A Survey of network traffic visualization in detecting network security threats

Xiaomei Liu[1,3], Yong Sun[1,3], Liang Fang[2,3], Junpeng Liu[1,3], and Lingjing Yu[1,3]

[1] Institute of Information Engineering, Chinese Academy of Science, Beijing, China
[2] Beijing University of Posts and Telecommunications
[3] National Engineering Laboratory for Information Security Technologies
{liuxiaomei,sunyong,liujunpeng,yulingjing}@iie.ac.cn
fangliang@nelmail.iie.ac.cn

Abstract. Analyzing network traffic to detect network security threats has drawn attentions from security researchers for decades. However, the new characteristics of network traffic, such as explosive growth, more diverse attack types and higher dimension, have brought us new challenges. Because of these challenges, traditional detecting technologies like log analysis cannot directly identify threats from traffic in time. Visualization can straightly and quickly display multi-dimensional information of large network traffic. It can be our powerful weapon to meet the challenges. In this paper, we classify the network traffic into four layers. According to different layer, we systematically survey several well-known network traffic visualization systems. Then we analyze the advantages and disadvantages for each system and give out the comparisons. We also introduce the future works for network traffic visualization.

Keywords: network traffic, network security, visualization

1 Introduction

The evolution of the network technologies has brought us more conveniences, while explosive growth of network traffic also comes up with it. According to the report by CISCO, the whole global network traffic will reach up to 1.3ZB in 2016, which is much larger than we generate today. Would the number be larger? The answer is definitely yes. Meanwhile, network attacks hidden in the large network traffic have caused severe disastrous consequences. For example, as early as 2012, more than 68,000 DNS servers were utilized in a single DDoS attack. The traffic of network attacks has become even larger. In 2014, nearly 1/3 of the DDoS attack traffic reaches to 20Gbps while the highest traffic can be over 200Gbps. It brings us a challenge on how to identify network attacks from large network traffic directly and timely. Visualization technology has taken the researchers attention for it can straightly display the multi-dimensional information of the large network traffic in time.

Visualization transfers the invisible, unexpressed and abstract large data into visual images[1]. It can convert massive high-dimensional data to images and establish the image communications between human and data. By visualizing the network traffic, researchers can identify network patterns immediately and understand network traffic deeply. Meanwhile, researchers can find the hidden security threats, such as advanced persistent threat, internal threats and network cheating behavior, in the large network traffic. The advantages of visualizing the large network traffic are listed as follow:

- *Real-time:* Researchers can capture network traffic in real-time, extract the required properties of network traffic for visualizing, and update the visual graphics according to the updated data. In this way, researchers can identify network security threats more quickly.
- *High integrality:* All relevant data in the network traffic can be visualized to describe the continuous change over time. This not only helps to understand integral network patterns but also detects network security threats hidden in the network traffic more accurately.
- *Strong interactivity:* By using visualization technologies such as focus+context and dynamic queries etc., researchers can get more detailed information of the suspicious nodes for further analyzing. Besides, researchers can predict network attacks that have not occurred according to image schemas so as to guarantee the network security.

Although lots of visualization researches have made contributes to detect network security threats from traffic, a comprehensive survey of visualization trends and threats detection is absent.

Contributes and outline:In this paper, we systematically survey several well-known network traffic visualization systems for detecting network security threats. Meanwhile, we highlight the development trend of visualization between systems. Firstly, we classify the network traffic into four layers and state the methods and steps for network traffic visualization. Secondly, we show how to visualize large network traffic for detecting security threats. Thirdly, we point out future works. We hope to provide research ideas and literature references for detecting network security threats by visualizing network traffic in the future.

2 Layers, methods and steps of network traffic visualization

2.1 Layers of network traffic visualization

It is difficult to visualize the entire network traffic because the traffic is abstract, complex and large. So we need to preprocess the network traffic to extract and visualize specific traffic attributes that are closely related to network attacks. By summarizing and analyzing the related researches, we divide the current objects into four layers from top network traffic to bottom:

L1: visualization of bandwidth: Visualizing bandwidth to get the locations that consuming most network resources. Maybe most bandwidth is possessed by several illegal hosts.

L2: visualization of propagation delay: Propagation delay is the measurement of network connectivity and network performance. It can reflect the working state of network nodes to help researchers analyze abnormal network situations.

L3: visualization of communication paths: Communication paths represent the nodes that data go through from source to destination. It can reflect the dynamic changes of network topologies and help to detect abnormal network topologies like botnets, etc.

L4: visualization of packet attributes: Packet attributes include IP addresses, ports and protocols, etc. They are the most frequent objects of visualizing for detecting network abnormal behavior.

2.2 Methods and Research Steps of Network traffic visualization

According to the specific application status, we should select and design reasonable visualization methods to help us understand the different traffic. At present, the basic visualization methods include charts, 2D/3D graphics, 2D/scatter/circular plot, trajectory drawing, 3D geometry, topology, tree, etc. Most visual designs integrate the methods mentioned above with human-computer interaction technologies such as dynamic queries etc.

We can visualize network traffic as following steps[2]:

- S1.data source selection
- S2. Visualization methods selection
- S3. Concurrent display of global and local information
- S4. Human-computer interaction design

3 Network traffic visualization

3.1 visualization based on bandwidth

In order to achieve high quality network services, network administrators must allocate reasonable bandwidth for network applications. Through visualizing bandwidth, we can both quickly perceive the locations of consuming most network resources and identify the hosts that occupy illegal bandwidth. Initially, researchers just use ordinary histograms or pie-charts to display bandwidth. But such technologies cannot display high-dimensional and dynamic network traffic vividly. With the development of visualization, researchers begin to use the novel visual technologies to display the bandwidth. Oetiker[3] introduced MRTG which uses RRDtool to display bandwidth in form of charts. However, OPENGL can display bandwidth in 3D space[4]. Samak provided CISCO NetFlow Analyzer which could analyze utilization of network bandwidth and would have alerts if it detects attack. Akamai shows the proportion of bandwidth of different areas around the world based on the global map and different colors represent different amount of bandwidth[5]. As is shown in Fig. 1, the redder color indicates the greater the traffic.

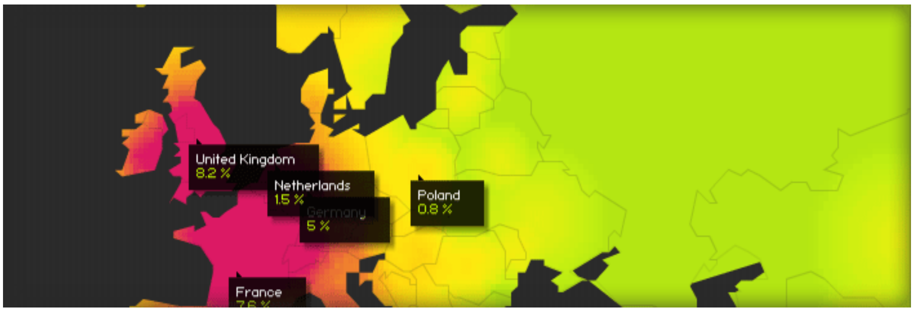
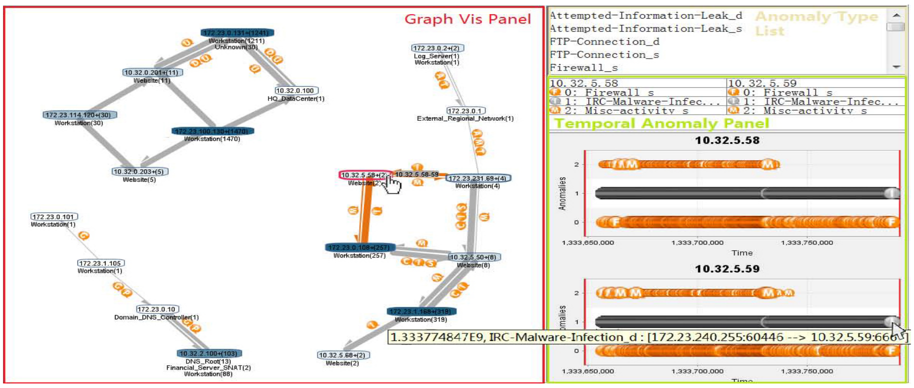


Fig. 1. Akamai - the proportion of traffic

3.2 visualization based on communication paths

As mentioned in section 2.1, through visualizing communication paths we can find weak routing nodes and abnormal network topologies. However, with the increase of the traffic scale, it is difficult to visualize paths in 2D. Skitter visualizes communication paths based on a 3D earth model[6]. IP addresses in different positions are represented in different colored nodes. For example, yellow nodes represent the source nodes, while green nodes represent intermediate nodes or destination nodes. Communications between nodes are linked with lines. Skitter is used to measure the IP forwarding paths, evaluate the performance and topologies of Internet. More importantly, it helps researchers identify the abnormal topologies like botnets etc. In order to prevent topologies being too large, Shi[7] uses the compressed technology which maps a set of IP addresses into a single node to reduce the display burden. As shown in Fig. 2, the main interface shows the compressed communication paths and the possible network anomalies, while the details of network anomalies are showed in the right.



Attempted-Information-Leak_d
Attempted-Information-Leak_s
FTP-Connection_d
FTP-Connection_s
Firewall_s

10. 32. 5. 58	10. 32. 5. 59
0: Firewall s	0: Firewall s
1: IRC-Malware-Infec...	1: IRC-Malware-Infec...
2: Misc-activity s	2: Misc-activity s

1.333774847E9, IRC-Malware-Infection_d : [172.23.240.255:60446 --> 10.32.5.59:666]

Fig. 2. Topologies display using compressed technology

3.3 Visualization based on propagation delay

Although the success rate of transferring packets from source to destination is important, real-time transmission of the packets is also important. Propagation delay can reflect the speed of processing and forwarding packets of intermediate nodes. Then researchers can get suspicious network nodes through propagation delay. Akamai is one of the largest CDNs in the world [5]. It can test and show the slowest web connections between some important cities through a few of network behavior such as download etc. As shown in Fig.3, every important city has two vertical bars. The bars represent both the current absolute time delay and the relative time delay compared with its historical average time delay. If the gap between current time delay and average time delay is very big, the web server may has been damaged by a malicious user. Then researchers should further analyze potential malicious network behavior.

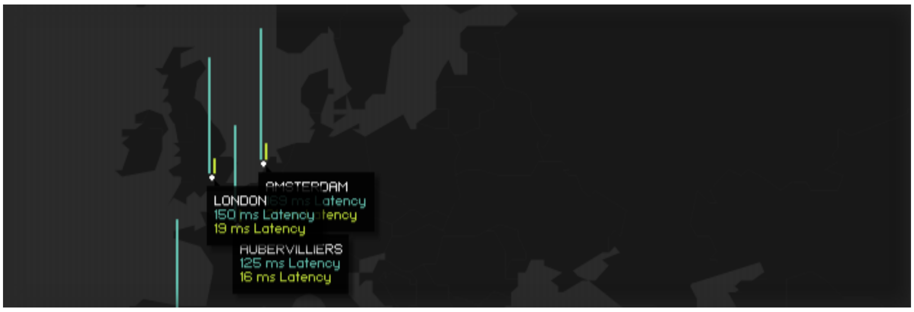
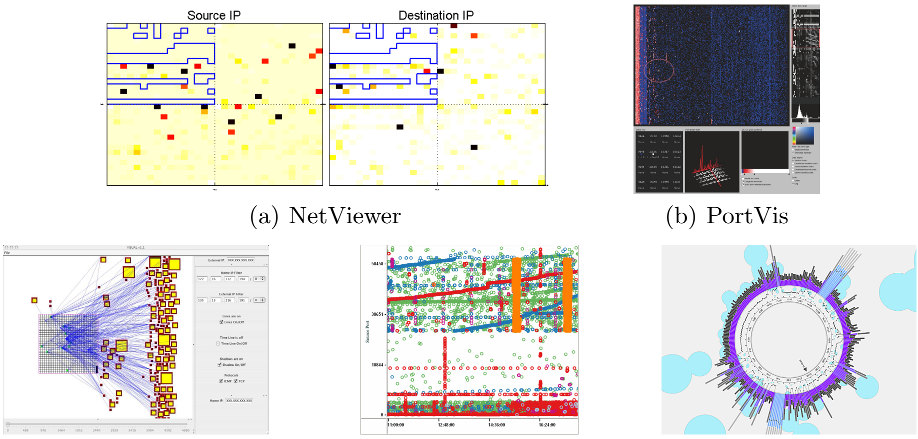


Fig. 3. Akamai-propagation delay visualization

3.4 Visualization based on packet attributes

Network traffic consists of packets that have basic attributes which are closely related to network attacks like source/destination IP address and port, protocol etc. Current network anomalies such as port scanning, worm attacks have obvious corresponding characteristics of one-to-many, many-to-one and many-to-many. We can quickly perceive them through visualizing IP address and port.

Both Spanning Cube of Potential Doom[8]and NetViewercite[9] map the entire IP addresses into the grid. The former shows the scanning activities in form of lines and detects malicious traffic by incomplete connections. The latter, however, uses different colors to represent the number of packets that go through the IP address over a period of time and uses polygons to classify IP addresses in the grid. NetViewer can monitor network anomalies such as the source IP-spoof, worm scan, etc. As shown in Fig.4.a), the IANA reserved IP addresses are in the blue polygon. As we all know, the reserved IP addresses cannot serve as the destination address. Then we can monitor the source IP-spoof by using NetViewer.



(a) NetViewer (b) PortVis

(c) VISUAL (d) 2D PLOT (e) LogTool

Fig. 4. Visualization based on packet attributes

The two tools are difficult for researchers to further analyze network patterns because of the lack of human-computer interaction technology. Improvably, as shown in Fig.4.b), PortVis[10] not only observe the information of all ports but also analyze specific ports in particular time by setting parameters and filtering details. VISUAL[11], VisFlowConnect[12] and NAV[13] visualize local hosts and remote hosts separately to determine the attack position accurately. VISUAL as shown in Fig.4.c), local hosts are in the grid, while remote hosts are around outside the grid. Communications between them are represented by lines. VisFlowConnect and NAV use the parallel technology to depict the relationships between internal and external network and display detailed traffic information of specified nodes. We can know which remote host was scanning local hosts by using them. The common disadvantage of these tools is that visualizing large traffic may cause visual clutter. To avoid the visual clutter, NFlowVis[14] uses hierarchical edge bundles to classify traffic and reveals large-scale distributed attacks. Meanwhile, it connects an alarm of an intrusion detection system and sends alerts in time if network attacks are detected. Destination ports correlate to network connections and network behavior. As IDGraph[15] and n-dimensional histogram [16] observed, they can detect TCP SYN flooding. The former uses the traffic aggregation methods to visualize the destination IP addresses and ports to reveal it. While the later detects it by scanning each port and building an n-dimensional histogram to display unsuccessful connection number of destination ports. Gerth uses 2D PLOT to visualize destination ports to detect attacks[17]. As shown in Fig.4.d), green represents mail, blue represents DNS, red represents scanning activities and orange represents SSH attacks, etc. Flying Term[18] can find abnormal DNS query patterns and observe DNS dynamics over time by visualizing the query frequency of port 53. LogTool[19] as shown in Fig.4.e), the time circle of a day is divided into 288 pieces. Each piece is a radial histogram that shows the proportions of uplink and downlink traffic. The dot-dash line inside the circle shows the number of HTTP(port 80) connection requests, so it can analyze users online surfing behavior.

4 Future works

Although visualization has gained great achievements in quickly identifying network attacks from large network traffic, constructing a complete and practical visualization system to detect network security threats still faces many challenges.

- ***Lack of a set of complete and systematic theories for the network traffic visualization.*** Meanwhile, it is also difficult to choose a unified evaluation standard for verifying and evaluating the result provided by those visualization products. Because different users may have different needs, there is certain subjective for the evaluation standard. Above all, we need to put more attention on the theoretical research of network traffic visualization.
- ***How to identify new network attacks by using network traffic visualization.*** The network structure is getting more complex. The security threats are more varied. How to predict the attacks that have not occurred and how to find the new attacks in the normal flow are becoming very important.
- ***How to design network traffic visualization system with high efficiency.*** Because the scale of network traffic is getting more large, many systems cannot process and display such large data in real-time. We need to improve or redesign the algorithms for processing and visualizing the network traffic.

5 Conclusions

As a relatively mature technology, visualization has played an important role in many research fields. Aiming at identifying network security threats from large network traffic, we summarize and analyze the current visualization technologies from four aspects that are closely related to network attacks:1) bandwidth, 2) propagation delay, 3) communication paths and 4) packet attributes. Though the existing works can partially handle the challenges brought by the new characteristics of the network traffic, there are still a lot of future works to be done. It will be our pleasure that our work could provide references for the future researches of network traffic visualization.

6 Acknowledgement

This work was supported by The National Science and Technology Support Program (Grant No. 2012BAH46B02); the National Natural Science Foundation (Grant No. 61402464,61402474).

References

1. MCCORMICK B H,DEFANTI T A,BROWN M D. Visualization in Scientific Computing[J].Computer Graphics,1987,21(6):1103-1109.
2. Lv Liang fu, ZHANG Jiawan, SUN Ji zhou, HE Pilian, SUN Ligang. Survey of network security visualization techniques. Computer Applications, 2008: 28(8)
3. Tobias Oetiker. Multi router traffic grapher. `http://oss.oetiker.ch/mrtg/`
4. LIU Yi, WU Ni, ZHANG Han. Network Traffic Statistic Analysis and Visualization System. Microelectronics Computer, 2007:24(6)
5. Fabian Popa.Network Traffic Visualization.Seminar Innovative Internet-Technologien und Mobilkommunikation, WS 2008/2009 Institut fr Informatik, Lehrstuhl Netzarchitekturen und Netzdienste Technische Universit?t Mnchen.
6. CAIDA Skitter `http://www.caida.org/tools/measurement/skitter/`
7. Lei Shi,Qi Liao,Chunxin Yang."Investigating Network Traffic Through Compressed Graph Visualization" VAST 2012 Mini Challenge 2 Award:Good Adaptation of Graph Analysis Techniques
8. Stephen Lau. The Spinning Cube of Potential Doom. Commun. ACM, 47(6):25-26, 2004.
9. LISA '05 Paper NetViewer: A Network Traffic Visualiza-tion and Analysis Tool Seong Soo Kim and A. L. Narasimha Reddy - Texas A M Uni-versity
10. Jonathan McPherson, Kwan-Liu Ma, Paul Krystosk, Tony Bartoletti,and Marvin Christensen. Portvis: a tool for port-based detection of security events. In VizSEC/DMSEC 04: Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security, pages 73-81. ACM Press, 2004
11. Robert Ball, Glenn A. Fink, and Chris North. Home-centric visualization of network traffic for security administration. In VizSEC/DMSEC04: Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security, pages 55-64. ACM Press, 2004.
12. Xiaoxin,Yin,William,Yurcik,Yifan Li,Kiran Lakkaraju.VisFlowConnect:Providing Security Situational Awareness by Visualizing Networks Traffic Flow. Cristina Abad.Proceedings of the IEEE 2004
13. M. Allen, P. McLachlan, NAV Network Analysis Visualization, University of British Columbia, [Online, 29 May 2009].
14. Fabian Fischer, Florian Mansmann, Daniel A. Keim, Stephan Pietzko, and Marcel Waldvogel.Large-Scale Network Monitoring for Visual Analysis of Attacks. 5th International Workshop, VizSec 2008, Cambridge, MA, USA, September 15, 2008. Proceedings. pp 111-118.
15. Pin Ren,Yan Gao,Zhichun Li,Yan Chen,Benjamin Watson. IDGraph :Intrusion Detection and analysis using Stream Compositing. Published in IEEE Computer Graphics and Applications, pages 28-39, 2006
16. E. Wes Bethel,Scott Campbell,Eli Dart. Accelerating Network Traffic Analytics Using Query-Driven Visualization. IEEE Symposium on Visual Analytics Science and Technology 2006
17. Ling Xiao,John Gerth,Pat Hanrahan. Enhancing Visual Analysis of Network Traffic Using Knowledge Representation. Proceedings of the IEEE Symposium on Visual Analytics Science and Technology 2006
18. Pin Ren,John Kristoff,sBruce Gooch. Visualizing DNS traffic.VizSEC06 Proceedings of the 3rd international workshop on Visualization for computer security Pages 23-30.
19. `http://infosthetics.com/archives/2010/10/logtool_revealing_the_hidden_patterns_of_online_surfing_behavior.html`