Radviz is its superior interactivity. The DAs on the circumference can be moved, inserted, or reordered either manually or algorithmically to help users explore meaningful visual clustering results.

The time series of traffic features in the timeline view are actually multidimensional data. Each time span is a record with multiple traffic features. In Figure 1b, NCon, NVol, and four entropy-based traffic features represent six DAs that are uniformly arranged on the circumference of a circle. Each dimension's data value is normalized in the interval [0, 1] for comparison in a uniform space. Time spans with similar traffic features will get together in the Radviz through the six spring tensions from the six DAs.

Rich interactions are provided for the free exploration of data, including the choice of a single record or multiple records, increasing or decreasing the DAs on circles, reordering dimensions, and displaying detailed information about the records. When a record in Radviz view is selected, the relevant time span in the timeline view will be highlighted. Detailed traffic distribution information on IPs and ports in this time span will also be updated in the matrix view. Figure 2 shows the different visual clustering effects on the 140 time spans in the week that result from changing the number of DAs or reordering the DAs in Radviz.

Matrix View

Entropy-based traffic features manifest the overall traffic distribution within a time span. However, users cannot easily locate the abnormal hosts and ports because of the lack of detailed distribution information of IPs and ports. The matrix view is designed to arrange all source/destination IPs/ports in the same interface to express clearly the

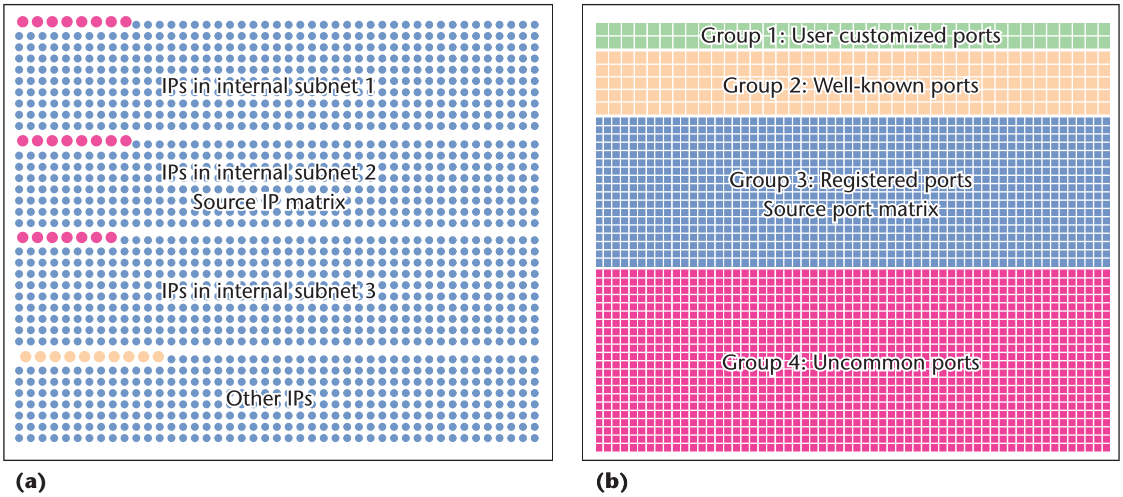

Figure 3. Matrix view with subnets. The grouping and layout of (a) an IP matrix and (b) port matrix.

complete traffic distribution information of IPs and ports. Therefore, the entropy-based metrics combine in the matrix view range from an overview of the traffic distribution analysis to a more detailed analysis.

The matrix view consists of four submatrix diagrams (Figure 1c). The top two diagrams focus on source IPs and destination IPs, whereas the bottom two focus on source ports and destination ports. In the IP matrix view, arranging all IPv4 addresses in a limited matrix space is a challenge. In our method, we provide the customized service of an IP layout according to the network architecture. The IP matrix view in Figure 3 illustrates an enterprise network with three subnets. Thousands of dots representing hosts and servers are arranged in groups by the internal subnet, with the bigger dots representing the servers and the smaller dots representing the general hosts. The external and broadcast IP addresses that often appear are located in the last part of the IP matrix view.

In the port matrix view, each port is represented as a grid. The ports are grouped on the basis of port number division. The first group allows users to select the ports that they mostly concerned about, such as port 80 for Web service. The second group comprises the well-known ports. The third group comprises registered ports that are occupied by some famous applications. The fourth group comprises dynamic ports. Unlike in the first three groups, every 100 continuous ports in the fourth group are placed in one grid cell.

The activity of a host or port is color coded from cool tones to warm tones to indicate its activity. Cool tones, such as navy and blue, indicate low network traffic, whereas warm tones, such as orange and red, indicate high network traffic.

In the matrix view, users can visualize the detailed traffic distributions of any time span that is highlighted in the timeline or Radviz views. Two types of traffic distribution, NCon and NVol, are provided. One of these distributions can be selected for display in the matrix view. If an IP address or a port is selected, detailed text messages will be shown in the state box in the control panel. Furthermore, the matrix view provides the traffic filter on the basis of the selected IP or port.

Case Studies

Here, we illustrate how our visualization tool operates using three cases: overall network analysis, DDoS attack detection, and port scan analysis.

In our case studies, we use a Netflow dataset provided by an enterprise network (see www.vacommunity.org/VAST+Challenge+2013). A network flow is an abstraction of a sequence of packets between two terminals. A typical Netflow record is identified by timestamp, source IP address, destination IP address, source port, destination port, protocol type, amount of traffic, and so forth. The enterprise network consists of three different branches, with each branch possessing around 400 workstations and a group of servers. Approximately 45 million Netflow records were collected between 1–7 April 2013. The data preparation includes data clearance; aggregated calculations on each IP, each port, and the whole network for every 5 and 60 minute interval; and the calculations of four entropy-based metrics.

Overall Analysis of the Week

In the first case, we analyzed the overall traffic patterns for the entire week. Therefore, the time span is set to 60 minutes. Figure 4a shows that 140 points in Radviz can be divided into three clusters labeled A, B, and C.

The points of cluster B in Radviz are near the DAs of EntroDIP and EntroDPort because the entropy values of their destination IP/port are large and the other four traffic features are relatively small. In the timeline view, cluster B is mainly distributed in two consecutive periods, namely,

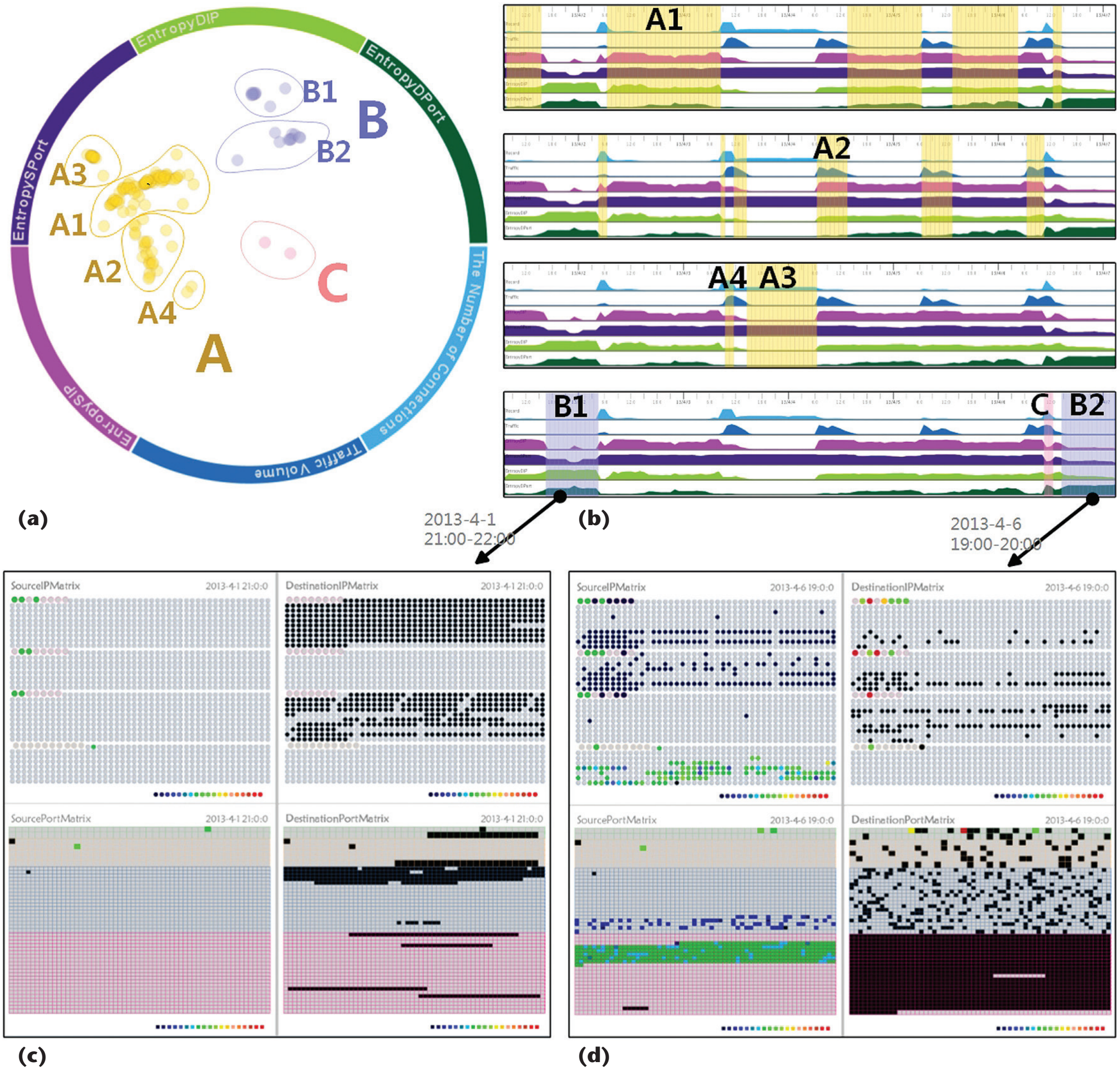

Figure 4. Overall analysis of traffic patterns in a week. (a) All time spans are divided into three clusters in Radviz. (b) Temporal distributions of the three clusters in the timeline view. (c) Detailed traffic information from 21:00 to 22:00 on 1 April 2013 in the matrix view. (d) Detailed traffic information from 19:00 to 20:00 on 6 April 2013 in the matrix view.

from 17:00 on 1 April to 5:00 on 2 April and from 16:00 on 6 April 6 to 5:00 on 7 April. Therefore, cluster B can be considered two subclusters, B1 and B2. The main difference between these two subclusters is that the EntroDIP value of B1 is larger than that of B2. Therefore, the high pull tension from the DA of EntroDIP makes the B1 points closer to the DA. We also select each time span from B1 (22:00 on 1 April) and B2 (18:00 on 6 April) to observe the traffic volume distribution in the matrix view, (Figures 4c and 4d). The traffic volume is calm in the two time spans. Figure 4c shows a few source IPs with a similar probability of visiting many destination IPs. Figure 4d shows a large number of dynamic destination ports that have been visited.

Cluster A occupies most of the time spans in the timeline view. The common characteristic of cluster A is that the values of EntroSIP and EntroSPort are large. This result indicates that a large number of source IPs/ports are involved in the network activities. Cluster A can also be divided into four subclusters according to the values of NCon and NVol. In the next case, we will analyze the abnormal events in clusters A3 and A4.

DDoS Attack Analysis

DDoS is usually a destructive attack that consists of a surge in traffic volume. We will demonstrate the process of analyzing the patterns of a DDoS attack, locating the involved hosts and assessing the damage by using ENTVis.

Two time spans in cluster A4 are considered abnormal because the values of NCon and NVol suddenly became large and the values of EntroDIP and EntroDPort suddenly became small. This finding indicates that a large amount of traffic has been sent to a few destination hosts and ports. This type of anomaly belongs to the typical traffic distribution features of a DDoS attack. The detailed NCon information in the matrix view can help us validate our speculation and locate the attackers and victims. The deepest red dot in the destination matrix is Web server 172.30.0.4,

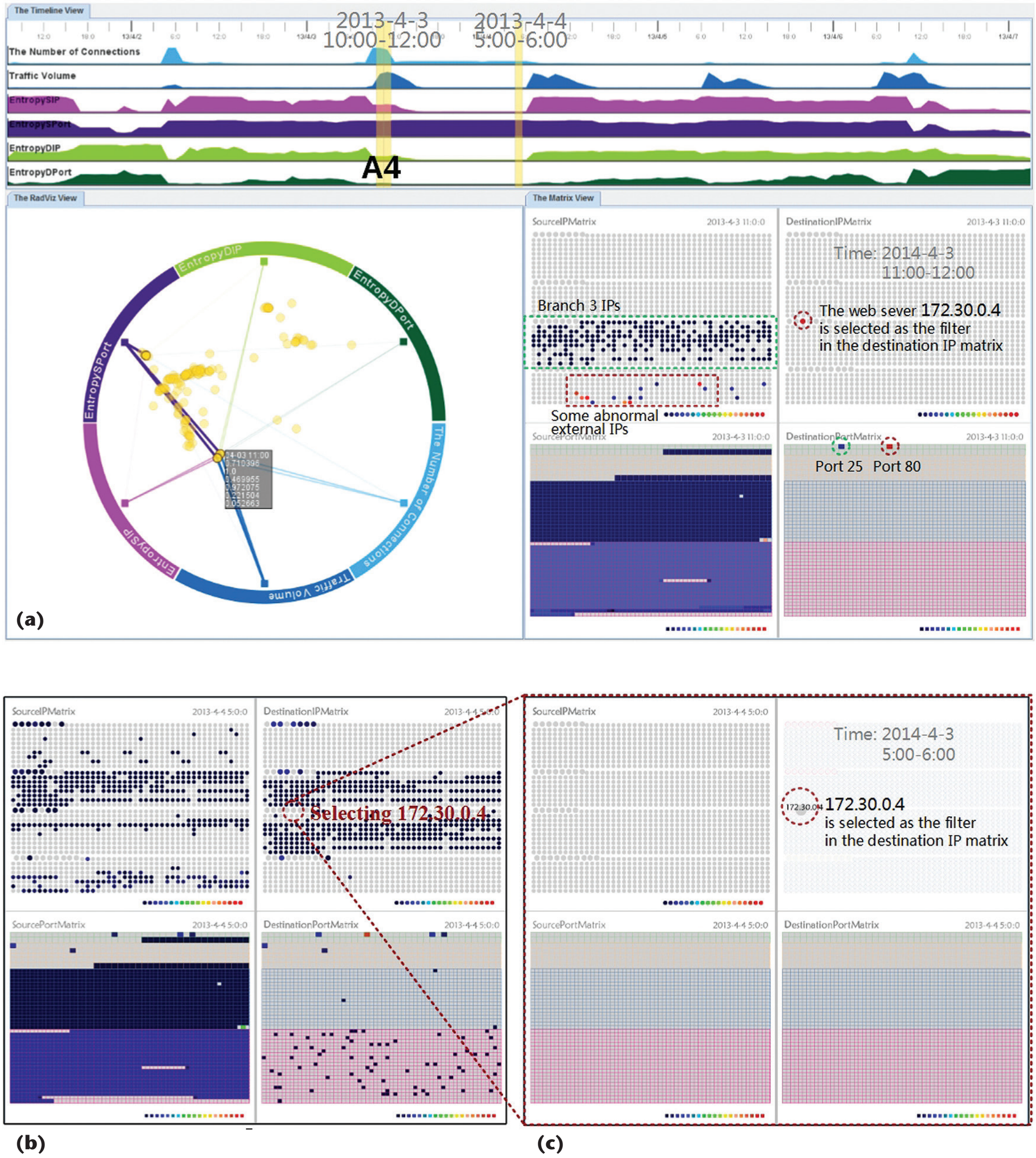

Figure 5. Analysis of a distributed denial of service (DDoS) attack. (a) The DDoS attack occurred at noon on 3 April 2013. (b) The matrix view shows the traffic distributions after the DDoS attack. (c) The suspicious outage appears on server 172.30.0.4.

which we select as the filtering option to update the other three matrixes. As Figure 5a shows, during the normal behavior, most of the hosts in the third subnet have visited 172.30.0.4 with low traffic, because 172.30.0.4 is a Web server in this subnet. The behavior became abnormal when some external IPs (in red) maliciously visited port 80 of 172.30.0.4 through more than 60,000 source ports. Thus, we can determine that 172.30.0.4 suffered from a DDoS attack at noon on 3 April.

The time spans of cluster A3 follow closely after the DDoS attack. All traffic features are unexpectedly harmonious in the hours of cluster A3. Figure 5b shows a large amount of active source ports from 5:00 to 6:00 on 4 April. With the Web server 172.30.0.4 still selected as the filtering option, Figure 5c shows that it had no related traffic. Therefore, we can speculate that the Web server 172.30.0.4 was paralyzed after the DDoS attack.

Port Scan Analysis

In the third case study, we analyze the abnormal events in cluster C by zooming in on the time span, switching from 60 minute to 5 minute intervals.

Cluster C only includes two successive time spans, namely, from 10:00 to 12:00 on 6 April. In the Radviz view, these two points are isolated from the other points (Figure 4a). In the timeline view, their values of NCon and NVol are large, whereas the values of four entropy measures suddenly change. We zoom into these two hours with bins of five minutes to explore what happened.

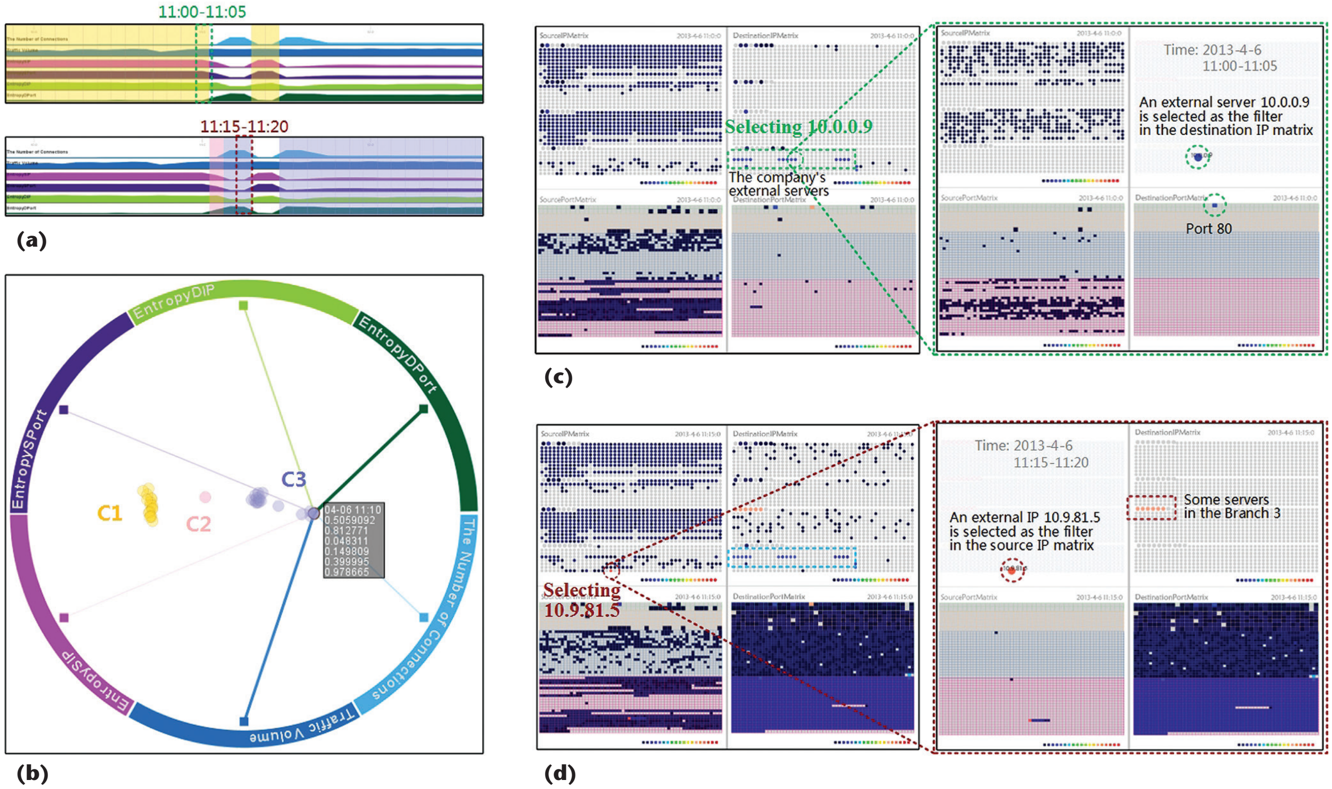
11:00-11:05
11:15-11:20
(a)
EntropyDIP
EntropyDPort
EntropySPort
The Number of Connections
Traffic Volume
EntropySIP
C1
C2
C3
04-06 11:10
0.5059092
0.812771
0.048311
0.149809
0.399995
0.978665
(b)
Selecting 10.0.0.9
The company's external servers
Time: 2013-4-6 11:00-11:05
An external server 10.0.0.9 is selected as the filter in the destination IP matrix
Port 80
(c)
Selecting 10.9.81.5
Time: 2013-4-6 11:15-11:20
An external IP 10.9.81.5 is selected as the filter in the source IP matrix
Some servers in the Branch 3
(d)

Figure 6. Analysis of a port scan. (a) Time series from 10:00 to 12:00 on 4 April 2013 in 5 minute samples. (b) All time spans in these two hours are divided into three clusters in Radviz. (c) Normal traffic patterns from 11:00 to 11:05. (d) Malicious port scan at 11:15.

Figures 6a and 6b provide an overview of traffic patterns from 10:00 to 12:00 on 6 April. The time spans are clustered into three groups in yellow (C1), red (C2), and blue (C3). Importantly, the yellow and blue groups have opposite traffic features for the NCon, EntroSIP, EntroSPort, and EntroDPort values. For example, the values of EntroDPort on the yellow points are small, and the values of EntroDPort on the blue points are large. The switches from the yellow group to the blue group occurred several times within 30 minutes after 11:00. The only red time span in Figure 6a is a transitional time span between the yellow and blue groups.

We selected two time spans in the borders between the yellow and blue groups to explore the traffic features in the matrix view. Figures 6c and 6d show the detailed traffic distributions in the time spans from 11:00 to 11:05 and from 11:15 to 11:20. In the destination IP matrix view in Figure 6c, some highly visited external destination IPs can be observed. There IPs are the corporation's external Web servers. We selected one of these IPs (namely, 10.0.0.6) as the filter and found that many internal IPs used the scattered ports to visit port 80. At this moment, the traffic was peaceful and the network was healthy. Figure 6d shows an obvious abnormal event in the destination port matrix because almost all destination ports were visited at this moment. We selected the highest traffic source IP (namely, 10.10.11.15) with the deepest red color as the filter. Surprisingly, IP 10.10.11.15 scanned all server ports in the third subnet. This abnormal behavior is clearly a malicious port scan that scans for any vulnerable port in the servers, thus explaining why the traffic features suddenly changed. This attack lasted throughout the whole blue time span.

Expert Review and Discussion

To evaluate our approach further, we had two potential users (network administrators in a university) perform an expert review. We designed a test and an interview for our expert review. The test included two parts. Part one implemented entropy-based anomaly detection without our prototype system, whereas part two involved working with our prototype system. The experts were given one hour for each part. During the interview, we talked about the user experience, and the advantages and limitations of the visualization framework.

In part one of the test, expert A almost ignored the entropy-based traffic features and queried the top-*N* of traffic volume to look for clues. He could not easily accept entropy theory in such a short time and preferred to use his accustomed method. Expert B tried to follow our direction by marking out many suspect time spans according to the entropy-based traffic features. Thereafter, he spent much of his time checking which moments were the real risks. In part two of the test, both experts became familiar with our tool quickly and identified many abnormal events by using the tool. The experts were also able to develop a comprehensive understanding of the overall network situation using the tool.

Both experts believe that our interface could be a useful tool for quickly understanding and using entropy-based traffic features for anomaly detection. Expert A commented, "It will take much time for me to use entropy in practice without this tool." Expert B added, "We really need tools like this to check the anomalies that are detected by automated alarming tools because the number of false positives emitted by them is truly staggering."

Based on the fruitful comments we received from the experts, we identified four limitations of our approach as well as potential methods to improve it. First, the IP layout method in IP-matrix can adapt to small-sized enterprise networks, but it cannot expand to show all IPs in a large-scale network. Second, our tool is supported by a large number of aggregated computations. Thus, we plan to optimize our data processing strategy to improve our response to large-scale data and real-time analysis. Third, if more traffic features were added into our analysis cycle, the visual clutters in the timeline view would increase, and the demand of providing the optimal placements of DAs would be urgent in the Radviz view. Lastly, the experts suggested that the high-risk period that was detected by some automated methods should be marked in the timeline view to help users to initiate visual analysis quickly. They also suggested that we consider monitoring the dynamic assigned IPs for the temporary hosts.

This analytical method can be applied to many entropy-based domains of data analysis, such as understanding the human mobility features in urban computing and exploring the communication patterns of people on social media.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under grant 61402540. We thank the anonymous reviewers for their comments. Ying Zhao is the corresponding author of this work.

References

1. A. Lakhina, M. Crovella, and C. Diot, "Mining Anomalies Using Traffic Feature Distributions," *ACM Sigcomm Computer Comm. Rev.*, vol. 35, no. 4, 2005, pp. 217–228.
2. G. Nychis et al., "An Empirical Evaluation of Entropy-Based Traffic Anomaly Detection," *Proc. 8th ACM Sigcomm Conf. Internet Measurement*, 2008, pp. 151–156.
3. F. Zhou et al., "Extending Dimensions in Radviz Based on Mean Shift," *Proc. IEEE Pacific Visualization Symp.* (PacificVis), 2015, pp. 111–115.

***Fangfang Zhou** is an associate professor in the School of Information Science and Engineering at Central South University, China. Her research interests include visualization and VR. Zhou has a PhD in control science and control engineering from Central South University. Contact her at zff@csu.edu.cn.*

***Wei Huang** is a graduate student in the School of Software at Central South University. Her research interests are visual analytics and data mining. Huang has an MS in software engineering Central South University. Contact her at huangwei_grace@csu.edu.cn.*

***Ying Zhao** (corresponding author) is a lecturer in the School of Information Science and Engineering at Central South University. His research interests include visual analytics and information security. Zhao has a PhD in in computer science and technology from Central South University. Contact him at zhaoying@csu.edu.cn.*

***Yang Shi** is a PhD student in the School of Information Science and Engineering at Central South University. Her research interests are information visualization and human-computer interaction. Shi has an MS in entertainment technology from Carnegie Mellon University. Contact her at shiyang@csu.edu.cn.*

***Xing Liang** is a PhD student in the School of Computing Informatics and Decision Systems Engineering at Arizona State University. His research interests include information visualization and visual analytics. Liang has a BS in computer science and technology Central South University. Contact him at csushin1004@gmail.com.*

***Xiaoping Fan** is a professor at in the School of Information Science and Engineering at Central South University and in the Laboratory of Networked Systems at the Hunan University of Finance and Economics. His research interests include intelligent controls and networked system. Fan has a PhD in control science and control engineering from the South China University Of Technology. Contact him at xpfan@csu.edu.cn.*