INVITED PAPER *Special Section on Information and Communication System Security*

Practical Correlation Analysis between Scan and Malware Profiles against Zero-Day Attacks Based on Darknet Monitoring

Koji NAKAO[†a)], **Daisuke INOUE**[†], **Masashi ETO**[†], ***and*** **Katsunari YOSHIOKA**[††], ***Members***

SUMMARY Considering rapid increase of recent highly organized and sophisticated malwares, practical solutions for the countermeasures against malwares especially related to zero-day attacks should be effectively developed in an urgent manner. Several research activities have been already carried out focusing on statistic calculation of network events by means of global network sensors (so-called macroscopic approach) as well as on direct malware analysis such as code analysis (so-called microscopic approach). However, in the current research activities, it is not clear at all how to inter-correlate between network behaviors obtained from macroscopic approach and malware behaviors obtained from microscopic approach. In this paper, in one side, network behaviors observed from darknet are strictly analyzed to produce scan profiles, and in the other side, malware behaviors obtained from honeypots are correctly analyzed so as to produce a set of profiles containing malware characteristics. To this end, inter-relationship between above two types of profiles is practically discussed and studied so that frequently observed malwares behaviors can be finally identified in view of *scan-malware* chain.

key words: *network monitoring, darknet malware analysis, sandbox, correlation analysis*

1. Introduction

Considering rapid increase of recent highly organized and sophisticated malwares, practical solutions for the countermeasures against malwares especially related to zero-day attacks should be effectively developed in an urgent manner. Various commercial, academic, or government-backed projects are ongoing to research and develop the countermeasure technologies [1]–[11]. Many of these projects are concentrating on events analysis providing statistical data, such as rapid increase of accesses on certain port numbers, by using network events monitoring. Particularly, it is getting popular and easier to monitor a dark address space, which is a set of globally announced unused IP addresses [1], [2], [12]. One can set up honeypots [13]–[16] on these addresses to masquerade as an vulnerable hosts in order to monitor and record the malicious activities or listen quietly (black hole monitoring) to the incoming packets, which often contain great amount of malware scans, DDoS backscatter, etc. This paper calls these events global observations over the Internet in a macroscopic view "Macro Analysis". That is, Macro Analysis can be applied to efficiently grasp the macroscopic behaviors (such as global scans) which are the first stage of malware activities over the Internet. However, since it is based on "events (scans) observations" in the macroscopic level and is performed without any explicit information regarding the attacker's behavior, its results often leave certain level of uncertainty on the attack caused by the malware.

On the other hand, apart from the macroscopic view, analyzing an actual malware executable has been another challenge. Reverse engineering techniques are applied to disassemble a malware executable in order for the analyst to understand its structure [17], [18]. Also, sandbox analysis, in which a malware code is actually executed in closed (or access-controlled) experimental environment, is capable to observe its behavior [18]–[21]. We call these direct malware analyses in a microscopic view "Micro Analysis". Micro Analysis reveals detailed structures and behaviors of malwares although it does not provide any information on their activities in real networks simply because it is performed in the closed experimental environment.

Even though the above Macro Analysis and Micro Analysis have been studied and deployed in various analysis systems, the knowledge obtained from these activities has not been effectively and efficiently linked, which is making the identification of the root causes of security incidents more difficult. To achieve the link between Macro and Micro Analysis in real time basis will provide a strong countermeasure against a new malware (zero-day virus) which is previously-unknown to the public and to the product vendor without any signature, since current signature-based approaches are not effective against zero-day viruses/attacks.

In order to come up to the above expectation of the linkage, we have been developing and researching Network Incident Analysis Center for Tactical Emergency Response (nicter) [18], [22]–[26]. The nicter realizes a practical implementation of Macro-Micro Correlation Analysis, in which the observations "in the darknet" by Macro Analysis and malware analysis "in the lab" by Micro Analysis are correlated to bind the observed attacks (mainly scans) with their possible root causes, namely malwares based on the fundamental propagation steps of malwares such as *scan* → *exploit code* → *malware download*. However, the link between the Macro Analysis and the Micro Analysis have not been strong enough to ensure the precise identification of detailed attacks' behavior since the correlation had been mainly made only by scan/exploit code behaviors of malwares that are observed by means of black hole monitoring.

In this paper, as the current practical correlation

Manuscript received December 23, 2008.

[†]The authors are with National Institute of Information and Communications Technology, Koganei-shi, 184–8795 Japan.

[††]The author is with Yokohama National University, Yokohama-shi, 240–8501 Japan.

a) E-mail: ko-nakao@nict.go.jp

DOI: 10.1587/transinf.E92.D.787

analysis system, the nicter is firstly presented. For further investigation of accurate and practical correlation analysis activities as an extension of the nicter, several extended considerations are made not only to cover external injections which are based on the attacks from remote hosts by means of *scan* → *exploit code* → *malware download*, but also to consider internal injections which are based on the malware download from malicious web-sites without any scans.

The organization of this paper is as follows: In Sect. 2, we detail the current nicter as a practical implementation for the Macro-Micro correlation analysis. In Sect. 3, we provide practical experimental results on the correlation obtained from the nicter described in Sect. 2. Moreover, in Sect. 4, further considerations of the correlation analysis as an extension of the nicter, including discussion and future actions to be required for network security technologies against malwares. Finally, in Sect. 5, we give conclusions.

2. Practical Implementation of Correlation Analysis System

This section provides a nicter which is a practical experimental case study to realize correlation analysis system against zero-day attacks.

2.1 Basic Concept of nicter

Observing the internet by using darknet, we are evidently receiving large quantities of scans from the internet everyday. The production of the nicter was triggered to answer what types of activities were connected to the observed millions of scans. The nicter has been a grand challenge to bind the latest scan activities with malwares behaviors by means of correlation analysis technology so as to identify the root cause of the scan activity in real time basis with the following characteristics.

(1) Based on darknet

Darknet is a set IP addresses that are not used in organizations. IP addresses in darknet are not assigned to the any operational servers/PC systems. Since IP addresses of the darknet are public, but are not assigned to legitimate hosts, all incoming traffic belonging to darknet IP domains may be inferred as a consequence of either malicious activities, or that of a mis-configuration. By using monitored packets in the darknet IP domains, we could observe emerging network attacks, including malware-initiated network scanning, malware infection behavior and DDoS Backscatters. Therefore, our research has been carried out based on the darknet which is so easy to extend monitoring coverage without any problems on the privacy issues discussed in the real network monitoring. Based on the darknet, two methods are used to observe malicious activities related traffics on the Internet, namely, black hole monitoring, and low interaction monitoring. The detailed implementation can be seen in the later section.

(2) Based on real-time analysis

Lately, we observe a zero-day (or zero-hour) attack which is a computer threat that tries to exploit unknown, undisclosed or patch-free computer application vulnerabilities to the public and to the product vendor. In this paper, a zero-day malware, parts of zero-day attacks, for which specific anti-virus signatures are not yet available is our major concern. That is, the nicter should work in real time basis in order to inform related entities of the latest prevailing malware (including a zero-day malware) in the internet by means of correlation analysis for stopping its further spreading to the wide-area of users.

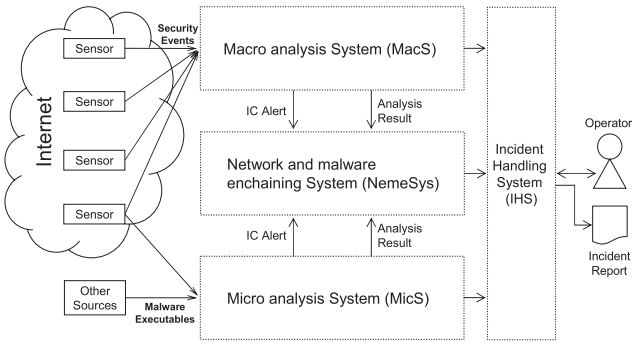

Fig. 1 The overview of nicter.

(3) Based on the global trend analysis

Target attacks are often discussed nowadays as a serious threat which specifically target government, organizations, or individuals to attack for several reasons such as extortions, intimidation. Such target attacks are out of our scope because the attacks may not be so visible in the global darknet observation and should be appropriately taken care by targeted sides. Therefore, our research is based on the global trend analysis focusing on the latest prevailing malware.

We briefly describe the overview of ongoing nicter system as depicted in Fig. 1. The nicter consists of four subsystems, namely Macro analysis System (MacS), Micro analysis System (MicS), Network and malware enchaining System (NemeSys) and Incident Handling System (IHS).

2.2 Macro Analysis (MacS)

MacS consists of widely distributed sensors, various visualization and analysis engines. The sensors monitor the network traffic and generate security events for the further analysis in the analysis engines. The results from the engines are a collective set of attributes such as sensor ID, analysis engine ID, timestamp, and other analyzer-specific attributes for the correlation of the analysis later on.

2.2.1 Sensors

We presently have several /16 and /24 darknets for observations, in which we are deploying wide range of black hole sensors that only listen to the incoming packets, a number of sensors that respond to certain incoming packets such as TCP SYN packet and ICMP echo request as low interaction sensors. The latter sensors are often configured to disguise

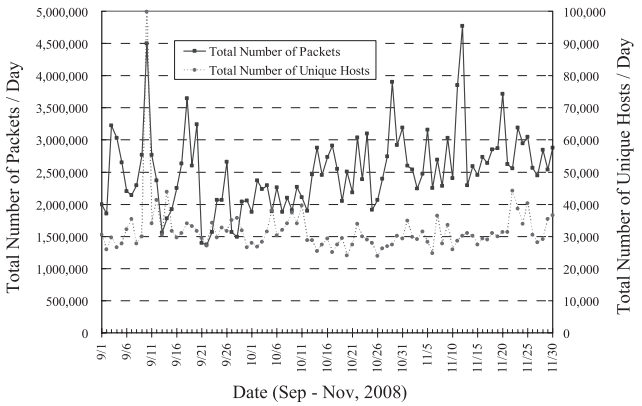

Fig. 2 Number of packets and unique hosts per day in a /16 darknet.

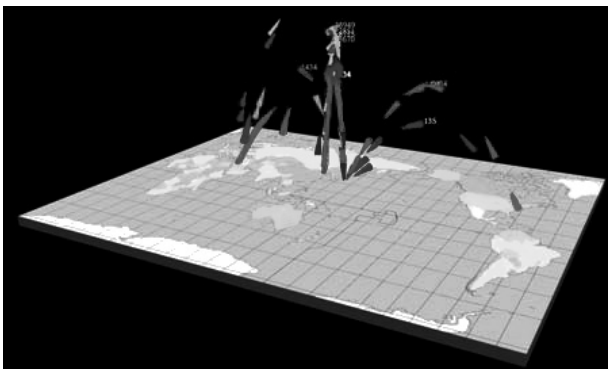

Fig. 3 Traffic visualization on Atlas.

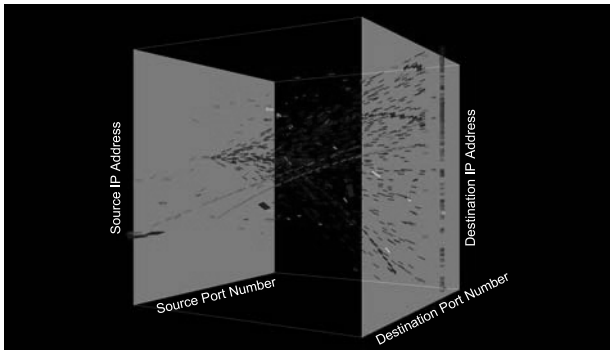

Fig. 4 Traffic visualization on cube.

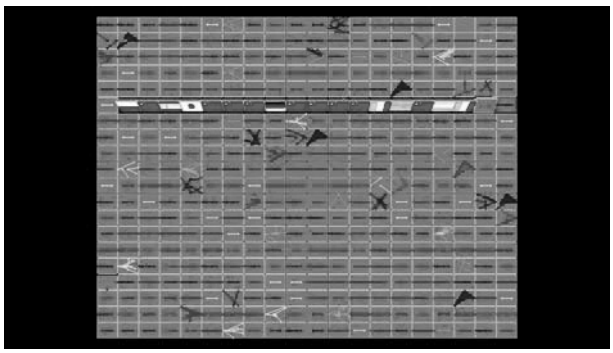

Fig. 5 Host behavior visualization on TAP view.

themselves as systems with unfixed vulnerabilities to attract attacks, namely they are deployed as the honeypots.

In Fig. 2, we show the number of incoming packets and the number of unique IP addresses observed by one of our /16 black hole sensors in Sep–Nov 2008. In average, over 2.5 millions packets and nearly 32 thousands unique IP addresses are observed per day.

2.2.2 Visualization Engines

Traffic visualization is important for system operators to grasp comprehensive trends of the monitored networks in real time. The nicter deploys several visualization engines as follows.

Atlas, a geographical traffic visualization engine, determines geographical positions of a packet's source and destination from the IP addresses and plots each packet on a world map as shown in Fig. 3. The packet is represented by animation effects as if a missile traverses from source to destination. The color of the missile indicates the type of packet as TCP SYN (blue), TCP SYN-ACK (yellow), other types of TCP (green), UDP (red), and ICMP (white). The altitude of the missile is in proportion to its port number.

Cube, a 3D traffic visualization engine, shows comprehensive traffic animation inside a cube. It consists of two planes, attacking plane (source) and darknet plane (destination) as in Fig. 4. Each incoming packet to the darknet is represented by a thin rectangle (its color has the identical meaning with the Atlas) animated from source to destination in about six seconds. It is placed on the plane by using the source IP address and source port number of the packet. The position it reaches on the other side is according to the destination IP address and port number of the packet as in Fig. 4. The entire darknet traffic can be visualized in the same manner and consequently, many interesting attack patterns, such as sequential network scanning or distributed attack, can be visualized for further analysis hereafter.

TAP[†] **View**, a host behavior visualization engine, represents the characteristic attack behaviors of attacking hosts by many tiles as shown in Fig. 5. For packets from an analyzed host, a tile is drawn according to the source and destination port numbers, the destination IP address and the time when the packet arrived at the sensor as shown in Fig. 6. The national flag in which the analyzed host is located is shown on the reverse side of the tile. The input parameters for this visualization are from the results of TAP analyzer described in 2.2.3.

2.2.3 Macro Analysis Engines

There are various real-time automated analysis engines in MacS. Some analysis engines are capable of detecting an incident candidate (IC) while others are for providing deep insights regarding the status of monitored networks. We now

[†]TAP stands for Traffic Analysis and Profiling.

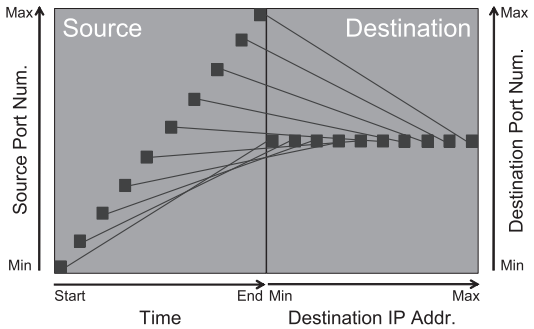

Fig. 6 Visualization method of a tile.

describe some of the Macro Analysis engines to detect new types of attack behavior which are hopefully connected with a zero-day attack as follows.

TAP analyzer takes security events from darknet, focuses on the short-term and long-term behaviors of individual attacking hosts [27]. The short-term behaviors (e.g. 30 sec.) are automatically classified by the analyzer and a new attacking behavior is reported as an IC based on the relationships between the following five parameters: (1) number of unique source port numbers, (2) the destination port numbers that the host is using, (3) number of unique destination IP addresses to which the host is attacking, (4) the total number of packets from the host, and (5) the randomness of destination IP addresses. This is carried out in real-time basis and its result is visualized on the TAP View. On the other hand, the long-term behavior is accumulation of the short-term behaviors for a long term to be used for further analysis of long-term attack trends and characteristics.

Change point detector (CPD) is a time series analysis method based on Auto-Regressive (AR) model, which is specially designed to achieve low complexity for real-time analysis [28]. It takes various time series data such as the number of accesses to particular destination ports per unit time, the number of IDS alerts per unit time, the number of a certain access pattern classified by the TAP analyzer, etc. For each time point, it calculates a score that indicates the likelihood for the point to be a change point and if the score exceeds the threshold, it sends an IC alert to the IHS. It is expected to detect comprehensive network incidents such as a propagation of new worms or a large scale DDoS attacks. Presently, the nicter can handle more than 20000 parallel change point detection processes including static monitoring and dynamic monitoring. In the static monitoring mode it continuously checks certain time series data including number of scans on certain ports with well-known vulnerability. In the dynamic monitoring mode, it starts analyzing by a trigger from TAP analyzer so that a change point for the new attack behavior from TAP can be efficiently detected.

Exploit code detector finds buffer overflow exploit codes which may result into system hackings. By utilizing low interaction sensors, it disassembles a binary code contained in the payload of attack packets to obtain an assembly

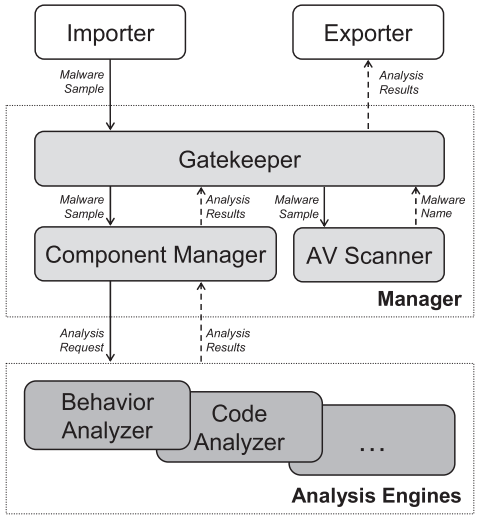

Fig. 7 MicS architecture.

code. It then examines the assembly code whether it is in characteristic structures that are essential to exploitation. As it depends solely on the algorithmic verification process and not on any signatures, it can detect any malicious exploit codes in the above-mentioned structures even though they are unknown. Also, the algorithm is extremely fast to be applied for real-time analysis. Consequently, it is deployed in conjunction with IDS for detecting unknown attacks among real network with live traffic.

2.3 Micro Analysis (MicS)

The purpose of the Micro Analysis is to conduct automated in-depth examinations of malwares in order to grasp their characteristics and activities.

As mentioned in Sect. 2.2.1, the nicter has several honeypots to collect malware executables in the wild. These collected executables are input into the MicS. Consequently, analysis results are stored in the MNOP (Malware kNOwledge Pool). The architecture of MicS is shown in Fig. 7.

Manager section of the MicS consists of several components: importer, exporter, gatekeeper, component manager, Anti Virus (AV) scanner. All malware samples captured by the capturers are automatically submitted to the importer. The gatekeeper periodically (e.g., every few seconds) downloads the submitted malwares from the importer. When a new malware identified by MD5 hash values is submitted, the gatekeeper passes it to the component manager to start analyzing the new.

In the meantime, it sends the sample to the AV scanner to obtain its names (if known). Once the component manager has received the submission, it sends a request to all the analysis engines to start analyzing the submitted sample. After finishing the analysis, each analysis engine returns the analysis results to the component manager. After obtaining results from all analysis engines, the component manager

sends them to the gatekeeper. Finally, all the analysis results along with their names defined by the AV scanner are output through the exporter to the MNOP (Malware kNOwledge Pool) for further correlation analysis. The processes of the MicS are fully automated from the submission of malware samples to the output of analysis results.

A single set of analysis engine can handle 150 to 250 malware samples per day: a rate of approximately six to nine minutes to analyze one malware sample. Presently, the MicS has several sets of analysis engines in parallel; consequently the system can analyze more than a thousand malware samples per day in total.

2.3.1 Micro Analysis Engines

There mainly exist two approaches in malware analysis: static analysis and dynamic analysis. The MicS deploys two analysis engines: the code analyzer, which is based on the static analysis, and behavior analyzer, which is based on the dynamic analysis.

Code analyzer, which can extract the internal characteristics of malwares, mainly takes advantage of the static approach. The approach has a potential to provide a whole view of the malware executables by disassembling them, however, it is sometimes ineffective to detect obfuscated malwares, since the disassembling may not be successful due to the obfuscation. Therefore, our code analyzer also applies the dynamic approach to the beginning of the examination. The analyzer first executes a given malware executable on an insulated victim machine in order to load unobfuscated code on the memory. Since the analyzer observes the process list of the victim machine, it can detect newly created process or process that changes its own memory size after running the malware. The analyzer dumps the detected process and disassembles it to obtain an assembly code of the malware. Then the analyzer examines the assembly code and extracts its characteristics such as a list of APIs and their arguments, files and registries to be created or modified, URLs or IP addresses and ports to be accessed, etc. Eventually, these characteristics of the malware are semantically classified and stored in MNOP as an XML file, which will be transformed into a human readable HTML file.

Behavior analyzer, which can observe the external activities of malwares as well as the internal activities, is based on the dynamic approach. Figure 8 shows the overview of the behavior analyzer, which consists of a sandbox and a data analyzer. In the sandbox, we use a real machine called victim host, which is infected by the malware sample. The victim host is not connected to the real Internet but to an isolated miniature network, called an Internet emulator. Since the sandbox is totally isolated from outside networks, it does not cause any unwanted infection and incident. It collects the API calls in the victim host, server logs in the Internet emulator, and packets transmitted between the victim host and the Internet emulator.

The sandbox operates on various modes such as black hole mode and low-interaction mode. In the black hole mode, the internet emulator provides minimum internet services to create a similar situation to monitoring with the black hole sensor in MacS. In the low-interaction mode, the internet emulator is configured to properly reply to requests by the victim host to create a similar situation to monitoring with the low-interaction sensor to observe exploitations [25].

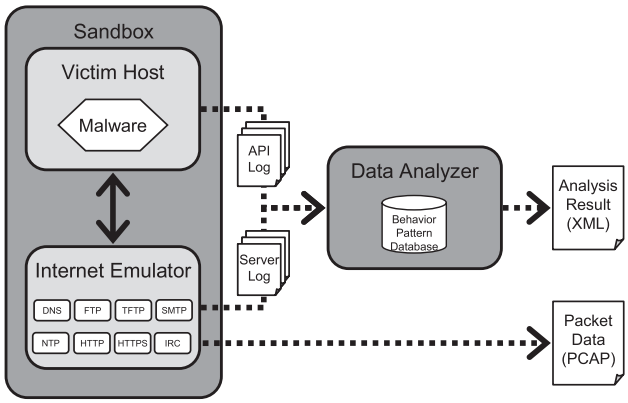

Fig. 8 Behavior analyzer.

All the collected information is then analyzed by the data analyzer and finally high-level descriptions of the observed behavior are output in XML format, and a human readable HTML.

The packet data observed in the sandbox is categorized into scans on global addresses, scans on local addresses, accesses to certain servers, etc. Since the scans on global addresses can only be monitored by black hole sensors even in the real internet environment, the sliced traffic of global scans from the analyzer can be utilized to obtain a scan profile for the correlation analysis in the NemeSys. Finally, all analysis results of the malwares are stored into the MNOP.

2.4 Correlation Analysis (NemeSys)

NemeSys correlates the results from MacS and MicS to identify the observed attacks in more accurate level. The NemeSys is based on an approach called network behavior profiling, in which network behaviors of captured malwares in MicS and attacking hosts observed by the darknet in MacS are summarized into profiles for fast and diverse correlation.

NemeSys has two main sub-components, profiler and correlator, which are controlled by correlation manager. We explain the analysis flow of the correlation analysis in NemeSys below as depicted in Fig. 9.

(1) The TAP analyzer in MacS detects an attacking host whose attack pattern is new, then issues an IC alert.
(2) Triggered by the IC alert, the correlation manager queries the MacS for packet data connected to the IC alert (namely, all packet data from the attacking host in certain period of time).
(3) The correlation manager sends the packet data to the

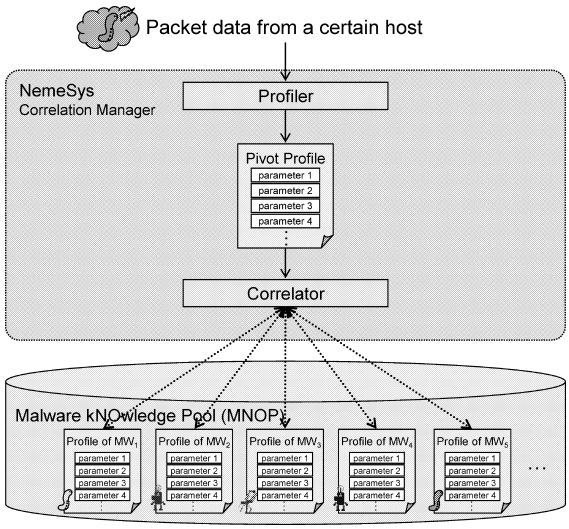

Fig. 9 Analysis flow of NemeSys.

profiler.

(4) The profiler generates a profile of the attacking host and sends it to the correlator as a pivot profile.
(5) The correlator calculates similarities between the pivot profile and each of all malware profiles in the MNOP and outputs the list of malwares whose profile is consequently correlated to the pivot profile.

2.4.1 Parameters for Making Scan Profile

In order to make profiles from observed scans on the darknet and malware scans in the sandbox, we focus on the following parameters:

Destination port: Destination ports of scan packets are the fundamental parameter for malware distinction as they tell us which services the malware attempts to attack or exploit the vulnerability of. Therefore, the profiler includes the set of scanned destination ports in the profile.

Source port: Compare to destination ports, source ports of scan packets can be selected freely by malwares. Some malwares use fixed source port while others change it for each packet. The profiler examines if each malware uses single or multiple source ports. If it uses multiple ports, the profiler also checks whether the source ports are changed for each packet or not. The profiler finally includes such parameters regarding source ports in the profile.

Destination IP address: The purpose of scans by malwares is to find their targets. There are many different tactics they can take to search the targets over the networks. Typical scans are the network scan that searches through certain network such as /16 or /24. Another frequently observed scans are random scans. There are also combinations of the two. For example, the first and second octets of the target IP addresses are randomly decided by malware and then the third and fourth octets are sequentially scanned. Thus, the way they seek for their targets can be another important parameter. The profiler checks whether the scan is random or sequential and includes the information into the profile.

Protocol and flag: As malwares take various tactics for scan, the protocol and flags (if TCP) of scan packets can differ among them. As malwares utilize various types of scan utilities, the scans created by them can be of varieties such as ping sweep, TCP SYN scan, TCP FIN scan, TCP Null scan, TCP Xmas scan, TCP Maimon scan, TCP ACK Scan, UDP scan, etc. Therefore, the profiler includes in the profile the information regarding protocols and flags of the scan packets.

Time related parameters: Some malwares make scans intensively while others tend to do it slowly. We refer to the average number of attack packets per unit time. It is effective to look into the arrival timing of the packets as well. However, external factors such as network delay and power of the infected machine may cause larger uncertainty for this type of timing analysis. The profiler includes this time related information into the profile.

Payload: Some attack packets, such as UDP exploitation, have payloads. It is effective to make signature (digest) of payloads to compare with each other. We are also considering applying the exploit code detector so that we can distinguish between normal payloads and exploit codes. The profiler includes the payload signature into the profile in case the scan packets have payloads.

Other parameters: There are other parameters in attack packets such as TTL, identification, and sequence number.

By using the above parameters, the correlator calculates a similarity between a pivot scan profile and those of pre-analyzed malwares in the MNOP. Finally, the NemeSys lists malwares having similar scan to an IC observed in MacS.

2.5 Incident Handling System (IHS)

IHS provides interfaces to provide the various analysis results to the human system operators. Basically, it is capable of managing the incident candidate (IC) alerts issued by the several analyzers in the MacS and MicS so that the operators can recognize the possible/assumed incidents as candidates and then move to further deep analysis in order to obtain the real incidents in manual manner. It also provides a platform for the operators to make an incident report to be issued as a final output of the nicter system.

Figure 10 shows a graphical user interface of the Workbench for the operators in the IHS. The Workbench effectively integrates the visualization engines, namely, Cube, Atlas and TAP View. It enables operators to conduct prompt and detailed investigation of the incident candidates. The IHS also provides a centralized web interface, called nicter web, on which every alert issued by the analysis engines in the MacS, visual snapshots captured on the Workbench, and analysis results of malwares from the MicS are concentrated for the operators.

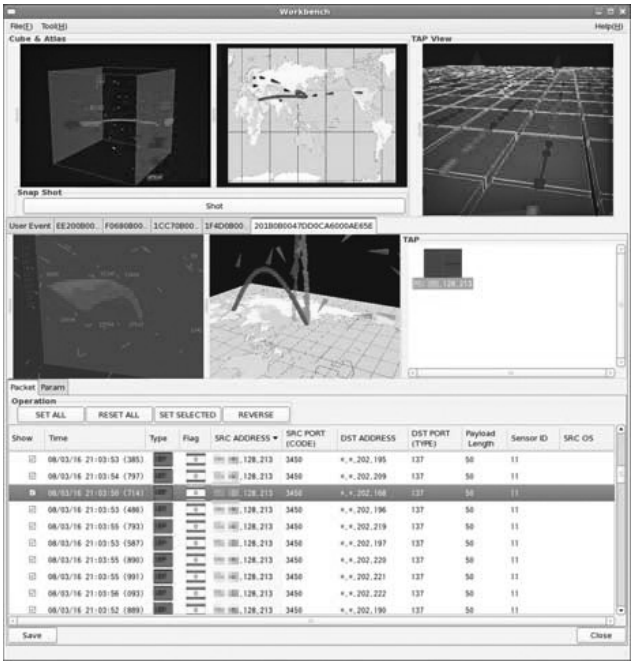

Fig. 10 Graphical user interface of Workbench.

3. Experimental Results

In this section, first we present some experimental results of the global trend analysis based on the black hole (darknet) monitoring. Second we provide a correlation experimental case of making macro and micro profiles for a scan behavior. Finally, based on the above correlation study, we conduct an inspection of the global trend for a specific certain scan by use of the micro profile.

3.1 Global Trend Analysis

The global trend analysis, which is a part of the MacS, is the first step to grasp the macroscopic behaviors in the Internet. As mentioned in Sect. 2.2.1, one of our /16 black hole sensors observed over 2.5 millions packets and nearly 32 thousands unique IP addresses per day in Sep–Nov 2008. Here, we conduct some close inspections of the darknet traffic in this period of time.

3.1.1 Number of Packets Sliced by Protocol

Figure 11 shows the number of packets per day sliced by the protocol (i.e., TCP, UDP and other protocols) in the IP header of each packet. Most of the packets using other protocols are ICMP packets. In average, we observed nearly 1.5 million TCP packets, 0.8 million UDP packets, and 0.3 million other packets per day.

The highest peak of TCP packets was observed on Nov 12th. On the day we observed nearly 2.5 million backscatter (TCP SYN-ACK) packets from a certain host in China, which means that the host probably suffered a large-scale DoS attack from many spoofed IP addresses.

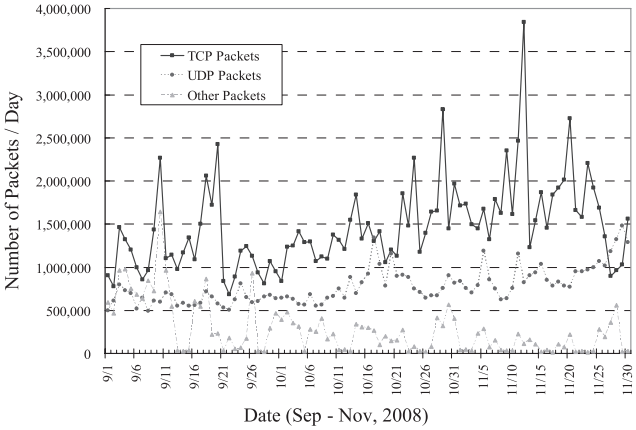

Fig. 11 Number of packets per day sliced by protocol.

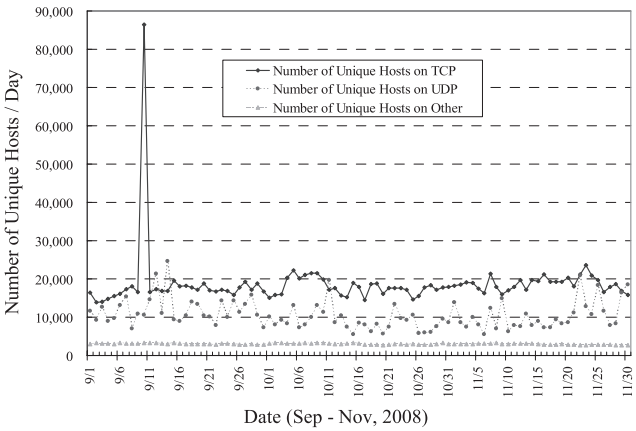

Fig. 12 Number of unique hosts per day sliced by protocol.

3.1.2 Number of Unique Hosts Sliced by Protocol

Figure 12 shows the number of unique hosts per day sliced by the protocol. In average, we observed over 18 thousands hosts on TCP, 10 thousands hosts on UDP, and 3 thousands hosts on other protocols per day.

On Sep 10th, we observed the highest peak of the unique hosts on TCP. The peak might be caused by a rather large scan activity of a botnet. The botnet contained over 70 thousand unique hosts in total and continuously executed the SYN scan on 1433/tcp for fourteen hours.

3.1.3 Trend in 445/tcp

Here we extract the packets toward 445/tcp from the darknet traffic to have a grasp of their trend. This port is widely used for the server service of Windows OS family, while critical vulnerability has been discovered frequently. Figure 13 shows the number of packets and unique hosts on 445/tcp per day. From the middle of September, both of packets and unique hosts started to increase gradually. The visualization engines (mentioned in Sect. 2.2.2) also told us a symptom of

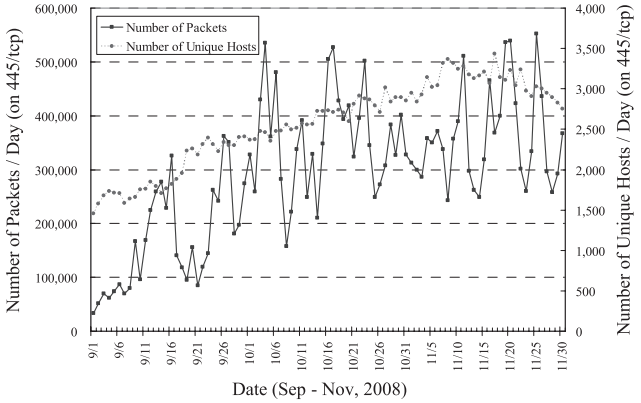

Fig. 13 Number of packets and unique hosts on 445/tcp.

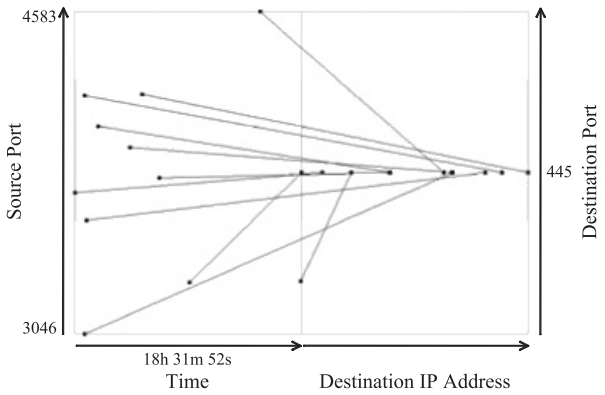

Fig. 14 Scan behavior in MacS.

a pandemic. After several weeks, on Oct 23rd, the Microsoft published a security bulletin of MS08-067, which was about a new vulnerability in the server service. Subsequently, several anti-virus vendors reported new types of malwares that exploit the vulnerability.

The global trend analysis based on the large-scale darknet monitoring can be a powerful tool to promptly grasp the malicious climate in the Internet.

3.2 Macro Profiling

In order to proceed to a deeper inspection, we drill down through the darknet traffic, and analyze the scan behavior of each attacking host to make their macro profiles. Here we show an example for making a macro profile. The /16 black hole sensor, also used in Sect. 3.1, observed an attacking host that was scanning on 445/tcp during the following term.

Time: 2:15:38am to 20:47:29pm, Oct 3, 2008 (18h 31m 52s)
Source IP: xxx.xxx.234.231

The scan behavior of the host is visualized as Fig. 14 in similar manner to Fig. 6.

The scan behavior of the attacking host can be translated into a macro profile as follows.

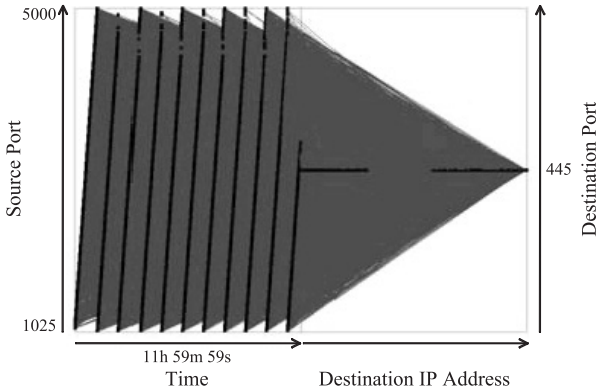

Fig. 15 Scan behavior in MicS.

Protocol: TCP
TCP flag: SYN
Destination port: Single (445)
Source port: Multiple (3046-4583)
Destination IP Address: Multiple (11 addresses)
Scan type: Random scan
Number of packets: 33 packets (1.78 packets/hour)

3.3 Micro Profiling

Meanwhile, the attacking host exploited one of our honeypots, and then a malware sample could be captured. Following is the detailed information of the capture.

Time: 13:36:12pm, Oct 3, 2008
Source IP: xxx.xxx.234.231
MD 5 value of sample: df17a625eec94cdcd4b1b7998c099d87
Symantec name: W32.Ifbo.A

We analyzed the captured malware sample in the MicS to observe its scan behavior and make a micro profile. The malware was executed in the sandbox with black hole mode for about twelve hours. As a result, the malware carried out a massive random scan on 445/tcp; the total number of packets was 117,220. The scan behavior of the malware is visualized as Fig. 15.

Then, in order to adjust the size of monitored network in the sandbox to the /16 black hole sensor in the MacS, scan packets that accessed one of /16 networks were extracted from all the observed packets. The resultant scan behavior is visualized as Fig. 16.

The scan behavior of the sample can be translated into a micro profile as follows.

Protocol: TCP
TCP flag: SYN
Destination port: Single (445)
Source port: Multiple (1126-4552)
Destination IP: Multiple (7 addresses)
Scan type: Random scan
Number of packets: 21 packets (2.09 packets/hour)

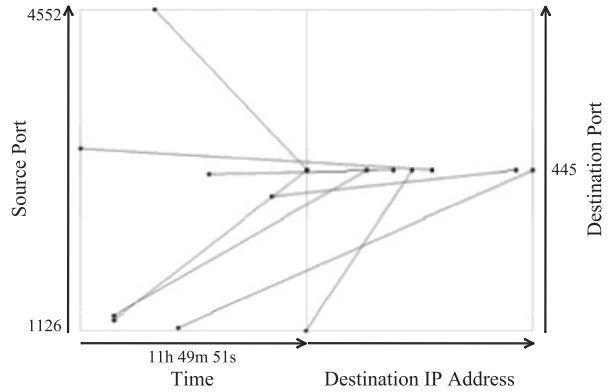

Fig. 16 Scan behavior in MicS sliced by a /16 network.

Consequently, in line with the approach described in NemeSys (Sect. 2.4), it is obvious that the micro profile has a strong likeness to the macro profile.

3.4 Feedback to Global Trend

Finally, we conducted an inspection of the global trend in 445/tcp again by use of the micro profile derived in Sect. 3.3 in order to reveal the distribution of attacking hosts that have similar scan behavior to the W32.Ifbo.A.

Here we go back to Oct 3rd, 2008 as an example. In the day, the /16 black hole sensor observed 28,243 attacking hosts in total, and 2,478 attacking hosts sending one or more packets on 445/tcp. We made macro profiles of the 2,478 hosts and conducted the correlation analysis between the macro profiles and the micro profile of the malware. As a result, 797 out of 2,478 hosts (32.2%) had the similar scan behavior with the W32.Ifbo.A.

By expanding this methodology through many types of scan behavior observed in the Internet, we believe that we will be able to correlate the actual scans behaviors with the specific malwares and effectively clarify the distribution of the malicious activities using the scan.

4. Further Considerations

Sections 2 and 3 provide a case study on the practical implementation experience to achieve Macro-Micro correlation analysis, in which the observations "in the darknet" by Macro Analysis and malware analysis "in the lab" by Micro Analysis are correlated to bind the observed attacks (mainly scans) with their possible root causes (malwares). Although the current nicter work on the correlation is beneficial for the early warnings against the zero-day attacks to some extent, the Macro analysis based on the global observations over the internet in the nicter is only limiting on those of scan behaviors of malwares and does not cover other types of global behaviors such as spreading exploit codes and/or SPAM which could be further correlated with the malwares activities.

Taking into account the above concerns based on the current practical achievement in the nicter, the following further considerations are strongly required to provide more sophisticated and effective solutions against zero-day attacks.

4.1 Stronger Binding of Network Attacks and Malwares by Multi-Layer Observation

We have discussed the correlation analysis based on scan behaviors of malwares. However, scan is one of the network behaviors in the course of malware propagation. For example, after finding their targets by scan, malware tries to take control of the targets by sending exploit codes. If the attempt is successful, the malware finally sends the actual executables (or get it downloaded) to the targets to complete its propagation. This whole activity can be observed if the deployed sensors are interactive enough. However, such high-interaction sensors are expensive to deploy in many locations. Therefore, we have proposed deploying a mixture of sensors, black-hole sensors, low-interaction sensors, and high-interaction sensors [25]. The black hole sensors does not reply to any scan packets and therefore easy to deploy and maintain. The role of black-hole sensors is to grab global trends of scan activities. Contrastingly, the high-interaction sensors can only be deployed to certain observation points although they can observe the network attacks in depth. The low-interaction sensors are in between the two sensors, that is, to observe network attacks in certain depth such as observing exploit codes while still achieving certain coverage. When correlating such variety of network attacks, the method of malware analysis needs to be improved as well. Namely, malware behavior analyzer should be able to observe scans, exploit codes, and download of malware executables.

Therefore, we proposed a configurable sandbox that can suitably change its responses to the malware sample. Namely, when correlating scan behavior the sandbox does not respond to any scan packets from malware so that it can observe scans that could have been observed by the black hole sensor (i.e. scan layer correlation). Likewise, when correlating exploit codes the sandbox responds as a low-interaction sensor so that it can observe exploit codes that could have been observed by the low-interaction sensor (i.e. exploit code layer correlation). Such multi-layer correlation can increase the preciseness of correlation.

In the following, we show some possibility to enhance the scan-based correlation analysis explained in Sect. 3 under the nicter. We analyzed the malware sample (i.e., W32.Ifbo.A) in the sandbox with low-interaction mode. That is, we deployed the low-interaction honeypot Nepenthes [29] in the sandbox and redirected the scans from malware to Nepenthes. In the sandbox, we were able to experimentally observe 8,089 TCP sessions within the execution of 30 seconds. Out of them, 2,001 sessions contained exploit codes according to our exploit code analyzer. The 2,001 extracted exploit codes were almost identical. Figure 17 illustrates a hexdump of an extracted exploit code. As one can see, it contains an IP address of the infected host

```
eb 45 68 74 74 70 3a 2f  2f 31 39 32 2e 31 36 38  |.Ehttp://192.168|
2e 32 30 2e 32 31 3a 38  30 2f 78 78 78 78 78 78  |.20.21:80/xxxxxx|
78 78 ae ae 62 ae 5d 33  c9 66 b9 a0 01 8d 75 05  |xx..b.]3.f....u.|
8b fe 8a 06 3c 99 75 05  46 8a 06 2c 30 46 34 99  |....<.u.F..,0F4.|
88 07 47 e2 ed eb 0a e8  da ff ff ff              |..G.........|
```

Fig. 17 Hexdump of an exploit code obtained by sandbox analysis.

```
eb 58 68 74 74 70 3a 2f  2f 37 34 2e 37 35 2e 32  |.Xhttp://74.75.2|
33 34 2e 32 33 31 3a 38  35 34 2f 78 2e 65 78 65  |34.231:854/x.exe|
df df df df df df df df  df df df df 4d 6f 7a     |............Moz|
69 6c 6c 61 2f 34 2e 30  df 5d 33 c9 66 b9 ee 01  |illa/4.0.]3.f...|
8d 75 05 8b fe 8a 06 3c  99 75 05 46 8a 06 2c 30  |.u.....<.u.F..,0|
46 34 99 88 07 47 e2 ed  eb 0a e8 da ff ff ff     |F4...G.........|
```

Fig. 18 Hexdump of an exploit code observed by network monitoring.

who sent this exploit code. Meantime, we checked if any of our low-interaction sensors in MacS have observed similar exploitation. We then found an attacker who seems to be utilizing very similar exploit codes as the malware sample. Figure 18 illustrates the exploit codes extracted from the attack packets observed by MacS.

One can confirm that the binary string following the URL (the underlined parts) are almost identical, indicating these two exploitations seem to be closely related. Such correlation based on exploit codes can also powerfully support the results of scan correlation.

4.2 Extensively Applied Macro/Micro Correlation to SPAM Messages

According to the current global observations over the internet, although millions of scan behaviors be observed recently, it is apparently increasing the number of SPAM which often contain malicious attachments with malwares or URLs navigating users to malicious web-sites for the purpose of downloading malwares. In the case of SPAM attacks, it is not necessary for the attackers to utilize the system vulnerabilities, but malwares can be easily downloaded just by users' click on malicious attachments or suspicious URLs without any searching any vulnerable systems. Therefore, it is getting so important for Macro Analysis to globally observe the SPAM activities in an extensive manner and for Micro Analysis to deeply examine the malwares connected with the SPAM.

To realize the Macro/Micro Analysis and the correlation analysis focusing on global observations of SPAM, we have to widely collect any types of SPAM messages to examine the content of the messages in order to obtain the characteristics of SPAM (SPAM profiles) which may consist of several specific parameters such as source address, subject, content, URL, attachment. These activities on the collections and examinations could be categorized in the Macro Analysis for SPAM messages.

On the other hand, SPAM messages are sometimes connected with malwares which are hidden in the attachments to the messages or are downloaded from suspicious URLs in the messages. We need to investigate how to obtain the malwares in the attachments or how to download the malwares based on the suspicious URLs by using web crawler and search robot.

If the attachment to the SPAM message is an executable program, then it should be instantly applied to the Micro Analysis (MicS) discussed in Sect. 2.3 to obtain its behavior. However, if it does not look like an executable program, we need to study to carefully inspect the attachment whether it contains malicious executable program in concealment or not. Furthermore, the malicious executable program is sometimes hidden in a human readable file and should be correctly extracted from the attachment with special techniques.

If the suspicious URLs are detected in the message, then we need to research the automatic and efficient mechanisms how to correctly download the malwares in order to collect malwares related to the SPAM messages. The need of the special mechanism is caused by the fact that we could frequently meet malwares which cannot be easily downloaded from the URLs. For example, according to observations of botnet malwares infection, a single click of the URL may not be effective enough to download the malware but may provide a downloader to move to the next step. After the several steps, we could finally obtain the malware related to the initial URL in the message. In this case, it is necessary for us to record a sequence of URLs together with the malware finally obtained. As for the downloaded malwares, they should be also applied to Micro Analysis (MicS) to obtain their further behaviors in depth. The above activities on the inspections of the attachment and research of the downloading mechanisms could be also categorized in the part of the Micro Analysis for SPAM messages.

As for the correlation between Macro and Micro Analysis focusing on SPAM messages, it can be carried out to bind SPAM profiles with malwares extracted from the attachments or downloaded from URLs in the messages. If the Macro and Micro Analysis are successfully executed on a SPAM message, then we could automatically obtain the root cause (malware) of the prevailing SPAM message over the Internet without any difficulties, because Macro and Micro Analysis be carried out by using the same SPAM message. Otherwise, the correlation is not able to complete for the SPAM message. In any cases, it would be extremely valuable to store any results of the analysis in a specific DB for further extensive use.

5. Conclusion

Protecting against the most sensitive security incidents caused by zero-day attacks (especially zero-day malwares), we provided a practical solution of the nicter focusing on the correlation analysis between scan and malware profiles based on darknet monitoring. Although the nicter provides a certain level of the correlation capabilities in practices, further considerations should be extensively carried out to effectively achieve correctness of the correlation, to enrich the coverage of malwares attacks including internal injections

which are based on the malware download from malicious web-sites independent of any scans. We expect the above significant concept of correlation analysis to be globally studied and utilized for network security researchers and administrators.

Acknowledgements

We would like to thank Mr. Shunsuke Baba and Mr. Kazuya Suzuki of Yokogawa Electric Corp. for their great contributions on designing and implementing the nicter as the core members of our project. We also would like to thank Mr. Jumpei Shimamura of ForSchooner Inc., Mr. Yaichiro Takagi, Mr. Mio Suzuki and Mr. Yoshinori Hashimoto of NICT for their continuous efforts on providing various analysis results and observation data.

References

[1] M. Bailey, E. Cooke, F. Jahanian, J. Nazario, and D. Watson, "The internet motion sensor: A distributed blackhole monitoring system," 12th Annual Network and Distributed System Security Symposium (NDSS05), 2005.
[2] D. Moore, "Network telescopes: tracking denial-of-service attacks and internet worms around the globe," 17th Large Installation Systems Administration Conference (LISA'03), USENIX, 2003.
[3] V. Yegneswaran, P. Barford, and D. Plonka, "On the design and use of Internet sinks for network abuse monitoring," 7th International Symposium on Recent Advances in Intrusion Detection (RAID 2004), LNCS 3224, pp.146–165, 2004.
[4] SANS Internet Storm Center, http://isc.sans.org/
[5] REN-ISAC: Research and Education Networking Information Sharing and Analysis Center, http://www.ren-isac.net/
[6] Leurrecom.org Honeypot project, http://www.leurrecom.org/
[7] National Cyber Security Center, Korea, http://www.ncsc.go.kr/eng/
[8] Telecom Information Sharing and Analysis Center, Japan, https://www.telecom-isac.jp/
[9] IT Security Center, Information-Technology Promotion Agency, Japan, https://www.ipa.go.jp/security/index-e.html
[10] Japan Computer Emergency Response Team Coordination Center, http://jpcert.jp/isdas/index-en.html
[11] @police, http://www.cyberpolice.go.jp/english/obs_e.html
[12] M. Bailey, E. Cooke, F. Jahanian, A. Myrick, and S. Sinha, "Practical darknet measurement," 2006 Conference on Information Sciences and Systems (CISS '06), pp.1496–1501, 2006.
[13] N. Provos, "Honeyd: A virtual honeypot daemon," 10th DFN-CERT Workshop, 2003.
[14] N. Provos, "A virtual honeypot framework," 13th USENIX Security Symposium, pp.1–14, 2004.
[15] E. Alata, V. Nicomette, M. Kaaniche, and M. Dacier, "Lessons learned from the deployment of a high-interaction honeypot," 6th European Dependable Computing Conference (EDCC-6), pp.39–44, 2006.
[16] C. Leita, M. Dacier, and F. Massicotte, "Automatic handling of protocol dependencies and reaction to 0-day attacks with ScriptGen based honeypots," 9th International Symposium on Recent Advances in Intrusion Detection (RAID2006), pp.185–205, 2006.
[17] R. Isawa, S. Ichikawa, Y. Shiraishi, M. Mohri, and M. Morii, "A virus analysis supporting system; for automatic grasping virus behavior by code-analysis result," Computer Security Symposium 2005 (CSS2005), vol.1, pp.169–174, 2006.
[18] D. Inoue, M. Eto, K. Yoshioka, Y. Hoshizawa, R. Isawa, M. Morii, and K. Nakao, "Micro analysis system for analyzing malware code and its behavior on nicter," 2007 Symposium on Cryptography and Information Security (SCIS2007), 2F2-1, 2007.
[19] Y. Hoshizawa, M. Morii, and K. Nakao, "A proposal of automated malware behavior analysis system, Information and Communication System Security," IEICE Technical Report, ICSS2006-07, 2006.
[20] C. Willems, T. Holz, and F. Freiling, "Toward automated dynamic malware analysis using CWSandbox," IEEE Security & Privacy Magazine, vol.5, no.2, pp.32–39, 2007.
[21] NORMAN Sandbox Information Center, http://www.norman.com/microsites/nsic/
[22] K. Nakao, F. Matsumoto, D. Inoue, S. Baba, K. Suzuki, M. Eto, K. Yoshioka, K. Rikitake, and Y. Hori, "Visualization technologies of nicter incident analysis system," IEICE Technical Report, ISEC-176, 2006.
[23] K. Nakao, K. Yoshioka, D. Inoue, M. Eto, and K. Rikitake, "nicter: An incident analysis system using correlation between network monitoring and malware analysis," 1st Joint Workshop on Information Security (JWIS06), pp.363–377, 2006.
[24] K. Yoshioka, M. Eto, D. Inoue, and K. Nakao, "Macro-micro correlation analysis for binding darknet traffic and malwares," 2007 Symposium on Cryptography and Information Security (SCIS2007), 2F2-2, 2007.
[25] K. Nakao, K. Yoshioka, D. Inoue, and M. Eto, "A novel concept of network incident analysis based on multi-layer observations of malware activities," 2nd Joint Workshop on Information Security (JWIS07), pp.267–279, 2007.
[26] D. Inoue, M. Eto, K. Yoshioka, S. Baba, K. Suzuki, J. Nakazato, K. Ohtaka, and K. Nakao, "nicter: An incident analysis system toward binding network monitoring with malware analysis," WOMBAT Workshop on Information Security Threats Data Collection and Sharing (WISTDCS 2008), pp.58–66, 2008.
[27] K. Suzuki, S. Baba, and H. Takakura, "Analyzing traffic directed to unused IP address blocks," IEICE Technical Report, IA2005-23, Jan. 2006.
[28] J. Takeuchi and K. Yamanishi, "A unifying framework for detecting outliers and change points from non-stationary time series data," IEEE Trans. Knowledge Data Eng., vol.18, no.4, pp.482–489, 2006.
[29] P. Baecher, M. Koetter, T. Holz, M. Dornseif, and F.C. Freiling, "The nepenthes platform: an efficient approach to collect malware," 9th International Symposium on Recent Advances in Intrusion Detection, (RAID 2006), pp.165–184, 2006.

Koji Nakao is the Information Security Fellow in KDDI, Japan. Since joining KDDI in 1979, he has been engaged in the research on multimedia communications, communication protocol, secure communicating system and information security technology for the telecommunications network. He is also an active member of Japan ISMS user group, which was established in the 1st Quarter of 2004. He is the board member of Japan Information Security Audit Association (JASA) and that of Telecom-ISAC Japan, and concurrently, a Technical Group Chairs (ICSS: information communication system security) of the Institute of Electronics, Information and Communication Engineers. He received the B.E. degree of Mathematics from Waseda University, in Japan, in 1979. He received the IPSJ Research Award in 1992, METI Ministry Award and KPMG Security Award in 2006, and Contribution Award (Japan ITU), NICT Research Award, Best Paper Award (JWIS) and MIC Bureau Award in 2007. He is a member of IPJS. He has also been a part-time instructor in Waseda University since 2002.

Daisuke Inoue received his B.E. and M.E. degrees in electrical and computer engineering from Yokohama National University in 1998 and 2000, respectively, and Ph.D. degree in engineering from Yokohama National University in 2003. He joined the Communications Research Laboratory (CRL), Japan, in 2003. The CRL was relaunched as the National Institute of Information and Communications Technology (NICT) in 2004, where he is a senior researcher of the Information Security Research Center. His research interests include security and privacy technologies in wired and wireless networks, incident handling based on network monitoring and malware analysis. He received the Best Paper Award at the 2002 Symposium on Cryptography and Information Security (SCIS 2002).

Masashi Eto received LL.B degree from Keio University in 1999, received the M.E. and Ph.D. degrees from Nara Institute of Science and Technology (NAIST) in 2003, 2005, respectively. From 1999 to 2003, he was a system engineer at Nihon Unisys, Ltd., Japan. He is currently a researcher at National Institute of Information and Communications Technology (NICT), Japan. His research interests include network monitoring, intrusion detection, malware analysis and auto-configuration of the Internetworking. He received the Best Paper Award at the 2007 Symposium on Cryptography and Information Security (SCIS 2007).

Katsunari Yoshioka received the B.E., M.E. and Ph.D. degrees in Computer Engineering from Yokohama National University in 2000, 2002, and 2005, respectively. From 2005 to 2007, he was a Researcher at the National Institute of Information and Communications Technology, Japan. Currently, he is an Assistant Professor at the Interdisciplinary Research Center, Yokohama National University. His research interest covers wide range of information security, including malware analysis, network monitoring, intrusion detection, and information hiding. He received the Best Paper Award at the 3rd Joint Workshop on Information Security (JWIS 2008).