4-2 Rapid Analysis Technologies for Live Networks

Ichiro SHIMADA and Yu TSUDA

In targeted cyberattacks, the attackers intrude into the internal network of an organization, after breaking through border-defense security measures. The attackers then commit data theft, destructive actions, or other malicious activities that advance their objectives. Responding to such attacks thus requires not only border-defense security measures but also methods for rapidly detecting malicious activities once attackers have successfully intruded. In this paper, we propose a method for rapid detection of scanning activities or communications with malicious hosts in the context of massive live network communications within the internal network of an organization.

1 Introduction

In recent years, targeted cyberattack incidents have been brought to light one after another — starting from the case of a Japanese major heavy industry manufacturer to similar incidents such as breaking into the networks of the House of Representatives and the House of Councillors as well as other ministerial organizations and agencies — signaling the pressing challenge for establishing sweeping measures and technologies against targeted cyberattacks. The attacker's objective is to steal information inside an organization and send it to a malicious host located outside the organization. For this purpose, the attacker first tries to break through the border-defense security measures by sending a targeted email, paving the way for malware to intrude into the organization. As a preventive measure against such targeted attacks, an expeditious method to detect suspicious communications among intra-organization live network communications is strongly needed. The authors studied an early detection method to single out suspicious communications attempting to link with a malware host, among the vast sea of live network communications. In this paper, the authors report two technologies — blacklist-based suspicious communication detection, and slow scan detection using Bayesian decision making. The blacklist used in the former technology was constructed based on the darknet observation information of NICTER[1] (a database containing a vast amount of malicious host information acquired through observation).

2 Related technologies

2.1 Network Incident analysis Center for Tactical Emergency Response (NICTER)

NICTER is a project under research and development at the National Institute of Information and Communications Technology (NICT) that implements a macroscopic analysis system to detect and analyze incidents, whereby a large-scale darknet observation network is used for analyzing events. The macroscopic analysis system uses sensors distributed within and outside of Japan for monitoring the darknet. The information gained through darknet monitoring includes a huge amount of malicious host information, necessarily including those related to C&C servers. The system is designed to store the packets gathered from the sensors to the MacS DB[2] database system in real-time, which enables the packet data in the storage to be used as a list of IP addresses that includes those of malicious attackers (hereafter referred to as black list IPs).

2.2 NIRVANA: Live network traffic visualization system

NIRVANA[3] is an attempt to apply the real-time visualization technology of NICTER — especially the packets gathered by darknet monitoring — to traffic in the internal network of a specified organization: network- and transport-layer header information from the vast amount of packets flowing through the intra-organization network is collected and aggregated, and then is sent to the visualization terminal for visual representation of the state of live network communication. In this study, the functions provided by NIRVANA — collecting and aggregating header information from intra-organization traffic — are applied to monitor live network traffic for the detection of black list IPs and attacks as described in Sections **3** and **4**. Figure 1 illustrates the volume of traffic observed during the period from the August 24 to 31, 2014.

2.3 System organization

Figure 2 shows a schematic diagram of the system constructed for this study. In step ①, the system creates an IP address list (IPLIST in Fig.2) to be referenced in the analysis. This IP address list not only serves as a blacklist IP list in blacklist search, but also as a whitelist IP address list in low speed scan detection. Next, packets are extracted, in step ②, from the Livenet Databus used by NIRVANA, which are used by the Analysis Module to perform checkups against the IP address list. If an instance of improper communication is detected in the checkup process ③, an alarm is sent to the visualization system.

3 Blacklist detection

3.1 Overview

The objective of this study to develop a method capable of detecting communications that are linked to malicious hosts (C&C servers) quickly, among the vast sea of normal communications, by utilizing the darknet monitoring information provided by NICTER as a blacklist. To realize this objective, the live network communication and blacklist of malicious hosts must be quickly analyzed in real-time. In general, the characteristics of blacklist based detection methods — i.e. based on the source information of past cyberattack incidents, or predefined signatures — makes it difficult to address unknown attacks in real-time. However, as the darknet observation data provided by NICTER is real-time attack information, it can be defined as a real-time blacklist IP for detecting irregular communications.

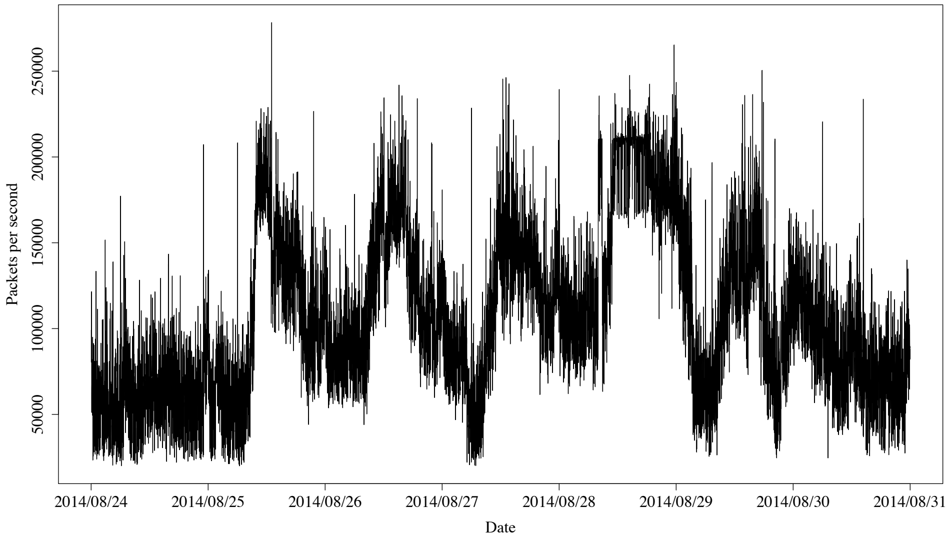

Fig. 1 Observed live network traffic within NICT

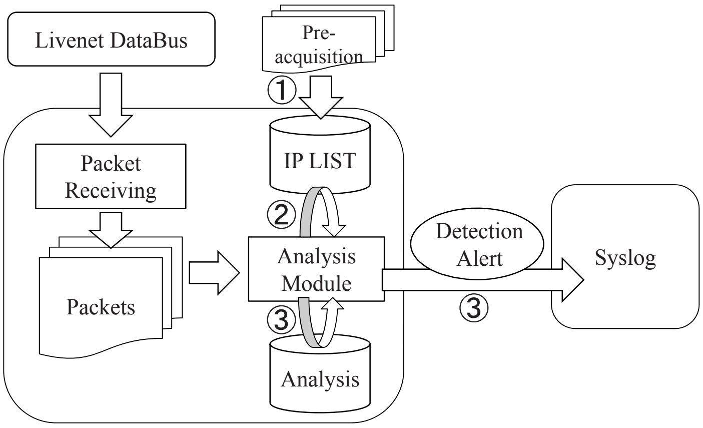

Fig. 2 System overview

3.2 Real-time detection

The blacklist IP is compiled in reference to the MacS DB. The MacS DB is a database that stores the header information of the network- and transport layer of all the packets sent to darknet. The data in the MacS DB for the past one week is used to create blacklist IPs which are saved in a blacklist database. To rule out backscatter, two types of data are selected for storage: the source IP addresses of external malicious hosts that send a TCP SYN packet to darknet, and its time of receipt.

To keep the blacklist IP list always up-to-date, the latest data in the MacS DB is retrieved once an hour to update. Then, the TCP packets are read from the live network, and their source/destination IP addresses are examined against the blacklist IP list. If a hit is found and the host is responding to a TCP SYN packet with a TCP SYN-ACK packet, an alarm is generated assuming that a session is being connected.

In the matching process against the blacklist IPs, conventional simple linear search tends to require a longer time as the number of blacklist IPs increases. To avoid this problem, all of the IPv4 IP addresses are mapped into memory on a bit-by-bit basis, enabling cross-check by turning a bit on or off for searching blacklist IPs. This approach enables fast search irrespective of the number of blacklist IPs. The approach has been verified to have a sufficient matching speed, and completed the process in a few steps irrespective of the number of blacklist IPs [4].

3.3 Results of observation collected in NICT network environment

To identify the state of communications with blacklist IPs more clearly, they are classified into Inbound and Outbound communication based on their direction. Figure 3 shows a graphical representation of the communication status. The observation results under the NICT network environment indicate that the accesses from the blacklist IPs to external open servers — such as the hosts on DMZs, honey pots and web servers — were successfully detected. In Outbound communications, those from internal hosts via proxy servers were detected. In addition, the external host's scans that could be detected by other security appliances can also be detected using the blacklist detection method. This indicates that the host scanning the address space on the internet, including darknet, can be detected by this method, promising its effective use to judge the importance of blacklist IPs. Better accuracy for issuing alerts is the major challenge for the future. Possible methods to achieve this goal include the effective utilization of other irregular communication observation data, in addition to the information acquired from darknet observation.

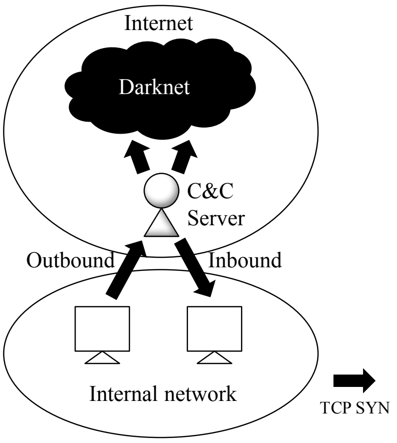

Fig. 3 How communication is made with blacklist IPs

4 Slow scan detection

4.1 Overview

As the attackers typically do not possess detailed knowledge of the targeted network, they usually send packets to the hosts within the organization they have sneaked into, with the intention of checking the existence of the hosts and gathering vulnerability information. The attacker tends to send out a large number of packets in a short period of time to achieve their goal efficiently, and the network IDS tries to take advantage of this trait to capture the scan. To evade detection by the network IDS, the attacker sometimes performs slow scans intentionally. It is a tough task for a conventional network IDS to capture such slow scans. As shown in reference [5], the authors made a study on detecting slow SYN stealth scans out of the vast sea of Livenet communications. This detection method consists basically of the following procedures.

First, the current status of connection attempts being made by scan packets is detected using the framework of Threshold Random Walk (TRW) [6], in which a Bayesian decision making approach was applied. Then, to facilitate scan packet extraction, normal traffic was removed through reference against a whitelist.

4.2 Detection method

Among the set of information that a packet of live network traffic carries, NIRVANA only makes use of header information of the network- and transport layer.

Therefore, the applicability of TRW must be judged using such header information, which provides knowledge on if the connection attempt through TCP has succeeded or failed. To judge success/failure of connection attempts, the authors used a decision method based on the following items (see reference [5]): 5-tuple, TCP flag, Timestamp, and Sequence number. This approach has an effect of reducing erroneous judgement — typically caused by packet loss — and improving the judgement accuracy of successful connection. To reduce False Positives (FP) while trying to detect slow scans, the authors introduced filtering by means of a whitelist (i.e. exclusion of normal traffic). The whitelist was constructed based on the success/failure information of connection attempts, as judged based on the criteria described above, and included the source/target IP address, target port number, timestamp, and the number of successful/failed attempts.

To facilitate the observer's understanding of current scan conditions, the authors tried to represent the connection attempt situation using random walk of degree of belief (changes of subjective probabilities). In this process, the connection probabilities — for both benign and scanner hosts — are first estimated, followed by Bayesian decision making based on the amount of information contained in a connection attempt. The latter step makes use of the TRW framework that detects scans using successive hypothesis testing based on the information contained in successful/failed connection attempts.

4.3 Results of observation collected in NICT network environment

Slow scan experiments were conducted for evaluation under the network environment of NICT. The interval between scan packet transmissions was set to a variety of lengths, from several minutes to one day.

All the hosts that carried out slow scans were successfully detected.

The results of detection attempts depend on the parameter that indicates success/failure probability of connection with the assumed malicious host, but the slow scans were detected within a relatively small number of attempts. Figure 4 shows a graph of failed connection attempts and source hosts, extracted from the experimental results. The graph was produced using a graph mining tool [7]. A node in the graph represents a host, and the edges represent the direction of connection attempts. As is apparent from the graph, the host 1-3 (used in the experiment) showed the highest out-degree. From the viewpoint of rapid response to incidents, future challenges include the development of a method that provides an easier grasp of the scan situation for the observer. The authors consider that the following two aspects are of special importance for further study: comparison with other detection methods, and validation of detection accuracy.

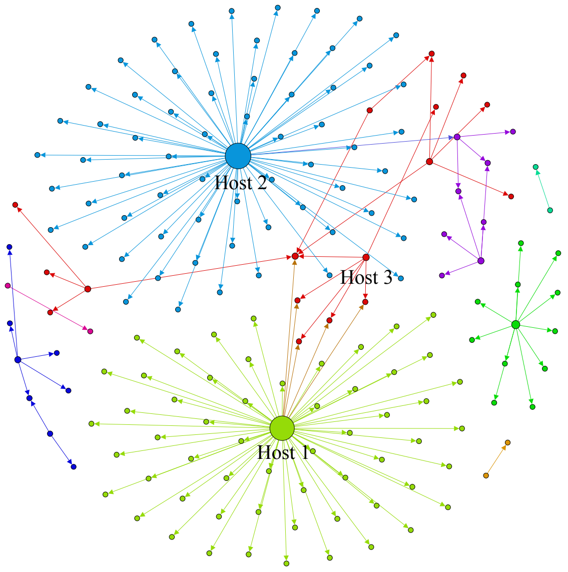

Fig. 4 Failed connection attempts and source hosts

5 Concluding remarks

The authors conducted a study, using the network environment within NICT, to verify the validity of an improper communication detection method that utilizes darknet monitoring information as a blacklist. The authors also proposed a slow scan detection method, and demonstrated its validity. The authors are planning to pursue further verification studies within the framework of the NICT network environment, aiming at future implementation of the results in society.

Acknowledgments

The authors express deep appreciation to the members of the Information Security Unit in NICT (a National Research and Development Agency) for their kind provision of traffic data within NICT and various types of security compliance data, which played an integral role in our study.

References

1 D. Inoue, M. Eto, K. Yoshioka, S. Baba, K. Suzuki, J. Nakazato, K. Ohtaka, 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.

2 M. Eto, Y. Takagi, “An Incident Analysis Center “nicter” and its Social Commitment,” Journal of the National Institute of Information and Communications Technology, vol.58, pp.17-26, 2011.

3 K. Suzuki, M. Eto, D. Inoue, “Development and Evaluation of NIRVANA: Real Network Traffic Visualization System,” Journal of the National Institute of Information and Communications Technology, vol.58, pp.61-77, 2011.

4 I. Shimada, Y. Tsuda, M. Eto, D. Inoue, “A Slow-Scan Detection Method for Live Network Environments, ” Computer Security Symposium 2014 (CSS2014), 2014.

5 I. Shimada, Y. Tsuda, M. Eto, D. Inoue, “Using Bayesian Decision Making to Detect Slow Scans, ” Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2015), 2015.

6 J. Jung, V. Paxson, A. W. Berger, H. Balakrishnan, “Fast portscan detection using sequential hypothesis testing,” IEEE Symp. Sec. and Priv, 2004.

7 M. Bastian, S. Heymann, M. Jacomy, “Gephi: An open source software for exploring and manipulating networks,” International AAAI Conference on Weblogs and Social Media, 2009.

Ichiro SHIMADA

KOZO KEIKAKU ENGINEERING Inc./
Former: Research Expert, Cybersecurity Laboratory, Network Security Research Institute
Network Security, Network Traffic Analysis

Yu TSUDA, Ph.D.

Researcher, Cybersecurity Laboratory,
Cybersecurity Research Institute
Cybersecurity, Countermeasure against APT