2009 Fifth International Joint Conference on INC, IMS and IDC

A Stepwise Methodology for Tracing Computer Usage

SeungBong Lee, Jewan Bang, KyungSoo Lim*, Jongsung Kim, and Sangjin Lee
Center for Information Security Technologies, Korea University
Seoul, Republic of Korea
{*fdc629, jwbang, lukelim, joshep, sangjin*}*@korea.ac.kr*

***Abstract*—In digital forensics investigation, a general method of investigating the suspect's computer was to duplicate storage media or image and then obtain the case-related data from these. However, the increase in the capacity of storage media made this method take much longer time. Also, this implies that more data can exist in the suspect's computer so that finding relevant data will take a lot of time and efforts. Moreover, in case where imaging of the entire disk is not possible due to legal matters, selective acquisition of data is needed. In this paper, we propose methods for selective acquisition of file system metadata, registry & prefetch files, web browser files, specific document files without duplicating or imaging the storage media. Furthermore, we suggest a method to analyze the acquired data stepwise and quickly and effectively trace the use of computer in the crime scene.**

***Keywords*-selectively acquisition; pre-investigation; PIM;**

I. Introduction

Digital forensics refers to the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations [1]. Digital forensic investigations have been focused on disk images. Recently, however, some limitations have arisen in this method of investigating evidence.

Firstly, the time for disk imaging has increased. In general forensic investigations, data integrity will be damaged when the hard disk of the target system is directly investigated. Thus, disk imaging is needed to make a copy of it. That is, disk imaging is done to every disk on the target system to produce copies, which in turn are used for the investigation. However, recent development of storage technology has abruptly increased the use of mass storage, making the time elapsed for the duplication of original hard disks much longer. The storage of hard disks on the market is already exceeding terabyte, where imaging of a 250-gigabyte hard disk by hardware device takes more than 90 minutes [2]. If a person is using mass storage device or multiple devices at once, disk imaging will take a considerable amount of time. This implies that more time will be elapsed to find the case-related data from the evidence collected by an investigator and acquiring evidence have become more difficult [3].

Secondly, there are cases where the suspect's hard disk cannot be confiscated or duplicated. As digital forensics have recently gained much recognition, legal circles are objecting to the disk imaging as it can violate the suspect's privacy or cause secret leakage when investigating a company [4], [5]. Since indiscreet disk imaging investigation will make all information of the target system open to the investigator, selective methods that only confiscate case-related data are of uppermost importance. Moreover, law firms or accounting firms that are brought at a civil lawsuit with other corporations frequently deal with cases where business secrets are leaked, which makes it impossible to force the selection of investigation target, confiscate and duplicate hard disks. Having said that, methods for selective confiscation of case-related data and evidence investigation are necessary.

In this paper, we suggest a new investigational methodology for selecting the target and investigating the relevant evidences only. PIM (Phased Investigation Methodology for Tracing Computer Usage), which is the name of our methodology, utilizes stepwise approaches based on the crime scene and is aimed at efficiently selecting and investigating the system, enabling one to overcome the limitations of conventional methodologies.

This methodology consists of 4 steps: selection and pre-investigation of the target system, tracing of the recent computer usage, analysis of computer usage pattern and investigation of user files. In step 1, image of the system is not obtained and only the file system metadata are acquired and analyzed to judge if a forensic investigation is needed for that system. In step 2, the recent work history from cache files of applications, registry and web browser files are investigated to judge if the suspect has run a relevant task or if the computer was used in the case. In step 3, statistics of internet connection, files and applications are used to investigate hourly and daily task history and analyze the computer usage pattern. In step 4, actual files are acquired based on the information analyzed in the previous steps to acquire files that can serve as evidences. In this way, PIM enables us to promptly select the target system and effectively trace the history of computer usage.

*Corresponding author.

978-0-7695-3769-6/09 $26.00 © 2009 IEEE
DOI 10.1109/NCM.2009.246

In section2, we briefly explain about the data that will be dealt with in this methodology. Finally, in section3, the methodology will be described in detail.

II. A RELEVANT STUDY - TARGET SYSTEM DATA

As previously mentioned, digital forensic investigation that is based on the well-known post-mortem paradigm hard disk image has the following limitations.

- Duplicating or imaging a disk takes a lot of time
- Difficulties arise in finding the relevant data from the acquired disk or image
- Cases where a copy or image of hard disk cannot be acquired due to privacy or secret leakage concerns

Even if a forensic image is acquired, we cannot know what data exist in the suspect's computer, what they were used for, and which of them is relevant to the case before a thorough investigation of the image has been done. Thus, the investigator can promptly react to the case by acquiring only those data that will enable him/her to grasp the suspect's usage history from the target system. Also, methodologies for selectively acquiring case-related data from the crime scene and analyzing them are crucial. Furthermore, in cases where large-scale system should be investigated, such as a civil lawsuit between corporations, it is virtually impossible to acquire the images of every hard disk and then investigate them. This calls for yet another approach.

PIM acquires only non-volatile data that are needed for tracing computer usage history. When data are acquired at the application level of the live system, integrity of the data is damaged due to the change in metadata, such as the time information. Thus, in order to guarantee the integrity of the target system as much as possible, all data acquisition should be done by physically approaching the disk. In case of systems that are turned off, general disk browsing tool can be used for this purpose. For the result of this methodology to be reasonable evidences, every acquisition/analysis process of data should be able to guarantee integrity [6]. This can be implemented by acquiring each file on the physical level, compute its hash value and record every process.

In PIM, focus is on the computer usage history and the target system is selected stepwise for effective forensic investigation. Before describing this methodology in detail, we classify important data that are needed for tracing computer usage history and briefly explain about them. Five categories of data were defined for tracing the suspect's system usage history. Those are metadata, prefetch file, registry, web browser file and specific document file, which are shown in the Table 1 below.

Table I
ACQUIRED DATA

Item	Description
File System Metadata	All file and folder lists
Prefetch	Number of application usage and time of the last running
registry	List of executed commands, search keywords, last accessed folder, recently executed files and application usage
Web Browser File	Visited URL/time, downloaded files, search keywords
Specific Document File	Encrypted files, file name and files with modified extension

A. The File-System Metadata

File system metadata contains information of all files and directories that exist on the hard disk [7]. In case of NTFS which is the basic file system of Windows XP, these metadata exist in $MFT file. Information that can be read from $MFT file are: file and directory names, file extension, creation time, modification time, access time, file size, and file location and its possessor. Files of size smaller than 700 bytes can also contain evidence [8]. In most cases, file system metadata are extremely small in size that very short acquisition time is needed.

Since we can obtain the name, type and time of the file stored on the target system by acquiring and analyzing this limited set of information, this it can be effectively used in selecting a potential target system by file or keyword search. If computer usage history is traced by file system metadata acquired as such, this method is more efficient than to trace the whole disk image.

B. The Prefetch File

Prefetch file is a cache file that is used to shorten the running time of applications [9]. A specific process of an application uses a prefetch file to load necessary data files on memory before launching the application, thus improving the running speed. The information that can be read from a prefetch file is: a file name, number of running programs, time of the last running, and the list of reference files needed to launch the program. This information lets us know what program the suspect has recently run and what program he/she frequently runs.

C. The Registry

The registry is a central hierarchical database used in Microsoft Windows 98, Windows CE, Windows NT, and Windows 2000 used to store information that is necessary to configure the system for single or multiple users, applications and hardware devices [10]. Also being used in Windows XP and Windows Vista, the registry is very important in forensics as it has the majority information concerning the computer usage and configuration. In order to investigate the trace and purpose of the system usage, which is the focus of this paper, we have categorized information according to executed commands, search keywords, last accessed folders, and application use logs. Table 2 shows

Table II
INFORMATION EXTRACTED FROM THE REGISTRY

Item	Description
Executed Command	Commands that were executed by ”Start + Run”
Search Keyword	Keywords used in Windows search
Last Accessed Folder	Last folder accessed by each application and the corresponding time
Recently Executed Files	Recently executed files and folders
Application Log	Last execution time of application and the number of running

the type of the information extracted from registry and its description.

The executed command is the list of commands used by “Start + Run(R)” in Windows, where search keyword shows the list of keywords used by search feature of Windows. The last accessed folder contains the file accessed by “Open” of each application; recently executed files show the files and folders the user has recently opened. All these items provide the time information of the last task only. The application logs provide the running path, last execution time and the number of running of a given application.

D. Web Browser File

The web browser records most of its Internet usage history on a web browser log file. For instance, Internet Explorer saves it on index.dat, Firefox2 on history.dat, Firefox3 on places.sqlite, and Google Chrome on sqlite database file named History, etc. This web browser file provides us with the address of visited sites and time of visit. The keyword information can be extracted from these URL’s [11]. By extracting the search list of a specific time, this web usage history can be helpful to trace the computer usage history.

E. Specific Document File

A specific document refers to files with modified file/extension name. It is likely that a suspect might have encrypted the corresponding file or modified the file/extension name to cover up a crime. If encrypted files, files with modified file/extension name are found in the suspect’s system, these can be very important evidences relevant to the case.

File encryption can be categorized into that by NTFS-EFS, by application’s encryption feature and by encryption algorithm per se. If NTFS-EFS was used the corresponding file will have $LOGGED_UTILITY_STREAM attribute in $MFT, which can be checked to see if the file is encrypted [12]. EFS creates a temporary file named EFS0.TMP during the encryption and deletes it when the encryption is complete. Thus, if EFS0.TMP file can be restored, the content can be viewed without decrypting the encrypted file. In case of encryption by applications, most files have a flag that can tell if the file was encrypted and this can be used to extract encrypted files. If the file itself is encrypted by encryption algorithm, frequency test can be used to detect the encryption [13].

As for the modified file/extension name, it can be assumed that these were modified if the creation, last modified time, last accessed time, entry modification time of $FILE_NAME attribute in $MFT are different. Especially, the modification of the extension can be detected by comparing the signature of the general file format and the actual hexadecimal code.

III. METHODOLOGY & STRATEGY

PIM (Phased Investigation Methodology for tracing computer usage) presented in this paper breaks from the conventional disk image-based forensics and focuses on the selection of investigation target and tracing the use of target system. By dividing forensics investigation into each step and applying it stepwise, PIM enables a prompt reaction to the case. PIM is divided into 4 steps: selection of the target system and pre-investigation, tracing recent computer usage, analysis of compute usage pattern and investigation of user file contents. Data analysis is performed only with those data that are needed in each step. Figure 1 describes the overall procedure of PIM.

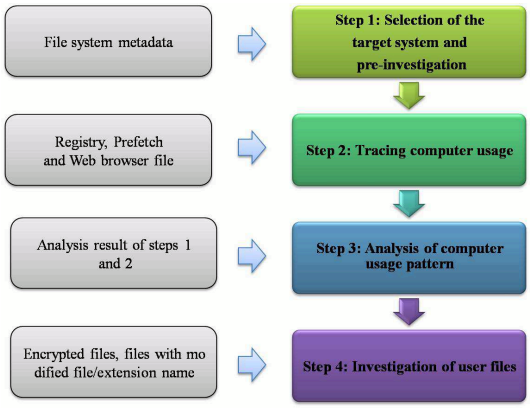

Figure 3. Overall procedure of PIM

If PIM is applied to large-scale investigations such as a civil lawsuit, a significant amount of investigation time can be saved. This is because PIM does not acquire image of every computer in the crime scene but effectively selects those systems that need to be investigated. Thus, in step 1, file system metadata are firstly acquired and analyzed so that we can judge if this system needs to be investigation and specify the target system. In other words, case-related keywords can be used to investigate the file list and judge the relevance. In step 2, the recent task information from registry, application cache file, and web browser file are investigated to specify if the suspect has performed relevant tasks or if the computer was used in the case. Even though the system was selected in step 1, it can be excluded from the target if no recent tasks or traces dating back to when

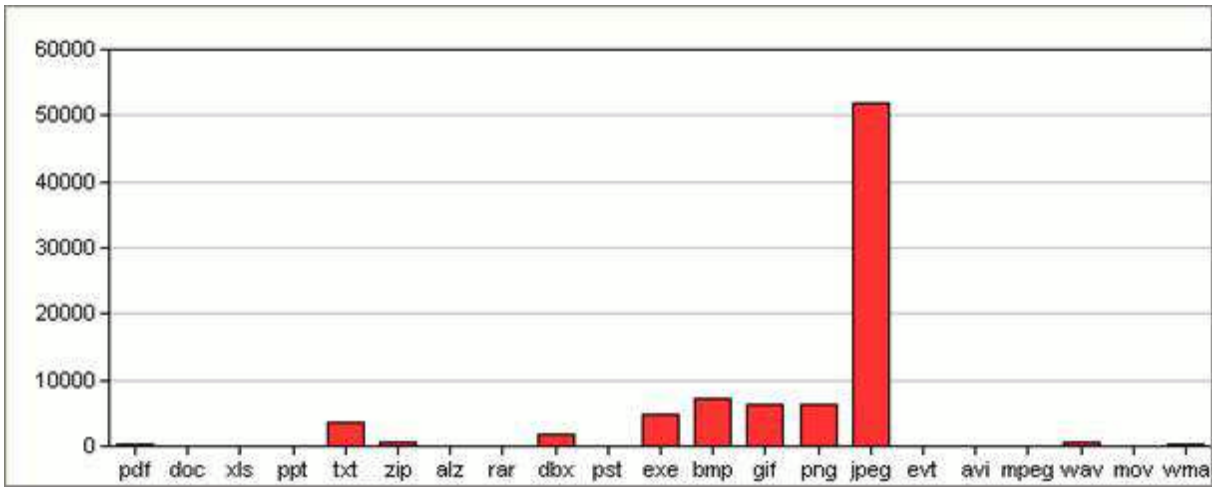

Figure 1. Statistics based on the extensions of file

	FIle Name	Extension	Create Time	Modified Time	Accessed Time
F	sex084.jpg	JPG	2009-01-13 00:54:28	2002-12-05 16:45:10	2009-01-13 01:03:43
F	sex085.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:19:48	2009-01-13 01:03:43
F	sex086.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:20:18	2009-01-13 01:03:44
F	sex087.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:20:44	2009-01-13 01:03:44
F	sex088.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:21:10	2009-01-13 01:03:44
F	sex089.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:21:52	2009-01-13 01:03:45
F	sex090.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:22:26	2009-01-13 01:03:45
F	sex091.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:22:54	2009-01-13 01:03:45
F	sex092.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:23:24	2009-01-13 01:03:46
F	sex093.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:24:06	2009-01-13 01:03:46
F	sex094.jpg	JPG	2009-01-13 00:54:28	2002-12-05 15:27:20	2009-01-13 01:03:47
F	sex095.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:24:34	2009-01-13 01:03:47
F	sex096.jpg	JPG	2009-01-13 00:54:28	2002-12-05 15:32:16	2009-01-13 01:03:48
F	sex097.jpg	JPG	2009-01-13 00:54:28	2002-12-05 15:34:52	2009-01-13 01:03:48
F	sex098.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:25:12	2009-01-13 01:03:48
F	sex099.jpg	JPG	2009-01-13 00:54:28	2002-12-23 13:25:36	2009-01-13 01:03:49

Figure 2. Extension search of .jpeg files

the case occurred are found. In step 3, statistics of Internet connection, file and application usage are used to analyze the computer usage pattern and investigate the hourly and daily task information. In step 4, actual files are acquired based on the previously analyzed information to investigate their contents and obtain files that can serve as evidences. In sum, PIM enables a prompt selection of the investigation target system and effective tracing of computer usage. Table 3 shows the data in each step and the information that can be extracted from them.

Table III
TARGET DATA AND EXTRACTED INFORMATION

Step	Target Data	Extracted Information
1	File system metadata	File list and statistics
2	Registry & prefetch, web browser file	Usage history of application and file, search keyword, Internet history
3	Analysis result of steps 1 and 2	Computer usage pattern
3	Specific document files	Encrypted files, files with modified file/extension name

A. Step 1: Selection of the target system and pre-investigation

The pre-investigation of the target system is helpful to grasp the characteristics of the system. Also, if there is a large number of systems to be investigated, pre-investigation can serve as an effective means of reacting to the case by reducing the overall investigation time.

File system metadata are mostly investigated in this step. As mentioned in section 2, these data contain name, attribute and location of all files stored in the disk's file system. In particular, the location information enables a physical extraction of the file to be acquired and thus guarantees the

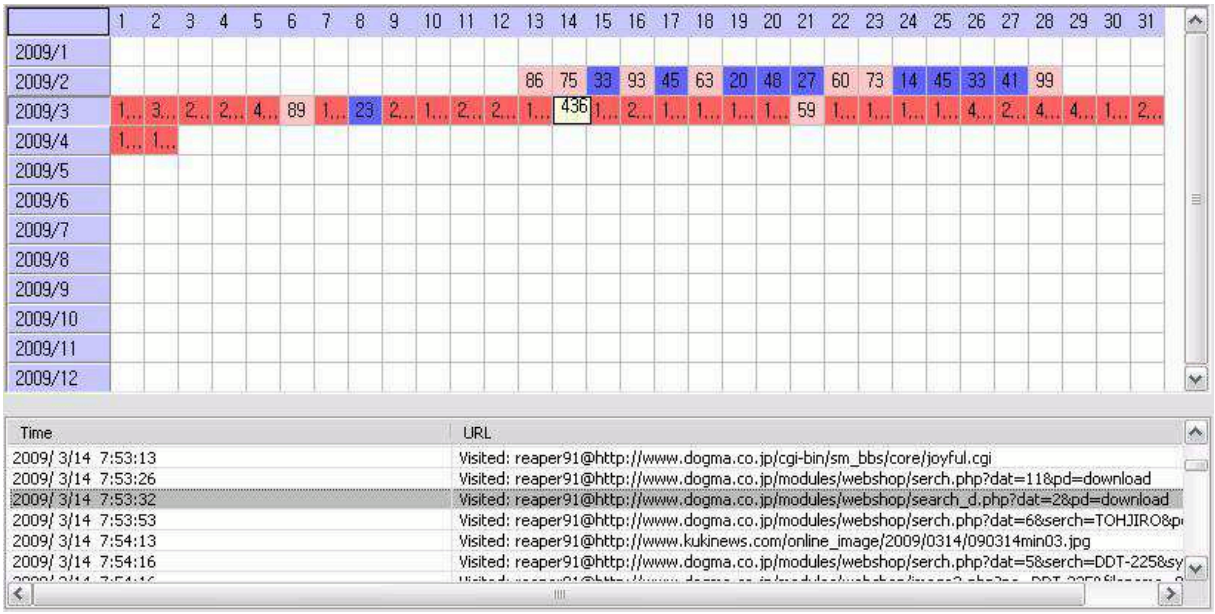

Figure 4. A time line of the suspect's Internet usage

file's integrity.

The investigator can use some information from the metadata to do a forensic analysis. First of all, every list of files can be derived read from the metadata. From the file name and extension, the investigator can look into the existence of case-related files through keyword search or filtering. For example, in a child pornography-related case, the relevant data would be image and movie files. Thus, the investigator can search for either the extension of these file types or relevant keywords to find those files, whose contents will then be extracted in the last step.

Moreover, statistics based on the extension and time information of files can serve as a clue to find out the main purpose of the computer usage. If extensions such as .xls and .doc are relatively abundant, it can be assumed that the computer is usually used for word processing. One should note that temporary files of web browser and default files that are installed with the system should be excluded from the statistics. These files contain a lot of picture files, HTML files and executables, which may lead to an inaccurate statistics. It is much more effective to analyze temporary files later in the web browser file.

Figure 2 shows statistics based on the extensions of file list that exists in the suspect's metadata. It can be confirmed that there are quite a lot of JPEG files in the computer. After performing an extensive search on these in the metadata, many files that were suspected to be adult contents were found, as is shown Figure 3. As a consequence, we could grasp the main purpose of the computer usage in this step and perform a pre-investigation to search for case-related files through keyword search and filtering.

B. Step 2: Tracing computer usage

Once the target selection and pre-investigation have been completed in step 1, a more detailed tracing of computer usage is performed in step 2. The target data that are acquired and analyzed in this step are the registry, prefetch and web history files. The acquisition is done using the location information of each file, which was analyzed in step 1.

The registry and prefetch files are suitable for tracing the suspect's usage of applications and files. The prefetch file can be used to determine which applications were often used recently, where the recently accessed file list can be used to see whether the case-related task was recently performed. Moreover, in case where the recently accessed file does not exist in the actual file system, it can be assumed that the suspect deleted it in order to destroy evidence.

The registry can offer information such as executed commands, search keyword, last accessed folder, last executed file, and application logs. The investigator can perform a registry analysis to extract files and data related to these. For example, if a suspect used office programs frequently, the investigator can extract the file with the name of the corresponding program, the file extension that is made by the program, and files that were created, modified and accessed at time of the last execution of office program, from the file list made by the file system metadata analysis.

The web browser file is a powerful tool for tracing the suspect's Internet usage. The time information can show the URL addresses and the time of the sites the suspect has visited. Analyzing the URL enables us to track the downloaded files / location and searched results, and the

temporary files can even show the content of web mail. Thus, by tracing Internet usage, the investigator can expect to improve the probability of extracting case-related data. Figure 4 shows a time line of the suspect's Internet usage. It can be seen that he/she accessed 436 Internet sites on 14th March.

C. Step 3: Analysis of computer usage pattern

Based on what had been analyzed so far, the investigator can analyze the suspect's computer usage pattern. The analysis of usage pattern in this step can provide us with clues about when the suspect often used computer and what kind of files or applications were used by reorganizing the case according to data used in the previous steps. Firstly, The MAC time of the file system metadata can be used to see when the suspect created, modified and accessed files. The file and application usage logs can be also used to trace when files and applications were used. The MAC time can be used to predict the usage pattern of applications. The web browser file can be used to investigate when the suspect accessed a certain website. This is supported by additional checking of keyword search webmail to grasp the main interest of the suspect. The web browser usage also enables to see, through a timeline as in figure 4, the overall flow of sites that the suspect has visited.

D. Step 4: Investigation of user files

In this final step, relevant files and data are extracted and evidence files are investigated according to data analysis results and the overall understanding of the case. That is, files that were confirmed by the file system metadata in step 1 can actually be extracted and analyzed to judge whether the suspect has committed a crime.

Step 4, focuses on investigating whether the suspect deleted, encrypted or modified the name/extension of the file to destroy evidence. If he/she has recognized the possibility of investigation, he/she will be likely to try these things. As this attempt of destroying evidence can imply a possible collusion, a detailed analysis is crucial.

IV. CONCLUSION & FUTURE WORK

The traditional digital forensics investigation is based on collecting evidence from a hard disk. Acquiring the entire image take a lot of time, and problems arise whenever this is not possible. In this case, the investigator should extract the case-related data without imaging the storage device. This calls for selective collection of data from the target system and the analysis of the suspect's system usage. In this paper, we have presented a methodology to trace the computer usage and extract relevant data by collecting file system metadata, prefetch and registry files, web browser files and specific document files and analyzing them stepwise. This method is very useful for tracing the suspect's system usage and extracting case-related files and data. If too many systems are to be investigated, a part of the systems can be excluded from the target or have priority over other tasks for the sake of prompt and effective investigation.

ACKNOWLEDGMENT

This work was supported by the IT R&D program of MKE/IITA. [2007-S019-03, Development of Digital Forensic System for Information Transparency]

REFERENCES

[1] Gary Palmer, A Road Map for Digital Forensic Research. Technical Report DTR-T0010-01, Report from the First Digital Forensic Research Workshop, Novmber 2001

[2] Jack Riley, David Dampier and Rayford Vaughn, Time Analysis Of Hard Drive Imaging Tools, Advances in Digital Forensics IV, Springer, pp. 335-344, 2008.

[3] Mark Pollitt and Authony Whitledge, Exploring Big HayStacks: Data Mining and Knowledge Management, Advances in Digital Forensics , Springer, pp. 67-76, 2006.

[4] Ricci Ieong, How to Balance Privilege and Digital Forensics Investigation, The Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kaohsiung Taiwan, November 2007.

[5] Ricci Ieong, Live-Privileged data aware Live Forensics Scheme, Proceedings of 4th Annual IFIP WG 11.9 International Conference on Digital Forensics(ICDF 2008) - Short Papers-, Kyoto, Japan, pp. 33 - 47, January, 2008.

[6] Renico Koen and Martin Oliver, An Evidence Acquisition Tool for Live System, Advances in Digital Forensics IV, Springer, pp. 325-334, 2008.

[7] Brian Carrier, File System Forensic Analysis, Addison-Wesly, pp. 186-198, 2005.

[8] Brian Carrier, File System Forensic Analysis, Addison-Wesly, pp. 273-299, 2005.

[9] Harlan Carvey, Windows Forensic Analysis, Syngress, pp. 226-229, 2007.

[10] Microsoft Corp,. Windows registry information for advanced user, http://suport.microsoft.com/kb/256986, 2008.

[11] SeungBong Lee, HyukDon Kwon, KyungSoo Lim and Sangjin Lee, The Design of Web Browser File Analyzer Tool, Journal of Digital Forensic, Volume2, Issue1, pp. 47-69, June 2008.

[12] Brian Carrier, File System Forensic Analysis, Addison-Wesly, pp. 209-211, 2005.

[13] Bora Park and Sangjin Lee, Determinant Whether the Data Fragment in Unallocated Space is Compress or Not Decompressing of Compressed Data Fragment, Journal of The Korea Institute of Information Security & Cryptology, Volume 18, Issue 4, pp. 175-186, August 2008.