Journal of Information Security, 2012, 3, 196-201
http://dx.doi.org/10.4236/jis.2012.33024 Published Online July 2012 (http://www.SciRP.org/journal/jis)

Digital Forensics and Cyber Crime Datamining

K. K. Sindhu[1], B. B. Meshram[2]
[1]Computer Engineering Department, Shah and Anchor Kutchhi Engineering College, Mumbai, India
[2]Computer Engineering Department, Veermata Jijabai Technological Institute, Mumbai, India
Email: kksindu@gmail.com, bbmeshram@vjti.org.in

Received February 6, 2012; revised March 13, 2012; accepted April 10, 2012

ABSTRACT

Digital forensics is the science of identifying, extracting, analyzing and presenting the digital evidence that has been stored in the digital devices. Various digital tools and techniques are being used to achieve this. Our paper explains forensic analysis steps in the storage media, hidden data analysis in the file system, network forensic methods and cyber crime data mining. This paper proposes a new tool which is the combination of digital forensic investigation and crime data mining. The proposed system is designed for finding motive, pattern of cyber attacks and counts of attacks types happened during a period. Hence the proposed tool enables the system administrators to minimize the system vulnerability.

Keywords: Cyber Forensic; Digital Forensic Tool; Network Forensic Tool; Crime Data Mining

1. Introduction

Computer forensics is the process that applies computer science and technology to collect and analyze evidence which is crucial and admissible to cyber investigations. Network forensics is used to find out attackers' behaviours and trace them by collecting and analyzing log and status information.

A digital forensic investigation is an inquiry into the unfamiliar or questionable activities in the Cyber space or digital world. The investigation process is as follows (As per National Institute of Standards and Technology) [1]. **Figure 1** shows the complete phases of Digital Forensic investigation processes.

Collection phase: The first step in the forensic process is to identify potential sources of data and acquire forensic data from them. Major sources of data are desktops, storage media, Routers, Cell Phones, Digital Camera etc. A plan is developed to acquire data according to their importance, volatility and amount of effort to collect [2].

Examination: Once data has been collected, the next phase is to examine it, which involves assessing and extracting the relevant pieces of information from the collected data [2].

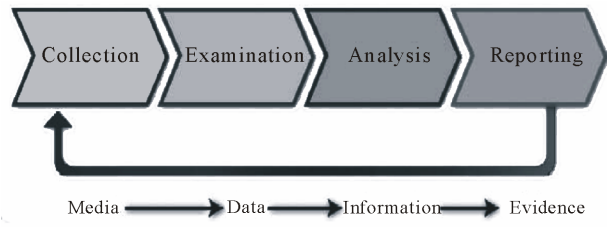

Figure 1. The digital forensic investigation processes [1].

Analysis: Extracted and relevant data has been analysed to draw conclusions. If additional data is sought for detail investigation will call for in depth data collection.

Reporting: This is the process of preparing and presenting the outcome of the Analysis phase.

Digital Forensic Science covers Computer forensics, Disk forensics, Network forensics, Firewall forensics, Device forensics, Database forensics, Mobile device forensics, Software forensics, live systems forensics etc.

2. File System Forensics

The File system investigation is the identification, collection and analysis of the evidence from the storage media. File systems or file management systems is a part of operating system which organize and locate sectors for file storage [3,4].

2.1. Basic Steps in Storage Media Investigation

1) Replication of forensic image: Nonintrusive acquisition of a replicated image of data extracted from the questioned device.

2) For integrity perform Hash value calculation.

3) Conducting a file-fragment recovery procedure to recover files and folders to a new location.

4) Examine all files especially deleted files.

5) Reviewing typical evidentiary objects such as:

a) Analyse free spaces, slack spaces and bad sectors.

b) Application software file.

c) Digital camera, printer and ancillary devices.
d) E-mails, Games & Graphics images.
e) Internet chat logs & Network activity logs.
f) Recycle folders.
g) System and file date/time objects.
h) User-created directories, folders, and files.
i) Latent data extraction from page, temp, and registry space.

6) Copy the content of the evidentiary object into text files.

7) Searching for key-term strings.

8) Reviewing file notations.

9) Scrutinize applications or indications of as file eradications, file encryption, file compressors or file hiding utilities.

10) Preparing evidence summaries, exhibits, reports, and expert findings based on evidentiary extracts and investigative analysis.

2.2. Hidden Evidence Analysis in the File System

Suspects can hide their sensitive data in various areas of the file system such as Volume slack; file slack, bad clusters, deleted file spaces [5].

1) *Hard Disk*: The maintenance track/Protected Area on ATA disks are used to hide information. The evidence collection tools can copy the above contents.

2) *File System Tables*: A file allocation table in FAT and Master File Table (MFT) in NTFS are used to keep track of files. **Figure 2** shows MFT structure. MFT entries are manipulated to hide vital and sensitive information [5].

3) *File Deletion*: When a file is deleted, the record of the file is removed from the table, thereby making it appear that it does not exist anymore. The clusters used by the deleted file are marked as being free and can now be used to store other data. However, although the record is gone, the data may still reside in the clusters of the hard disk. That data we can recover by calculate starting and end of the file in Hex format and copy it into a text file and save with corresponding extension.

Recover a JPEG file

a) Open file in the hex format.
b) Check the file signature.
c) Copy From starting signature upto ending signature.
d) For example (JPEG/JPG/JPE/JFIF file starting signature is FF D8 FF E1 XX XX 45 78 69 66 00 (EXIF in ascii Exchangeable image file format trailer is FF D9).

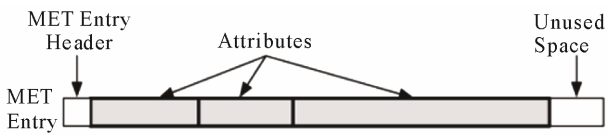

Figure 2. MFT structure [5].

e) Open the file with corresponding application.

4) *Partition Tables*: Information about how partitions are set up on a machine is stored in a partition table, which is a part of the Master Boot Record (MBR). When the computer is booted, the partition table allows the computer to understand how the hard disk is organized and then passes this information to the operating system. When a partition is deleted, the entry in the partition table is removed, making the data inaccessible. However, even though the partition entry has been removed, the data still resides on the hard disk.

5) *Slack Space*: A file system may not use an entire partition. The space after the end of the volume called *volume slack* that can be used to hide data. The space between Partitions is also vulnerable for hiding data, *file slack* space is another hidden storage. **Figure 3** shows slack spaces in a Disk.

When a file does not end on a sector boundary, operating systems prior to Windows 95 a fill the rest of the sector with data from RAM, giving it the name *RAM slack.* When a file is deleted, its entry in the file system is updated to indicate its deleted status and the clusters that were previously allocated to storing are *unallocated* and can be reused to store a new file. However, the data are left on the disk and it is often possible to retrieve a file immediately after it has been deleted. The data will remain on the disk until a new file overwrites them however, if the new file does not take up the entire cluster, a portion of the old file might remain in the slack space. In this case, a portion of a file can be retrieved long after it has been deleted and partially overwritten.

6) *Free Space*: However, when a file is moved from one hard disk or partition to another, it is actually a multistep process of copying and deleting the file. First, a new copy of the file is created on the target partition. After the file has been copied, the original file is then deleted. This process also requires some housekeeping in the FAT or MFT tables. A new entry is created in the table on the partition where it has been copied, whereas the record for the deleted file is removed from the table on its partition. When a file get deleted, that space considered as free space, there also criminal can hide sensitive information [6].

7) *Faked Bad Clusters*: Clusters marked as bad may be used to hide data. In NFTS, bad clusters are marked in metadata file called $BadClus, which is in MFT entry 8.

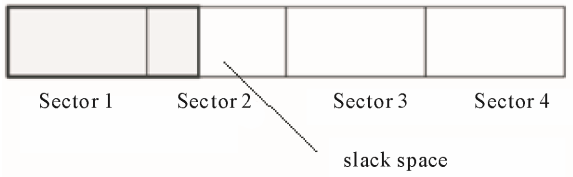

Figure 3. File slack [5].

Originally, $BadClus is a sparse file which file size is set to the size of entire file system. When bad clusters are detected, they will be allocated to this file. The size of data that can be hidden with this technique is unlimited. Suspects can simply allocate more clusters [6].

3. Network Forensic Analysis

Network forensics is capturing, recording and analysis of network events in order to discover the source of cyber attacks. In network forensics there are two major types of investigation [7,8] *i.e.* Network Traffic Analysis & Log Files Analysis.

3.1. Network Traffic Analysis

Network traffic analysis can be used to reconstruct and analyse network-based attacks, inappropriate network usage. The content of communications carried over networks, such as e-mail, chat etc can also support of an investigation. A Packet Sniffer tool is used for capturing network traffic. The header information encapsulated in the captured packet can be analysed by the forensic analyst [8,9].

This is very important when an investigation conducting on active network intrusions or attacks. Some cases evidences are available only in running processes or RAM.

Procedure for Network Live Acquisition

1) Create a bootable forensic CD.

2) Perform Remote access to the suspected machine or insert bootable CD in suspects machine directly.

3) Record or keep a log of all the actions of forensic investigator.

4) If need to take out away the evidence then use USB.

5) Next, Take a copy of the physical memory using a forensic tool example memfetch.

6) Create an image of the drive.

7) For Intrusion first check Root kit is installed or not, for that root kit revealers are available.

8) Perform hash value of the created image for integrity checking.

3.2. Network Investigation Tools

There is a powerful windows tools available at Sysinternals:

- Filemon shows file system activity.
- RegMon shows all Registry data in real time.
- Process Explorer shows what files, registry keys and dynamic link libraries (DLLs) are loaded at a specific time.
- Pstools is a suite created by SysInternals that includes the following tools.
- PsExec—Run processes remotely.
- PsGetSid—Displays the security identifier of a computer.
- PsKill—Kills processes by name or processes ID.
- PsList—Lists detailed information about processes.
- PsLoggedOn—Displays who's logged on locally.
- PsPassword—Allows user to change account passwords.
- PsService—Enables to view and control services.
- PsShutDown—Shut down and optionally restarts a computer.
- PsSuspend—Allows to suspend processes.
- Tcpdump and Ethereal—Packet sniffers.

3.3. Log Files Analysis

During investigation to recognize malicious activities by mining user log files. Access logs can contain vast amount of data regarding each user activities [10].

Analysis steps:

1) Input a server log file;

2) Identify each sessions;

3) Log file parser converts dump file into formatted order;

4) Using a Search function find the required data. Or Data mining algorithms give relations or sequential patterns.

4. Data Mining for Digital Forensics

Cyber Crime Data mining is the extraction of Computer crime related data to determine crime patterns. With the growing sizes of databases, law enforcement and intelligence agencies face the challenge of analysing large volumes of data involved in criminal and terrorist activities. Thus, a suitable scientific method for digital forensics is data mining. Crime data mining is classified as follows [11,12].

1) *Entity extraction* has been used to automatically identify person, login ID, Password, ID no, IP of the system, and personal properties from reports or logs.

2) *Clustering techniques* such as "concept space" have been used to automatically associate different objects (such as persons, organizations, hardware systems) in crime records [12].

3) *Deviation detection* has been applied in fraud detection, network intrusion detection, and other crime analyses that involve tracing abnormal activities [12].

4) *Association* rule has been applied to finding aassociations and sequential patterns between web transactions are based on the Apriori Algorithm.

Mining results shows motive, pattern and counts of similar types of attacks happened during a period.

4.1. Crime Data Mining Algorithm

1) Identify variables/itemsets from a case report (our proposed system stores these variables as attributes of tables, filesystem table, network table).

2) Item sets I = {I1, I2, I3 ⋯ Im}.

3) Set of actions D = {t1, t2, t3 ⋯ tn}.

4) Find frequent item sets by using Apriori algorithm.

Employs an iterative level to find set of frequent item sets.

E.g. if an attacker attacked database, login attempt results a data loss/Data tampering and case report show actions like Data deleted, Login attempt, attack type = SQL injection, If these item sets are frequent then we can set a rule " motive of attack is Data theft".

5) Make Association Rules

i.e. It is a rule in the form X → Y showing an association between X and Y that if X occurs then Y will occur.

If the attacker accessed operating system files then we can say motive of attack is system Crash.

If the attacker attacked Database login and Password steel then we can say criminal motive for data theft/data change.

This maximum frequent item sets also shows attack patterns.

Finding other signs of evidence Correlation, contingences (Consider these values while making rule sets).

6) Set SQL queries according to the rules.

7) Retrieve data.

4.2. Proposed Digital Forensic Tool

Our proposed model is the combination of digital forensics and data mining. Our proposed system helps to increase the security of the organization. When an incident reported, it investigates and report is saved in the database. Using crime data mining tool the nature of the attack is identified and alert administrator about similar attacks in future. Proactive measures can be initiated to prevent future cyber attacks. **Figure 4** shows the Block diagram of our proposed tool.

4.2.1. Block Diagram of the Proposed System

Graphical User Interface: It is used by the forensic investigator to enter case details and apply tools (File system, Network) to collect evidences. Investigators can input their queries in the system. This also displays the result of the query in the form of Bar chart or report. It is the presentation layer of our three tier architecture.

File System Analyser: This tool Collects evidence from the file system, it recovers all files, searches data in the free space, slack spaces and deleted spaces.

Network Analyser: This tool collects data from the network traffics and server log files.

Database: Database loader collects evidences from the above tools and loader loads into the database as attributes of the tables. OLTP (Online Transaction Processing): Set relations between the tables of the detected crime attributes. This applies data mining and extracts of required data. OLAP (Online Analytical processing) apply analytical queries and retrieves the output/decisions. Database server helps to store and retrieve crime attributes and results.

Decision Making System: This module applies data mining algorithm and also SQL queries into the database and generates reports.

Log file analyser module parses the web server logs,

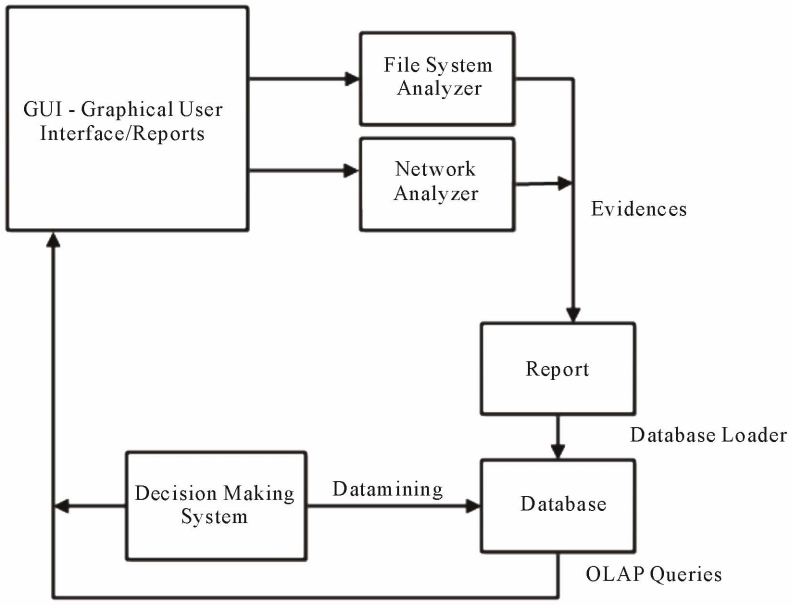

Figure 4. Block diagram of proposed system.

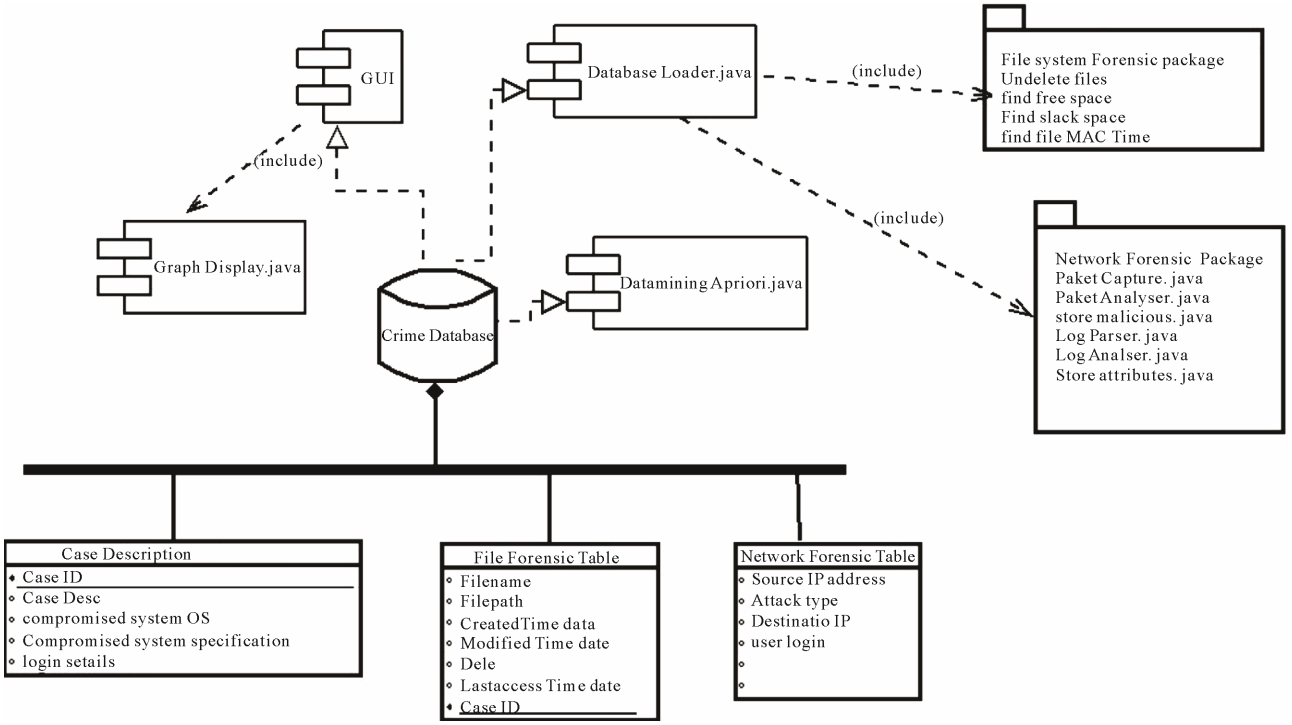

Figure 5. Software architecture of proposed tool.

syslogs and searches required keywords and patterns, which helps investigator to detect attacks like SQL injection, Brute Force attack for the login attempt.

4.2.2. Software Design of the System

Network forensic module is equipped with a traffic monitoring tool for data/evidence collection. A packet analyzer provides live forensic information about an attack. Java has API Jpcap captures information from the live network. The Network Analysis module analyse different types of packets ICMP, TCP, UDP.

File system analyzer module finding the evidence from the deleted files, free spaces (File slack, Volume slack).

The above modules give the output to flat file or CSV file.

A Java program module (File converter/Database Loader) converts as Table format and loads into the database.

Apply an Association mining (Apriori Algorithm) finding relation between these item sets of Crime Data and generate a prediction.

Graphical visualization module generates the required results in the form of Bar Charts or Graphs. **Figure 5** shows the Software Architecture of the proposed System.

5. Conclusion

This paper explains the hidden evidence acquisition from file system. Second section explains investigation on the Network. There are two types of investigation in network, live data acquisition (Packet capturing and analysis) and log file analysis. Third section explains crime data mining. On the basis we propose a new system with Digital forensic tool for decision making in the computer security domain.

REFERENCES

[1] K. Kent, S. Chevaller, T. Grance and H. Dang, "Guide to Integrating Forensic Techniques into Incident Response," NIST SP800-86 Notes, 2006.

[2] S. K. Brannon and T. Song, "Computer Forensics: Digital Forensic Analysis Methodology," *Computer Forensics Journal*, Vol. 56, No. 1, 2008, pp. 1-8.

[3] D. Klieiman, K. Timothy and M. Cross, "The Official CHFI Study Guide for Forensic Investigators," 2007.

[4] B. Carrier, "File System Forensic Analysis," Addison Wesley Professional, 2005.

[5] C. Kaiwee, "Analysis of Hidden Data in NTFS File System," Whitepaper.

[6] M. Alazab, S. Venktraman and P. Watters, "Effective Digital Forensic Analysis of the NTFS Disk Image," *Ubiquitous Computing and Communication Journal*, Vol. 4, No. 3, 2009, pp. 551-558.

[7] N. Meghanathan, S. R. Allam and L. A. Moore, "Tools and Techniques for Network Forensics," *International Journal of Network Security & Its Applications*, Vol. 1, No. 1, 2009, pp. 14-25.

[8] E. Casey, "Network Traffic as a Source of Evidence: Tool Strengths, Weaknesses, and Future Needs," *Journal of Digital Investigation*, Vol. 1, No. 1, 2004, pp. 28-43.

doi:10.1016/j.diin.2003.12.002

[9] H. Achi, A. Hellany and M. Nagrial, "Network Security Approach for Digital Forensics Analysis," *International Conference on Computer Engineering & Systems*, 25-27 November 2008, pp. 263-267.

[10] A. R. Arasteh, M. Debbabi, A. Sakha and M. Saleh, "Analyzing Multiple Logs for Forensic Evidence," *Digital Investigation*, Vol. 4S, 2007, pp. S82-S91.

doi:10.1016/j.diin.2007.06.013

[11] H. Chen, W. Chung, Y. Qin, M. Chau, J. J. Xu, G. Wang, R. Zheng and H. Atabakhsh, "Crime Data Mining: An Overview and Case Studies," *Proceeding of ACM International Conference*, Vol. 130, 2003, pp. 1-5.

[12] V. Justickis, "Criminal Datamining," Security Handbook of Electronic Security and Digital Forensics, 2010.

