ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 11, November 2016

Digital Forensics,Cyber Crime and Datamining

[1] Anirudhan C [2] Deepa M.P [3] Kodeeshwaran [4] Komal Shetty K [5] Dr. B.Shadaksharappa
Department of Computer Science Engineering,
Sri Sairam College of Engineering, Bengaluru.

***Abstract:* -- The world has evolved from using one's brain to cloning one's brain. A person is not assured whether the data with him is safe or leaked.People have moved on in their formal and informal behavior to being digitalized.Profits and loss in the commercial markets today are decided in matters of software. Hacking and electronic crimes sophistication has grown at an exponential rate in recent years. In fact, recent reports have indicated that cyber crime already surpasses the illegal drug trade!Source code theft, Online banking frauds, Online share trading fraud,Virus attacks, Cyber sabotage, Phishing attacks, E-mail hijacking are the most common threats in the cyber world today.In this paper we are discussing about the how factor of cyber crime and analyzing the methods to retrieve data, tools to find the culprit.The detailed process of detecting hacking attacks and properly extracting evidence to report the crime and conduct audits to prevent future attacks is also discussed**

***Keywords*: - Software Architecture,Digital Forensic tools, Control design, Crime pattern detection**

I. INTRODUCTION

Analyzing, investigating and accumulating the digital evidence and cyber trails are better known as Cyber Crime Investigation. You'll fall across them in computer hard disks, cell phones, CDs, DVDs, floppies, computer networks, the internet etc. Digital evidence and cyber trails can be derived from pictures, (stangnography), encrypted files, password protected files, deleted files, formatted hard disks, deleted emails, chat transcripts etc., The technical nature of cyber crimes demands for a specialized discipline to investigate for such sophis-ticated crimes. The volatile nature of digital evidence further adds a layer of complexity to the entire process of cyber crime Cyber crime Cyber Law Consultancies. It is a dream shared and brought up by two computer geniuses to make the society upgraded and making them cognizant about the cyber crimes that curb the innocence of environment. Thus, commencing with a rebellion in favour of cyber security.

Computer forensics is the practice of collecting, analysing and reporting on digital data in a way that is legally admissible. It can be used in the detection and prevention of crime and in any dispute where evidence is stored digitally. Computer forensics follows a similar process to other forensic disciplines, and faces similar issues.

Computer forensics- the preservation, identification, extraction, interpretation, and documentation of computer evidence, to include the rules of evidence, legal processes, integrity of evidence, factual reporting of the information found, and providing expert opinion in a court of law or other legal and/or administrative proceeding as to what was found.

Cyber crime- any illegal act that involves a computer, its systems, or its applications. Forensic investigator- an investigator who helps organizations and law enforcement agencies in investigating cyber crimes and prosecuting the perpetrators of those crimes.

Data mining is part of the interdisciplinary field of knowledge discovery in databases. Research on data mining began in the 1980s and grew rapidly in the 1990s.Specific techniques that have been developed within disciplines such as artificial intelligence, machine learning and pattern recognition have been successfully employed in data mining. Data mining has been successfully introduced in many different fields. An important application area for data mining techniques is the World Wide Web Recently, data mining techniques have also being applied to the field of criminal forensics nothing but Digital forensics. Examples include detecting deceptive criminal identities, identifying groups of criminals who are engaging in various illegal activities and many more. Data mining techniques typically aim to produce insight from large volumes of data.

II. 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. File systems are classified into the following four categories:

- ♣ Disk file system: A disk file system is used for stor-ing and recovering the files on a storage device, such as a hard disk, that is directly or indirectly connected to a computer. A few examples of disk file systems are FAT16, FAT32, NTFS, ext2, ISO 9660, ODS-5, and UDF.

- ♣ Network file system: A network file system is a type of file system that provides access to files on other computers on a network. The file system is transparent to the user. A few examples of network file systems are NFS, CIFS, and GFS

- ♣ Database file system: Earlier file systems use a hie-rarchical management structure, but in the database file system, files are identified by their characteristics, like the name, type, topic, and author of the file, or similar metadata. Therefore, a file can be easily searched using SQL queries or text searches. For example, if a user needs to find the documents written by ABC, then the search string “documents written by ABC” will show the results.

- ♣ Special purpose file system: A special purpose file system is a file system where the files are organized by software during runtime. This type of file system is used for various purposes, such as communication between computer processes or temporary file space. Special purpose file systems are used by filecentric operating systems such as UNIX. One example in UNIX is the /proc file system, which can beused to access information about processes and other operating system features.

III. HARD DISKS

Data is organized on a hard disk in a method similar to that of a filing cabinet. The user can easily access the data and programs. When a computer uses a program or data, the program or data is copied from its location to a temporary location. When a user makes changes to a file, the computer saves the file by replacing the older file with the new file.journal, so please be sure to refer to the correct journal when seeking information.

Lost Clusters A lost cluster is a FAT file system error that results from how the FAT file system allocates space and chains files together. It is mainly the result of a logical structure error and not a physical disk error.They usually occur because of interrupted file activities; thus, the clusters involved never get correctly linked to a file. Operating systems mark these clusters as being used in the FAT, even though they are not assigned to any file. Disk-checking programs can scan an entire disk volume for lost clusters. The programs can then either clear the lost clusters or save them as files. In the latter case, artificial files are generated and linked to these clusters. These newly formed files are considered damaged, but some orphaned data can be seen and recovered. Disk-checking programs, such as ScanDisk, can find lost clusters using the following procedure:

- ♣ Create a memory copy of the FAT, noting all of the clusters marked as being in use.

- ♣ Trace the clusters starting from the root directory, and mark each cluster used by a file as being accounted for. Continue through all of the directories on the disk.

- ♣ When the scanning process is finished, any clusters that are in use but not accounted for are orphans,or lost clusters.

IV. 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

3) File Deletion: When a file is deleted, the record of the file is removed from the table, thereby making it ap-pear 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 signa-ture 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.
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 re-main 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 consi-dered as free space, there also criminal can hide sensitive information.It is crucial since the content is not recreated, but rather converted into the final published version.

V. DATA MINING AND CRIME PATTERNS

We will look at how to convert crime information into a data-mining problem , such that it can help the detectives in solving crimes faster. We have seen that in crime terminology a cluster is a group of crimes in a geographical region or a hot spot of crime. Whereas, in data mining terminology a cluster is group of similar data points – a possible crime pattern. Thus appropriate clusters or a subset of the cluster will have a one-to-one correspondence to crime patterns. Thus clustering algorithms in data mining are equivalent to the task of identifying groups of records that are similar between themselves but different from the rest of the data. In our case some of these clusters will useful for identifying a crime spree committed by one or same group of suspects. Given this information, the next challenge is to find the variables providing the best clustering. These clusters will then be presented to the detectives to drill down using their domain expertise. The automated detection of crime patterns, allows the detectives to focus on crime sprees first and solving one of these crimes results in solving the whole "spree" or in some cases if the groups of incidents are suspected to be one spree, the complete evidence can be built from the different bits of information from each of the crime incidents. For instance, one crime site reveals that suspect has black hair, the next incident/witness re-veals that suspect is middle aged and third one reveals there is tattoo on left arm, all together it will give a much more complete picture than any one of those alone. Without a suspected crime pattern, the detective is less likely to build the complete picture from bits of information from different crime incidents. Today most of it is manually done with the help of multiple spreadsheet reports that the detectives usually get from the computer data analysts and their own crime logs. We choose to use clustering technique over any supervised technique such as classification, since crimes vary in nature widely and crime database often contains several unsolved crimes. Therefore, classification technique that will rely on the existing and known solved crimes, will not give good predictive quality for future crimes. Also nature of crimes change over time, such as Internet based cyber crimes or crimes using cell-phones were uncommon not too long ago. Thus, in order to be able to detect newer and unknown patterns in future, clustering techniques work better

VI. DIGITAL FORENSIC TOOLS

Restricted Access Tools

Users can access the following tools after they register and are vetted. Live View LE allows forensic investigators to take a physical device or an image file of a disk or partition and automatically transform it into a virtual machine. CCFinder is a suite of utilities designed to facilitate the discovery, organization, and query of financial data and related personally identifiable information in large-scale investigations. CryptHunter alerts law enforcement if active encryp-tion is running on a system so that investigators can act to preserve evidence that would be lost if the system were shut down.

ADIA is a VMware-based appliance used for digital investigation and acquisition. Unrestricted Access Tools Users can access the following tools for free; no secondary access is required. AfterLife permits the collection of physical memory contents from a system after a warm or cold reboot. Live View (public version) is a Java-based graphical forensics tool that creates a VMware virtual machine out of a raw (dd-style) disk image or physical disk.

DINO is a lightweight front end for network visualization and utilizes the open source network monitoring tools SiLK and SNORT to create an easy-to-use dashboard for situational awareness.

LATK is a collection of command line and web-based tools for use in incident response and long-term analysis of web server and proxy server log data.
CERT Linux Forensics Tools Repository houses pack-ages for Linux distributions. The repository provides useful tools for cyber forensics acquisition and analysis practitioners and is currently offering Fedora and Cen-tos/RHEL.

Information Only

Users can access information and perhaps more about the following tools; requests are handled on a case-by-case basis.

C-CAP is a state-of-the-art forensics analysis environ-ment that provides a broad set of tools for host-based and network investigations. MCARTA is a completed incident analysis framework in respect to run-time analysis with automated log and pocket data correlation.

VII. CONCLUSIONS AND FUTURE DIRECTION

We looked at the use of data mining for identifying crime patterns crime pattern using the clustering techniques. Our contribution here was to formulate crime pattern detection as machine learning task and to thereby use data mining to support police detectives in solving crimes. We identified the significant attributes; using expert based semi-supervised learning method and developed the scheme for weighting the significant attributes. Our modeling technique was able to identify the crime patterns from a large number of crimes making the job for crime detectives easier. Some of the limitations of our study includes that crime pattern analysis can only help the detective, not replace them. Also data mining is sensitive to quality of input data that may be inaccurate, have missing information, be data entry error prone etc. Also mapping real data to data mining attributes is not always an easy task and often requires skilled data miner and crime data analyst with good domain knowledge. They need to work closely with a detective in the initial phases. As a future extension of this study we will create models for predicting the crime hot-spots [3] that will help in the deployment of police at most likely places of crime for any given window of time, to allow most effective utilization of police resources. We also plan to look into developing social link networks to link criminals, suspects, gangs and study their interrelationships. Additionally the ability to search suspect description in regional, FBI databases, to traffic violation databases from different states etc. to aid the crime pattern detection or more specifically counter terrorsim measures will also add value to this crime detection paradigm.

REFERENCES

1.Computer forensics: Evidence collection and preservation.

2. Computer Forensics: Investigating Hard Disks, File and Operating Systems

3.Crime pattern detection using data mining- Brown University

4. Common Phases Of Computer Forensics Investigation Models-College of Information Technology, Universiti Tenaga Nasional, Selangor, Ma-laysia

5.http://www.gfi.com/blog/top-20-free-digital-forensic-investigation-tools-for-sysadmins/