Visualizing Cyber Attacks using IP Matrix

Hideki Koike*
Graduate School of Information Systems
University of Electro-Communications

Kazuhiro Ohno†
Graduate School of Information Systems
University of Electro-Communications

Kanba Koizumi‡
Graduate School of Media and Governance
Keio University

Abstract

An Internet cyber threat monitoring system detects cyber threats using network sensors deployed at particular points on the Internet, statistically analyzes the time of attack, source of attack, and type of attack, and then visualizes the result of this analysis. Existing systems, however, simply visualize country-by-country statistics of attacks or hourly changes of attacks. Using these systems, it is difficult to understand the source of attack, the diffusion of the attack, or the relation between the target and the source of the attack.

This paper described a method for visualizing cyber threats by using 2-dimensional matrix representation of IP addresses. The advantages of this method are that: (1) the logical distance of IP addresses is represented intuitively, (2) Internet address space is visualized economically, (3) macroscopic information (Internet level) and microscopic information (local level) are visualized simultaneously. By using this visualization framework, propagation of the Welchia worm and the Sasser.D worm are visualized.

CR Categories: C.2.0 [Computer Communication Networks]: General—Security and Protection; C.2.3 [Computer-Communication Networks]: Network Operations—Network monitoring; H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces

Keywords: intrusion detection, information visualization, information security, computer virus, Internet worm, virus visualization, worm visualization, Internet forecasting

1 Introduction

Large scale cyber attacks such as Internet worms, automatic scans, or Distributed Denial of Service (DDoS) attacks are major concerns in our society. They can stop Internet services such as e-mail or web services, and as a result they would cause serious damage to our economic activity and public services.

In order to minimize the damage caused by cyber attacks or to prevent such attacks, the world-wide cyber attack monitoring is becoming more and more important. For example, DShield.org [18] analyzes huge amounts of log data collected from the network sensors deployed in the Internet. Then, it visualizes the statistical data such as the number of attacks in each continent as pie charts on a world map (Fig. 1). It also visualizes time-transition of the number of attacks as time-diagram. Although it provides people useful information about current trends in cyber space, it is, however, not appropriate for practical cyber threat monitoring because it lacks detailed information which administrators need to know, such as the attacker's IP address, the spatial diffusion of the attacks, the relation between the attacker and its victims, and so on.

*e-mail: koike@acm.org
†e-mail: ohno@vogue.is.uec.ac.jp
‡e-mail: kanba@sfc.keio.ac.jp

Workshop on Visualization for Computer Security
October 26, Minneapolis, MN, USA
0-7803-9477-1/05/$20.00 ©2005 IEEE.

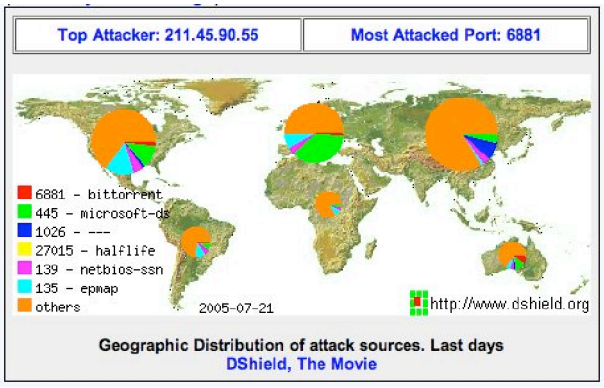


Figure 1: World map visualization at DShield.org. Although current trends of cyber attacks are visible, it lacks detailed information such as the attacker's IP address.

This paper presents a visualization framework and a system that is effectively used in a cyber attack monitoring system. Particularly, we are focusing on automated attacks such as Internet worms or network scans. The next section of this paper discusses how to effectively visualize the propagation of worms by focusing on algorithms of propagation. Section 3 describes the system we developed. Section 4 shows visualizations of the Welchia worm and the Sasser.D worm by using real data captured by our network sensors. Section 5 discusses advantages and limitations of our method. Section 6 describes related work. Section 7 contains our conclusion.

2 Discussion on Visualization Method

2.1 Requirements

In this section, we discuss the requirements for the cyber threat visualization system.

The first requirement is an ability to visualize attacker's IP address. In practical cyber threat monitoring, the source IP of the attack is very important. The administrators could recognize who is attacking their site and take appropriate countermeasures such as blocking the access from that site. Even if the source IP is spoofed, to know that spoofed address is still important.

As we described previously, cyber attack monitoring system often uses intrusion detection systems (IDS) as sensors to detect attacks. Since the IP address of the attacking site is recorded in IDS logs, we can identify the origin of the attack. However, the world map visualization used at current existing monitoring system [18] lacks such information.

When visualizing IP address space, another factor to be considered is the physical limitation of display space. For example, when we want to visualize all 32-bit IP addresses at one time, we have 2^{32} elements to be displayed. Since the minimum unit for displaying one element is a pixel, 2^{32} pixels are required. Even if these elements are laid out in a rectangle, which is the most economi-

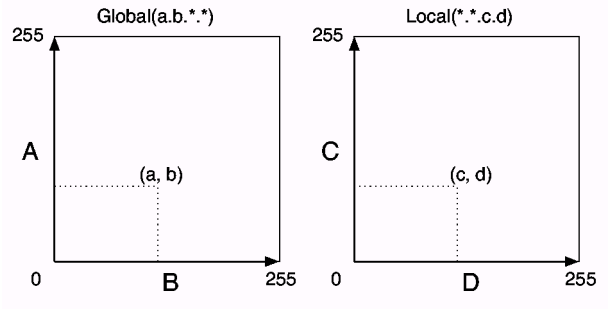


Figure 2: A matrix layout of IP address. In Internet-level monitoring (left), the vertical axis indicates the highest 8 bits of the IP address, and the horizontal axis indicates the next 8 bits. In local level monitoring (right), the vertical axis indicates the 3rd octet of the IP address, and the horizontal axis indicates the 4th octet.

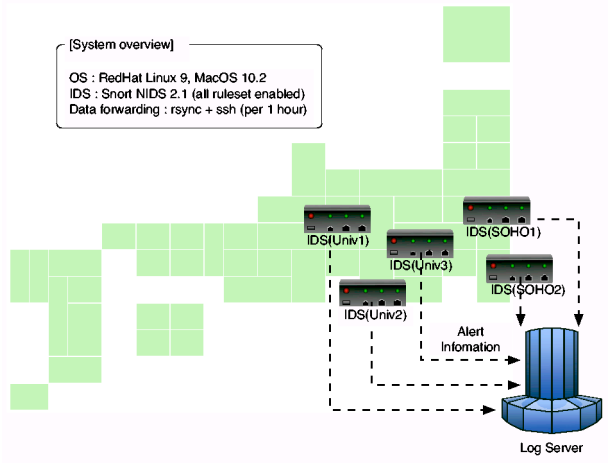


Figure 3: An overview of our cyber attack monitoring system. Currently five network sensors are deployed. Each log is sent to the central log server periodically.

cal layout in 2-D, $65536 \times 65536 > 4 \times 10^9$ resolution is required[1]. This is far beyond the resolution of current existing display monitors (e.g. $1024 \times 768 < 10^6$ in XGA).

Let's think about the worm's propagation algorithm. Recent large-scale worms such as Welchia and Sasser.D use an algorithm called *local scan* that fixes higher 8 bits or 16 bits and searches lower bits sequentially. The time that is necessary to scan every lower bit is quite short, and therefore the worm spreads inside this site very rapidly. That is, if a worm's attack from a certain site is observed, it is highly possible that the worm has already spread inside that site. Therefore, from the viewpoint of cyber attack monitoring, it is not necessary to visualize the exact IP address of the attacking computer, but is usually enough to visualize the higher 16 bits of the IP address, which represent a certain site. On the other hand, in order to know the infection situation inside a certain site, it is enough if the lower 16 bits are visualized.

2.2 IP Matrix

Based on the discussion above, we decided to use 2-D matrix representation of IP addresses as shown in Figure 2. In the figure, the vertical axis indicates the highest 8 bits of the IP address, and the horizontal axis indicates the next 8 bits. In the same way, we use another 2-D matrix where the vertical axis indicates the 3rd octet of the IP address, and the horizontal axis indicates the 4th octet. The former is used for Internet-level monitoring, and the latter is used for local-level monitoring.

Let us consider the case if the worm such as Welchia is observed at the site of IP address a.b.*.*. Welchia performs a local scan first, increments the second octet of the IP address, and performs a local scan again. Therefore, the computers in a.b+1.*.* (site A) have a high possibility of encountering the damage of Welchia. On the other hand, since a+1.b.*.* is not in a direct route of a local scan of Welchia, the possibility of encountering damage by Welchia will be lower than that of site A. Thus, IP Matrix can be used to judge IP address proximity intuitively.

3 Implementation

3.1 Log collection

Our research group started monitoring cyber attacks in May 2003. Currently some network sensors are deployed in the Internet (Fig. 3).

In order to capture cyber attacks, a network-based intrusion detection system Snort [10] is being used. For the signature base, all rules files in the latest version of Snort are included.

Figure 4 shows the detail of a central log server. The logs collected at each network sensor are automatically sent periodically (e.g. in every minute) to the log server.

First, Snort alerts obtained at each sensor are sorted and unified by time. Then, the information used in the visualization is extracted from the unified data. The information used in the visualization is: (1) date, (2) ID of network sensor, (3) alert ID, (4) alert name, (5) source IP address, (6) source port number, (7) destination IP address, (8) destination port number, (9) the number of times the same alerts are generated continuously.

In the log server, Apache Web server is running and communicating with the visualization system using CGI (Common Gateway Interface) programs. Each network sensor transmits the data to the log server using an HTTP protocol that we have defined. Therefore, log information can be obtained even in an environment equipped with a firewall, such as in a typical office network.

3.2 Visualization

3.2.1 System overview

Figure 5 shows a snapshot of IP Matrix's visualization of the log information collected by the log server. The main window is separated into two panes. The left pane is the Internet-level IP Matrix, and the right pane is the local-level IP Matrix.

In the Internet-level IP Matrix, the higher 16 bits of the source of attack is used and is displayed as one pixel. The local-level IP Matrix visualizes the situation of the local network where the network sensor is deployed.

Currently the system obtains attack information from the log server in every minute. The obtained information is soon displayed using pixels on the matrix. Each color of the pixel is decided based on the Snort alert number. Since the Snort has about 2500 alerts in its signature base, we divided them into 8 groups in the order of the alert number and assigned different colors to the group. For example, The alert whose number is between 1 to 400 is colored as red. The alert between 401 and 800 is colored as gray, and so on.

Since one pixel is sometimes too small to be realized when the

[1]This is an overestimation because there are many IP addresses that are not used

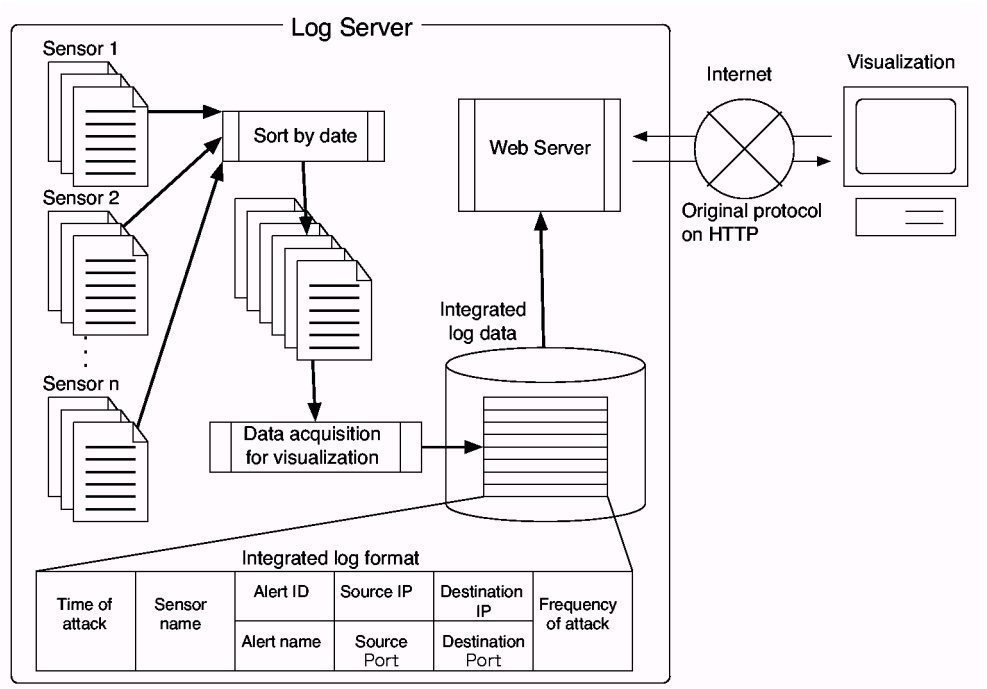


Figure 4: The detail of a central log server. Snort alerts are sorted and unified by time. In the log server, Apache Web server is running and communicating with the visualization system using CGI. Each sensor transmits the data to the server using an HTTP protocol so that the log can be obtained even in an environment with a firewall.

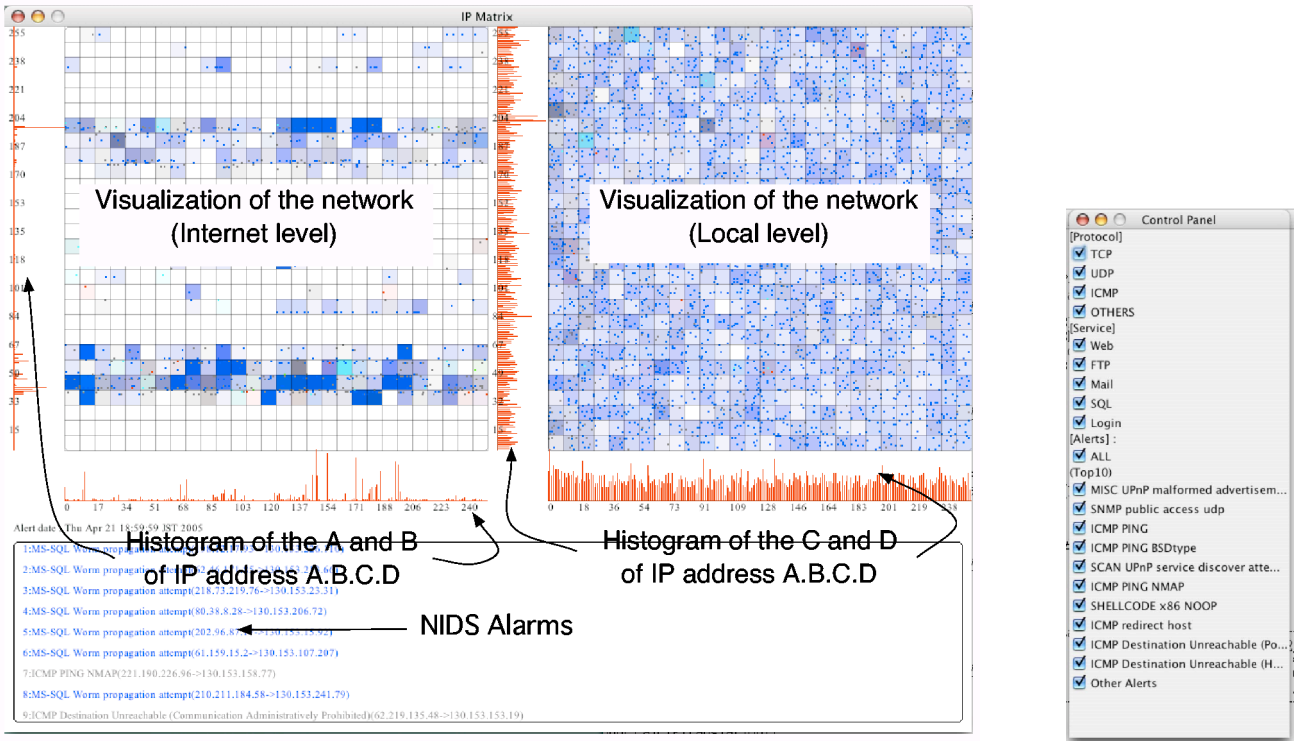


Figure 5: A snapshot of IP Matrix system and its control panel. In the main window (left), the left pane is the Internet-level IP Matrix, and the right pane is the local-level IP Matrix. The color of the pixel indicates the type of each attack. The color of each grid shows the most frequent attack to the sites in the grid. Vertical and horizontal histograms shows relative number of attacks in each address block. When the user click a pixel or drag an area in the matrix, the detail information of attacks of the selected pixel or in the selected area are shown in the bottom of the window.

attack occurs, we use a colored grid as seen in Figure 5. The color of each grid indicates the most frequent attack to the sites in the grid, but the saturation of the grid is much less than that of the pixel for the visibility of each attack.

In addition, there are two histograms shown on the left side and the bottom of each IP Matrix. These show the relative number of attacks occurring at each address block. For example, the bars at the vertical axis show the relative number of attacks at a.*.*.* and the bars at the horizontal axis shows that at *.b.*.*. The system counts the number of attacks and the length of bar is normalize to fit the space. Since we use a pixel for the site and its color is used to show the type of attacks, it is difficult to show the amount of attacks only by using the pixel. By adding these histograms, users can know how much attacks are observed from each site.

The users can also control the displayed information by selecting on the control panel (Fig. 5 right) a focusing attribute such as protocols, services, and Snort alerts.

3.2.2 Interactions

One of the important interactive capabilities of the system is animation. The system displays multiple attacks during a certain period of time simultaneously on the screen. This displayed information is updated periodically. Users can choose the update time such as one month, one day, one hour, and one minute. When the update time is set to one day, it is useful to analyze the worm's activity in a long range. On the other hand, when the update time is set to one minute, it is useful to monitor short-range scans by the worm in almost real time. The users can also play back and forward the animations.

Another interactive capability is the user's ability to display detailed information about the attack by using the mouse. When the user clicks a pixel that indicates a certain attack, the details of the attack are displayed at the bottom of the window (Fig. 5). Since it is sometimes hard to click one pixel by mouse, users can also select multiple pixels by dragging the mouse and can then see all the information regarding these attacks.

4 Examples

4.1 Welchia

Figure 6 shows the infection situation of Welchia, which appeared on August 18, 2003. The figures show how rapidly it spread over the world, resulting in 3,377 incidents the following day and reaching 148,825 incidents in the month of December.

Welchia was programmed to disappear on January 1, 2004, but Figure 6 shows that Welchia has appeared after that date and is still alive. Virus researchers have suggested that its failure to disappear is partially because some systems have not rebooted since January 1, 2004.

Welchia also uses a dictionary to select targets. The dictionary is a list of a popular first octets of IP addresses, such as 61, 202, 203, 210, 211, 218, 219, and 220. Welchia uses these numbers to decide the first octet of the target IP address.

4.2 Sasser.D

Figure 7 shows the distribution of Sasser.D, which appeared on May 3, 2004. We first observed the attack on May 7. It is visually clear that the worm was spreading over the Internet. Even though the infection method was similar to that of Welchia, the number of infected sites was smaller than in the case of Welchia. Virus researchers reported that one of the reasons for this was that people realized the importance of applying a security patch after the disaster of Welchia. Many systems had already had security patches applied, and therefore Sasser.D infected a smaller number of computers.

Another reason for the lower number of attacks was the difference in the port number used for intrusion. Welchia used port 135 for intrusion. Since this port is always listed in the Microsoft Windows environment, there was a high possibility of doing damage. On the other hand, Sasser.D used port 445. Since this port is not necessarily always open, there was less damage than in the case of Welchia.

5 Discussion

5.1 Advantages

One of the advantages of our visualization framework is that it is possible to visualize IP address space economically. In this framework, we used a matrix representation of the higher 16 bits of IP addresses and that of lower 16 bits. This helps to economically visualize Internet-level monitoring and local-level monitoring on a physically limited display space without showing unnecessary IP addresses.

The second advantage is that it is possible to visualize intuitively the proximity of an attack by comparing IP addresses. In many cyber attacks, the IP address is used for propagation. In our system, the distance between the attacker and the victim can be measured by seeing the position of the attack on the matrix. Therefore, the user might be able to prepare for the attack before the attack strikes. For example, the propagation of Welchia is represented as a horizontal line on the matrix. If administrators find their site is on the extension of the line, they might be able to take appropriate actions before the attack reaches that IP address.

5.2 Limitations

There are some issues to be solved in our system such as shown below.

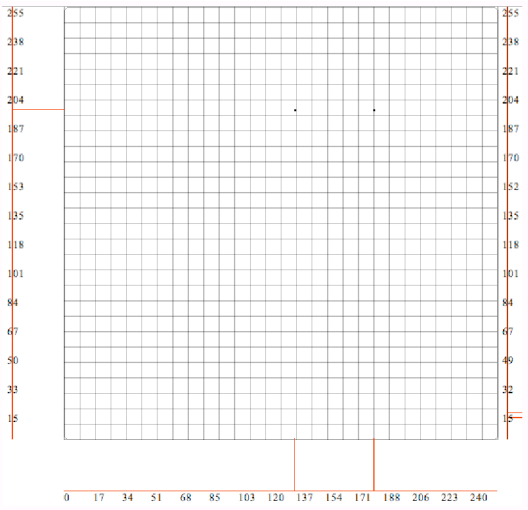
unused IP addresses: As we can see from the visualization, there are many unused IP addresses, and the display space is used wastefully. To use display space economically, some visualization techniques should be applied. However, it is not appropriate to eliminate non-existent IP addresses because there could be attacks from the spoofed IP addresses. Moreover, a 256×256 matrix representation represents the position of each IP address so that it can be intuitively understood.

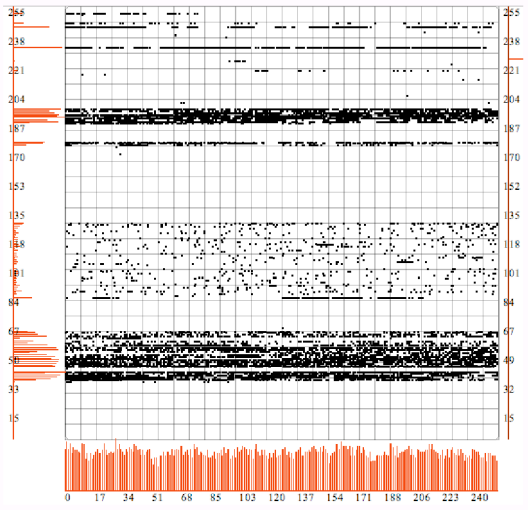
limitation of expressiveness: Currently our system represents one site by one pixel. It can just show existence of the attack and add some information using colors. To address this issue, displays with higher resolution could be used. For example, if 4×4 pixels are assigned to a site, 1024×1024 resolution is required. The currently available SXGA monitor has such resolution.

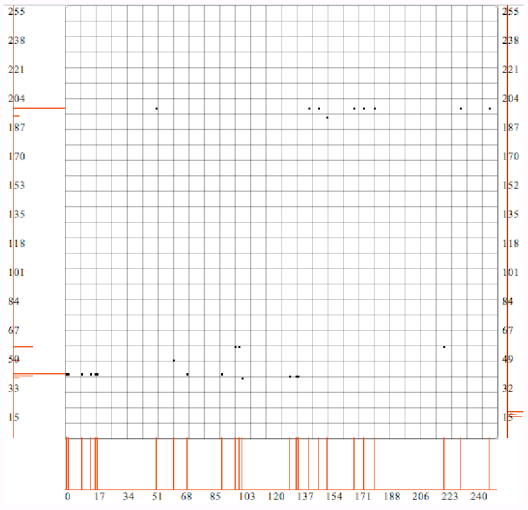
mail virus: This paper focuses on the worms using automatic scan. There are other type of viruses that use other channels for infection. For example, in the case of a mail virus that is attached to electronic mail, the next target is selected by using the user's address book, and these addresses have no relation to IP addresses. In this case, IP Matrix is not used for prediction, but it is still helpful for understanding security trends in cyber space.

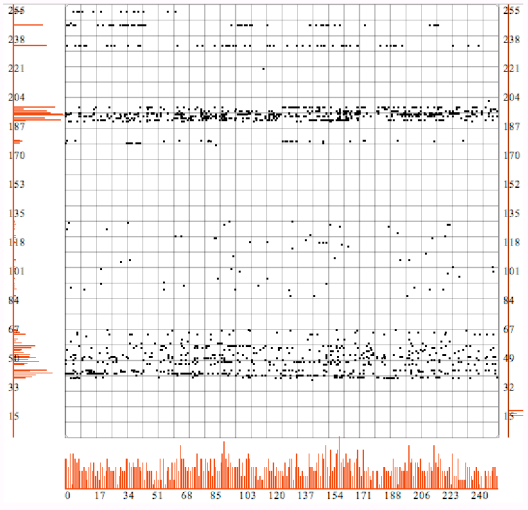
5.3 IP Matrix 3D

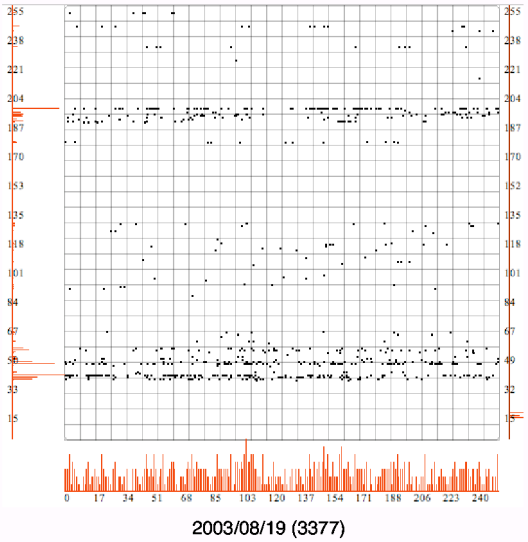
Since our system displays attack information on 2-D matrices, the number of attacks at each point is not clear. However, the number of











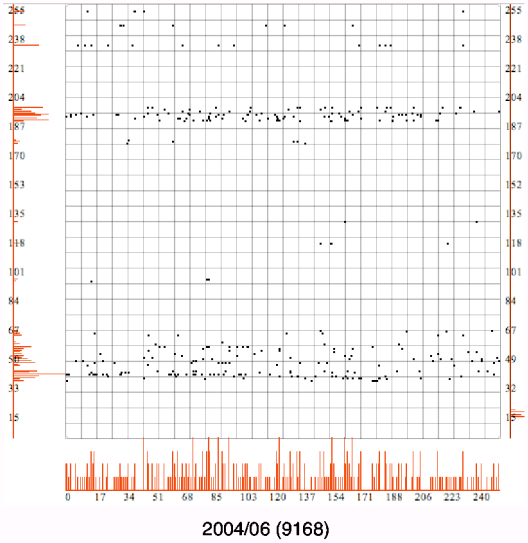


Figure 6: Infection of Welchia.

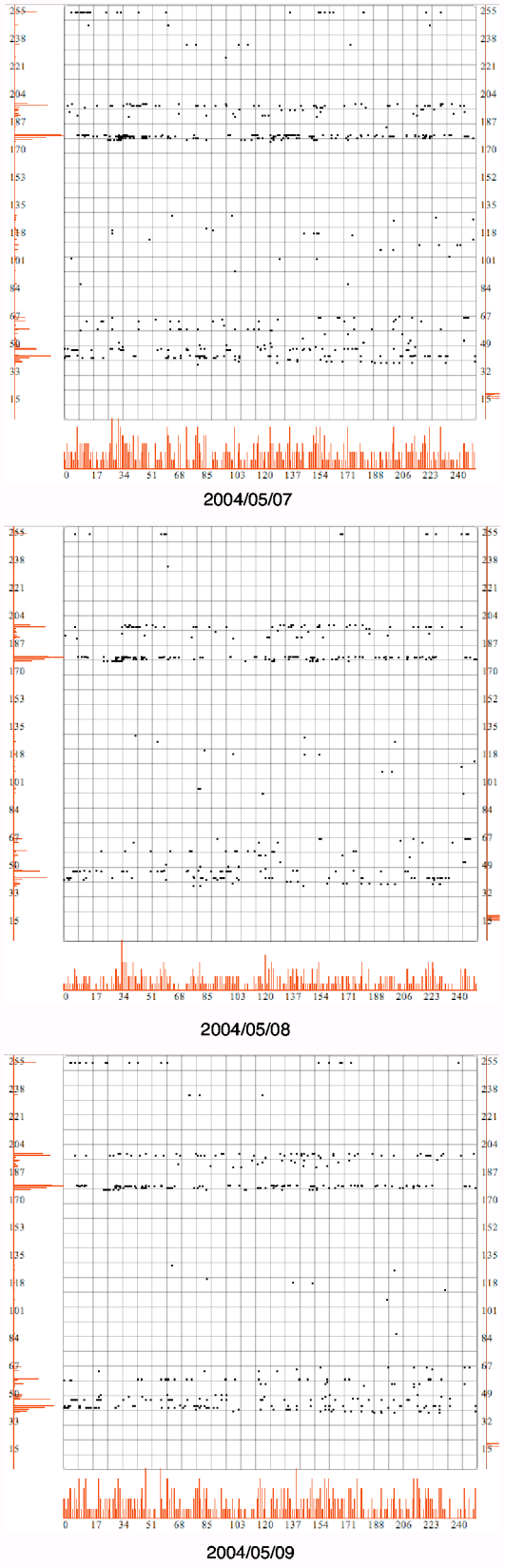


Figure 7: Infection of Sasser.D.

attacks is important and it should be visualized implicitly. In order to solve this issue, we experimentally developed the 3-dimensional version of the IP Matrix, which is called IP Matrix 3D.

Figure 8 shows a screen shot of IP Matrix 3D. This system arranges multiple 2-D IP Matrices in layers. In each layer, one of the following four attributes is assigned to the vertical axis. (1) the number of attacks, (2) degree of attack propagation, (3) the type of attacks (Snort alert ID), and (4) the operating system and its version. Moreover, the attack information of all layers is projected on the bottommost layer. Users can rotate, zoom in/out the visualization to get better position to observe. However, there is still an occlusion problem in such 3-D visualization.

There are two different kinds of arrangements in IP Matrix 3D as shown in Figure 9. In Figure 9(a), each layer illustrates the statistical data for the same attack at each network sensor. This is used to compare the same attack at different locations. In contrast, in Figure 9(b), each layer illustrates statistical data for different attacks at the same network sensor. This is used to compare different attacks at one location.

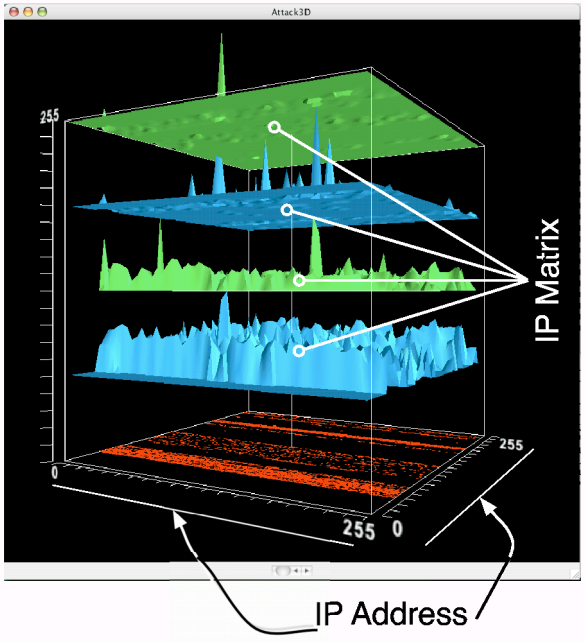


Figure 8: IP Matrix 3D. Multiple IP Matrices are stacked to height direction. In each IP Matrix, the height of each site represents the relative number of attacks. Users can rotate and zoom in/out the visualization using mouse.

5.4 Toward security forecast

One of the important challenges in cyber attack monitoring is prediction of attacks. If the attack is predicted, the damage of the attack could be minimized. Although IDS can detect attacks in real time, it cannot predict the attacks.

We believe our system might be used for prediction, because the user can understand the spatial diffusion of attacks as well as their transition in time. As we described previously, recent worms use a random scan and a local scan. When one site is infected by a worm, it is highly possible that its logical neighborhoods are the next targets. This infection pattern will be visually shown as distinct visual signatures in our system. By utilizing this visual pattern, administrators might be able to predict when their site will be attacked.

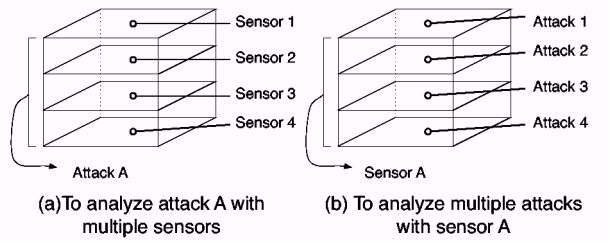


Figure 9: Two different arrangements of IP Matrix to height direction. (a) would be used to compare the same attack at different locations. (b) would be used to compare different attacks at one location.

6 Related Work

As we described previously, a variety of cyber attack monitoring systems are in service today [18, 11]. They are providing information on cyber attacks using country-by-country statistics visualization or time diagrams.

One Internet visualization project is being done in CAIDA (Cooperative Association for Internet Data Analysis) at UC San Diego Supercomputer Center. These researchers developed some interesting visualization systems such as Skitter [2]. Particularly their work on visualizing IPv4 address space utilization [3] should be noted. They visualized IPv4 address space as a 4096×4096 2-D matrix. Although their visualization looks very similar to ours, the layout algorithm is different. For example, they represent 24-bit address information as one pixel, while we represent 16-bit address information as one pixel. Also we used 2 matrices, Internet-level matrix and local-level matrix. What is more important, CAIDA's visualization focuses on the connectivity of CIDR (Classless Inter-Domain Routing) which are represented as pixels. In contrast, our system focuses on security information. Also, it is important for a monitoring system that the visualization fits in the screen space without scrolling.

The Spinning Cube of Potential Doom [8] presented the similar visualization approach to ours. It maps source IP addresses to Z axis (depth), destination IP addresses to X axis (horizontal), and port numbers to Y axis (vertical). By using three axis, it tried to provide as much information as possible. Although the system is good for getting impressions of what is going on in the Internet, it is not adequate to be used in practical monitoring task because the administrator cannot understand the exact 3-D coordinates of each plot. In addition, in order to let all source IP addresses be fit in Z axis, each source IP address has to share its Z coordinate with its neighbors (Otherwise it requires 2^{32} pixels along Z axis!). This is not adequate for monitoring the Internet.

The SIFT (Security Incident Fusion Tools) project visualizes security information. NVisionIP [6] visualizes both the transfer information and the connection information in a small-scale network. VisFlowConnect [9] visualizes traffic between an internal network and an outer network using a line. This system focuses on a local area network and is not appropriate for use in large-scale network monitoring.

Our research group also developed some visualization systems for computer security. Tudumi [16] visualizes connection to a particular server in one 3-D visualization by taking several logs (syslog, sulog, wtmpx, etc.) as input. In Tudumi, the connecting hosts are categorized by network domain and displayed as a stack of circles. It does not have real-time monitoring capability. SnortView [7] visualizes Snort logs using a time-diagram. It shows each attack as an icon whose shape and color indicate the type and priority of the attack. It overlays a histogram of the attack on each icon, which helps to avoid overwhelming the screen large numbers of the same icon.

7 Conclusion

This paper proposed a visualization framework for cyber attack monitoring. Based on the framework, the visualization system named IP Matrix was developed. IP Matrix can visualize a large number of IP addresses economically. Also, it can visualize the logical proximity of IP addresses, which is important in worm propagation algorithms.

For future work, we are planning to add more useful interaction capabilities to the system, such as zooming interface. Using the zooming interface, users click a particular pixel on the matrix and see much more detailed information near the site. In the zooming window, we can use much more pixels to show detailed information of each site. It would be also interesting when IP Matrix, which visualizes spatial information of cyber attacks, is combined with SnortViews [7], which visualizes temporal information of cyber attacks.

References

[1] Robert Ball, Glenn A.Fink, Chris North, Home-Centric Visualization of Network Traffic for Security Administration, Conference on Computer and Communications Security (CCS2004), Proc. of the 2004 ACM workshop on Visualization and data mining for computer security (VizSEC'04), pp.55-64, 2004.

[2] AS Internet graph, CAIDA, http://www.caida.org/analysis/topology/as_core_network/AS_Network.xml

[3] IPv4 Address Space Utilization, CAIDA, http://www.caida.org/outreach/resources/learn/ipv4space/

[4] Brent N. Chun, Jason Lee, Hakim Weatherspoon, Netbait: a Distributed Worm Detection Service, Intel Research Berkeley Technical Report IRB-TR-03-033, 2003.

[5] Gregory Conti, Kulsoom Abdullah, Passive Visual Fingerprinting of Network Attack Tools, Conference on Computer and Communications Security (CCS2004), Proc. of the 2004 ACM workshop on Visualization and data mining for computer security (VizSEC'04), pp.45-54, 2004.

[6] Kiran Lakkaraju, William Yurcik, Adam J.Lee, NVisionIP: NetFlow Visualization of System State for Security Situational Awareness, Conference on Computer and Communications Security (CCS2004), Proc. of the 2004 ACM workshop on Visualization and data mining for computer security (VizSEC'04), pp.65–72, 2004.

[7] Hideki Koike, Kazuhiro Ohno, SnortView: Visualization system of Snort Logs, Conference on Computer and Communications Security (CCS2004), Proc. of the 2004 ACM workshop on Visualization and data mining for computer security (VizSEC'04), pp.143-147, 2004.

[8] Stephen Lau, The Spinning Cube of Potential Doom, Communications of the ACM, Vol.47, Issue 6, pp.25-26, 2004.

[9] Xiaoxin Yin, William Yurcik, Yifan Li, Kiran Lakkaraju, Cristina Abad, VisFlowConnect: Providing Security Situational Awareness by Visualizing Network Traffic Flows, Workshop on Information Assurance (WIA04) held in conjunction with the 23rd IEEE International Performance Computing and Communications Conference(IPCCC 2004) , 2004.

[10] Martin Roesch, Snort - Lightweight Intrusion Detection for Networks, Proc. of the 1999 USENIX LISA conference, pp.229–238, 1999.

[11] Internet Storm Center, SANS, http://isc.sans.org/

[12] SecureScope, Secure Decisions, http://www.securedecisions.com/

[13] Security Incident Fusion Tools, http://www.ncassr.org/projects/sift/

[14] Stuart Staniford, Vern Paxson, Nicholas Weaver, How to Own the Internet in Your Spare Time, Proc. of the 11th USENIX Security Symposium, pp.149–167, 2002.

[15] Symantec Corp.,
http://www.symantec.co.jp/
[16] Tetsuji Takada, Hideki Koike, Tudumi: Information Visualization System for Monitoring and Auditing Computer Logs, Proc. on Information Visualization (IV2002), IEEE/CS, pp.570–576, 2002.
[17] Soon Tee Teoh, Kwan-Liu Ma, and Felix Wu, A visual exploration process for the analysis of Internet routing data, Proc. of IEEE Information Visualization, 2003.
[18] Distributed Intrusion Detection System, DShield.org
http://www.dshield.org/