

2 Network Security Incident Response Technology

2-1 Overview of R&D Activities on nictcr

NAKAO Koji and INOUE Daisuke

The Internet has faced various security threats ever since it became widespread. The malicious activities of malwares are spread all over the Internet and often lead to serious security incidents that can cause significant damages to both infrastructures and end users. There are two main approaches to fight against malwares: macroscopic (i.e., network monitoring) and microscopic (i.e., malware analysis) approaches. We have developed the Network Incident analysis Center for Tactical Emergency Response (nictcr), which integrates both the macroscopic and microscopic approaches in order to promptly grasp the malicious activities and their root causes. In this paper, we describe a whole overview of the nictcr and its sub-systems: macro analysis system, micro analysis system, and macro-micro correlation analysis system.

Keywords

Security incident, Network monitoring, Malware analysis, Correlation analysis

1 Research background

The Internet, which has turned into the worldwide social infrastructure, has largely contributed to our social and economical activities, and has brought such irreversible changes to every corner of modern society that it would be impossible to return to the times before its diffusion. On the other hand, as the Internet has been expanded, security threats regarding the Internet have been increasing. The variety of security incidents occurring on a daily basis includes intrusion attacks to web services, denial-of-service (DoS) attacks, leakage of personal information or organizations' confidential information, and fishing induced by a huge amount of spam mail. Malware^{*1} which infects users' machines plays a primary role in many of those incidents.

In the late 1980s to the early 1990s before the word malware had been commonly used, generally malicious programs such as viruses

and worms had been created and distributed for attackers' pleasure and/or ostentation, and those problems had caused such phenomena easily recognized by users as screen displays which notify users of infection, performance deterioration of the machine, and data destruction. However, since the late 1990s malware has been used as a tool for organized crime aiming at money exploitation, and accordingly its stealth capability has been improved and its functions have been further advanced. Since around 2004 "botnets" have appeared on the Internet, which is a wide-scale infected host group which can be controlled by attackers at will, due to the technological innovation which allows remote simultaneous operations through Internet Relay Chat (IRC). At present, botnets

*1 This is a generic term for such software that performs harmful activities (e.g., information leakage, data destruction and computer infection) as virus, worm, Trojan horse, spyware, and bot. It is a coined term from the combination of "malicious" and "software".

are a major cause of various security incidents including massive spam transmission, distributed denial-of-service (DDoS) attacks, and large-scale infection activities.

In order to deal with those security incidents caused by malware, security technologies resorting to a local approach have been introduced, such as antivirus software and personal firewalls by end users and intrusion detection systems (IDSes) and intrusion prevention systems (IPSeS) by organizations. However, the security of the Internet itself as social infrastructure cannot be ensured merely with a local approach, and security incidents must be approached from a panoramic viewpoint. That is, it has been necessary to establish a framework which swiftly and correctly recognizes the entire picture of security incidents which occur on the Internet, which is a vast network, identifies their causes, and finds effective solutions.

The Network Security Incident Response Group of the Information Security Research Center aims at early detection, cause investigation, and presentation of solutions in terms of security incidents which widely affect the Internet, and has conducted research regarding the Network Incident analysis Center for Tactical Emergency Response (nicter)[1]-[3]. The primary nictcr research is to make a real-time estimate of the types of malware which are currently present on the Internet, by collecting attack information through Internet monitoring at a number of points, analyzing the connected information (macro analysis), using analysis technologies (micro analysis) regarding malware specimens collected with honeypots or similar, and using correlation analysis technologies which merge these technologies. This is likely to provide measures for early solution in terms of diffusion of unknown malware by zero-day attacks*².

This paper outlines network monitoring and malware analysis in Chapter 2, describes in Chapter 3 the nictcr research which merges them and the functions of its subsystems, namely macro analysis system, micro analysis system, and correlation analysis system, and

presents conclusions in Chapter 4.

2 Network monitoring and malware analysis

Approaches for analysis of security incidents caused by malware can be broadly divided into macro approaches and micro approaches. A macro approach is used to analyze traffic information collected through network monitoring and recognize incidents' phenomena from a broad perspective. A micro approach is used to analyze detected malware and microscopically clarify the movements of malware which causes security incidents. Both the macro and micro approaches need sensors which collect data to be input, namely traffic information and malware specimens. Generally, these sensors are installed in IP address spaces called darknets.

This chapter first describes darknets and various sensors. Then it introduces domestic and overseas darknet monitoring projects as an example of macro approaches and domestic and overseas malware analysis projects as an example of micro approaches.

2.1 Darknets and sensors

A darknet refers to IP address space which can be reached and has not been used on the Internet. As for ordinary use of the Internet, packet transmission to IP addresses not used is not highly likely, but in practice, a large amount of packets reach the darknets. Many of those packets are attributed to devious activities on the Internet: e.g., scan to search for the next infection targets by malware which spreads infection through networks; UDP packets*³ with malware in their payload; rendezvous packets by which different malware programs establish a P2P network together;

*² This refers to an attack which exploits the vulnerabilities of OSes and applications before security patches for fixing the vulnerabilities are released. Even systems having the latest security patches applied cannot prevent zero-day attacks, and thus massive incidents can be resulted.

*³ For example, the data size of malware called SQLSlammer is so small (376 bytes) that it can be placed in the payload of a single UDP packet.

backscatter, which is a response (SYN-ACK) from a server that is under a DDoS attack with a feigned source IP address. Therefore, it is possible to recognize the trend of devious activities which are present on the Internet, through passive monitoring of packets which reach the darknet. It is also possible to catch more detailed attack information and malware specimens by sending appropriate responses to packets actively.

One advantage of darknet monitoring is to be able to regard and analyze all packets as illegal ones without the necessity of classifying traffic as legal or illegal. Another advantage is that communication privacy problems can be avoided because IP addresses which have been possessed by observers and have not been used are monitored at network terminations.

Darknet monitoring requires installation of a server machine called a sensor, for packet collection and response. Sensors are classified as below in accordance with the levels of responses to packet sources.

- **Black hole sensor:** Sensor which returns no response to packet sources. The maintenance of this sensor is easy, and the sensor is suited to large-scale darknet monitoring. Because it sends no response, detection of its presence from the outside is difficult. However, while it allows monitoring of scans performed at the initial stages of malware's infection activities, it does not allow monitoring of the subsequent movements.
- **Low-interactive sensor:** Sensor which returns a certain level of responses to packet sources. This type of sensor includes a sensor which returns SYN-ACK packets to SYN packets of TCP and a low-interactive honeypot which imitates an OS's known vulnerabilities. The presence of this type of sensor is easily detected due to factors such as listening ports and response trends, and thus it is not suited to use with a large-scale darknet with continuous addresses.
- **High-interactive sensor:** Actual machine or sensor (so called high-interactive honeypot) which returns responses that conform to it. This type of sensor allows acquisition of various information, such as the action history of illegal access attempted by attackers. However, it is not suited to large-scale operations because its costs for safe operation are very high.

2.2 Darknet monitoring project

This section outlines domestic and overseas major darknet monitoring projects, which have taken macro approaches for incident analysis.

- **Network telescope:** Darknet monitoring project by CAIDA (Cooperative Association for Internet Data Analysis) in the U.S. It has monitored darknets with more than 160,000 addresses and has released data sets of traffic by backscatter and worm.
- **Internet motion sensor:** Monitoring project of large-scale darknets with more than 17 million addresses, by the University of Michigan in the U.S. It attempts to establish TCP connections by returning SYN-ACK from a sensor to a part of the TCP SYN packets monitored and to collect and analyze the payload of the first packets after establishment of a connection.
- **Leurre.com:** Information acquisition and analysis project with distributed honeypots, by Eurecom in France. While the number of IP addresses which are monitored is relatively small, observation locations have been widely dispersed throughout the world. The first-generation, Leurre.com v1.0, used the low-interaction sensor Honeyd, but the second-generation, Leurre.com v2.0, uses SGNET to improve the information acquisition capability.
- **REN-ISAC:** Security information sharing and analysis project by Research and Education Networking (REN) in the U.S.

It has analyzed traffics monitored with Internet2 and released observation results.

The network monitoring projects currently in progress in Japan include ISDAS by JPCERT/CC, @police by the National Police Agency, MUSTAN by the Information-technology Promotion Agency, and WCLSCAN by Mitsubishi Research Institute and other institutions.

2.3 Malware analysis project

The malware analysis methods can be broadly divided into two approaches, dynamic analysis and static analysis. Dynamic analysis is also called black box analysis, and it executes malware specimens on machines which serve as sacrifices in order to analyze, for example, the machines' internal operations and network accesses. Static analysis is also called white box analysis, and it disassembles executable codes of malware and analyzes the malware's functions and characteristics at assembly level in detail. While automation of analysis is relatively easy for dynamic analysis, manual analysis by analysts having advanced skills is common for static analysis because recent malware is equipped with code obfuscation and anti-debug functions which impede disassembly.

The descriptions below are about projects which have automated dynamic analysis of malware and provided analysis services. All of these systems recognize movements of malware by monitoring malware's API calls and network accesses.

- **CWSandbox:** Dynamic analysis system by University of Mannheim in Germany. It executes malware on Windows XP on a virtual machine (VMware Server). It allows connection to the Internet by the malware being analyzed.
- **Anubis:** Dynamic analysis system by Vienna University of Technology in Austria. It executes malware on a PC emulator called QEMU. It allows connection

to the Internet by the malware being analyzed.

- **Norman Sandbox:** Dynamic analysis system by Norman Corporation in Norway. It executes malware on a Windows clone operation system. It does not allow connection to the Internet by the malware being analyzed, but it has prepared dummy DNS and Web servers in the analysis environment.

3 Network Incident analysis Center for Tactical Emergency Response (nicter)

As previously mentioned, darknet monitoring (macro approach) and malware analysis (micro approach) have been conducted by various institutions, in terms of R&D and practical operations. Pursuing the causes of security incidents certainly needs a combination of the both approaches. However, many of the darknet monitoring projects have focused on quantitative traffic analysis and many of the malware analysis projects have focused on the analysis of malware's functions, and there is such a large distance between the two approaches that it has made it difficult to identify the causes of phenomena monitored on the Internet at present.

Therefore, the Network Security Incident Response Group of the Information Security Research Center has conducted research regarding the Network Incident analysis Center for Tactical Emergency Response (nicter), aiming at early detection, prompt cause investigation and presentation of solutions in terms of security incidents which widely affect networks, by merging darknet monitoring and malware analysis.

The nicter has two analysis passes: a macro analysis system which analyzes events collected through extensive darknet monitoring and detects security incidents from among the events and a micro analysis system which collects and analyzes malware specimens and identifies their movements (Fig.1). With

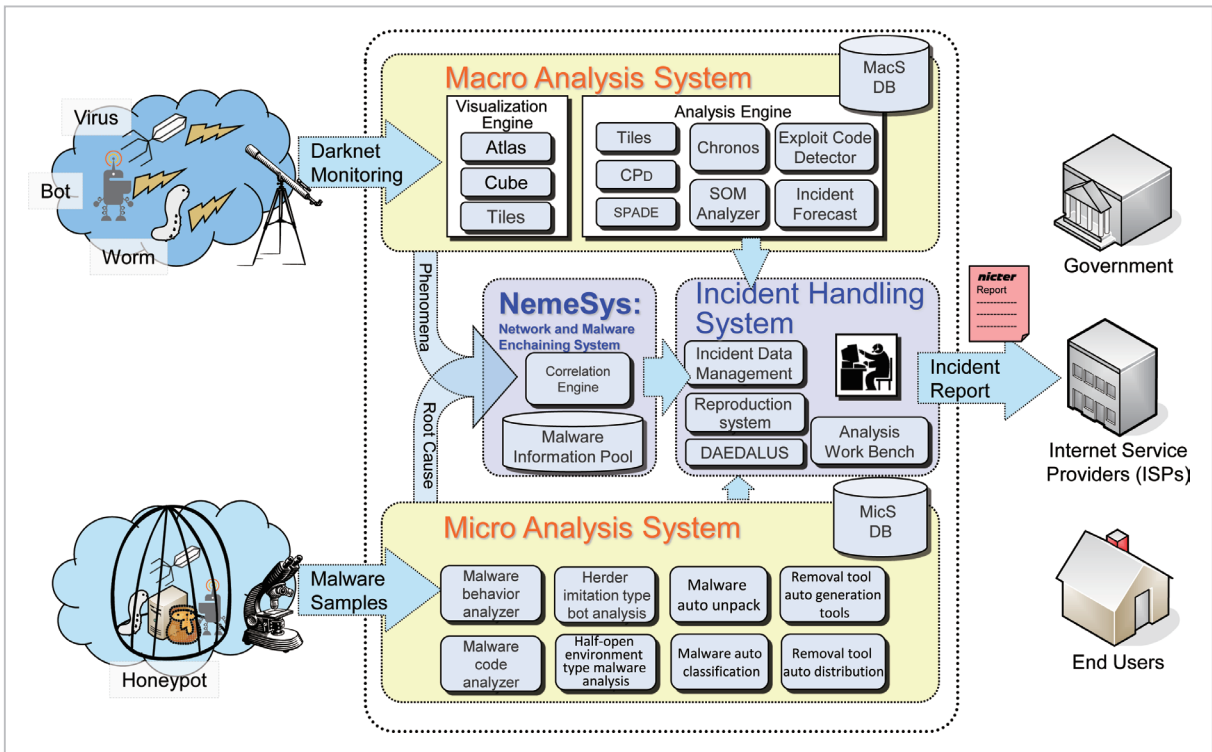


Fig.1 Entire view of the nictcr

respect to analysis results from the two systems, the correlation is analyzed with a correlation analysis system, and incidents' phenomena and causes are related. Because a macro analysis system can identify phenomena of incidents which have occurred on networks and a micro analysis system can recognize movements of malware which has possibly caused incidents, it is possible to identify the causes of incidents which are currently present by comparing the analysis results from both systems and to come up with measures for dealing with identified malware. Analysis results from a macro analysis system, micro analysis system, and correlation analysis system are sent to an incident handling system, which provides analysts with an integrated web interface and visualization interface, so that analysts will be able to make detailed reports about the incidents later.

In this way, through concretization of a concept which merges a macro approach and micro approach, it is possible not only to present statistical data about darknets of the traffic monitored but also to provide the government agencies, ISPs, and general users with effective

and instantaneous incident reports and alert information which show the causes of incidents and solutions. In 3.1 to 3.3, the nictcr's macro analysis system, micro analysis system, and correlation analysis system are described.

3.1 Macro analysis system

The main input of a macro analysis system is darknet traffic which is monitored with black hole sensors installed at multiple monitoring points. At present the nictcr monitors more than 140,000 IPv4 addresses which are not used. Figure 2 presents results (March 1 to 31, 2011) of monitoring of approximately 78,000 IPv4 addresses among the darknets possessed by the nictcr, and shows the number of packets which have reached the darknets and the number of sources' unique hosts (daily number of unique source IP addresses). The figure shows that an average of approximately 300,000 hosts per day transmitted an average of approximately 5.9 million packets a day to darknets.

In this way, collection and analysis of traffic which reaches darknets makes it possible to

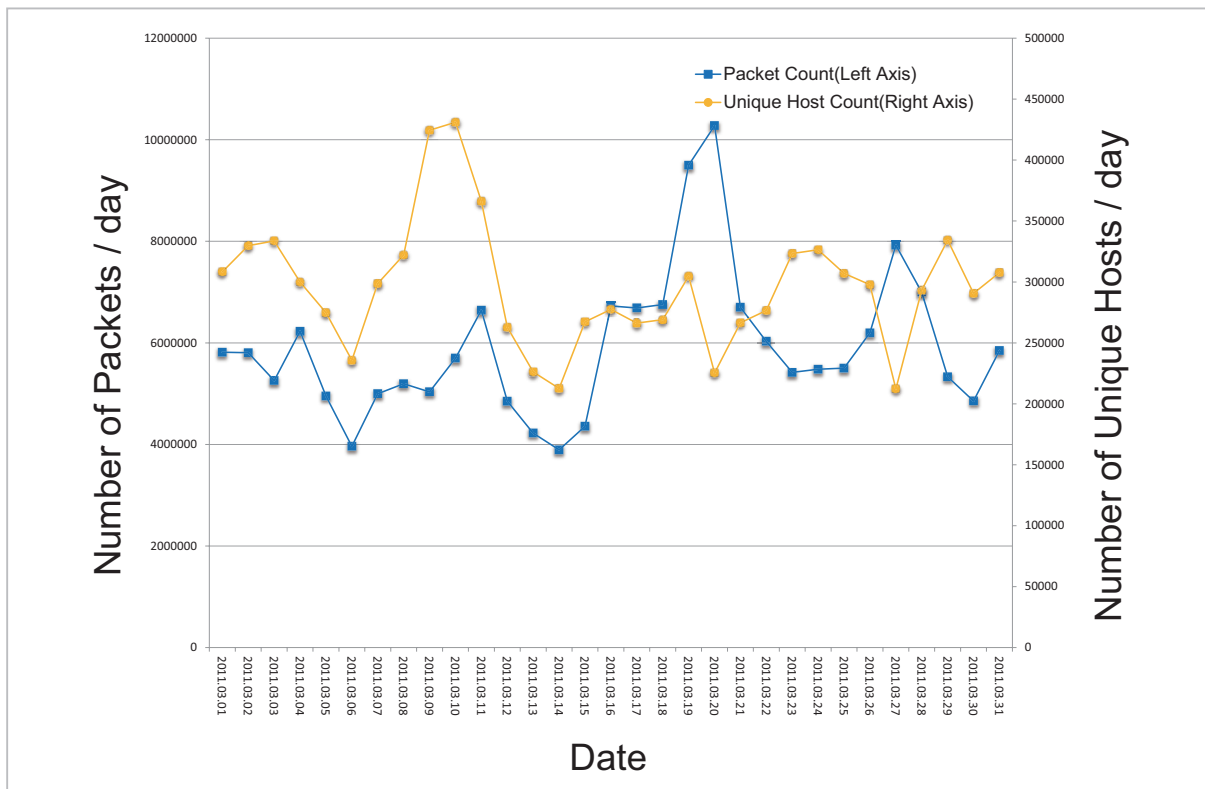


Fig.2 The nictet's darknet monitoring results (78,000 addresses)

recognize macro trends of attack activities on wide-area networks. A macro analysis system is composed of a visualization engine which helps analysts with intuitive incident analysis and an analysis engine which performs auto traffic analysis. Some of these engines are outlined below.

3.1.1 Visualization engine

(1) Atlas

Atlas (Fig.3) is a visualization engine which presents a real-time animated display of darknet traffic on a world map. It identifies countries*⁴ to which source and destination IP addresses belong, for each of the packets which reach darknets, and shows transmission of packets from the capitals of the sources' countries to the capitals of the destinations' countries with animation, so that global malware activity trends can be recognized instinctively. The colors of the individual packets represent packet types*⁵, and the packet trajectory height is proportional to the port number size (logarithmic axis). The engine allows analysts to perform interactive operations such

as viewpoint change and enlargement/reduction through mouse operations and display of detailed information (Fig. 4) by clicking packet objects.

(2) Cube

Cube (Fig.5) is a visualization engine which presents an animated display of packets reaching darknets, in cubes floating in three-dimensional space, based on various information of sources and destinations. Where the vertical axes of a cube represent the sources' and destinations' IP addresses, and the horizontal axes of a cube represent source and destination port numbers, packets are sent from sources (left plane of Fig.5) to destinations (right plane of Fig.5), so that movements such as scan and backscatter can be visualized. Like

*4 IP addresses and latitude/longitude are mapped using MaxMind's GeoIP City Database.

*5 Blue (TCP SYN); yellow (TCP SYN-ACK); green (TCP ACK); pink (TCP FIN); purple (TCP RST); orange (TCP PUSH); light blue (TCP OTHER); red (UDP); white (ICMP) (This also applies to the colors for Cube and Tiles described later.)

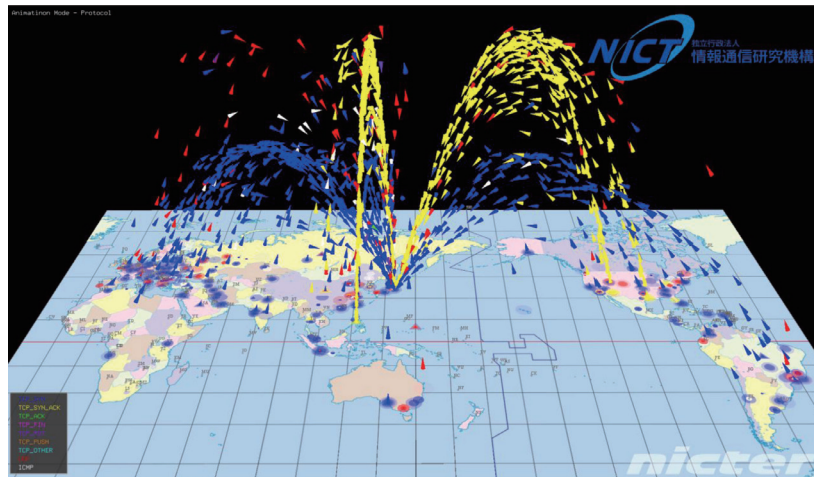


Fig.3 Atlas

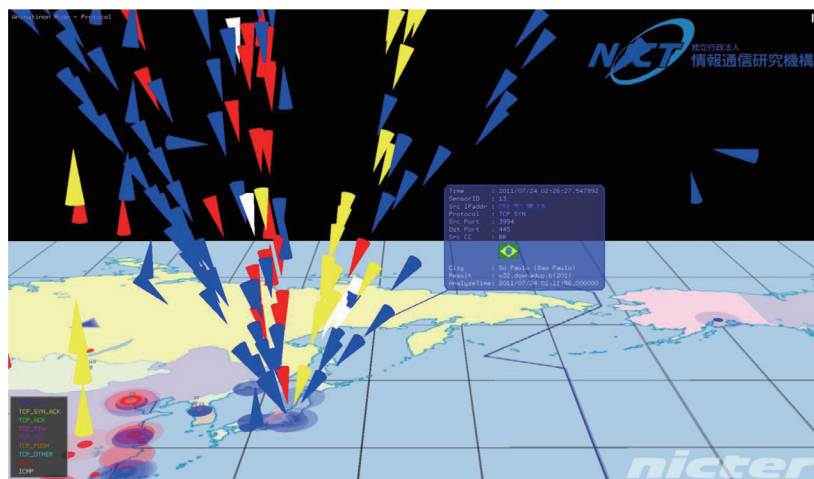


Fig.4 Atlas (Detailed packet information display)

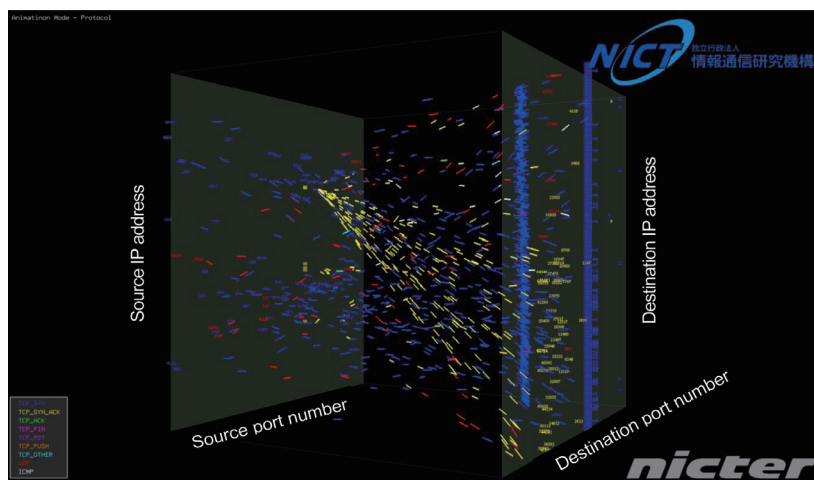


Fig.5 Cube

Atlas, Cube allows operations such as viewpoint change and enlargement/reduction with a mouse and display (Fig. 6) of detailed information of packets. It enables analysts to recognize attacks from source hosts in real-time and allows them to use triggers for starting detailed analysis.

(3) Tiles

Tiles (Fig. 7) is a visualization engine which presents real-time display of analysis results of a behavior analysis engine described later. Each of the tiny tiles of Fig. 7 represents an individual source hosts' movements and is updated to the latest analysis result. The back of the tiles shows the national flags of coun-

tries to which source hosts belong. Individual tiles are visualized as shown in Fig. 8, using time of packets sent from source hosts for 30 seconds, source/destination port numbers, and destination IP addresses. Each of the packets is expressed with a single line which connects between a source (left plane) and a destination (right plane). Figure 8 represents a typical pattern of network scan, incrementing the source port number while transmitting TCP SYN packets to a single destination port of multiple destination IP addresses. When a tile is clicked, tiles having the same pattern as the tile are highlighted in white (Fig. 9). The results of auto classification of scan patterns by

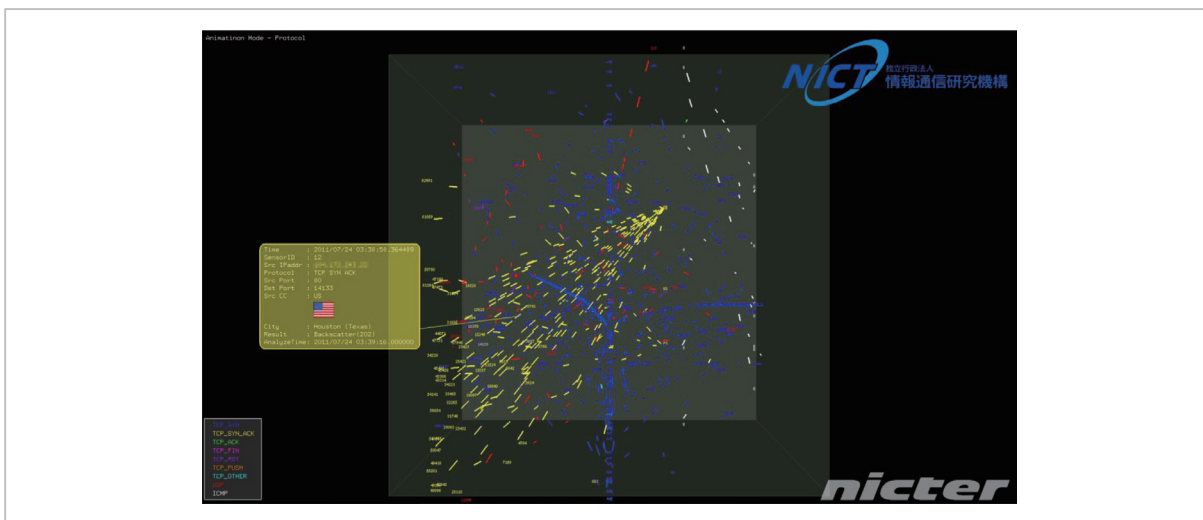


Fig.6 Cube (Detailed packet information display)



Fig.7 Tiles

a behavior analysis engine are reflected.

3.1.2 Analysis engine

(1) Change point detection engine[4]

A change point detection engine is an analysis engine which applies 2-stage online discounting learning to time-series data such as the number of packets going to specific ports per unit of time and the number of unique hosts, and swiftly detects rapid changes of such time-series data. The engine calculates the degree of changes of time-series data models as change point scores rather than setting simple thresholds to time-series data, making early detection of security incidents possible; e.g., detection of very small changes at the early stage of a large-scale worm infection. Figure 10 shows an example where a change point detection engine detected a scan of tcp/135 by MSBlast which caused large-

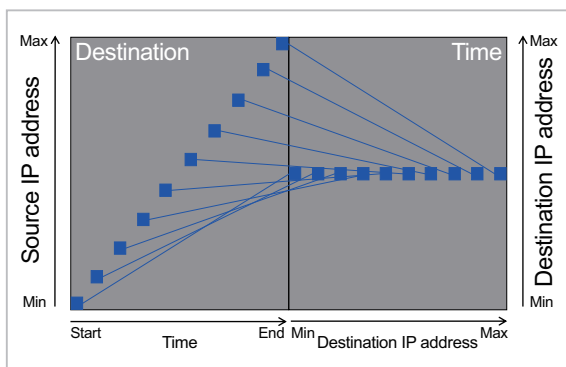


Fig.8 Expression format of individual tiles

scale infection in August 2003. It indicates that change point scores largely fluctuated around August 12, the early infection period. Figure 11 represents a change point detection engine's web interface. The change points of specific ports were detected and alerts (red ! in the figure) were automatically issued.

(2) Behavior analysis engine

A behavior analysis engine segments dark-net traffics according to each source host, and analyzes and classifies the movements of individual hosts for short period of time (30 seconds). The parameters to be used for the classification includes the number of packets, number of source and destination ports, pairs of destination port numbers, number of destination IP addresses, and scan types (sequential/random). Accumulation of classification history makes it possible to judge in real-time whether movements of any source host are known scan patterns or new scan patterns. Results of analysis and classification are visualized with the aforementioned Tiles.

As for macro analysis systems, research and development have been performed with various analysis engines (besides the analysis engines above) including a long-term behavior analysis engine, which analyses longitudinal movements of source hosts, SPADE analysis engine, which classifies scan patterns with spectrum analysis, exploit code detection

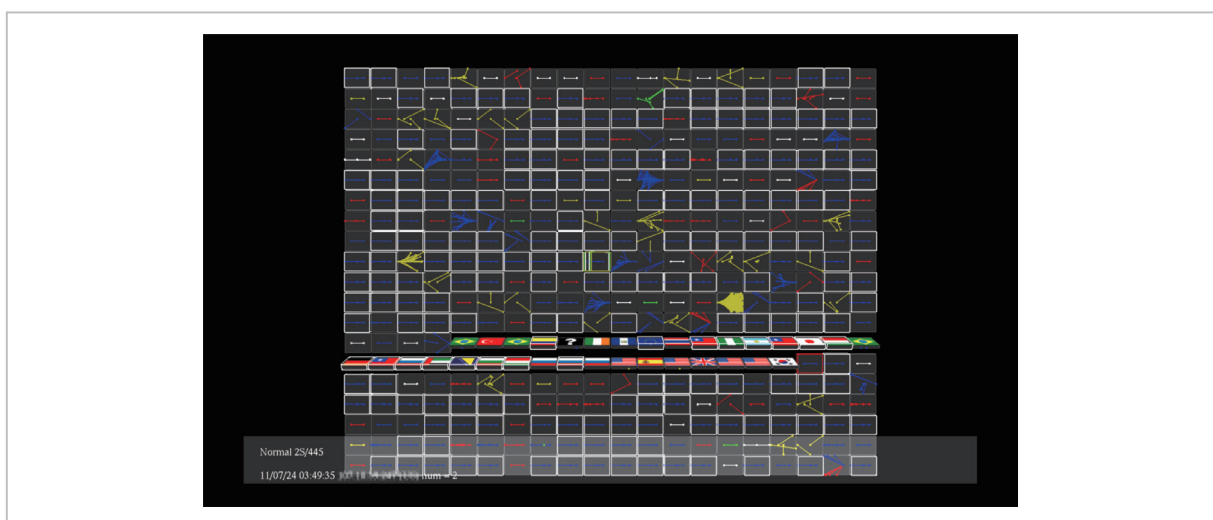


Fig.9 Highlight display of tiles having the same patterns

engine, which detects attack codes, incident forecast engine, which forecasts the change of darknet traffic, and spam analysis engine, which analyzes spam mail sources and URL link destinations included in text besides darknet traffic.

3.2 Micro analysis system [5][6]

The inputs of a micro analysis system are

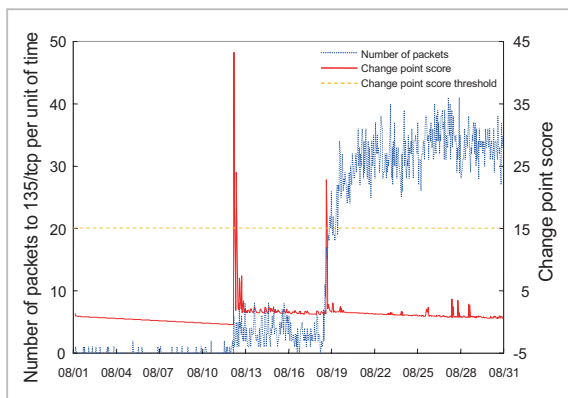


Fig.10 Example of MSBlast detection by a change point detection engine

malware specimens captured with honeypots, web crawlers and others. The nicter has implemented automation of malware analysis, has materialized high-speed dynamic analysis (6 to 9 minutes per sample), and has made it possible to analyze up to 2,000 samples per day through parallelization of analysis. Static and dynamic analysis engines, which are primary engines of a micro analysis system, are outlined below.

3.2.1 Static analysis engine

Static analysis is a method by which executable codes of malware are disassembled and malware's functions and characteristics are analyzed in detail at assembly levels. Many of recent malware programs have undergone code obfuscation which impedes disassembly, making static analysis difficult. Therefore, the nicter's static analysis engine executes malware having code obfuscation processed on a machine which serves as a sacrifice (hereafter, sacrifice host) and dumps self-decrypting codes to memory to eliminate the effects of



Fig.11 Change point detection engine's web interface

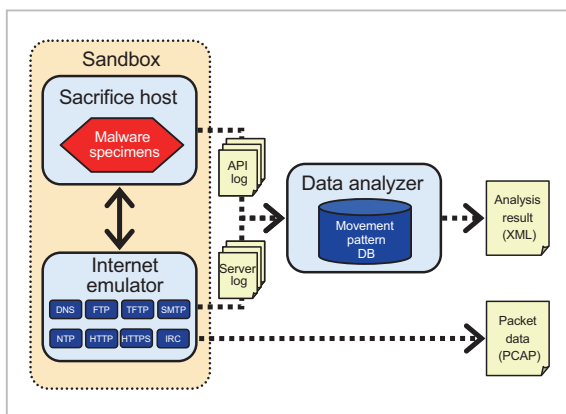


Fig.12 Malware dynamic-analysis engine

code obfuscation. Auto analysis of assembly obtained in this way makes it possible to extract various information such as a list of APIs included in executable codes of malware and character strings of private IRC messages to be used by bots.

3.2.2 Dynamic analysis engine

Dynamic analysis is a method by which malware is placed into execution status and the movements of, for example, APIs used by the malware at that time and network access are analyzed. To deal with such analysis, many recent malware programs, for example, monitor surrounding network environments and stop working or delete themselves upon detecting that they are under isolated environments, making dynamic analysis difficult. Therefore, in part of the malware dynamic analysis projects described in 2.3, a sacrifice host is allowed to connect to the Internet at the time of analysis, inducing a chance of causing damage to the outside. The nictcr's dynamic analysis engine completely isolates a sacrifice host in a sandbox environment, and places an internet emulator composed of many dummy servers such as DNS and IRC at the opposite sides, to enable safe dynamic analysis. Furthermore, in order to deal with virtual-machine detection which many malware programs perform to disturb analysis, a sacrifice host is composed of an OS auto recovery mechanism and an actual machine having an API hook function.

As a result of dynamic analysis in a sandbox environment like this, API logs are output

from a sacrifice host and server logs are output from the Internet emulator, followed by extraction of movements of malware from these logs. Figure 13 shows an example of dynamic-analysis results of malware. Traffic information from a sacrifice host is recorded as packet data, and a scan included in the packet data is used as a key for correlation analysis described later.

As for micro analysis systems, research and development has been performed for systems (besides static and dynamic analysis engines) such as herder imitation type bot analysis engines, which imitate herders (attackers which sends instructions to bots) in a sandbox environment and enable bot control, half-open type malware analysis engines, which connect only those judged safe among malware communications to the actual Internet and perform dynamic analysis, malware auto unpack engines, which enable auto cancel of obfuscation through auto detection of original entry points (OEP) of obfuscated malware, malware auto unpack classification engines, which classifies malware, based on dynamic-analysis results of malware, removal tool auto creation engines, which automatically create simple removal tools, based on dynamic-analysis results of malware, and systems, which automatically distribute the removal tools.

3.3 Correlation analysis system [1]-[3]

A correlation analysis system profiles scans monitored with a macro analysis system, based on various characteristics*6, and compares with profiles of scans extracted from malware with a micro analysis system, to find out malware candidates having similar profiles. The results of macro analysis and micro analysis are accumulated in a malware information pool named Malware kNOWLEDGE Pool (MNOP), and real-time comparison is performed by a correlation analysis engine.

Figure 14 shows correlation analysis results

*6 These include packet protocols, TCP flags, sender port numbers and their changes, addressee port setting, addressee IP address transition (sequential/random), number of packets per unit time, and payload length.

visualized with a visualization engine, Atlas. Malware names (or backscatters) listed as the first candidates through correlation analysis are shown above the individual packet objects. Malware names (in Fig. 14, w32.downadup.b) are also included in the detailed packet information. Furthermore, accumulating the total of correlation analysis results makes it possible to find out the global trend of malware. The box at the bottom left corner of Fig. 14 shows the accumulated total of correlation analysis results (number of unique hosts for each malware name); it is estimated that as of 2011 more than 70 percents of hosts have been infected with w32.downadup.b (or malware

having similar scan engines).

4 Conclusion

This paper describes the Network Incident analysis Center for Tactical Emergency Response (nicter), which aims at early detection, cause investigation, and solution presentation in terms of security incidents, through combinations of network monitoring and malware analysis. R&D activities by the nicter have made it possible to recognize activity trends of malware, from a panoramic viewpoint, which spreads infection through networks (so called remote exploit types) and



Fig.13 Malware dynamic-analysis result

-
- 3 K. Nakao, D. Inoue, M. Eto, and K. Yoshioka, "Practical Correlation Analysis between Scan and Malware Profiles against Zero-Day Attacks based on Darknet Monitoring," *IEICE Trans. Information and Systems*, Vol. E92-D, No. 5, pp. 787–798, 2009.
 - 4 D. Inoue, K. Yoshioka, M. Eto, M. Yamagata, E. Nishino, J. Takeuchi, K. Ohkouchi, and K. Nakao, "An Incident Analysis System NICTER and Its Analysis Engines Based on Data Mining Techniques," 15th International Conference on Neuro- Information Processing of the Asia Pacific Neural Network Assembly (ICONIP 2008), 2008.
 - 5 D. Inoue, K. Yoshioka, M. Eto, Y. Hoshizawa, and K. Nakao, "Malware Behavior Analysis in Isolated Miniature Network for Revealing Malware's Network Activity," *IEEE International Conference on Communications (ICC 2008)*, pp. 1715–1721, 2008.
 - 6 D. Inoue, K. Yoshioka, M. Eto, Y. Hoshizawa, and K. Nakao, "Automated Malware Analysis System and its Sandbox for Revealing Malware's Internal and External Activities," *IEICE Trans. Information and Systems*, Vol. E92-D, No. 5, pp. 945–954, 2009.

(Accepted June 15, 2011)



NAKAO Koji

Distinguished Researcher, Network Security Research Institute

Security Technologies, Security Management



INOUE Daisuke, Ph.D.

Director, Cybersecurity Laboratory, Network Security Research Institute

Network Security, Information Security