2011

Forensic investigation method and tool based on the user behaviour analysis

Namheun Son
Korea University, Seoul, Korea

Sangjin Lee
Korea University, Seoul, Korea

DOI: 10.4225/75/57b2c4e440cf4
Originally published in the Proceedings of the 9th Australian Digital Forensics Conference, Edith Cowan University, Perth Western Australia, 5th -7th December 2011

FORENSIC INVESTIGATION METHOD AND TOOL BASED ON THE USER BEHAVIOUR ANALYSIS

Namheun Son, Sangjin Lee
Center for Information Security Technologies
Korea University, Seoul, Korea
pida2@korea.ac.kr, sangjin@korea.ac.kr

Abstract

Today, people use a variety of digital devices, and events taking place in them are stored in specific forms mostly including data indicating when each event took place. So far, different methods have been constantly researched and developed to analyse various events, most of which analyse event data unnecessary for a forensic investigation. As a result, investigators should carry out additional work to select data needed for an actual investigation, making the process of analysis more difficult and longer. Besides, since the capacity of storage media gets higher and events become more diversified, such a phenomenon seems gradually worsened. Thus, this paper suggests a timeline-based method of checking 'users' behaviour patterns' at a look by analysing, interpreting and visualizing various user behaviour-based events in a short time, since time information exists in digital devices. Moreover, the range of analyses can be widened since investigators can analyse events through computer and smartphone used most out of all the digital devices, not simply through a single system.

Keywords

User Behaviour, Smartphone Forensics, Event based, Timeline-based, Visualization

INTRODUCTION

Recently, the capacity and kinds of storage media used for digital devices have become massive, various and such a change makes it more difficult to carry out the classical investigation methods like imaging and analysing the entire data. To supplement the existing methods, new active forensic methods are developed to investigate and analyse data in an active state. However, since the existing forensic investigation methods have weak functions of collecting data selectively and automatically, it takes long to collect data, further prolonging the period of investigation. Besides, out of all the data, investigators should select and analyse data they really need once again, which requires other tools. Information analysed has different output modes by different tools, and it is needed to provide meaningful data for an investigation by collecting all the outputs coming through different tools. The problem is that such work should be processed once or more than once, and such a processing stage requires extra hours. In case of multiple systems, even experienced forensic investigators should carry out repetitive work with different tools, which prolongs the period of investigation exponentially. Thanks to a rapid increase of smartphone use, new smart-phone forensic methods are developed, but since such a method is actually far away from the existing computer forensic investigation, such a phenomenon will be worsened.

This paper attempts to solve investigation time-related problems that existing forensic investigation methods have and to supplement their efficiency. Unlike the existing analytical methods, this new method makes it possible to complete an investigation earlier, based on events indicating what behaviour users did at a specific time. Besides, it expresses and visualizes data into timeline-based data showing certain behaviour so that investigators may instantly know at what specific time each event took place.

As for the composition of this paper, Section 2 explains the previous researches conducted on the timeline of events. Section 3 introduces a new method of collecting and analysing user behaviour-based events and explains the method by applying it to computer with Microsoft Windows system and smartphone. Section 4 explains the main functions and process of this new tool based on the suggested methodology. In the last Section, case study will be done in order to verify the method's effectiveness.

RELATED WORK

Although the previous researches were conducted on some forensic tools and methods supporting the timeline function, they are still insufficient in showing or visualizing users' behaviour at a look. This chapter will discuss the existing researches on events and timeline.

Brian D. Carrier suggested a framework for the process of a digital investigation based on events, which consists of 4 stages. However, it is not a framework based on timeline.

There is a tool called 'Zeitline' researched and designed by Florian Buchholz at Purdue University in 2004, by focusing on the timeline function. Zeitline supports analysing syslog files of Unix/Linux and shows timeline based on the MAC time of each file. However, before analysing data with this tool, it is required for investigators to manually add and arrange items they need, and even the range of analysis targets is limited. However, the function of visualization based on timeline doesn't even exist in this tool.

In CyberForensics TimeLab Viewer introduced by a paper titled 'Computer Forensic Timeline Visualization Tool' in 2009, the timeline function is provided for events having occurred in the registry, event log and file system, and even a function of visualizing events is also provided. However, this function is only to list events, and since unnecessary results are included, investigators should reanalyse detailed events once again.

In Event-based Computer Profiling for the Forensics' provided by QUT Digital Repository in 2007, linkage analyses are conducted on the basis of Windows event logs (evt), but since this tool focuses on event logs, there exist many restrictions.

There is another tool named 'Oxygen Forensic Suite', providing timeline by collecting and analysing data related to smartphone, but it doesn't provide a function of visualization and an integrated analysis with a computer system.

At present, there exist powerful forensic tools providing various analysis functions for evidence data. The tool of this paper, however, doesn't provide all the functions that the existing tools have. It is because this paper aims to visualize data for investigators to shorten investigation time and carry out an efficient process of investigation by automating the creation and analysis of timeline based on users' behaviour.

As mentioned earlier, some of the existing tools provide the timeline function while having analysis functions, but such an analysis function is simply to analyse a single event without a tool mainly focusing on users' behaviour. Besides, such a tool apparently has difficult parts for forensic investigators, who are not tool experts, to deal with. In fact, no one but experts can understand exactly how to handle such a tool. That's why this paper attempts to analyse “users' behaviour patterns” by selecting only events related to users' behaviour.

A METHOD OF COLLECTING AND ANALYZING USER BEHAVIOUR-BASED EVENTS

When a computer turns on, the booting gets completed and the operating system screen appears. Then, when a user starts web-surfing, using web-mail and working on documents with word programs, they copy or move files into USB storage devices. They even communicate with others by driving a messenger commonly used recently. When they've done all the work, they turn off the computer.

Many users use smartphone in addition to computer, and even when they turn off the computer, they still use smartphone. In a smartphone, there exist almost the same events as those in a computer, such as Call History, SMS/MMS, SNS(Social Network Service), internet and e-mail. The information can be usefully used as analysis data in addition to service details of the related computer.

We've just briefed the service details of a computer, but the same kinds of events take place in all kinds of digital devices, and these events are recorded in a specific log form which mostly has time information. Such a log recorded can be very useful information for an investigation on users' behaviour patterns.

In this Section, we suggest a method of creating data by collecting events from computer and smartphone, where time stamp exists and which is possible infer the service history of users' systems. We also suggest a new paradigm that investigators can carry out rapid forensic investigations by automating this method.

Overall Framework

Before analysing a smartphone and a computer at the same time, there are some things to be checked. First of all, it is required to determine whether backup data of the smartphone remains in the computer, or whether the smartphone is connected to the computer. Then, investigators can carry out analysis with the smartphone together. The framework of analysis for smartphone is the same as that of computer.

As for the existing forensic investigation, when events wanted exist, analyses are carried out by accessing or extracting files based on the file system and acquiring information with tools likely to analyse the related events.

The present tools have great performance, but since they extract information from only an event after all, the information may be insufficient, or there may be a plenty of unnecessary information. Especially, out of all the information stored in the registry of Windows, only some can be used for an actual investigation, so it is very important to filter data in advance. After the process of filtering, investigators can acquire right information by analysing events.

Despite the process of filtering, there still remains a possibility that information acquired from a single event is insufficient. In this case, finding and using linkage relations with other events will help acquire much more information. For example, in case of the service history in USB storage devices of Windows Operating System, using device information from the registry leads to acquiring the latest access time, but because the very first access time remains in the SetupAPI log, it is possible to acquire the latest access time and the first access time from the USB storage device by finding linkage relations and using the serial number of the same USB storage device.

The next stage is to analyse events and convert data, connecting the linkage relation of each event, into the uniform marking-pattern. Each event has a different output mode, so such a process is essential. The basic output mode consists of time an event took place, contents used for the event, behaviour the event performed and even detailed statements and kinds of devices. [Table 1] shows how certain information can be expressed: 'Microsoft Office Professional Plus 2010' was installed in the folder, 'C:\Program Files\Microsoft Office', at 2011-07-14 12:29:25 in the computer, and a SMS message, ‘See you’, was received at 2011-07-14 13:10:12 in the smartphone.

Time	Content	Action	Detail	Device
2011-07-14 12:29:25	Microsoft Office Professional Plus 2010	Program Installed	C:\Program Files\Microsoft office\	Computer
2011-07-14 13:10:12	See you.	SMS(Received)		Smartphone

Table1- Event Display

When all the data change into an expression mode in the same form, most of the event history can be marked at once regardless of kinds of devices. The next one is to arrange data into a single form in order of time. As shown in [Table 1], since converting all the events into a single behaviour form helps understand intuitionally, investigators don't have difficulty carrying out an investigation although they don't have detailed event information. At this point, a statistical technique can be used, and with an event related to files or folders, it is possible to find out what the most-used extension is by compiling statistics with the extension. Likewise, for web browsers, it is possible to find the most-visited URL or most-used search word by collecting statistics. Data acquired by using statistics indicates users' concerns.

The next stage is to visualize data marked in order of time so that investigators may analyse and understand the data rapidly. Showing data in order of time helps save time, but too much data marked at once puts limits on understanding with the eyes. Therefore, marking different events with different colors by the time each event took place in a graph will make it possible to carry out analysis much faster. As for additional functions, a function of highlighting words and URL used for an event by creating an interest word list or an interest URL list will make an intuitional analysis possible. Similarly, using searching and filtering functions can help acquire behaviour investigators look for.

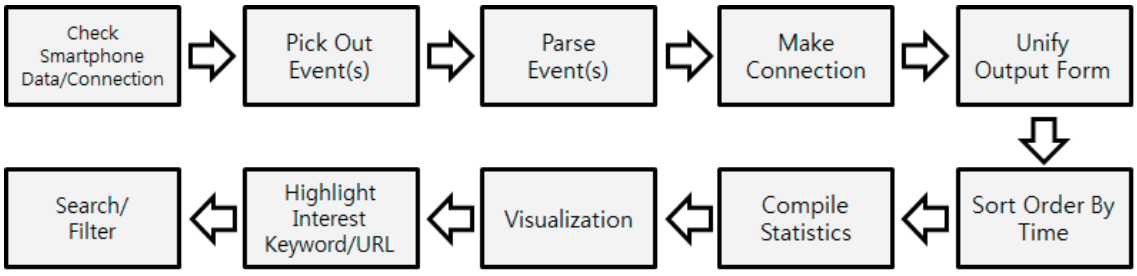

Figure1- Analysis Methodology

User Behaviour Events on the Windows System

In this passage, we show how to select events, indicating users' behaviour based on timeline from the Windows environment, on the ground of the explanation in above passage.

• System OnOFF - Using the event log of Windows helps find out the time users started driving and shutting down the system. With the system on/off time, it is likely to know the range of users' main activity time. Generally, an event or more should take place between the system drive time and the shut-down time, but if there is no record about it, it is clear that the time has been manipulated or the computer was shut down forcibly.

• USB Storage Device History - Using the registry and the log file, SetupAPI.log (Windows XP), will lead to knowing the very first access history and the latest connection/disconnection history for USB storage devices. Since the history of using USB is widely used for incidents of leaking classified documents or to find out the course malicious codes came through, it is very important information.

• Web Use History - As a programs used most, web browsers will be a good source to find out sites and search words users have visited and used. All the sites visited are classified into 14 kinds of behaviour, such as WebMail, Blog, Weather, Map, Dictionary, News, Video, Music, Stock, Bank, Shopping, Knowledge, Game and SNS, which helps grasp what behaviour users did on the web.

• Recent Document History - Analysing link files (.lnk) automatically created by the Windows system will help understand the details of documents recently used. In addition, using meta-data stored inside the system can lead to grasping the time documents were modified or created and whether files used still exist in the present system. Information about the recent documents can be usefully used for an incident of leaking classified documents since it has information about the location of documents recently executed, whether they were executed through external devices, serial number and volume name, if they were executed in the USB storage device.

• Program Execution History - When an execution file is executed in the Windows system, it is possible to know the time programs were last executed, the courses of programs and even the frequency of program execution, which indicates which program was used most often.

• Program Installation List - It is acquired by using information of the Windows registry, such as the program installation time, name and version. It helps judge not only users' behaviour patterns but whether an anti-forensic tool is installed.

• Execution Process List - It is acquired only in an active state by using Windows API. Through this list, process executed times can be found out, and can be also used for additional analysis.

• Messenger - Using values stored in the registry or configuration files in each messenger, the time of messenger installation can be found out, and when the account information exists, acquiring the related information helps grasp the actual conditions of using the messenger. We analysed MSN, Yahoo!, and Skype.

• Recent Command Execution - It informs us of the recent list executed from the start of Windows to the execution.

• Recent Search List - A list searched through the explorer and My Computer can be acquired.

User Behaviour Events on Smartphone

This passage is almost the same as before Passage, but one difference is the target is a smartphone. Since data of a smartphone is mostly stored in files in a form of plist or sqlite, it is more convenient to carry out analysis than computer.

There are lots of events taking place in the smartphone, but here we mention only 5 kinds of users' behaviour most commonly taking place through smartphone.

• Call History – Possible to analyse when and with whom users had a phone call.

• SMS/MMS – Possible to know when, about what and with whom users talked. Unlike Call History, it is possible to know the contents of their conversation, so it can be used as very important data.

• SNS – People use SNS services like Skype, Facebook and Twitter through their smartphone, and since such service includes meaningful data, such as sending text messages and free-charge calls. It is inevitable to analyse SNS data.

• Geo Location – When an application using map information is used, the application comes to include data indicating the present location, which can be used to detect the user's location. For instance, picture information has history of using the Google map or history of EXIF location information with pictures.

• Web Use History –The basic contents are the same as those of a computer, and web browsers used vary by the kinds of smartphone.

Unlike a computer, events acquired from a smartphone include other people's phone numbers and e-mail addresses, so it is possible to analyse human relations and can be used for a crime investigation.

DFRC USER BEHAVIOUR ANALYZER (DFRC UBA)

The tool of this paper was designed on the basis of the methodology mentioned above. Based on a simple interface, it can be driven in either an active-stated or an off-line system, and raw images can be received through input. The target systems are a Windows system and a smartphone. The time of analysis is less than 10 minutes, and this tool converts data into timeline-based one so that investigators who are not tool experts can easily understand the data. Besides, this tool visualizes data based on timeline for an efficient analysis. Through such a method, it is possible for tool amateurs to analyse data in a short time, even for multiple systems. The Windows support OS versions of Microsoft are XP, Vista and Seven, and the smartphone supports iPhone of Apple.

The basic composition of this tool is as shown in [Figure 2]. In the beginning, it is needed to select an active mode after determining whether to read Rawimage, whether the system is a live system or a dead system. Then, the tool starts collecting data and reinterprets and visualizes the data. [Figure 3] shows the visualization graph. X-axis indicates date, y-axis indicates time and each event has each colour for identification.

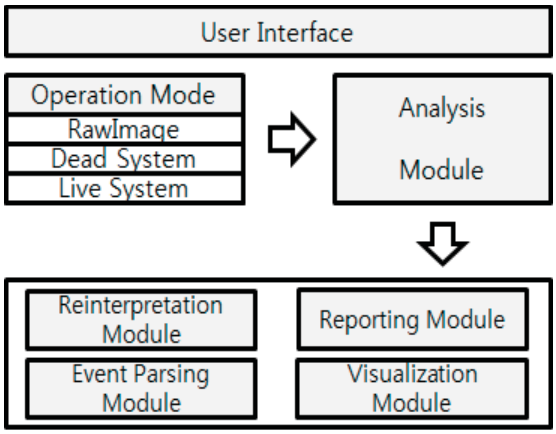

Figure 2 – UBA Architecture

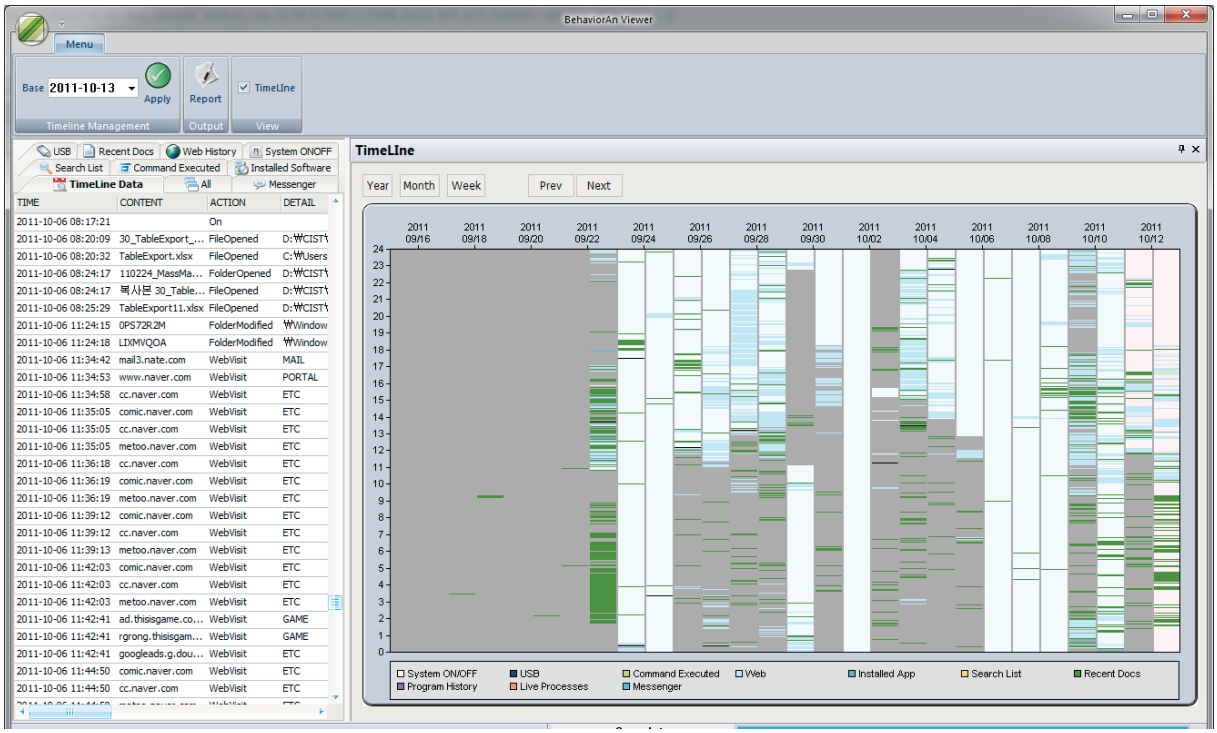

Figure 3 – Visualization Graph

Main Functions

This tool has five main functions, such as event analysis, statistics, visualization, file system access and report creation, and the detailed particulars are as blow.

The event analysis function is to automatically analyse events to infer users' behaviour. If a word from the list of interest words exists in the timeline data created, the related event will be marked red, to be found out immediately.

The statistic function is to create data likely to compile statistics into a list called Top 10, such as web visit, web search and extension use out of all the users' behaviour and to show the top 10 data most used. By using the top 10 data, it is likely to grasp users' tendencies in the system investigated.

The visualization function is literally to visualize data based on timeline. With the data visualized, it is possible to intuitionally judge what behaviour users did and when they did it. If behaviour at a specific time should be known in detail, dragging and dropping the time range wanted from the graph will lead to creating a new window showing behaviour at the target time. Units of period marked are year, month and week, and the data corresponding to periods visualized can be compared by being marked in a list form.

With the function of accessing the file system, it is possible to search files used during the period relevant to various standards, such as time file name, extension and compressed files while providing a file extraction function as well. Since it is possible to extract suspicious files right away in an active or off-line state, it helps conduct an investigation promptly. As a files system supporting this process, we have NTFS, FAT16 and FAT32 mostly used in the Window system.

Finally, the report function is to prepare a report based on the data about users' entire behaviour, data marked in the visualization graph, data during the period designated by the user, or data relevant to the wanted period dragged and dropped from the visualized data. The contents of the report are about the investigation, incident number, computer name and report issue data, etc. Besides, users' behaviour information selected by different time zones are marked differently on the basis of system start time and shut-down time.

CASE STUDY

To diagnose whether this method has utility, compared to the other ones, we created an imaginary scenario about an incident of malicious code infection. With the scenario, we compared the time of 'DFRC UBA' tool designed with the new methodology with that of the existing forensic investigation methods.

Participants were divided into two different group, Group A using only 'DFRC UBA' and Group B carrying out an investigation without any restrictions. Each group consisted of 10 persons, all of whom were researchers taking doctor's or master's courses and majoring in digital forensics.

Case

The imaginary scenario about malicious code infection was made simply for our experiment. The contents of this case were as below.

An ordinary user's PC is infected by a malicious code through a specific USB, and then the malicious is executed right away. When it is executed, the infected PC is conquered by the attacker, who can manipulate many main functions, such as screen control, file management, process management. Through this process, he acquires iPhone back-up data from the infected system. From the iPhone back-up data, the attacker finds out the user's cell phone number and sends a great many of spam text messages to the number.

To analyse the scenario, it is needed to find out the history of accessing the USB from the computer system, so the analysis should be conducted by interlocking SetupAPI and the registry. To find out the history of executing the malicious code, analysis of prefetch files should be done simultaneously. For smartphone, it is also required to find out information that a great many of spam text messages were sent to the cell phone after a certain point of time.

The case is given disc images, and the imaging was done by using Encase. The histories of the malicious code infection are as below.

Computer system with Windows 7 OS installed

Connected to USB A at 11:12:53 on Aug. 8 (Mon), 2011

Connected to USB B at 10:10:23 on Aug. 10 (Wed), 2011

Connected to USB A at 23:20:11 on Aug. 12 (Fri), 2011

Malicious code executed at 23:20:20 on Aug. 12 (Fri), 2011

Many spam text messages received on Aug. 13 (Sat), 2011

Since the PC was infected with a malicious code when right after connected to USB, it is apparent that USB A was infected with the malicious code.

Experiment Result

With the graph of analysis time, we could find the result of our experiment interesting. In the group having used DFRC UBA, 10 participants all solved the given problem within 10 minutes, 7 in 5 minutes and 3 between 6 and 10 minutes. In the other group, however, one person solved the problem between 21 and 25 minutes, which was the fastest, 2 between 26 and 30 minutes and the rest over 30 minutes. As for the average time of submitting the assignment, the group having used DFRC UBA showed about 4 minutes and the other group showed about 36 minutes, making about a 9 time-difference, which indicates that the automatic analysis using DFRC UBA apparently is faster.

It shows that DFRC UBA makes much easier to investigate than previous method of digital forensics.

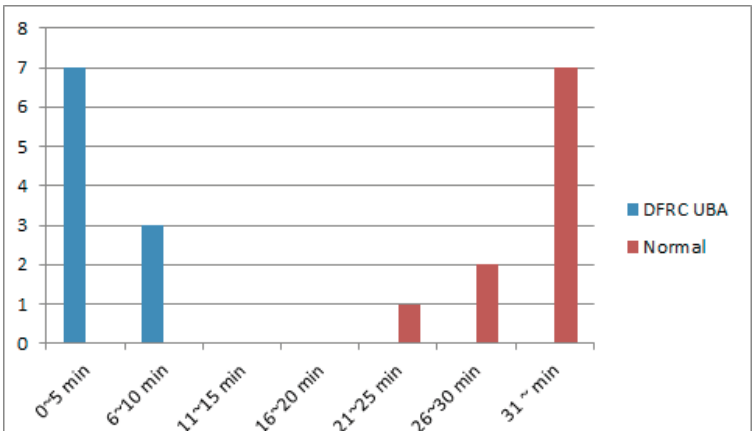

Figure 4 – Analysis Time

CONCLUSION

In this paper, we discussed user behaviour event-based forensic investigation methodology and tools designed on the basis of the method so that even non-experts can easily use it. Based on users' behaviour such as access history for USB storage devices, use history for link files, use history for web browsers, start and shut-down of systems, SMS/MMS of smartphone, call history and Geo Location, this study designed a tool to use for off-line, raw images and an active state, and this tool help complete analysis in a short time. Since this tool creates timeline, investigators can understand data at a look, and it visualizes data in order to grasp users' behaviour more easily. It is also possible to search and extract files by accessing the file system. Finally, investigators can prepare a report on the basis of timeline collected.

To verify the usability of this tool, we carried out an experiment through a case. As a result, we found it is 9 times faster than normal.

Overall, with such a method and tool, it is expected to save much time with more accuracy when carrying out a real forensic investigation for multiple systems in limited time.

Automated evidence collecting and analysing will play a more important role in digital forensics. In this paper we made a tool just for windows operating system except other operating systems. This is part of future research in the field of digital forensics.

ACKNOWLEDGEMENTS

This work was supported by the IT R&D program of MKE/KEIT

[10035157, Development of Digital Forensic Technologies for Real-Time Analysis].

REFERENCES

Simson L. Garfinkel, Digital forensics research: The next 10 years, Digital Investigation Volume 7, , Supplement, August 2010, Pages S64-S73

Ellick Chan, Shivaram Venkataraman, Francis David, Amey Chaugule, Roy Campbell, “Forenscope: a framework for live forensics”, ACSAC ’10 Dec. 6-10, 2010, pages pp. 307-316

Brian D. Carrier, "An Event-Based Digital Forensic Investigation Framework", DFRWS(Digital Forensic Research Work) 2004

Florian Buchholz, "Design and Implementation of Zeitline: a Forensic Timeline Editor", DFRWS(Digital Forensic Research Work) 2005

Jens Olsson, "Computer Forensic Timeline Visualization Tool", Digital Investigation, Volume 6, September 2009

Carvey, Harlan. "The Windows Registry as a forensic resource.", Digital Investigation, Volume 2, April 2005

Junghoon Oh, Seungbong Lee, Sangjin Lee, “Advanced Evidence Collection and Analysis of Web Browser Activity”, Digital Investigation Volume 8, Supplement, August 2011

Harry Parsonage, "The meaning of Link Files in Forensic Examinations", "Computer Forensic misCellany"

Marrington, Andrew D. and Mohay, George M. and Clark, Andrew J. and Morarji, Hasmukh L. (2007) “Event-based Computer Profiling for the Forensic”,. Proceedings AusCERT Asia Pacific Information Technology Security Conference’ (AusCERT2007): Refereed R&D Stream, pages pp. 71-87, Gold Coast, Qld.

Mona Bader, Ibrahim Baggili, PhD, , “iPhone 3GS Forensics: Logical analysis using Apple iTunes Backup Utility”, SMALL SCALE DIGITAL DEVICE FORENSICS JOURNAL VOL. 4, NO.1, SEPTEMBER 2010, ISSN# 1941-6164