Online Acquisition of Digital Forensic Evidence

Mark Scanlon and Mohand-Tahar Kechadi

UCD Centre for Cybercrime Investigation,
School of Computer Science and Informatics,
University College Dublin, Belfield, Dublin 4, Ireland.
{mark.scanlon,tahar.kechadi}@ucd.ie

Abstract. Providing the ability to any law enforcement officer to remotely transfer an image from any suspect computer directly to a forensic laboratory for analysis, can only help to greatly reduce the time wasted by forensic investigators in conducting on-site collection of computer equipment. RAFT (Remote Acquisition Forensic Tool) is a system designed to facilitate forensic investigators by remotely gathering digital evidence. This is achieved through the implementation of a secure, verifiable client/server imaging architecture. The RAFT system is designed to be relatively easy to use, requiring minimal technical knowledge on behalf of the user. One of the key focuses of RAFT is to ensure that the evidence it gathers remotely is court admissible. This is achieved by ensuring that the image taken using RAFT is verified to be identical to the original evidence on a suspect computer.

Key words: Digital Forensics, Evidence, Remote, Hard Drive Acquisition, Imaging, Internet, Verification.

1 Introduction

Current trends in technology are putting computers with high-bandwidth Internet connections into the hands of regular criminals. As this phenomenon continues, an increasing number of traditional crimes are being aided by computers, e.g., fraud, identity theft, terrorism, etc. As a result, digital forensic investigators are becoming overwhelmed with the number of cases they have to deal with. Traditional digital forensic investigations commence with the investigators leaving their laboratory to visit the crime scene, where they collect all the relevant evidence, and bring it back to the forensic laboratory for secure storage and analysis. This evidence may then lay untouched for extended periods while the investigating team deals with the backlog of cases.

In this paper, we introduce a solution to reduce the time taken to acquire the necessary evidence. We propose RAFT (Remote Acquisition Forensic Tool), a remote forensic hard drive imaging tool, that is designed to boot off a Linux Live CD or USB memory stick. A brief overview of the RAFT scenario is depicted in Fig. 1. The suspect computer is booted using a customised Linux Live distribution and any hard drives or removeable media connected to the computer are able to be securely imaged over an internet connection directly to the

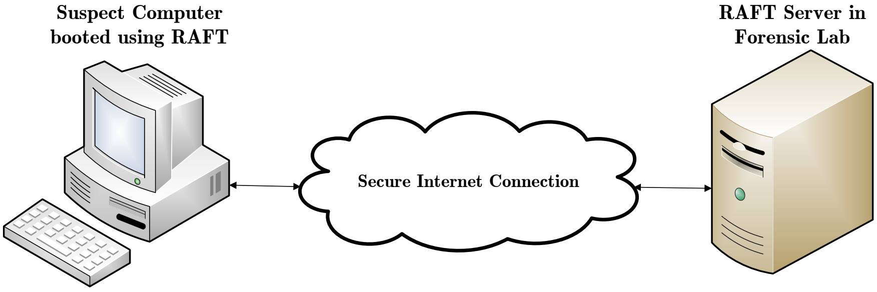


Fig. 1. Fundamental design of the RAFT system.

RAFT Server. This system is designed to equip any law enforcement or investigating officers with the ability to easily perform digital evidence acquisition, which would traditionally require the expertise of an on-site forensic investigator. One key objective of the RAFT system is to ensure that the evidence it gathers is court admissible. This is achieved by ensuring that the image taken using RAFT is forensically verified to be identical to the original evidence.

2 System Architecture

The RAFT System is based on a client/server architecture as can be seen in Fig. 2. The client side of the RAFT System is designed to be easy to use and to require minimal training in booting from the CD or USB memory stick. The RAFT Server is a multi-threaded server which can accept multiple connections simultaneously.

2.1 RAFT Client

The RAFT Client is installed on a customised lightweight copy of the Ubuntu Live Linux distribution [11]. Ubuntu was chosen for a number of reasons:

1. The standard Ubuntu install disk comes packaged with a live linux distribution. This live disk is bootable on any modern computer, regardless of manufacturer or operating system, e.g., Windows, *nix or MAC OS.
2. The compatability of the live disk to read numerous different drive formats, e.g., FAT, FAT16, FAT32, ext, ext2, ext3, HFS, HFS+, etc.
3. The ability for the live distribution to be fully customised removing any unnecessary software, while having the ability to easily include the RAFT Client and associated software (such as dcfldd [2] and sshfs [10]).
4. The ability to configure the auto mounting script to automatically mount all attached drives and storage as read-only.

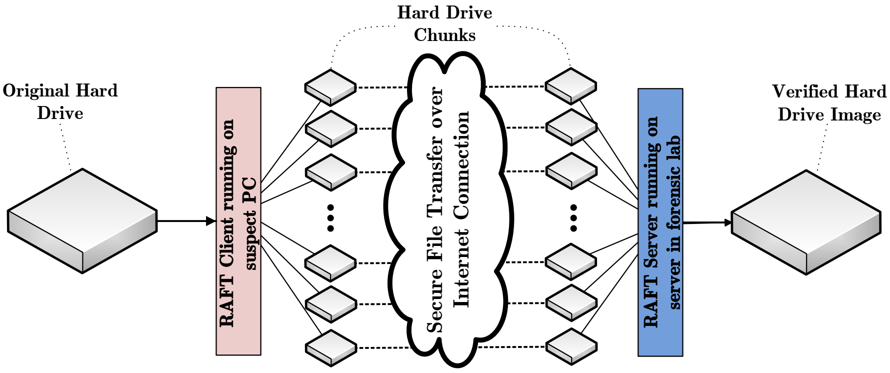

Hard Drive Chunks
Original Hard Drive
RAFT Client running on suspect PC
Secure File Transfer over Internet Connection
RAFT Server running on server in forensic lab
Verified Hard Drive Image

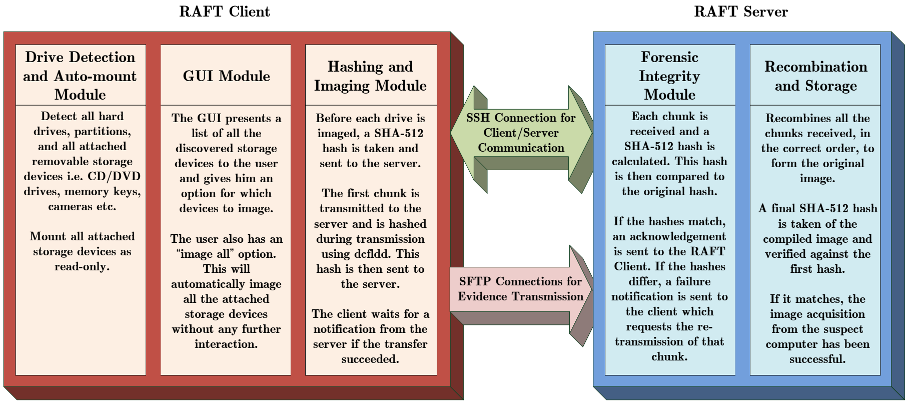
Fig. 2. Overview of RAFT imaging architecture. The left side of the figure represents the suspect computer, while the right represents the server-side of the system in a forensics laboratory.

When the Ubuntu Live CD is booted on the suspect computer, all the attached hard drives and removeable drives, e.g., USB memory sticks, external hard drives, cameras, CDs etc., which are currently connected are automatically mounted as read-only. The user is then prompted within the graphical user interface whether s/he would like to take an image of a specific drive or partition, or if s/he would like all the attached media to be imaged automatically. To ensure forensic verifiability, a hash is taken of the entire drive before it is imaged. This initial hash is taken to verify the untouched drive to the evidence ultimately collected from it.

2.2 RAFT Server

The RAFT Server is a multi-threaded system. When the server is running, it listens for a connection on any of its preconfigured ports. When a new client is connected, it creates a new space on the server where it stores all relevant files. Each new drive or volume imaged from the suspect computer is then stored within that space. Each chunk, when successfully transferred, is hashed server-side and compared against the original hash. If these hashes differ, a failure notification is sent to the RAFT Client which will result in that particular chunk being re-transmitted.

Upon the successful transmission of all the chunks belonging to a particular drive, they are recompiled back into a single file, hashed and verified against the original hash value taken by the RAFT Client before the imaging process commenced. When these two hashes match, a "successfully completed" transfer notification is then sent to the RAFT Client.


RAFT Client
Drive Detection and Auto-mount Module
GUI Module
Hashing and Imaging Module
SSH Connection for Client/Server Communication
SFTP Connections for Evidence Transmission
RAFT Server
Forensic Integrity Module
Recombination and Storage

Detect all hard drives, partitions, and all attached removable storage devices i.e. CD/DVD drives, memory keys, cameras etc.

Mount all attached storage devices as read-only.

The GUI presents a list of all the discovered storage devices to the user and gives him an option for which devices to image.

The user also has an "image all" option. This will automatically image all the attached storage devices without any further interaction.

Before each drive is imaged, a SHA-512 hash is taken and sent to the server.

The first chunk is transmitted to the server and is hashed during transmission using dcfldd. This hash is then sent to the server.

The client waits for a notification from the server if the transfer succeeded.

Each chunk is received and a SHA-512 hash is calculated. This hash is then compared to the original hash.

If the hashes match, an acknowledgement is sent to the RAFT Client. If the hashes differ, a failure notification is sent to the client which requests the re-transmission of that chunk.

Recombines all the chunks received, in the correct order, to form the original image.

A final SHA-512 hash is taken of the compiled image and verified against the first hash.

If it matches, the image acquisition from the suspect computer has been successful.

Fig. 3. Overview of the steps involved, both client and server side, in verifiable image acquisition using RAFT.

2.3 Forensic Integrity

Due to the sensitive nature of digital evidence collection, it is imperative that the data collected by any forensic tool is completely verifiable and identical to the original source. This integrity is insured in the RAFT system by the implementation of regular hash checking on the data being transferred using SHA-512, a 512-bit secure hashing algorithm. Once the RAFT Client is booted and a drive is selected for imaging, the first step is to calculate a hash value for the original drive. During the imaging process, the integrity of each of the chunks being transfered is maintained due to a SHA-512 hash being computed as the chunk is being transmitted. Server-side, once the transmission is completed, a SHA-512 hash is taken on the chunk and verified against the original. If these hashes do not match, i.e., the integrity of that chunk has been compromised in transmission, a failure notification is sent to the client, which queues that chunk up again for transmission.

3 Advantages

- Compatibility – One obvious advantage of using the RAFT system is that it is irrelevant what configuration the suspect PC has, i.e., RAFT is compatible with whatever interface or formatting the suspect hard drive or media might have. Take netbooks as an example: netbooks come in many various storage configurations, even within the same brand. Some netbooks use regular 2½" IDE or SATA laptop hard drives whereas some use flash storage. These flash drives can be soldered directly to the motherboard, connected via a regular IDE or SATA connection or connected via a mini-PCI/mini-PCIe connection. RAFT has no limitation on which configuration is used, the system will automatically mount and securely image any configuration.

- Cost – The cost involved in running the RAFT system is mainly on the server side. The RAFT Server would need to run on a high-end computer with a very high-speed internet connection. It would also be required to have a large amount of available storage, be it local storage or a connected NAS (network attached storage). However, once the initial outlay is spent in setting up the RAFT Server, the cost for using and re-producing the RAFT Client is minimal. For example, in a law enforcement scenario, the customised RAFT ubuntu image can be burnt to CD or a bootable USB key can be created as many times as required, e.g., one of each per police station.
- Automated Acquisition – This feature of the RAFT system results in users requiring little technological knowledge to operate the client side of the system. Due to the ease of use of the system, it will ultimately result in forensic acquisition being possible in more places at once, e.g., in the law enforcement scenario outlined above, each police station would have the capability to image a computer without the need to have a digital forensic specialist.
- Speed – While each individual image acquisition can take some time, multiple acquisitions can take place simultaneously. This results in an overall decrease in the time taken for multiple computers to be imaged at once to the same server.

The combined advantage of the above points results in the forensic investigator being able to spend more time in the laboratory analysing the evidence collected. Using RAFT in combination with more intelligent forensic analysis tools, e.g., a distributed digital forensic system [3], [6], [7], the investigator will be better armed to deal with an increasing amount of casework.

4 Potential Limitations

While the RAFT system has several advantages, such as those outlined above, there are also some potential limitations:

- Firewall – The RAFT Client has to have the ability to communicate to the server, for the transmission of the evidence. One obvious potential limitation of the system is that a hardware firewall may be filtering the suspect computer's internet connection, e.g., blocking specific port ranges, etc. This could potentially render the RAFT Client inoperable. One solution to this is to employ the use of a USB mobile broadband connection, connected to the suspect computer. Current 3G wireless broadband networks are capable of upstream speeds of up to 10 Mbps, with plans for 3G LTE (Long-Term Evolution) to increase the upstream speeds to over 50 Mbps [1]. These potential upload speed are set to improve even further when 4G mobile broadband networks become mainstream in the coming years. 4G networks will be capable of upload speeds of over 100 Mbps [4].
- Transfer Speed – The time taken to take an image of a hard drive over the Internet will take longer than the time required if the investigator had physical

access to the device and the imaging was conducted in a laboratory. RAFT can improve on this time required for traditional hard drive image acquisition if the time wasted by the investigation in travelling, transportation and storage of the suspect computer is taken into consideration. While high-speed broadband internet access is becoming more and more common place on both residential and commercial levels, it would be unrealistic to assume that every suspect computer would have an internet connection with a favourable upload speed, e.g., many asymmetric broadband connections are weighted towards download speed resulting in significantly slower upstream speeds. This limitation could again be overcome through the use of a mobile internet connection.

- Non-functional/no CD Drive – It is possible that some suspect computers have a non-functional CD drive. For that matter, some modern computing equipment do not feature a CD drive, e.g., most netbooks and small laptops. This issue can be overcome as the customised Ubuntu operating system containing RAFT can also be configured to boot of a USB memory key.
- Live System – Forensic investigators are increasingly concerned with the analysis of live systems, e.g., collection of evidence of current processes, memory and other state information. In its current from, the RAFT system is unable to collect evidence from a live system. However, an extended version of the RAFT client could be provided to run on a live system. The downside of executing RAFT on a live system is that there will be an unavoidable, yet predictable, change of state of the live system.
- Boot Passwords – Should the suspect PC have a CMOS boot password, before the user has the opportunity to boot up the RAFT Client, he must insert this password. In the quite likely event that this password is unknown, it will be necessary to reference documentation to retrieve the BIOS manufacturer's backdoor CMOS password. A sample list of common BIOS manufacturer's and their associated backdoor passwords can are given in table 1 [12]:

Table 1. Backdoor BIOS passwords for common motherboard manufacturers

Manufacturer	Commonly Used Passwords
AWARD	01322222, 589589, 589721, 595595, 598598 , ALFAROME, ALLY, ALLy, aLLY, aLLy, aPAf, award, AWARD PW, AWARD SW, *AWARD?SW*, *AWARD_PW*, *AWARD_SW*, AWKWARD, awkward, BIOSTAR, CONCAT, CONDO, Condo, condo, d8on, djonet, HLT, J256, J262, j262, j322, j332, J64, KDD, LKWPETER, Lkwpeter, PINT, pint, SER, *SKY_FOX*, SYXZ, syxz, TTPTHA, ZAAAADA, ZAAADA, ZBAAACA, ZJAAADC
AMI	AMI, AAAMMMIII, BIOS, PASSWORD, HEWITT RAND, *AMI?SW*, *AMI_SW*, LKWPETER, A.M.I., CONDO
PHOENIX	BIOS, CMOS, phoenix, PHOENIX, Phoenix

While such a list may seem combersome, a database of computers and motherboards and their corresponding backdoor passwords could be created to quickly access the correct password. If the password remains undocumented, a reset of the CMOS could be performed by removing the CMOS battery from the motherboard. After a short period of time, e.g., less than ten minutes, the BIOS will be reset to its factory state, with no boot password. Some motherboards also incorporate a jumper which, when removed, enables the user to bypass the CMOS password. Passwords further in the regular boot process of the suspect system, e.g., an operating system login password, will have no effect on the operation of the RAFT system as it will be the customised Live Ubuntu operating system which is booted immediately after the BIOS.

5 Results

To evaluate the performance of the RAFT system, numerous "real-world" scenarios were tested. For the purpose of this paper, we will discuss two of these scenarios. During the testing of the RAFT system, the performance of the imaging process tended to be linear. As a result, all of the results discussed below have been averaged to reflect the performance for each gigabyte of digital evidence collected. The "dcfldd" tool used in the RAFT system has the ability to compute the hash values at the same time as transmitting the chunk. The four values displayed in Figures 4 and 5 show the impact of the various hashing options on the overall performance.

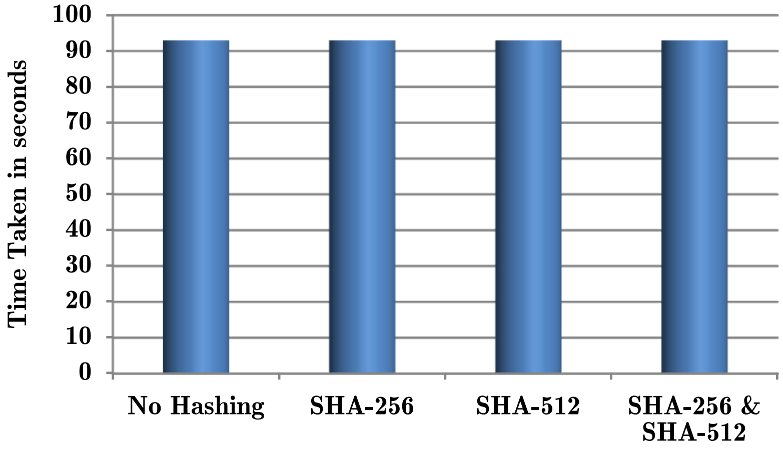


Fig. 4. Comparison of imaging times per gigabyte from a high-speed internet connection with a 83.26Mbps upload speed as tested using speedtest.net.

The first scenario concerns imaging a suspect computer with a very high-speed internet connection with a 87.62Mbps downlink and a 83.26Mbps upload streams (connection speed tested using Speedtest [9]). The suspect computer in this scenario was a Dell Optiplex 745 with a 2.66Ghz Intel Core 2 Duo processor, 2GB 667Mhz memory and a 250GB $3^1/_2$" 7200rpm hard drive. As can be seen in Fig. 4, the average time required per gigabyte was 92 seconds. The variance in the experiments conducted was ± 1 second or 1.09%.

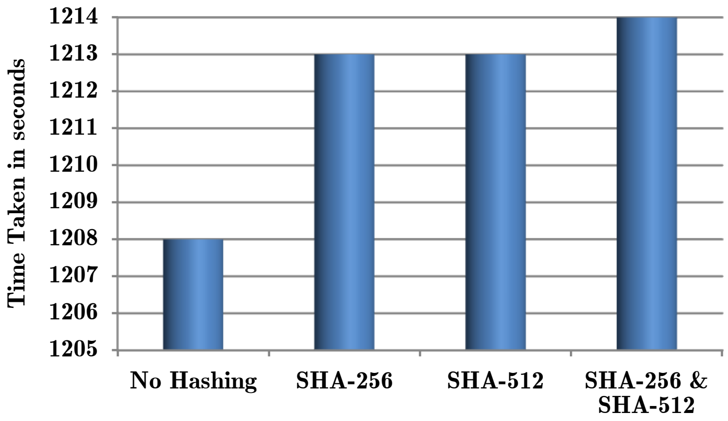


Fig. 5. Comparison of imaging per gigabyte times from a residential broadband connection with a 8.14Mbps upload speed as tested using speedtest.net.

The second scenario involved imaging a suspect computer from a residential broadband connection, with obviously low bandwidth speeds. The suspect computer in this scenario was a Dell XPS M1330 with a 2.5Ghz Intel Core 2 Duo processor, 4GB 667Mhz memory and a 320GB $2^1/_2$" 7200rpm hard drive. The broadband connection had a download speed of 23.82Mbps and an upload speed of 8.14Mbps (connection speed tested using Speedtest [9]). The result from these tests was that the average time to acquire a 320GB hard drive image was approximately 20 minutes per gigabyte, as can be seen in Fig. 5. The variance in the experiments conducted was ± 8 seconds or 0.0066%.

One requirement of the performance evaluation of the RAFT system was to quantify the overhead added through the secure hashing of each chunk. It was found that the cost for the hashing of each chunk averaged at 5.3 seconds per gigabyte (or a 0.41% increase in the time taken) as can be seen in Fig. 5.

The time taken for the server to verify each of the hard drive chunks received is approximately 20 seconds per gigabyte using the SHA-256 hashing algorithm,

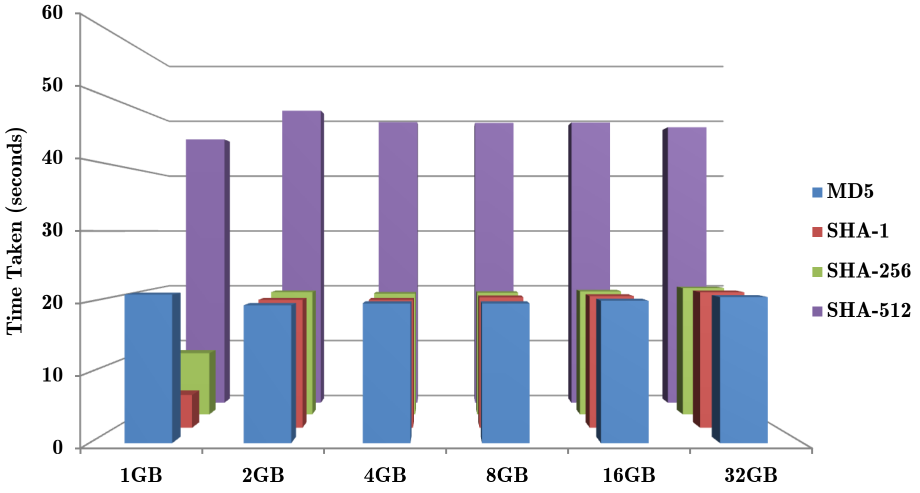


Fig. 6. Server side comparison of selected common hashing algorithms on various file sizes (normalised to a per-gigabyte value).

as can be seen in Fig. 6. This figure also displayed the server-side hashing times as a comparison of three other common hashing algorithms. One can notice that there is no additional overhead involved in choosing to use the 256-bit algorithm (SHA-256) as opposed to the 128-bit algorithm (MD5) and the 160-bit algorithm (SHA-1). However, there is a substantial cost of over double the computational time in using the more collision resistant 512-bit algorithm (SHA-512). The extra time required to compute the SHA-512 hash will only impact the overall imaging time twice; once on the RAFT Client before the imaging process begins, and again server-side when the image is completed.

6 Conclusion and Future Work

While the time taken to image a suspect computer over the Internet is substantially longer than the time taken using traditional forensic methods (with direct physical access to the hard drive [5]) the traditional approach does not factor in the time wasted by forensic professionals in the collection of this evidence. Using RAFT could give forensic investigators the power to remotely conduct investigations in more places at once.

One significant plus of using the RAFT system is that it is extremely cost effective to distribute the client side of the system over many locations. In the law enforcement scenario, this could result in every police station having a copy of the RAFT Client. This would result in granting regular police officers the ability to quickly image a suspect computer, after receiving any necessary warrents.

The RAFT System could be improved upon by giving optional total control of the RAFT Client to a remote investigator (after a suspect computer is booted). The investigator could remotely browse files on the suspect PC without the requirement to take an entire hard drive image. This would enable the investigator to determine if the suspect computer is relevant to the case and could help focus the investigation quickly on the computer(s) relevant to the crime being investigated. The imaging process would also be streamlined, focusing onto the necessary hard drives or partitions in the suspect computer.

A version of the RAFT Client could be created with the specific purpose of imaging a live system. This would have the ability to collect additional evidence from a live system, e.g., evidence located in memory, running process information and other state information. While executing any program on a live system will alter its state, this alteration would be predictable and should not interfere with the original evidence.

The system could also be improved upon by eliminating unnecessary transfer size, thus improving on the time required to collect all necessary evidence, e.g., through the employment of a lossless forensically sound compression algorithm.

References

1. Dahlman, E., Ekström, H., Furuskär, A., Jading, Y., Karlsson, J., Lundevall, M., Parkvall, S.: The 3G Long-Term Evolution – Radio Interface Concepts and Performance Evaluation. In: 63^{rd} IEEE Vehicular Technology Conference, Melbourne, Australia (May 2006)
2. dcfldd (Department of Defence Computer Lab Dataset Definition), April 2009, http://dcfldd.sourceforge.net
3. Gao, Y., Richard III, G.G., Roussev, V.: Bluepipe: A Scalable Architecture for On-the-Spot Digital Forensics. In: International Journal of Digital Evidence, vol. 3 no. 1 (2004)
4. Govil, J., Govil, J., An Empirical Feasibility Study of 4G's Key Technologies. In: Proceedings IEEE International Conference on Electro/Information Technology (EIT), pp. 267–270 (2008)
5. Ray, I., Shenoi, S., Advantages in Digital Forensics IV, Chapter 26: Time Analysis of Hard Drive Imaging Tools, pp. 340 – 343 (2008)
6. Richard III, G.G., Roussev, V., Next-Generation Digital Forensics. Communications of the ACM, vol. 49 no. 2, pp. 76 – 80 (2006)
7. Roussev, V., Richard III, G.G., Breaking the performance wall: the case for distributed digital forensics. In: Proceedings of the 2004 Digital Forensics Research Workshop, Baltimore, Maryland, USA (DFRWS August 2004)
8. Sealey, P., Remote Forensics. In: Digital Investigation vol. 1 no. 4, pp. 261–265. Elsevier (2004)
9. Speedtest Mini, Downloaded April 2009, http://www.speedtest.net/mini.php
10. SSH Filesystem, Downloaded April 2009 http://fuse.sourceforge.net/sshfs.html
11. Ubuntu Linux 8.04, November 2008, http://www.ubuntu.com
12. Wang, S-J., Measures of retaining digital evidence to prosecute computer-based cyber-crimes. In: Computer Standards & Interfaces, vol. 29 no. 2, pp. 216-223 (February 2007)