Designing a Fast Log-Tracing Scheme for Targeted Attack Prevention

Takuya Suzuki, Koki Ikeda, Pichaya Kietkaroon,
Junichi Murayama
School of Information and Telecommunication Engineering
Tokai University
2-3-23 Takanawa, Minato-ku, Tokyo, 108-8619, Japan

Takahiro Hamada,
Yuichi Murata

NTT Secure Platform Laboratories
3-9-11 Midori-cho, Musashino-shi, Tokyo 180-8585, Japan

***Abstract*—In this paper, we design a fast log-tracing scheme for preventing targeted attacks to enterprise information networks. In these attacks, confidential data leak through application gateways. In order to detect such leakage, a network management server collects multiple logs. Then a gateway traces them to check whether the forwarding data is confidential or not. In the conventional basic scheme, this check will require long processing time if log volume becomes large. In our proposed scheme, at first, multiple logs are preprocessed offline to form a black list. A gateway checks a file to be forwarded online using this black list. The evaluation results show that the tracing time can be shortened to one severalth by means of our proposed scheme.**

***Keywords*—targeted attack, information leakage, SIEM, log analysis, enterprise information network**

I. Introduction

Recently, targeted attack to enterprise information networks is becoming serious. In this attack, first, malware intrudes within an enterprise network. Then, it downloads confidential data from an enterprise file server. Finally, it sends the downloaded data to outside of the network via an application gateway (APGW).

In order to prevent such information leakage, server log and host log in the network should be analyzed rapidly. A Security Information and Event Management (SIEM) system is useful for this analysis [1]. This system alerts network operators when anomaly is detected through log analysis. In addition, it may trace several logs to improve anomaly detection probability [2].

On the other hand, an attacking method is becoming complicated more and more [3]. For example, filename of the downloaded data is changed in the middle of the attack [4]. In other case, operation interval within a single attack is extended to some months. These attack operations increase both the number of log-analyzing processes and the volume of log itself. Consequently, the processing time of anomaly detection is increased more and more [5]. This is an important problem for preventing information leakage by means of online processing.

In order to solve this problem, we first evaluated processing time of the basic log-tracing scheme. The results showed that several tens of seconds is required for log tracing in the typical condition. We think it is too large for online anomaly detection. Then, in this paper, we design a novel log-tracing scheme. In this scheme, at the initial offline processing, a black list is created from logs to be traced. In the online processing during the communication, only this black list is analyzed. We also evaluated our scheme and the results showed that the log-tracing period could be shortened to one severalth of the basic scheme. Although long period is required to create the black list, it does not spoil the fast online processing.

The rest of this paper is organized as follows: Section II shows the assumed network model and attacking model as preconditions on the study. Section III introduces the basic log-tracing scheme. Section IV represents the proposed log-tracing scheme. Section V shows the evaluation and their results. Finally, Section VI concludes this paper with brief summary.

II. Preconditions

A. Enterprise Network Model

An enterprise network is configured as an intranet as shown in Fig.1. This network comprises a file server, hosts and application gateway (APGW). A file server provides intranet hosts with files including confidential ones. APGW connects the Intranet with the Internet. Typical APGW examples are a web proxy and mail server.

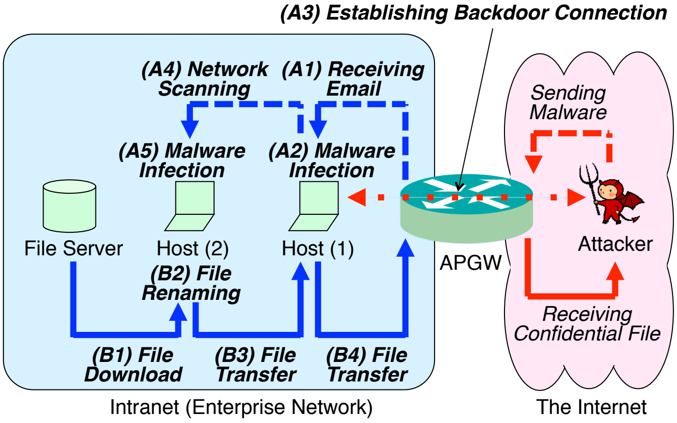

Fig. 1. Typical Flow of Targeted Atacks

B. *Targeted Attacking Model*

The targeted attacking model assumed in this paper is based on the literature [2]. This model is also depicted in Fig. 1. An intrusion flow is expressed as the dotted line at the top of this figure. The outline is as follows:

(A1) The host A receives an email including a malware.

(A2) The received malware infects the host A.

(A3) The malware establishes a backdoor connection with the attacker outside of the intranet.

(A4) It collects intranet configuration information by means of such as scanning.

(A5) It spreads and also infects the host B.

After the intrusion, the malware leaks confidential information. This leakage flow is expressed as the solid line at the bottom of Fig.1. The outline is as follows:

(B1) The host B downloads a confidential file from the file server.

(B2) It renames or splits the downloaded file.

(B3) The host B sends the processed file to the host A.

(B4) The host A sends the file to outside via APGW.

III. Basic Log Tracing Schemes

In order to detect information leakage, APGW should monitor all the forwarding data toward outside of the intranet. If the data has been encrypted, they should be decoded once to check. An example of such appliance is PacketBlackhole (PBH) [6]. When APGW detects a file in the forwarding data, it traces logs using SIEM to determine whether it is confidential or not. SIEM collects logs using a management server of the intranet. An example log-tracing platform is TRX [7]. SIEM can be deployed on this platform. If APGW detects a confidential file to be forwarded outside, it discards the file and disconnects the communication.

Logs are collected as shown in Fig.2. A management server and SIEM server are deployed in the intranet. They are logically located in the management plane because they handle only management data and do not forward user data. A management server collects logs from the file server and hosts. In this figure, the Log (1) is a filename list of confidential files stored in the file server. The Log (2) is a filename list of downloaded files from the file server. This log includes IP address of accessed host as attribution information. The log (3) is a filename list of operated files in the hosts. When a file is renamed or split, a new filename is also recorded in this log. The management server sends all collected logs to the SIEM server. In order to detect anomaly, the SIEM server analyzes those logs.

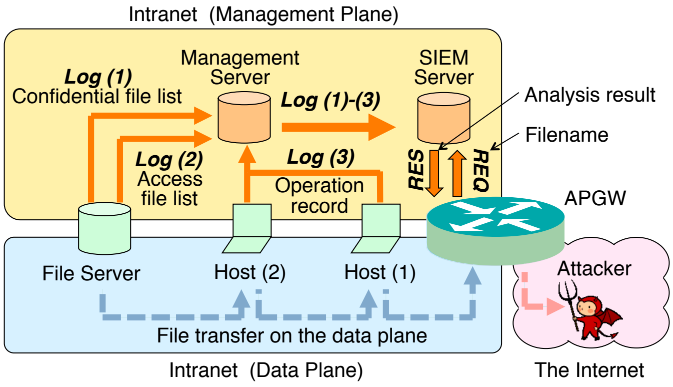

Fig. 2. Logs to be Traced for Anomaly Detection

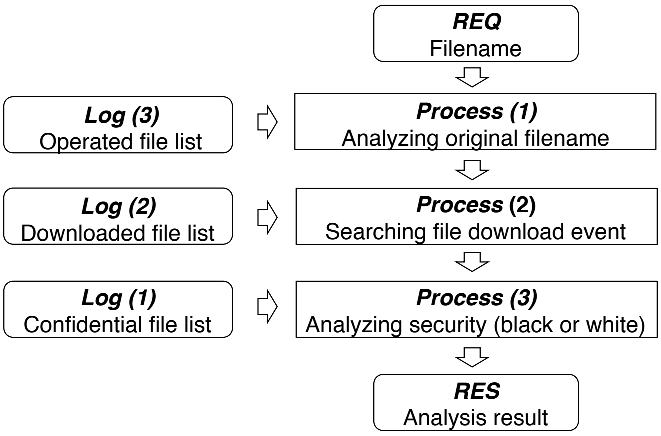

Fig. 3. Flow of Basic Log Tracing

The collected logs are traced in the SIEM server according to the flow depicted in Fig.3. When APGW transfers a file toward outside of the intranet, it requests SIEM server to check whether the file is a targeted one or not. Here, a targeted file is defined as a confidential one that has been downloaded from the file server by an attacker. Then, the SIEM server starts the log analysis.

At the initial process (1), it analyzes original filename of the requested file. Here, Log (3) that is a hosts' operated file list is searched using the requested filename as the search key. Then, at the next process (2), it checks whether the requested file is downloaded from the file server or not. In this process, Log (2) that is a server's downloaded file list is searched using the specified original filename as the search key. Finally, at the process (3), it checks whether the requested file is a targeted one or not. In this process, Log (1) that is a server's confidential file list is searched by using the specified downloaded filename as the search key. The analysis result is responded to the APGW from the SIEM server.

IV. Proposed Log Tracing Scheme

As the intranet scale or file access frequency increases, the volume of logs becomes large. In the basic log-tracing scheme, processing time becomes long as the log-volume increases. According to our estimation described in the next section, tracing log of 8 million lines requires 18 seconds. However, it should be shortened in order to prevent information leakage by means of online anomaly detection.

The weak point of the basic scheme is to perform all processing online. In order to solve this problem, we design a novel scheme to introduce offline processes in advance of online processes. A log-tracing flow of the proposed scheme is depicted in Fig.4.

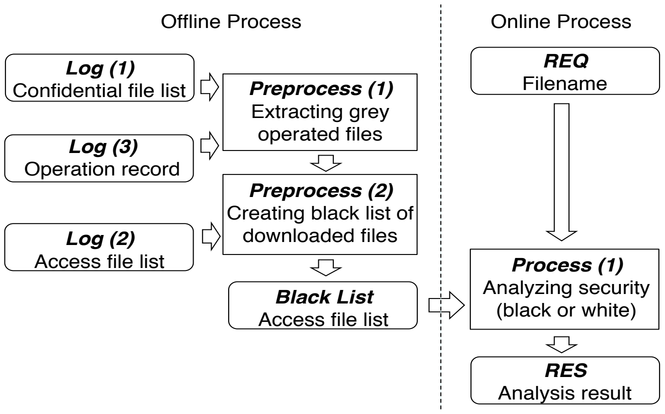

Fig. 4. Flow of Propose Log Tracing

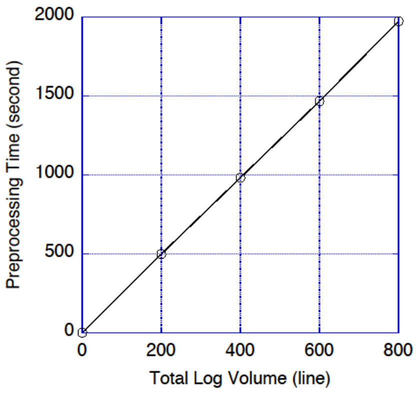

Fig. 6. Evaluation of Offline Preprocessing Time

As the initial offline preprocess (1), filenames that appear in both Log (1) and Log (3) are extracted to form a new list. This list shows the confidential files that might have been operated by hosts. Then, at the next offline preprocess (2), filenames that appear in both this created list and Log (2) are extracted to form a black list. This black list shows the confidential files that have been downloaded from the file server and have been operated by hosts.

After the offline preprocessing, APGW requests the SIEM server to check filenames. As the online process (1), the SIEM server searches the black list using the requested filename as the search key. The result is responded to the APGW.

V. Evaluation

Online log-tracing time of the network model shown in Fig.2 was evaluated by means of comparison between the proposed scheme and basic scheme. This comparison was performed using a prototype implementation of the emulated SIEM server. The results are depicted in Fig.5 In this figure, the horizontal and vertical axes show that the volume of logs (lines) and log-tracing time (seconds), respectively. Although the tracing time at 8 million lines was 18.6 seconds in the basic scheme, it was reduced to 2.3 seconds in the proposed scheme.

Offline preprocessing time was also evaluated. The results are depicted in Fig.6. In this figure, the horizontal and vertical axes show that the volume of logs (lines) and preprocessing time (seconds), respectively. The preprocessing time at 8 million lines was 1,972 seconds in the proposed scheme. In the evaluation conditions, the log volume to be searched online was reduced to one-twentieth by means of offline preprocessing.

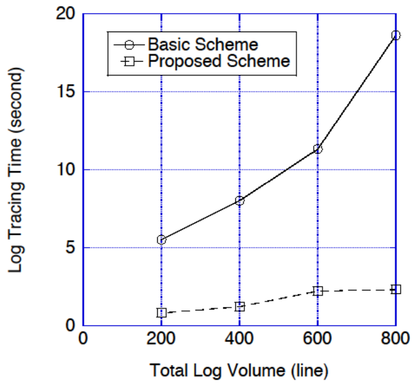

Fig. 5. Comparison of Online Log Tracing Time

Consequently, the proposed scheme needs long preparation time of offline preprocessing. However, it can reduce online tracing time almost one-tenth of the basic scheme.

VI. Conclusion

Online log tracing is attractive for enterprise networks to prevent leakage of confidential information. In the basic log-tracing scheme, processing time becomes long as the log-volume increases. In order to solve this problem, we design a novel scheme to introduce offline processes in advance of online processes. In our scheme, a black list is created at the offline preparation processing. Then online log-trace can be processed rapidly using the black list. The evaluation results show that after the long-time offline preprocessing, the proposed scheme can reduce online tracing time to almost one-tenth of the basic scheme.

References

[1] M. Futagi, M. Sato, F. Yamasaki and K. Uchida, "Consideration about Targeted Cyber Attack And Advanced Persistent Threat," IPSJ SIG Technical Reports, 2012-CSEC-56, February 2012. (In Japanese)

[2] T. Tomine, Y. Tsuda, M. Kamizono, K. Sugiura, D. Inoue and K. Nakao, "Timeline-Based Event Log Viewer over Multi-Host Environment," IEICE Technical Report, ICSS2013-79, March 2014. (In Japanese)

[3] H. Sakakibara and S. Sakurai, "Information Leakage Monitoring By Log Analysis," IPSJ SIG Technical Report, 2011-CSEC-52, March 2011. (In Japanese)

[4] S. Kitazawa, S. Sakurai, "A Detection Technique of a Targeted Attack Using a Decoy," IPSJ SIG Technical Report,2013-DPS-154, March 2013. (In Japanese)

[5] Y. Ishii, K. Ikeda, Y. Watanabe, B. Hu, Y. Murata and J. Murayama, "An Evaluation Method for a Targeted Attack Protection Scheme Comparing Multiple-Logs," IEICE General Conference, B-7-69, March 2014. (In Japanese)

[6] Net Agent, "Packet Black Hole," <http://www.packetblackhole.jp/>, January 2015 (Access) (in Japanese)

[7] T. Motoda, T. Nagayoshi, J. Akiba and K. Takeuchi, "The TRX Traceability Platform," NTT Technical Review, Vol.12 No.7, July 2014.