File Multi-analyses for Real-time Attack Source and Spread Site Trace

DaeHee Seo[1], SangWoo Lee[1], ByungDoo Kim[1], ByungGil Lee[1] and JangMi Baek[2]

[1]*Electronics and Telecommunications Research Institute*
[2]*SoonChunHyang University*
[1]*{dhseo, ttomlee, bdkim, bglee}@etri.re.kr,* [2]*bjm1453@sch.ac.kr*

Abstract

Recently, the illegal users with malicious intention utilize the file sharing site by making normal user's computer a zombie computer, which is a preliminary process for network intrusion attack. The propose scheme is divided into the method for real-time analysis for real-time tracking and the method of cooperative analysis method for non-real time analysis. By allowing the supervisors to choose the relevant analysis method selectively, we can conduct variable analysis depending on the network threat.

Keywords: *Cyber-attack, IP Trace, File analysis, Trace log, Global Cooperation*

1. Introduction

While the rapid advancement of the internet provides diversified communication and services the social elements, there is also an increasing number of cyber-attacks ranging from DDoS (Distributed Denial of Service) attack to the illegal leakage of user privacy against certain social infrastructure, leading to the social chaos [2, 9]. Especially, in the process that many users use the file sharing system universally, the illegal user with malicious purpose distributes files, making normal user's zombie PC, therefore utilizing it as network invasion attack. Thus, tracking the attack source and spread site is utmost needed in social chaotic situation, however, there exists limitations to the trace technology that is reliable and in real time, it is problematic to the social integration. In particular, known malignant code can be blocked by existing network security system, on the other hand, it is required that we need to conduct a practical analysis on suspicious unknown malignant file [6, 10]. Therefore, this paper proposes the method that can analyze independently each type of files that are uploaded to the file sharing site. Also, we distinguished between the real-time analysis and non-real-time analysis, thus enabling detailed analysis. For this, the second section of this paper describes the summary of file analysis and the third section suggests the security requirement after analyzing the vulnerability of existing file analysis methods. The fourth section proposes file multi analysis for tracking the attack source and spread site that suffices the security requirement that is suggested in the third section. Henceforth, in the fifth section, we analyze the proposed scheme by comparing it with the existing method. Finally, the sixth section derives the conclusion and provides a direction to the follow-up studies.

2. Overview on Technology

The classical file sharing method consisted of server-client structure. However, with the rapid advancement of the internet and personal devices, file sharing sites is generalized that sells digital information to the users by constructing common file sharing site based on simple form of server-client structure [1, 4]. On the other hand, while file sharing site provides generalized digital information services, there is an adverse effect to the social/national problems that it is utilized as a mean of DDoS attack, by uploading illegal files and making user's zombie PC. Especially, like 3.4 DDoS attack, for making numerous zombies PC in advance, the attacker stimulates user's curiosity or distributes the files disguised as certain file, making users install or download such files without users realizing that their computer becomes to zombie [11-12]. Therefore, different from existing file analysis of network security or anti-virus product, the method for confirming the validity of the files that are downloaded or uploaded through the file sharing site is separately needed. Based on this, it is needed that by making trace log, the study researches tracking the attack source and spread site by tracing the process of initial uploader and downloader and the site where the file is uploaded or downloaded when the network invasion attack is occurred by that file.

3. Analyzing the Existing Schemes

We can divide the existing methods into the method analyzing directly the source coda and the method verifying file monitoring and signature. We can distinguish direct source code analysis between ITS 4 and emulator (SASS) and sort the verification method of file monitoring and signature into virtual reality method and time stamp method. ITS4 method is the method that involves scanning device that finds C+ and C++ source code security vulnerability. it splits it into strum in receiving source file input, judging the accuracy by comparing the database it has and marks the indicator. So depending on the amount of the database it has, it can enhance the accuracy of pattern matching and has a high level of extraction rate to known attacks [3]. But this method has a poor accuracy since it involves heuristic method when searching for the indicator. Emulator (SASS) method, which is the method that searches for static assembly code, consists of input section search, summary DB and output. This method, targeting assembly code, traces the control stream according to abstract execution and sets virtual link according to function summarization. After setting up the virtual link, it verifies variable durability according to virtual memory cell, tracing the control stream by abstract execution [7-8]. Thus, it provides assembly code search method approaching it as new methods including buffer overflow, at the same time, it has vulnerabilities such as difficulties in real-time application and a great amount of analysis time is needed by playing the role as an interpreter that emulates abstract execution while reading assembly code step-by-step.

Through the behavior analysis method, which is suggested by Ahn Lab in 2011, and continuous monitoring, we verify suspicious files and conduct the process whether they are malignant code or not by executing it in virtual space. Therefore, this method minimizes the inconveniency of users through the process of executing directly the suspicious files and is able to judge it in a concrete form of suspicious file by direct execution [13-14]. However, that it is hard to analyze it in real-time according to direct execution and the increase of the amount of data and the efficiency in terms of management since it is applied in virtual space are problematic in this method.

The method involving time stamp is suggested by R. Koen, *et al*., in 2008. It uses digital forensics and conducts behavior analysis on data moving and change that involve the feature

of time change. We can infer the behavior and time of users based on time information acquired by measuring different time change in this method managing it and allocating data strum to the cluster [5]. Therefore, it is possible to analyze and predict the behavior of users in pace with the time flow, on the other hand, it is hard to analyze the behavior of data and users in real-time flow and there exists a vulnerability that the analysis result is different according to the measuring method of certain behavior and time.

4. File Multi-analysis for Real-time attack source and spread site trace

4.1. Assumptions

For the file multi analysis method to trace the web-based real-time attack source and spread site, we make the assumptions as the following:

- Web-based file sharing site has the information on the known attack source or black list (IP, ID, MAC) of spread site registered as Rule in advance.

- Database on Header signature, Footer signature and BFD to known files is constructed in advance.

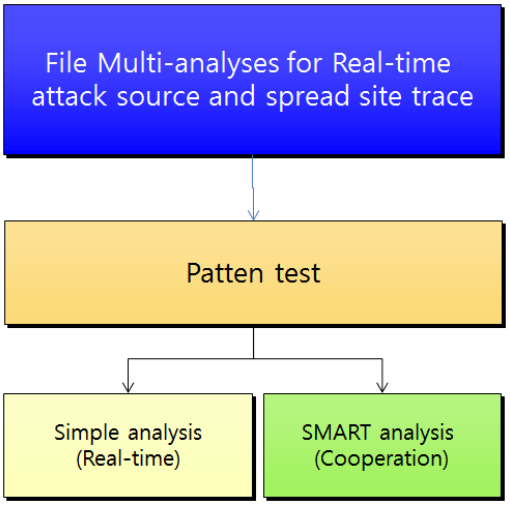

Figure 1. File Multi-Analyses Scheme

4.2. File Multi-analysis for real-time attack source and spread site trace

1) Pattern test

Pattern test consists of Simple test and SMART test as such:

A) Simple test

① Based on known Rule, simple test registers IP, ID or MAC of previously known black list in advance and executes black list-based Rule test

User ID	Network Address	NIC Address	Global Position	Trace log	Option field
ID_1	IP_1	MAC_1	GP_1	TL_1	....
ID_2	IP_2	MAC_2	GP_2	TL_2	....
....	....	....	....	....	....
ID_n	IP_n	MAC_n	GP_n	TL_n	....

②In case the new file is uploaded, vaccine program installed at file sharing site detects it on the basis of reputation and notifies reputation-based detected1 files.

③If the file that needs reputation-based detection is uploaded, fuzzy theory is used for Rule-based comparing and searching as such:

Ⓐ If unknown file is uploaded, we define the fitness on R_k as the following:

$$u_k(P_i) = u_{k_1}(P_1) \cdot u_{k_2}(P_2) \cdot \ldots \cdot u_{k_n}(P_{i_m})$$

We decide result class C_k and certainty gc_k on a fuzzy set M_{k_j}. According to the fitness of fuzzy rule R_k,

$$\beta_c = \sum_{P_i \in Class\ C} u_k(P_i), (C = 1,2,\ldots,n)$$

We decide a class (c') that has the maximum amount of the sum of fitness on result class C_k of fuzzy rule R_k, then we calculate $\beta_{c'} = \max\{\beta_1, \beta_2, \ldots, \beta_n\}$. If we cannot decide c', we define C_k as dummy class(empty class)

Ⓑ The certainty of all dummies is set as $g_{c_k} = 0$ and the certainty on other rules is defined as such :

$$g_{C_k} = \frac{\beta_{c'} - \bar{\beta}}{\sum_{c=1}^{n} \beta_c}, \bar{\beta} = \frac{\sum_{c \neq c'} \beta_c}{n-1}$$

Ⓒ We can guarantee the certainty of Rule based on current network information by defining g_{C_k} in the option field

Ⓓ After defining g_{C_k} in the option field, data on each of option value $\sigma(N)$ are calculated as the following. We calculate the data $\sigma(N)$ on each of option value as the following and decide whether we generate the trace log on uploaded file according to $\sigma(N)$.

$$\sigma(N) = \prod_{p|N}\left(1 - \frac{1}{(p-1)^2}\right)\prod_{p|N}\left(1 + \frac{1}{(p-1)^3}\right) > \frac{1}{2}, \begin{cases}\sigma(N) > \frac{1}{2} \\ \sigma(N) \leq \frac{1}{2}\end{cases}$$

If $\sigma(N) > 1/2$, then we do not generate trace log. If we do not generate trace log, post uploaded file to the server and $\sigma(N) \leq 1/2$, we send the attack source and spread site for generating trace log on this.

B) SMART test

[1] Through the user's reputation information, the technology confirms the credibility of unknown files and detects the dangerous elements, providing the reputation information after mining data and evaluating the reputation. Nowadays, mainly anti-virus programmed adopts that technology and utilizes it in the analysis and detection of virus.

SMART test analyze behavior analysis process based on unique format that each of file has as following:

①The uploader uploads Sample.jpg file to the file sharing site.

②File sharing site checks Header signature and Footer signature of uploaded Sample.jpg file.

JPEG's Header signature: "FF D8 FF ED"

③If the signature of uploaded file and known signature are the same, we upload and post that file to the board. If it is not the same Header signature, we use BFA (Byte Frequency Analysis) algorithm to extract the specific pattern of BFD (Byte Frequency Distribution) depending on types of the same file

④The frequency distribution for extracting a specific BFD pattern is as such:

$$\text{NFPS} = \frac{(OFPS \times PNF) + NFS}{PNF + 1}$$

- NFPS(New File Print Score), OFPS(Old File Print Score), PNF(Previous Number of Files), NFS(New File Score)

⑤After calculating the frequency distribution for extracting a specific pattern of that uploaded file, we decide whether to generate trace log on that file depending on types of file through the comparison of identification DB of known BFD.

5. Analysis of the proposed scheme

5.1. Security Analysis

The proposed scheme minimizes the overhead that can be occurred in self-monitoring situation through the analysis on the file itself, not on source analysis on files through continuous monitoring and is able to cross-analyze the suspected malignant files by constructing security database on previously known blacklist.

- Information sharing between systems: Sharing the analysis result on file multi analysis for tracing the attack source and spread site consists of RID (Router-ID). Therefore, real-time sharing of file multi analysis result is possible so that we can reflect it to the security policy based on the relevant information.
- Verification on Rule addition: By allowing the manager not only to decide but also add Rule arbitrarily in static form, it is possible to manage each of network and system organically and to verify the certainty of the result on additional rule.
- Efficiency of the management: In this method, it is possible to utilize the real-time analysis method selectively depending on the network threat status in usual. If the network threat is high, by additionally conducting cooperation analysis, the organic response is possible according to the security level of that web-based file sharing site.

5.2. Implementation

The realization environment of the attack source and spread site to which global cooperation-based integrated security control system is applicable is shown in Table 1 and we developed each of systems depending on (Figure 1, 2)

Table 1. Implementation Environment

OS	Linux/Windows
Database	Oracle Enterprise 10g
Application	Java 1.6, Adobe Flex 4.0
WAS	Apache tomcat 6.0

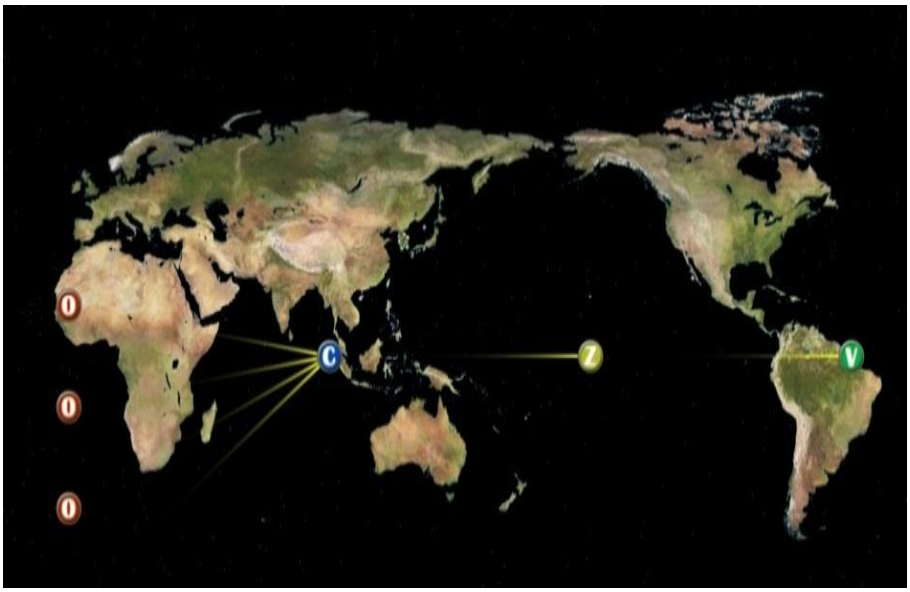

Figure 2. Real-time attack source and spread site trace system based on File Multi-analysis (3D)

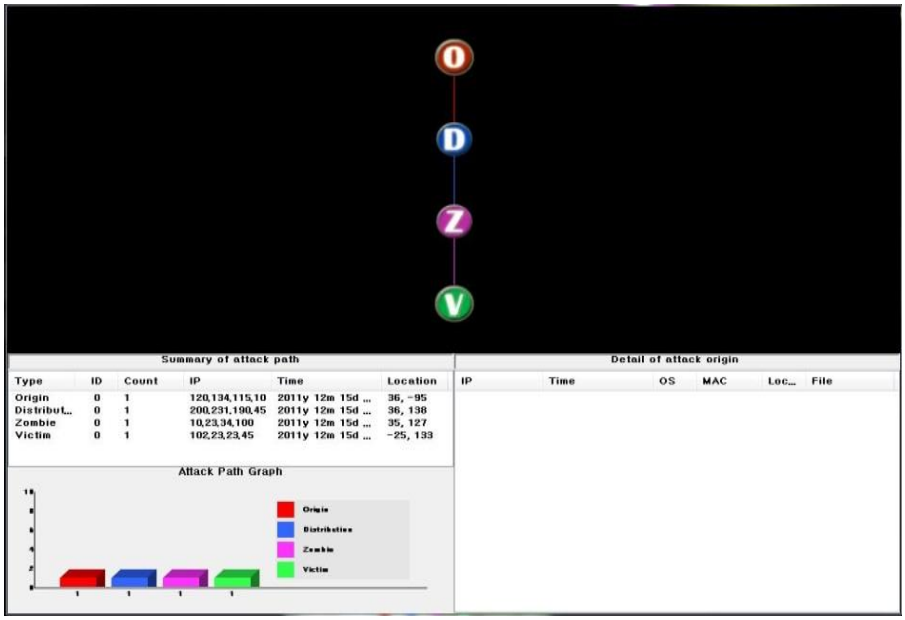

Figure 3. Real-time attack source and spread site trace system based on File Multi-analysis (2D)

6. Conclusions

The rapid advancement of the internet makes it possible the real-time information sharing of the data, supporting the diversity and universality of user approach and providing enhanced information service. Especially, web-based file sharing site has a strong point that the user can be provided with the information whatever and whenever he or she wants, however, the analysis of the uploaded file is needed since it is used as a attack spread site of network invasion attack. Therefore, proposed scheme suggests file multi analysis that can be utilized in the web-based file sharing site for constructing the trace system. To generate the trace log that is required in the trace system of the attack source and spread site, the proposed scheme suggests the method for real-time cooperative analysis on the uploaded file in the web-based file sharing site. Also, depending on the types of network security, it is possible not only to process it in real time but also to analyze it in detail, which enables the application on the attack source and spread site through the analysis on the file itself.

In the future, proposed scheme will departmentalize the proposed scheme and design the optimum analysis schedule for the real-time and non-real-time analysis and minimizing calculation overhead according to each of analysis method.

Acknowledgements

This work was supported by ETRI through Maritime Safety & Maritime Traffic Management R&D Program of the MLTM/KIMST (D10902411H360000110, Development of u-VTS for Maritime Safety).

References

[1] A. J. Gonzalez, R. M. Pozuelo, M. German, J. Alcober and F. Pinyol, "New Framework and Mechanisms of Context-Aware Service Composition in the Future Internet", ETRI Journal, vol. 34, no. 1, **(2013)**.
[2] D. Evans, LCLint User's Guide, University of Virginia, Department of Computer Science Publisher, Virginia **(2000)**.
[3] J. Viega, J. T. Bloch, Y. Kohno and G. McGraw, "ITS4: A Static Vulnerability Scanner for C and C++ Code", Proceedings 16th Annual Computer Security Applications Conference, New Orleans, Louisiana, US, **(2000)** December 11-15.
[4] M. Elsman, J. S. Foster and A. Aiken, "Carillon. A System to Find Y2K Problems in C Programs", **(1999)** http://bane.cs.berkeley.edu/carillon.
[5] R. Koen and M. S. Olivier, "The user of file timestamps in digital forensics", Proceedings of the Information Security South Africa Conference, Johannesburg, South Africa, **(2008)** July 7-9.
[6] R. Joanna, "Advanced Windows 2000 Rootkit Detection Execution Path Analysis", Proceedings of Black Hat USA, Las Vegas, US, **(2003)** July 28-31.
[7] S. A. Taghanaki, M. R. Ansari, B. Z. Dehkordi and S. A. Mousavi, "Nonlinear Feature Transformation and Genetic Feature Selection: Improving System Security and Decreasing Computational Cost", ETRI Journal, vol. 34, no. 6, **(2012)**.
[8] W. K. Kim and W. Y. Soh, "Design and Implementation of the Detection Tool of API Hooking Based on Window XP Kernel", Journal of Security Engineering, vol. 7, no. 4, **(2010)**.
[9] Y. H. Shah, M. Raza and S. UlHaq. Communication Issues in GSD. IJAST, vol. 40, **(2012)**.
[10] Z. Al-Ameen, G. Sulong and Md. Gapar Md. Johar, "A Comprehensive Study on Fast image Deblurring Techniques", IJAST, vol. 44, **(2012)**.
[11] D. Wheeler. Flawnder, (http://www.dwheeler.com/awnder/), **(2004)**.
[12] TREND MICRO report, (http://kr.trendmicro.com/kr/support/threat-reports/2011/TR007/index.html), **(2011)**.
[13] Ahnlab (http://www.ahnlab.com).
[14] Ahnlab Threat Research & Response Center (http://core.ahnlab.com).

Authors

DaeHee Seo received his Ph.D. degree from Soonchunhyang University, Choongnam, Korea. He is currently a Senior Member of Engineering Staff with the Electronics and Telecommunications Research Institute, Daejeon, Korea. His research interests include key management, network management, wireless security, and ubiquitous computing.

SangWoo Lee received his BS, MS, and PhD degrees in electronics from Kyungpook National University, Daegu, Rep. of Korea, in 1999, 2001, and 2009 respectively. Since 2001, He has been a senior member of engineering staff in Electronics and Telecommunications Research Institute (ETRI). His research interests include information security based on cryptography and its applications.

ByungDoo Kim was born in GyeongBuk, Korea, in 1971. He received the M.S. and Ph.D. degrees from the Department of Electronics Engineering from Ajou University, Korea, in 1998 and 2007, respectively. In 2002, he was with MteQ Systems (Currently STX Engine), Korea, where he worked on the research of signal processing and tracking algorithm of sea surveillance radar systems. Since 2004, he has been with Electronics and Telecommunications Research Institute (ETRI), Korea, as a Senior Researcher. His research interests include multi-target tracking, signal processing of radar system, and navigation algorithm based on GNSS for Telematics application systems.

ByungGil Lee received a Ph.D degree in electrical engineering from the Kyungpook National University in 2003. In 2001, he joined the research member of ETRI in Korea and is currently a team leader of Convergence Security Research Team. His current research interests in wired/wireless network security and convergence security.

JangMi Baek received an M.S. and Ph.D. in Computer Engineering from SoonChunHyang University in South Korea, in 2001, and in 2006, respectively. She worked at the School of Information Studies of Howard University as a Post-Doc from 2006 to 2007. He has been working as a visiting professor at SoonChunHyang University in South Korea since 20011. She is interests include Multimedia Network System, Mobile System Development and Design, Ubiquitous Healthcare System Development, and Embedded System.