Tools and Technology for Computer Forensics: Research and Development in Hong Kong (Invited Paper)

Lucas C.K. Hui, K.P. Chow, and S.M. Yiu

Department of Computer Science
The University of Hong Kong
Hong Kong
{hui,chow,smyiu}@cs.hku.hk

Abstract. With the increased use of Internet and information technology all over the world, there is an increased amount of criminal activities that involve computing and digital data. These digital crimes (e-crimes) impose new challenges on prevention, detection, investigation, and prosecution of the corresponding offences. Computer forensics (also known as cyberforensics) is an emerging research area that applies computer investigation and analysis techniques to help detection of these crimes and gathering of digital evidence suitable for presentation in courts. This new area combines the knowledge of information technology, forensics science, and law and gives rise to a number of interesting and challenging problems related to computer security and cryptography that are yet to be solved. In this paper, we present and discuss some of these problems together with two successful cases of computer forensics technology developed in Hong Kong that enable the law enforcement departments to detect and investigate digital crimes more efficiently and effectively. We believe that computer forensics research is an important area in applying security and computer knowledge to build a better society.

Keywords: Computer forensics, digital crimes, forensics technology.

1 Introduction

The use of Internet and information technology has been increasing tremendously all over the world. In Hong Kong, according to the surveys conducted by Census and Statistics Department of the Government, the percentage of households with personal computers at home that are connected to Internet has increased by more than 75% from 2000 to 2005 (see Table 1) while for the business sector, the percentage of business receipts through electronic means has increased by almost four folds (see Table 2). As one may expect, the amount of criminal activities that involve computing and digital data (digital crimes or e-crimes) has also increased. From the statistics provided by the Hong Kong Police [5], the number of digital crimes in Hong Kong has increased more than double from 2001 to 2004.

E. Dawson and D.S. Wong (Eds.): ISPEC 2007, LNCS 4464, pp. 11–19, 2007.
© Springer-Verlag Berlin Heidelberg 2007

Table 1. Penetration of Information Technology in the Household Sector in HK [2]

	Year 2000	Year 2005
Households with personal computers at home	49.7%	70.1%
Households with personal computers at home connected to Internet	36.4%	64.6%

Table 2. Penetration of Information Technology in the Business Sector in HK [2]

	Year 2000	Year 2005
Establishments with personal computers	51.5%	60.5%
Establishments with Internet connection	37.3%	54.7%
Establishments with Webpage or Website	7.3%	15.5%
Business receipts through electronic means	0.17%	0.64%

These digital crimes (e-crimes) impose new challenges on prevention, detection, investigation, and prosecution of the corresponding offences. Computer forensics (also known as cyberforensics) is an emerging research area that applies computer investigation and analysis techniques to help detection of these crimes and gathering of digital evidence suitable for presentation in courts. While forensic techniques for analyzing paper documents are very well established, very few of these techniques can be applied to digital data and they were not designed for collecting evidence from computers and networks. This new area combines the knowledge of information technology, forensics science, and law and gives rise to many interesting and challenging problems related to computer security and cryptography that are yet to be solved.

Among other issues in collecting evidence from computers, one fundamental difference between paper documents and digital data is that electronic data can be easily copied and modified. A suspect may easily argue that the evidence found in his/her computer was implanted or modified by the law enforcement agency after the computer has been seized by the agency. It is very important to verify the *file system integrity* of the suspect's computer after it has been seized by the law enforement agency.

Another problem is that there are many different file formats, operating systems and file system structures. Electronic documents can be generated by various kinds of application programs such as word processors, spreadsheet software, database software, graphic editors, electronic mail systems. The documents can be stored as user files in user directories, or as fake system files in the system

directories, or hidden files. Sometimes, evidence can also be found in the *deleted* files. When a file is deleted, the operation system usually only removes the references to the file in the file allocation table (FAT). The actual content of the file is still physically stored on the disk until that area has been overwritten by another file. It is a time consuming task to inspect every possible storage area of the whole computer for potentially useful evidence. And it is also not possible to check every file using all available application programs manually. In this paper, we will briefly describe a cyber crime evidence collection tool [4], called *Digital Evidence Search Kit (DESK)* ich tries to handle the above problems. DESK is the product developed by our research team and the Hong Kong Police Force and several other law enforcement agencies of the Hong Kong Special Adiminstrative Region.

Besides the problem of evidence collection, e-crime detection is also very important. Intrusion detection (e.g. detection of distributed denial of service attack [9,13]) is one of the well-known examples. In this paper, we focus on another example - detection of copyright infringement through peer-to-peer (P2P) file sharing. According to a survey conducted by the Hong Kong Government in 2005 [7], the public awareness of IP (Intellectual Property) rights has significantly improved. Out of about 1200 respondents, only 15% admitted that they would often (0.7%) or sometimes (14.3%) buy pirated or counterfeit goods. This is already a remarkable improvement from the 24.7% in 1999. However, the percentage of respondents who admitted that they would illegally download and upload files to Internet for the purpose of sharing with others has increased from 3.5% in 2004 to 6.8% in 2005. This may indicate that the copyright infringement problem becomes more serious (at least in Hong Kong) as the peer-to-peer file-sharing protocols become more popular and mature.

In fact, this is not only a problem in Hong Kong. According to a third-party research, potential losses to the recording industry from P2P file-sharing was estimated at US$2.1 billion in 2004 [6]. Among the few successful P2P protocols in existence, BitTorrent (BT) has evolved into one of the most popular networks [8] and has managed to attract millions of users since inception. By the end of 2004, BitTorrent was accounting for as much as 50% of all P2P-related traffic [11]. Without doubt, P2P technology offers a wonderful platform for individuals and organizations to share their digital materials worldwide extremely easily. Unfortunately, its illegitimate use on unauthorised sharing of copyrighted files is increasingly rampant and is reaching an alarming height.

With the existence of the overwhelming private BitTorrent networks, it is difficult to gauge the actual numbers of BT users. What we are certain, however, is the tremendous loss to the media industries. Over the years, law enforcement agencies have set out operations to fight against these illegal activities. With much of their effort, the world's first conviction of piracy of BitTorrent user was sentenced in the fall of 2005. However, the outcome seems not to be an effective deterrent to average BT users. Although many individuals realize that what they are doing is a kind of online piracy and is illegal under recently enacted legislation, they still pursue the file sharing as before. One critical issue behind

this is the limited manpower and resources available to law enforcement agencies. BT users may feel that it is almost impossible to crack down every single member of the enormous BT user base. To tackle this problem, it is desirable to have an automated system for monitoring these increasingly rampant BT activities. In this paper, we will briefly describe a rule-based BT monitoring system (BTM [3]) which takes the first step towards solving this problem.

The rest of the paper is organized as follows. The DESK system will be described in Section 2. Section 3 will briefly talk about the BTM system. Section 4 concludes the paper by presenting a few other related problems in computer forensics.

2 The Digital Evidence Search Kit

DESK (The Digital Evidence Search Kit) is a general purpose computer forensics system focusing on integrity control of the digital data. There are two design objectives of this tool. One of the objectives is to ensure the validity and reliability of the digital evidence. Once it has been proved that the tool has been used properly and in compliance with the Evidence Ordinance [10], the digital evidence found in the suspect's computer can be presented and used in courts for prosecution. Another objective is to provide an efficient and automatic search function to search for digital contents that can be used as evidence for the e-crime. DESK is also specially designed to be used in the bilingual environment of Hong Kong, so is capable of searching word patterns in both English and Chinese (traditional and simplified chinese characters).

2.1 The Framework of DESK

The DESK system consists of a DESK machine which is typically a notebook computer with a serial port and a floppy diskette used to start up the suspect's machine (subject machine). The DESK machine will be connected to the subject machine using a serial (RS-232) cable. There are two software components of DESK: the DESK client that is installed on the DESK machine; and the DESK server that is contained on the floppy diskette to be executed by the subject machine. The DESK client is mainly used to provide a user interface for issuing commands to inspect the subject machine.

The DESK server component, installed on the floppy diskette, has additional functionalities which include the followings.

1. To start up the subject machine: Usually the file (e.g. system files) in the subject machine will be modified if it is booted up by its own operating system.
2. To *lock* the subject machine: This is to protect the subject machine from any accidental corruption by the interrupts of the machine. This step is very important as it can ensure that the contents found on the subject machine

cannot be modified, thus ensures the integrity integrity of the subject machine while various forensic operations are being performed.
3. To provide a simple user interface for simple search operations: The user interface is much less sophisticated than that of the DESK client running on the notebook due to the storage limitations of floppy diskettes.

There are two main operations of DESK: keyword search and file system integrity checker.

Keyword Search: A pre-defined text pattern file which contains important keywords that can be specific to a particular case, in Chinese and/or English, to be searched for on the subject machine, is used for three different types of search, namely physical search, logical search and deleted file search. Physical search performs a search of the patterns in each physical sector of the subject machine's storage system. E-crime evidence stored purposely in unused sectors can be discovered. Logical seach, on the other hand, makes use of the information about the file system, so patterns stored across different sectors can be located. Deleted file search will try to locate the deleted file provided it is not yet overwritten by another file and perform the pattern search on these files.

File System Integrity Checker: There are two functions in this checker. Firstly, it is to ensure the integrity of the file system of the subject machine. We compute a hash value of the whole file system (e.g. a hard disk) of the subject machine. By recording this hard disk hash value properly, the law enforcement agency can easily prove that the content of the hard disk has not been modified after the machine has been captured by the agency. Also, in order to reduce the possibility of causing accidental damage to the hard disk, usually exact copies of disks (also called clone images) are made for the subsequent analysis. The hash values of the clone images and the original hard disk can be compared to show that the clone images are exactly the same as the original hard disk.

Secondly, the suspect may store some crime evidence in standard files of common software applications (e.g. freecell.exe). A hash value database that contains fingerprints (hash values) of all files in a standard software distribution are used to compare with the hash values of the corresponding files in the subject machine. More details of the DESK system can be found in [4].

2.2 Other Issues

There are other issues related to this research. For examples, it is very likely that there may be missing/bad sectors in the hard disk which may corrupte the data files in the system. How this can be handled to make sure that the recovered portion of the files can still be presented in courts needs more investigation. Also, the integrity checker relies very much on the hash functions. With the recent cracking of some well-known hash functions such as SHA-1 and MD5, may be a more detailed study needs to be done to make sure that the validity of the digital evidence is not questionable.

3 A Rule-Based BT Monitoring System

In this section, we briefly discuss the design of a rule-based BitTorrent monitoring system (BTM). For details, please refer to [3].

3.1 Basics of BitTorrent (BT)

A BitTorrent network is basically made up of four types of entities.

- Tracker: A server that coordinates the distribution of files. It acts as an information exchange center from which peers obtain necessary information about other peers to which they can connect.
- Torrent file: A small file which contains metadata about the files, including the address of the tracker, to be shared.
- Peer: Anyone that participates a download.
- Seeder: A peer who has a complete copy of the file and offers it for download. All peers (including the seeders) sharing the same torrent, are considered as a unit, called a *swarm*.

Note that the idea of BT is to redistribute the cost of upload to downloaders. When the peers are downloading the same file at the same time, they upload part of the files to one another. To start a download, a torrent file is generated, registered with a tracker and made available somewhere in a website. The owner of the initial copy of the file is referred as the initial seeder. Initially, peers will contact the initial seeder to request the file, as more and more peers join in, some peers will share their pieces with newly joined peers to offload the initial seeder.

3.2 The Framework of BTM

To track down the activities of a swarm, the torrent file is the key. BTM consists of two main components, the torrent searcher and the torrent analyzer. To locate torrent files, the torrent searcher searches target websites (or public forums) specified by user-inputted URLs. The torrent files will then be passed to the torrent analyzer for detailed analysis. There are several issues to be resolved by the torrent searcher. For examples, the searcher needs to explore the websites level by level following the hyperlinks to reach the torrent files. Automatic keyword searching needs to be performed by the searcher in order to explore potential illegal downloading activities in public forums. To conclude, this torrent searcher can be configurated to work on updated topics (e.g. newly released movies) and on scheduled websites/forums. It makes the monitoring possible for 24 hours.

After obtaining the torrent files from the torrent searcher, the torrent analyzer can locate and connect to the tracker(s) and retrieve the lists of peers currently participating in the torrent. It can further connect to the peers and gather useful information about the download activity and analyze the information to, say identify potential seeders and to determine if any necessary action needs to be triggered. The core engine inside the torrent analyzer is a rule-based system. Some preliminary tests have been conducted in some real scenarios. The

results are promising, however, more detailed analysis and experiments need to be performed to confirm its effectiveness.

3.3 Other Issues

This system represents the first step towards an automated monitoring system for the detection of copyright infringement activities through peer-to-peer file sharing. There are many other concerns that need an in-depth study. For examples, the anonymity level of BT is continuously being improved, how these anonymity features of the upgraded version affect the effectiveness of BTM is certainly one of the main concerns. On the other hand, the scalability of the tool is also a major issue needs to be resolved since the BT protocol seems to be very scalable and the number of peers can be huge.

4 Conclusion and Other Problems

In the previous two sections, we had described two examples in computer forensics research and development. To conclude this paper, we describe a few other related problems in computer forensics. Actually, we are working on some of these problems and preliminary research results may appear soon.

We believe that computer forensics research is an important area in applying security and computer knowledge to build a better society.

4.1 Live Systems Forensics

Most of existing computer forensics techniques concentrate on efficient search of digital evidence inside an inactive computer. The emphasis is on whether a particular piece of evidence exists in the machine or not. Recently research in computer forensics attempts to collect digital evidence from a live running system (e.g.[1]). This type of evidence may contain information that is transient, e.g. network connection. On the other hand, the ultimate goal of computer forensics is to reconstruct the crime scene in the digital world. Therefore one research direction is to concentrates on how to make use of the digital evidence collected from a live running system, filter out irrelevant information, and reconstruct the crime scene. This will involve not only carry out digital evidence search based on the syntactic elements, but also interpreting the evidence in a semantically correct way.

4.2 Cryptographic Scheme Design to Enhance Computer Evidence Validity

Digital data in a computer system needs confidentiality, integrity, and authentication control. With the viewpoint that those digital data may be extracted as

evidence by a computer forensics exercise, it will be better to design advanced cryptographic schemes which, during the time the digital data is generated, will provide cryptographic objects (such as hash values and digital signatures) at the same time. One example requiring this functionality is multi-media data. When a video file is used as a piece of computer evidence, we need to prove that a video file is really produced by a certain camera, it is really being created on a particular date, and is not tampered afterward. In addition, if part of the video file is corrupted, we still want the uncorrupted part to be valid evidence. This is an important research direction since our society is generating more and more different types of digital data, including text, documents, video, file systems, and others.

4.3 Authentication Schemes Providing Better Evidence

While authentication and the related topic of access control are being studied for a long time, there are still a lot of rooms for improvement regarding the provision of evidence. For example, to provide evidence about a login process using password, we need to assume the validity of the log file [12]. As a lot of criminal activities involve the step of impersonation, the computer evidence about authentication is particularly important. This situation is also being complicated by the diversified techniques of authentication, including password, digital signature, hardware cryptographic tokens, biometrics, one-time password, time-synchronous tokens, and third-party credentials. Therefore, the study of authentication with emphasis on the evidence provided is greatly desired.

4.4 Digital Objects with Dual Meanings

With the combination of cryptography, steganography, and the complicated data formats for digital documents, it is possible to create a digital object which can be interpreted in two or more different meanings. For example, a computer file can show different contents when it is being opened by two different software viewers. With one viewer the content can be a normal text story, while with another viewer it can be a child pornographic picture. Following the same idea, a more elaborate example is that a file can contain two encrypted portions. The file can be decrypted with two different decryption keys to show two different contents. What is the motivation of a person if he is storing such a file with dual meaning? Although finding the motivation of a person is not a computer security technical problem, there are technical problems involved: If a file with dual meanings is stored in a suspect's computer, will the computer forensics process be able to identify the two different meanings? What are the different technologies available for providing files with two or multiple meanings? Besides computer files, are there other kind of digital objects that can also have dual meanings? All these are interesting research topics with great impact.

References

1. Frank Adelstein. Live forensics: Diagnosing your system without killing it first. *Communications of the ACM*, 49(2):63–66, 2006.
2. Census and The Government of Hong Kong Special Administrative Region Statistics Department. Hong kong in figures, 2006 edition, 2006.
3. K.P. Chow, K.Y. Cheng, L.Y. Man, K.Y. Lai, L.C.K. Hui, C.F. Chong, K.H. Pun, W.W. Tsang, and H.W. Chan. A rule-based bt monitoring scheme for detecting copyright infringement activities, 2007. manuscript in preparation.
4. K.P. Chow, C.F. Chong, K.Y. Lai, L.C.K. Hui, K.H. Pun, W.W. Tsang, and H.W. Chan. Digital evidence search kit. In *Proceedings of the First International Workshop on Systematic Approaches to Digital Forensic Engineering*, pages 187–194, 2005.
5. The Government of Hong Kong Special Administrative Region Hong Kong Police. Technology crime statistics in hong kong, 2005.
6. IFPI. Ifpi external press pack, 2005.
7. The Government of Hong Kong Special Administrative Region Intellectual Property Department. Awareness of protection of intellectual property rights increases, 2006.
8. T. Karagiannis, A. Broido, N. Brownlee, K. Claffy, and M. Faloutsos. Is P2P dying or just hiding? In *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM'04)*, volume 3, pages 1532–1538, 2004.
9. D. Moore, G.M. Voelker, and S. Savage. Inferring internet denial-of-service activity. In *Proceedings of the 10th USENIX Security Conference*, pages 9–22, 2001.
10. Hong Kong Ordinances. Evidence ordinance. Chapter 8.
11. A. Parker. Peer-to-peer in 2005, 2005. CacheLogic Research.
12. Bruce Schneier and John Kelsey. Secure audit logs to support computer forensics. *ACM Transactions on Information and System Security*, 2(2):159–176, 1999.
13. P. Vixie, G. Sneeringer, and M. Schleifer. Event report, events of 21-oct-2002, 2002.