A Visualization Technique for Monitoring of Network Flow Data

Manami KIKUCHI
Ochanomizu University
Graduate School of Humanitics and Sciences
Otsuka 2-1-1, Bunkyo-ku, Tokyo, JAPAPN
manami@itolab.is.ocha.ac.jp

Takayuki ITOH
Ochanomizu University
Graduate School of Humanitics and Sciences
Otsuka 2-1-1, Bunkyo-ku, Tokyo, JAPAN
itot@computer.org

Hiroki TAKAKURA
Kyoto University
Academic Center for Computing and Media Studies
Yoshoda-Honmachi, Sakyo-ku, Kyoto, JAPAN
takakura@media.kyoto-u.ac.jp

Abstract

Research and development of IDS (intrusion Detection System) is a hot topic for the purpose of security maintenance of computer network. We have already presented a technique for visualizing logs of IDS. However, the present IDS products detect only known suspicious accesses, and therefore we need an extended visualization technique if we would like to visualize the statistics of malicious accesses unknown by IDS products.

This report proposes a technique for visualizing statistics of suspicious accesses, including the accesses not detected as intrusions by IDS products. The technique first constructs hierarchy of computers according to their IP addresses, and represents the groups of computers by the information visualization technique "HeiankyoView". Simultaneously, it reads the special network flow log files, which records information of buffer overflows, shellcodes, and intrusions detected by IDS products. Finally, it represents the statistics of the suspicious accesses recorded in the logs. Against our previous technique visualized only intrusions recorded to logs of IDS products, the presented technique enables discovery of more various malicious attacks.

1. Introduction

Detection and analysis of malicious accesses on the Internet is an active research topic, due to serious damages by the malicious accesses. Recently, intrusion Detection System (IDS) is widely used to detect and record suspicious accesses, under the assumptions that several security mechanisms such as firewalls are working on the local area network (LAN). Here, larger network yields huge access log files, and the accesses recorded to the logs complicatedly related each other. It is therefore very difficult to discover various behaviors of suspicious accesses by simply reading the logs.

Visualization techniques are very effective to represent perspective of logs, as an idea to solve the above problem. Many studies on visualization of IDS log data have been presented, including our own previous work [3]. However, we may miss new attacks if we only monitor IDS logs, because IDS logs record only suspicious accesses known by IDS products.

Here we consider using network flow log data in addition to IDS information, and monitoring suspicious phenomena that IDS products do not detect. Usual network flow logs record IP addresses and port numbers of senders and receivers, and communication protocols. In addition to them, we assume to use a special network flow log data, which also record information of buffer overflows, shellcodes, and intrusions detected by IDS. We can discover more various malicious phenomena by visualizing such network flow log data. Following are examples of phenomena that we can discover by the above idea:

- We may discover a new attack, if a buffer overflow is observed in an access, though it is NOT detected as an intrusion by IDS.
- A computer may be intensively attacked, if the computer intensively receives a same shellcode, or accesses which are detected as the same intrusion by IDS.

This paper presents a technique for visualizing the above log data, for the purpose of discovery of more various intrusions and attacks. The technique applies "HeiankyoView" [4], a large-scale hierarchal data visualization technique, as our previous technique [3] has applied. It is suitable for representation of statistics of intrusions to a large-scale network, because HeiankyoView can display the distribution of computer groups in large-scale LAN containing thousands of computers. Here, our previous technique [3] only represented the statistics of senders and receivers of suspicious accesses, and the target of statistics is limited to attacks known by IDS products. The technique in this paper extends the previous technique, which reads the abovementioned special log data, and realizes the visualization that we can observe more various intrusions in detail.

This paper shows several examples of visualization of log data recorded on the real LAN, and discusses what kind of phenomena we can discover from the visualization results.

2. Related Work

2.1. Visualization for network security

Visualization of network security data is a recent hot topic. This section introduces several famous works in this area.

Many of security visualization technique focus on representation of temporal information of malicious accesses. Mie-Log [9] horizontally divides a display space; the left side of the display space provides bar charts of statistics of suspicious phenomena per a constant time, and the right side provides detailed information of suspicious phenomena of a user-selected time span. The technique simultaneously provides global and local views, and detail-on-demand operations, for time sequence of suspicious phenomena. SnortView [5] main tool is a two-dimensional time diagram, which focuses on dealing with false alerts. PortVis [7] produces visualization of network traffic using 2D plots with time and port number as axes, and summarizes the network activity at each location in the plot using color. ID-Graphs [8] uses flow records and represents the statistics of unsuccessful accesses overtime.

Many other works focus on representation of distribution of malicious accesses in IP address spaces. Ball et al. [2] presented an IP address oriented technique simply mapping values of each byte of IP addresses onto horizontal and vertical axes of display spaces. NVisionIP [6] visualizes network flow data in a 2D matrix with IP addresses on each axis, where each cell in the matrix represents the interaction between the corresponding network hosts. IDS RainStorm [1] main view visualization presents an overview of alarms related to each IP address. The color of an alarm represents its severity. The user can select a range of IP addresses and use the Zoom View that focuses on the selected hosts and provides additional information for each alarm.

Our previous technique applied "HeiankyoView" [3] to visualize IDS log data, while it hierarchically clusters computers according to their IP addresses. The technique can display distribution of computer groups structuring a large-scale network. The next section introduces the detail of HeiankyoView.

There are some other woks which widely represent various attributes of traffic. Yin et al. applied Parallel Coordinates to discover suspicious traffic patterns [11]. Tee Teoh et al. applied robust visual analysis technology to discover suspicious traffic from network log files, by the combination of various visualization technique [10].

2.2. HeiankyoView

"HeiankyoView" [4] is a visualization technique that represents whole large-scale hierarchal data in one display space. It represents leaf nodes of the hierarchal data as icons, and branch nodes as nested rectangular borders.

Figure 1 shows an example of visualization of hierarchal data by HeiankyoView. It has advantages against other hierarchical data visualization techniques as follows: It represents large-scale hierarchical data containing thousands of leaf nodes as equally sized, without overlapping each other, in a small display space.

The technique in this paper constructs groups of computers according to their IP addresses. Firstly it groups computers according to the first byte of their IP addresses. It then groups them according to the second byte of IP addresses, and finally groups according to the third byte of IP addresses. Consequently, the technique builds 4 level hierarchal data, as shown in Figure 2. It represents the distribution of computers in an IP address space by HeiankyoView.

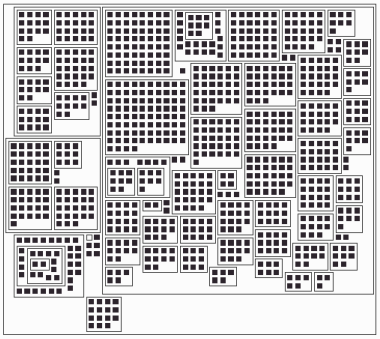

Figure 1. Example of visualization of hierarchal data by HeiankyoView.

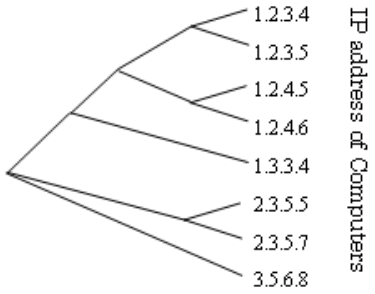

Figure 2. Illustration of hierarchy of 8 computers according to their IP addresses.

The technique can represent computers without overlapping each other on the display, and users can visually recognize the groups of computers according to their IP addresses (in many cases the groups mean the real organizations), by visualizing the computers as icons by HeiankyoView. This representation is very appropriate for the purpose of visualization of trends and statistics of suspicious accesses for each IP address, targeting large-scale network containing thousands of IP addresses. Here, our previous technique [3] just represents the statistics of senders and receivers of suspicious accesses as shown in Figure 3, and do not represent more detailed information. Therefore we consider the extension of our previous technique, to visualize information that we can get by using both IDS and network flow log data.

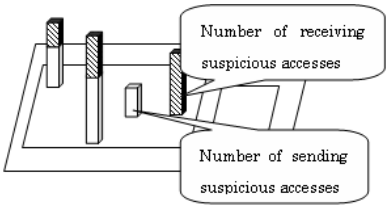

Figure 3. Example of statistics of senders and receivers of suspicious accesses.

3. Implementation

3.1. Supposed log data

As described in Section 1, we assume to use a special network flow data including information of buffer overflows, information of shellcodes and intrusions detected by IDS products. Following are the attributes of the network flow log data we assume in this paper:

- Sequential number.
- Date (Year - Month - Day).
- Time.
- Address of the sender.
- Address of the receiver.
- Protocol type.
- Data size.
- MD5 value of the payload.
- Information of buffer overflows, including:
  - MD5 value of the session data causing the buffer overflows,
  - the time when the system first observed the same MD5 value, and
  - total number of accesses observing the same MD5 value.
- Information of shellcode, including:
  - ID of the shellcode,
  - the time when the system first observed the same shellcode, and
  - total number of accesses observing the same shellcode.
- Signature IDs detected by IDS.

Here, buffer overflow means that oversized data input exceeding the allocated memory spaces cause runtime failure of software. Buffer overflow attack means that attackers, who intentionally send oversized data exceeding the tolerance of buffers to make systems blackout, or execute the overflowed data. Buffer overflow is one of the most typical security hole, and the problem of various operating systems and applications for tens of years. The technique proposed in this paper focuses on visualization of statistics of buffer overflows, because more than half of known security holes are related to buffer overflow.

3.2. Visualization of buffer overflows

We suppose that accesses may be malicious if they cause buffer overflows, because they are one of the typical methodology of abuse (i.e. Deny of Service). Based on the above assumption, the technique extracts accesses causing buffer overflows from network flow log data.

Here, it is possible to visualize suspicious accesses that cause buffer overflows, without applying IDS information.

However, we think we can realize more effective visualization, by associating information of buffer overflows with IDS information. Following is our assumptions of relationship between buffer overflows and IDS information:

- The accesses may be a kind of random attacks, if there are multiple signature IDs of IDS.
- The accesses may be concentrated attacks, if there is only one signature ID.
- The accesses may be a kind of new unknown attacks, if there is no signature ID.

Also, the technique also supports visualization of relationship between buffer overflows and shellcode information, in addition to IDS information. Following is our assumptions of relationship between buffer overflows and shellcode information:

- The accesses may be a kind of random attacks, if there are various kinds of shellcodes in a short time.
- The accesses may be concentrated attacks, if there is only one shellcode.
- The accesses may be attacks just targeting deny of services, if there is no shellcode.

Based on the above assumptions, we consider visualizing every access that cause buffer overflows, categorizing them into three groups according to the numbers of signature IDs of IDS or shellcodes (0, 1, and more than 1).

Figure 4 shows an example of the visualization [1]. In this example, gray bars denote computers sending the malicious accesses, and colored bars denote computers receiving the malicious accesses. Heights of bard denote numbers of sending/receiving accesses. The three colors denote the categories of accesses according to the number of shellcodes or signature IDs of IDS, as follows:

0: magenta,

1: green, and

more than 1: orange.

As developed in our previous technique [3], the technique supports a function to represent the communication between computers as yellow lines, as shown in Figure 5. When a user clicks a bar, the technique aggregates sending and receiving suspicious accesses of the computer, which is represented as the clicked bar. Then it extracts the computers received the suspicious accesses from the clicked computer, and draws the yellow lines between the clicked bar and the bars representing the computers received the suspicious accesses. Similarly, it extracts the computers sent the suspicious accesses to the clicked computer, and draws the yellow lines between the clicked bar and the bars representing the computers sent the suspicious accesses.

[1] We will publish the color visualization results at http://itolab.is.ocha.ac.jp/

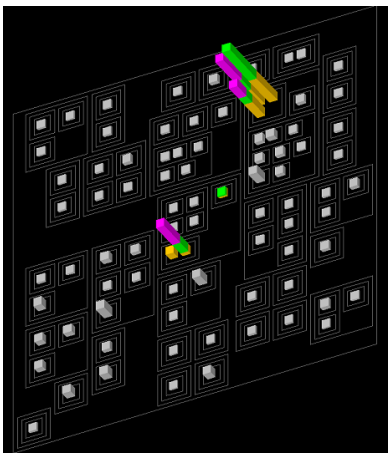

Figure 4. Visualization result of receiving accesses with IDS information.

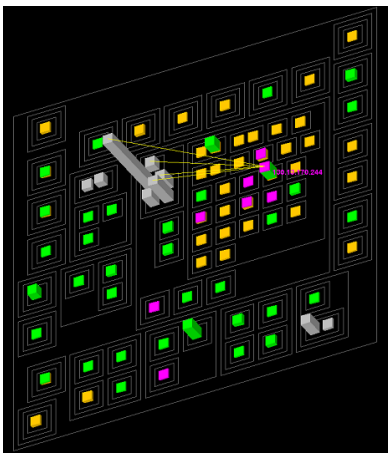

Figure 5. Yellow lines are drawn between pairs of senders and receivers.

4. Results

We implemented the presented technique on Java 1.5, and executed on HP dc5700 Small Form (CPU 2.8GHz, RAM 0.99GB) with Windows XP. We tested it using network flow log data of existing computer network, which we set a system so called ”honeypot” to aggressively collect suspicious accesses. The system detected buffer over-

flows and shellcodes by a non-public network security product. It also detected intrusions by Semantic IDS SNS-7120, and combined the information of intrusions to the logs of buffer overflows and shellcodes. We divided the log data by month, and experimented the visualization for each month of the log data.

Here, we introduce visualization results of buffer overflows with shellcode.

Figure 6 is a visualization result of received accesses that cause buffer overflows, with shellcode information. Here, colored bars denote computers receiving suspicious accesses, and gray bars denote computers sending them.

Figure 6(right) shows the result of one year after the result in Figure 6(left). This result denotes that same shellcodes are used during the year, because many shellcodes are used twice or more, and therefore many bars are painted as orange in Figure 6(right).

Figure 7 is a visualization result of sent accesses that cause buffer overflows, with shellcode information. Here, colored bars denote computers sending suspicious accesses, and gray bars denote computers receiving them. This result denotes that one computer might turn to a sender from a receiver, because there are bars that are partially painted colorful, and also partially painted as gray.

Next, we introduce visualization results of buffer overflows with IDS information.

Figure 8 is a visualization result of received accesses that cause buffer overflows, with IDS information. Here, colored bars denote computers receiving suspicious accesses, and gray bars denote computers sending them. Several bars in the left side of the figure are very high; these bars denote computers of the honeypot, which often receive suspicious accesses. It seems that most of the accesses may be a kind of random attacks, because the most parts of the bars are painted as orange. At the same time, center of the figure denotes that a specific computer is intensively attacked by computers whose first to third bytes of IP addresses are same, because we can observe that yellow lines are pulled by bars in the same cluster.

Figure 9 is a visualization result of sent accesses that cause buffer overflows, with IDS information. Here, colored bars denote computers sending suspicious accesses, and gray bars denote computers receiving them. This result denotes that several specific computers attempt various kinds of attacks, because we can observe that the bar corresponding to the computer is painted by three colors.

5. Conclusion

In this paper, we proposed a technique for visualizing the statistics of suspicious accesses brought from network flow log data with IDS and shellcode information. The technique is an extension of our previous work, which forms hierarchy of computers according to their IP addresses, and represents the hierarchy by HeiankyoView. The paper also introduced some examples of visualization by the proposed technique, which proofs the availability of the technique for the visualization of attacks unknown by IDS products.

As a future work in this study, we would like to extend the technique to represent a time change of accesses for visualization of more complicated suspicious accesses.

References

[1] K. Abdullah, C. Lee, G. Conti, J. A. Copeland, and j. Stasko: IDS RainStorm: Visualizing IDS Alarms, Workshop on visualization for Computer Security, pp. 1-10, 2005.

[2] R. Ball, G. A. Fink, and C. North: Home-Centric Visualization of Network Traffic for Security Administarion, ACM Workshop on Visualization and Data Mining for Computer Security, pp. 55-64, 2004.

[3] T. Itoh, H. Takakura, A. Sawada, and K. Koyamada: Hierarchical Visualization of Network Intrusion Detection Data in the IP Address Space, IEEE Computer Graphics and Applications, Vol. 26, No. 2, pp. 40-47, 2006.

[4] T. Itoh, Y. Yamaguchi, K. Koyamada: An Improvement of Nested-Rectangle-Based Hierarchical Data Visualization Technique, Transactions of the Visualization Society of Japan, Vol. 26, No. 6, pp. 51-61, 2006.

[5] H. Koike, and K. Ohno: SnortView: Visualization System of Snort Logs, ACM Workshop on Visualization and Data Mining for Computer Security, pp. 143-147, 2004.

[6] K. Lakkaraju, W. Yurcik, and A. J. Lee: NVisionIP: NetFlow Visualizations of System State for Security Situational Awareness, ACM Workshop on Visualization and Data Mining for Computer Security, pp65-72, 2004.

[7] J. McPherson, K.-L. Ma, P. Krystosk, T. Bartoletti, and M. Christensen: Portvis: A Tool for Port-Based Detection of Security Events, ACM visSEC/DBMSEC, pp. 73-81, 2004.

[8] P. Ren, Y. Geo, Z. Li, Y. Chen, and B. Watson: IDGraphs: intrusion Detection and Analysis Using Histographs, Workshop on Visualization for Computer Security, pp. 39-46, 2005.

[9] T. Takada, and H. Koike: MieLog: A Highly Interactive Visual log Browser Using Information Visualization and Statistical Analysis, LISA XVI Sixteenth Systems Administration Conferance, USENIX Association, pp. 133-144, 2002.

[10] S. T. Teoh, T. J. Jankun-Kelly, K.-L. Ma, S. F. Wu: Visual data Analysis for Detecting Flows and intruders in Computer Network Systems, IEEE Computer Graphics and Applications, Vol. 24, No. 5, pp. 27-35, 2004.

[11] X. Yin, W. Yurcik, M. Treaster, Y. Li, and K. Lakkaraju: VisFlowConnect: NetFlow Visualizations of Link Relationships fpr Security Situational Awareness, ACM Workshop on Visualization and Data Mining for Computer Security, pp. 26-34, 2004.

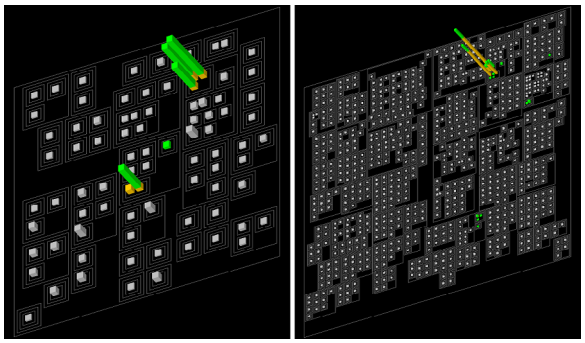

Figure 6. Visualization result of receivers with shellcode information: (Left) As of April in 2006. (Right) As of April in 2007. Same shellcodes are used during one year.

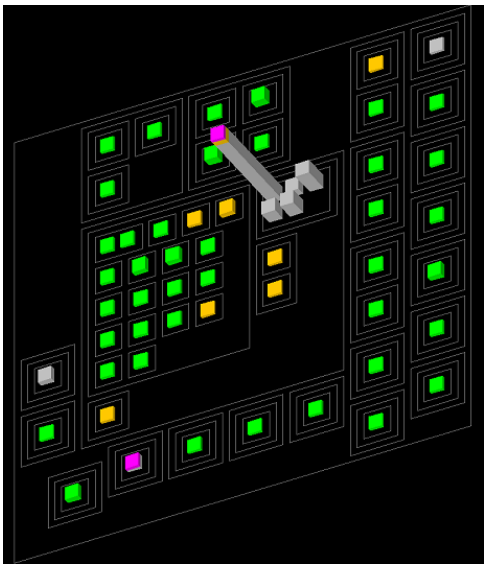

Figure 7. Visualization result of senders with shellcode information: A receiver turned into a sender.

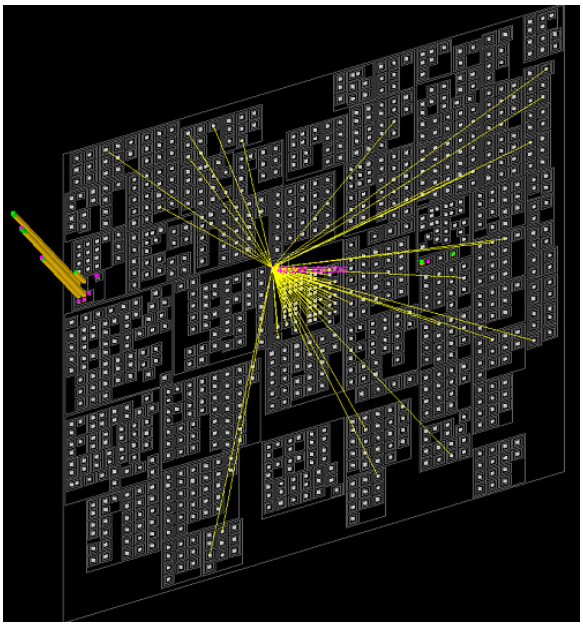

Figure 8. Visualization result of receivers with IDS information: Several computers (in the left) are often attacked. Another computer (in the center) is intensively attacked by computers whose first to third bytes of IP addresses are same.

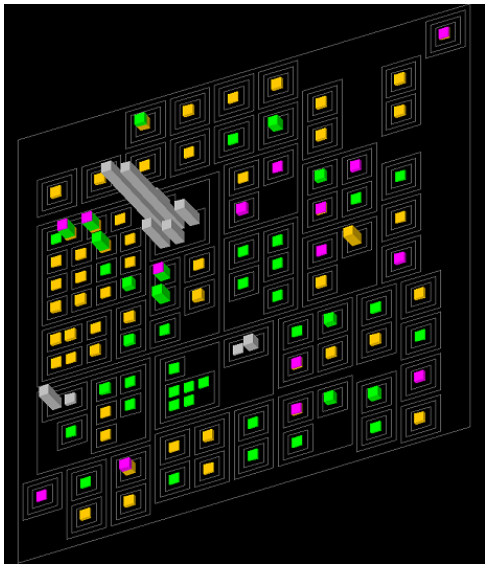

Figure 9. Visualization result of senders with IDS information: One computer attempts various kinds of attacks.