Proposal and Evaluation of Cyber Defense System using Blacklist Refined Based on Authentication Results

Hirofumi Nakakoji*†, Yasuhiro Fujii*, Yoshiaki Isobe*, Tomohiro Shigemoto*, Tetsuro Kito*,
Naoki Hayashi*, Nobutaka Kawaguchi*, Naoki Shimotsuma*, Hiroaki Kikuchi†

†Graduate School of Advanced Mathematical Sciences
Meiji University,
4-21-1 Nakano, Nakano-ku, Tokyo, Japan

*Hitachi, Ltd.,
292, Yoshida-cho, Totsuka-ku Yokohama-shi,
Kanagawa, Japan

***Abstract*— In recent years, the damage from cyber attacks caused by sophisticated malware has continuously increased. It is therefore becoming more difficult to take countermeasures using traditional approaches such as antivirus and firewall products. Against the intrusion of malware, we propose an automated countermeasure technology system named Autonomous Evolution of Defense, which mitigates the risk of actual damage by controlling the internet connection for malware, and in addition optimizes the system's operating conditions. The system takes countermeasures immediately to mitigate risk without causing disruptive effects on business. However, a graylist of malicious addresses generated by malware analysis systems contains many false-positive addresses and is very "noisy" for use in blocking access based on the list. We therefore propose a new technique for improving the accuracy of the unreliable graylist of addresses using image authentication. We report here on the implementation of our system and results of evaluation.**

Keywords— Malware; Proxy; CAPTCHA; Graylist; Blacklist

I. Introduction

Recent cyber attacks have been motivated by crimes such as a financial fraud, hacktivism and espionage. Attackers have acted in concert to carry out coordinated attacks each with their specialized tasks. Sophisticated methods for cyber attacks such as the zero-day and Watering Hole attacks[1], were used to target some critical infrastructures such as financial services, government and energy facilities. In June 2015, the Japan pension Service was compromised [2] using the malware EMDIVI [3] to exploit confidential information, and more than 44 organizations were involved in the same type of attack [4]. As knowledge about EMDIVI was not shared between these organizations, security incidents caused by the same malware occurred simultaneously. This means that there is a potential threat for massive cyber attacks.

In response to such sophisticated coordinated attacks, we propose the concept of "group defense" defined by sharing knowledge about threats and vulnerabilities with associated organizations. To put group defense into practice, government agencies such as Information Sharing and Analysis Center(ISAC) have been established [5][6], and FireEye [7] and ThreatConnect [8] have started new intelligence sharing services.

The naive way to solving the threat of attacks is to block the malicious connection to the suspicious addresses provided by the malware analysis system. However, the list of suspicious addresses extracted from the malware analysis contains some benign sites and URLs such as search engines. Hence, the "noisy" graylist cannot be adopted from the perspective of business continuity. Here, we propose a new system named Autonomous Evolution of Defense (AED) that adds an additional authentication to the proxy server when a client tries to access suspicious URLs provided from the dynamic malware analysis system. Therefore, even if the AED gives incorrect information concerning the URL of the benign site, the AED allows the internet connection for users who pass authentication without causing disruptive effects on business. Moreover, the AED blocks the internet connection that is accessed by a machine program (e.g. malware).

Here, we describe our proposed system that takes countermeasures immediately to mitigate risk based on an unreliable graylist of suspicious addresses and we evaluate the accuracy of improvement of the proposed method.

II. Related Work

A collaborative countermeasure that shares reports of malware analysis was proposed by Colajanni et al. [9]. Their proposed method distributes sensor nodes such as honeypots to multiple organizations, and then the collector server collects the malware from these sensor nodes. The collector server analyzes them using a malware dynamic analysis service and obtains the malware analysis report. Malware analysis reports (containing destination, port, and protocol) are shared by each organization and then used by organizations to block the malware activity. Further, a method of protection for malicious social networking sites using proxies was proposed by Tsai et al. [10]. Their approach scan the social networking site before a client accesses it; if the site is suspect, the proxy disconnects access and sends a warning message to the client.

These methods of countermeasure using a malware analysis system are similar to our proposed method. However,

2157-0426/16 $31.00 © 2016 IEEE
DOI 10.1109/NBiS.2016.25

if the malware partly involves benign activity (e.g. access to search engine services), the methods reported previously will take undesired countermeasures, which might then interrupt business activities. Our proposed method accepts intentional access by a human and prevents access by a machine program (e.g. malware), by providing additional authentication of countermeasures against unreliable information from malware analysis reports. Moreover, it has the advantage of improving the accuracy of countermeasures, such as automatically reducing the frequency of authentication by classifying unreliable information on suspicious addresses into highly reliable information using authentication results.

III. Proposed AED System

Fig. 1 shows an overview of the AED System, consisting of a graylist manager, a crawler, a risk-based proxy controller, and a multimodal malware analysis system (M3AS) [11]. The graylist manager manages a graylist of suspicious URLs. These are collected by the M3AS and the crawler, which crawls malicious URLs from the internet. The risk-based proxy controller provides an image authentication known as the "Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) before a user tries to access a suspicious URL included in the graylist. CAPTCHA is a type of challenge–response test used in computing as an attempt to ensure that the response is generated by a person. Therefore, when a user tries to access a benign site that is included in the graylist by mistake, the user can gain access by passing the CAPTCHA.

A. *Graylist Manager*

The graylist manager collects URLs from the following sources: M3AS and crawler (VirusTotal and ThreatConnect). It updates a whitelist of clearly benign addresses as well as the blacklist.

1) *M3AS*

Recent attacks were designed with malware to work only in specific environments and applications that have a vulnerability. We call such malware "an environment-dependent malware" (EDM). Existing malware analysis systems with a single sandbox fail to expose the behavior of EDMs because of environment mismatching. M3AS analyzes multiple environments to expose the malicious behavior of a given malware. The graylist manager stores the access URLs of a malware exposed by M3AS as a graylist.

2) *Crawler*

a) VirusTotal

The VirusTotal[12] service provides a summary of multiple reports of antivirus software. This service provides a malware's behavior database consisting of the malware's behavior observed by dynamic malware analysis. The crawler retrieves the URLs that the malware may have accessed from the VirusTotal service.

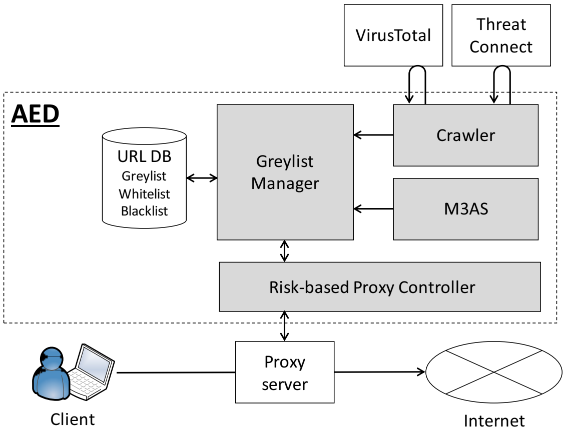

Fig. 1. Overview of AED

b) ThreatConnect

The ThreatConnect service provides a comprehensive threat intelligence platform, and this service shares the malicious URLs for the Security Operation Center (SOC) operator. The crawler retrieves malicious URLs from this service using the ThreatConnect API.

B. *Risk-based Proxy Controller*

Contemporary malware used in a targeted attack communicates with the attacker via a Command and Control (C&C) server. The risk-based proxy controller detects the connection to a suspicious destination and requires a CAPTCHA image authentication for the user. The controller uses the graylist to determine when the additional authentication is required. Fig. 2 shows the architecture of a risk-based proxy controller.

The proxy server offers a network service to allow clients to make indirect network connections to other hosts. We use a proxy server such as squid [13]. The Internet Content Adaptation Protocol (ICAP) [14] was used for additional authentication judgement at the proxy server. With this architecture, the AED can cooperate with the various proxy products that are installed already.

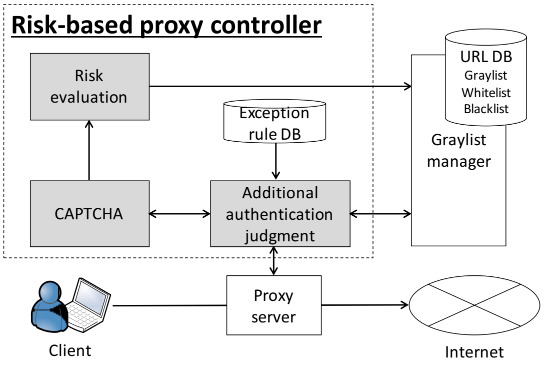

Fig. 2. Architecture of Risk-based Proxy Controller

The risk-based proxy controller consists of three subfunctions and a database.

1) Additional Authentication Judgment

The proxy server performs CAPTCHA before it allows a browser a network connection. If a given URL is included in the graylist, with CAPTCHA we have the user tested, to confirm that a human wishes to access the URL. If a given URL is included in the blacklist, we deny the connection request and warn the user with an alert message. If a given URL is included in the whitelist, we allow the connection.

Some web pages contain embedded videos and cascading style sheet (CSS) files that are hosted by another server. These materials are often classified as a graylist and involve a misclassification. To prevent potential attacks, we add CAPTCHA for a user who cannot authenticate successfully because the authentication process is performed in the background, and the web page will not be displayed correctly. To solve this problem, we create the exception rules shown in TABLE I.

2) CAPTCHA

Our CAPTCHA generates a distorted image text and asks the user to read the text correctly. Note that a simple malware fails to recognize the distorted text.

3) Risk Evaluation

We classify the authentication results into three statuses in TABLE II. URLs are classified into three lists, black, white, and gray based on the statistics of the authentication status. The blacklist, the whitelist and the graylist contain respectively clear malicious sites, clear benign sites, and suspicious sites. Each list uses Fully Qualified Domain Name (FQDN) as URL.

TABLE I. EXCEPTION RULES IN DB

Conditions	Actions
An http header includes a referrer tag	No CAPTCHA
Referrer tag indicates search engines: www.google.co.jp www.google.com www.bing.com www.yahoo.co.jp	Add CAPTCHA if http header includes a referrer tag

TABLE II. STATUS OF AUTHENTICATION

Status	Description
Success	CAPTCHA answer is correct
Failure	CAPTCHA answer is incorrect
No Try	No response within the specified time, e.g. 10 min

TABLE III. CLASSIFICATION RULES

Conditions	Actions
Number of “Success” results per URL is greater than 5	Classify the URL into the whitelist
Number of “Success” or “Failure” results per URL is 0 and the number of “No Try” results per URL is greater than 5	Classify the URL into the blacklist

In this paper, we simply classify the graylist into a whitelist and a blacklist according to the definitions in TABLE III. The graylist is classified automatically by repeating authentications. We identify users by means of the IP address or the user ID used in the BASIC authentication of the proxy server.

IV. EVALUATION

We evaluated the performance of the security risk mitigation. Accordingly, we conducted two experiments in this paper.

A. Focus of Evaluation

Experiment #1: Mitigation of security risk

In experiment #1, we evaluated the security risk mitigation of an infected client PC with malicious activities including a connection with C&C, and downloading of other malware.

Experiment #2: Adverse effects on business caused by adding CAPTCHA

In experiment #2, we evaluated the usability of our system that required additional CAPTCHA with the unreliable graylist.

B. Experiment Methods

We implemented AED, and the AED retrieved 52,653 suspicious URLs from M3AS and the crawler. M3AS analyzed 2,064 types of malware (received by our organization from March 2014 to January 2016) with 46 sandboxes. The crawler collected the malicious URLs (registered from October 1, 2015 to February 24, 2016, i.e. roughly 17 months) from the VirusTotal service and the ThreatConnect web page. We classified these URLs into graylist. The whitelist and the blacklist were initialized by blanks in both experiments.

In experiment #1, we assumed that all files attached to emails are analyzed by the M3AS before those are received by users.

We prepared a graylist provided by M3AS that analyzed malware samples, and tested whether the client could connect to the internet. We observed the behavior of sandboxes that were collected by the analysis environment of M3AS and infected with malware. Hence, the AED can have the graylist of all destination URLs reported by the M3AS. TABLE IV.

TABLE IV. NUMBER OF URLS

Information Source	count
M3AS	16,384
VirusTotal	31,937
ThreatConnect	6,257
Total	54,578
Union	52,653

TABLE V. EFFICIENCY OF COUNTERMEASURES AGAINST MALWARE

	Success	Failure	Accuracy
Malware	2,057	7	99.66%
URL	1,000	7	99.30%
Connection	212,094	43	99.98%

shows the statistics of suspicious URLs for each of sources. We use these URLs as graylist.

In experiment #2, to evaluate the convenience and adverse effects on business, we reproduced the graylist including a large amount of benign sites, and tested AED with subjects.

We evaluated the AED with two cases in experiment #2. All URLs on the internet were classified as a graylist in Case #1; the URLs in TABLE IV. were classified as a graylist in Case #2. To clarify the usability of our proposed AED, we conducted a survey questionnaire with 19 test subjects in our department. The subjects were asked to use the AED for daily business activities. Therefore, the experimental environment was clean without malicious access. In our experiments, 43 subjects participated for 16 days for Case #1, and 35 subjects joined for 9 days for Case #2.

C. *Experiment Results*

Experiment #1: Mitigation of security risk

TABLE V. shows the results of the experiment. More than 99% of malware connections to the internet were successfully blocked by the AED.

It failed to block seven malicious connections because the graylist included the URLs used by the malware which changed their behavior each time it ran. These malware usually use what has been called Domain Generation Algorithms (DGA)[15].

Experiment #2: Adverse effects on business caused by adding CAPTCHA

The total access numbers for the case #1 and #2 were 815,246 and 543,489, respectively. The unique numbers of FQDNs were 5,794 and 5,146 for Cases #1 and #2, respectively.

TABLE VI. shows the overview of classified lists. We illustrate classification of graylist in Fig. 3. The rate of successful additional authentications in Case #1 (No.3 in TABLE VI.) was 23%. In contrast, it was less than 1.0% in Case #2. The rates of failure (No.4 in TABLE VI.) were 7.0% and 3.2% in Cases #1 and #2, respectively. The results from the questionnaire showed 47% of the subjects found the AED inconvenient because too many authentications were required in Case #1, while no subjects felt that there were too many authentication in Case #2.

TABLE VI. RATIOS OF AUTHENTICATION AND NUMBER OF CLASSIFIED LISTS

No	Condition	Case #1	Case #2
1	Number of domains for which AED gave additional authentication	2,358 (100.0%)	186 (100.0%)
2	Number of domains for which the authentication status was "No Try"	1,034 (43.9%)	149 (80.1%)
3	Number of domains for which the authentication status was "Success"	1,324 (56.1%)	37 (19.9%)
4	Number of domains for which the authentication status was "Failure"	164 (7.0%)	6 (3.2%)
5	Number of domains classified in the whitelist	54 (0.9%)	7 (0.1%)
6	Number of domains classified in the blacklist	58 (2.5%)	46 (24.7%)

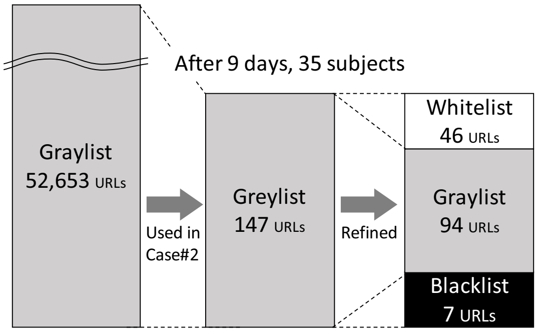

Fig. 3. Overview of Classified Lists

The ratio of URLs classified as either the blacklist or whitelist was respectively 3.4% for Case #1 and 24.8% for Case #2. The number of classified domains appears to depend on the quantities, such as the number of URLs in the graylist, the number of users, and the period of trial operation.

The whitelist contained benign sites such as www.adobe.com and www.google.com. However, benign certificate validation servers such as crl.verisign.com and ocsp.entrust.net were misclassified into the blacklist.

We expect that the usability of our system is going to be improved while the graylist are classified into either black or white. In fact, the more than half of subjects agreed the improvement of usability day by day.

V. CONCLUSIONS

We have proposed the AED method, which takes countermeasures to mitigate risk without disruption of business activities. Our system determines if a given URL is benign or malicious based on the suspicious noise in URLs reported by the malware analysis system. When users access a suspicious site included in the graylist, they need to pass an image authentication CAPTCHA test with the proxy server. Hence, our system does not disrupt business activities even when URLs are misclassified as belonging to the graylist. Moreover, based on the authentication results, our system could reduce the frequency of authentication by improving accuracy of the suspicious graylist, the whitelist, and the blacklist. We implemented the proposed AED and our experiment showed that our system succeeded in blocking malicious connection with more than 99% accuracy. According to the findings from a questionnaire, 47% of subjects found that the system had, at worst, adverse effects on their business.

Our experiment showed automated classification for validation of public key certificates failed. When those URLs were misclassified into a blacklist, users were not able to validate any certificates. Addressing these problems is a task for future work.

The system, products and service names used in this paper are generally the trademark or the registered trademarks of each organization.

REFERENCES

[1] Tech Target. (2016) watering hole attack. [Online]. Available:http://searchsecurity.techtarget.com/definition/watering-hole-attack

[2] Cybersecurity Strategic Headquarters. (2016) The report of cause investigation on personal information leakage incidents in the Japan pension service(in Japanese). [Online]. Available:http://www.nisc.go.jp/active/kihon/pdf/incident_report.pdf

[3] Symantec Corporation. (2016) Backdoor.Emdivi. [Online]. Available:https://www.symantec.com/security_response/writeup.jsp?docid=2014-101715-1341-99

[4] K. Kubo, Corresponding to the targeted attacks(in Japanese). (2015) Japan Computer Emergency Response Team Coordination Center. [Online]. Available:https://www.jpcert.or.jp/present/2015/JNSAWG20150630-apt.pdf

[5] Telecom-ISAC JAPAN. (2016) Telecom Information Sharing and Analysis Center Japan. [Online]. Available:https://www.telecom-isac.jp/

[6] Financials ISAC Japan. (2016) Information Sharing and Analysis Center. [Online]. Available:http://www.f-isac.jp/

[7] FireEye. (2016) FireEye Threat Intelligence. [Online]. Available:https://www.fireeye.jp/content/dam/fireeye-www/regional/ja_JP/products/pdfs/ds-threat-intelligence.pdf

[8] THREATCONNECT,INC. (2016) Enterprise Threat Intelligence Platform. [Online]. Available:https://www.threatconnect.com/

[9] Michele Colajanni, Daniele Gozzi, and Mirco Marchetti, “Collaborative architecture for malware detection and analysis, " IFIP International Federation for Information Processing, vol. 278, Proceedings of the IFIP TC 11 23rd International Information Security Conference, pp. 79-93, 2008.

[10] Dwen-Ren Tsai, Allen Y. Chang, Sheng-Chieh Chung, You Sheng Li, “A Proxy-based Real-time Protection Mechanism for Social Networking Sites," Security Technology (ICCST), pp. 30-34, 2010.

[11] Hirofumi Nakakoji, Tomohiro Shigemoto, Tetsuro Kito, Naoki Hayashi, Masato Terada, Hiroaki Kikuchi, “Proposal and Evaluation of Multimodal Malware Analysis System with Multiple Types of Sandboxes(in Japanese), " IPSJ Journal, Vol. 56, No. 9, pp. 1730-1744, 2015.

[12] Google. (2016) VirusTotal. [Online]. Available:https://www.virustotal.com/

[13] squid-cache.org. (2016) Squid. [Online]. Available:http://www.squid-cache.org/

[14] Elson, J. and A. Cerpa, Internet Content Adaptation Protocol (ICAP), RFC 3507, DOI 10.17487/RFC3507, http://www.rfc-editor.org/info/rfc3507, 2003.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, pp. 68-73, 1892.

[15] Manos Antonakakis, Roberto Perdisci, Yacin Nadji, Nikolaos Vasiloglou, Saeed Abu-Nimeh, Wenke Lee and David Dagon, “From Throw-Away Traffic to Bots:Detecting the Rise of DGA-Based Malware,” Security'12 Proceedings of the 21st USENIX conference on Security symposium, pp. 491-506, 2012.