Information Security Leakage: A Forensic Analysis of USB Storage Disks

A.Adam and N.L.Clarke

Centre for Information Security and Network Research,
University of Plymouth, Plymouth, United Kingdom
e-mail: info@cscan.org

Abstract

Mobile devices have become immensely popular and people are now frequently storing sensitive data on them. The cheap, handy and small USB Storage Disks are endorsed by people to carry, transfer and backup data. Unfortunately, the users of such devices are not fully aware of the importance to protect their data and how to securely wipe them prior to selling them second-hand. For this project, ten USB keys have been bought on an auction website. Then they have been forensically imaged, analysed and the amount of data has been classified according to its sensitivity. Six out of ten of the keys contained sensitive data such as names, addresses, invoices and health records. The sensitivity of the information retrieved and the potential threats they could have lead to seem to indicate that it is of importance to educate users about how to best protect their data.

Keywords

Information Leakage, Data Breach, Computer Forensic, Evidence Recovery, USB Storage Disks

1 Introduction

The reality and importance of the data leakage problem has been highlighted by several surveys. According to a study made by the company PGP (2007) on the cost of data breaches, the primary cause of a data breach is at 49% lost laptops and other mobile devices including USB storage disks. Due to their small size and their price regularly decreasing, USB keys are a device likely to be easily lost or stolen. Thus, the UK Ministry of Defence recently admitted that a hundred of their USB storage disks have been lost or stolen since 2004; devices that sometimes contained secret information (BBC, 2008). If those devices can be stolen or acquired following a loss, they can as well be freely bought second-hand for a small amount of money, generally between £1 and £10. Some academic studies (Jones et al., 2005) have highlighted the fact that people did not erase properly their data prior to disposing of their hard disk drives. If people have not taken the habit to erase securely their HDD, it can be wondered if the other computer devices they sell second-hand can become a source of data leakage due to a lack of secure erasing.

This project's objectives were three in number: to investigate whether it is possible to retrieve data from USB storage disks bought second-hand, to try to evaluate the

sensitivity of the retrieved data and finally to understand the consequences of a possible disclosure of the restored data.

The second section of this paper will give some background literature. Academic research projects and studies done by companies specialised in security will be detailed in this section. The third section will focus on the methodology that has been built for the project. In the fourth section, the results obtained during the examination of the keys following the steps described in the methodology will be detailed. A discussion on these results will be given in the fifth section of the paper. Among other topic, the discussion will focus on the threats the data retrieved during the keys' examination could have lead to. The sixth and final section will conclude this paper and give some thoughts about possible future work.

2 Background literature

In this section on the background literature, four studies will be described. Two of them were conducted by academic researchers while the two others were leaded by companies specialised in security. Most of these studies have been focusing on analysing data leakage coming from hard disk drives sold second-hand. However two of them pushed their study a little further towards mobile devices by analysing laptops and flash memory devices.

The University of Glamorgan in association with the Australian University Edith Cowan conducted a study (Jones et al., 2005) whose purpose was to determine whether hard disk drives sold second hand were efficiently wiped prior to their selling. After having bought a hundred HDD from Australia, Germany, North America and the UK, those were imaged using either EnCase Forensic or Linux based Knoppix software prior to be examined. The examination of the disks was done in two steps. They firstly determined the presence of data, and then, they tried to find data likely to tie the disk either to an individual or to a company. Following their investigation, 57% of the disks revealed to which company they belonged and 53% contained one or more identifiable usernames.

The Canadian University of Ottawa conducted a similar study (El Emam et al., 2007) buying 60 hard disk drives coming from several Canadian vendors. However their interest was in a particular kind of data: private health information. To recover the files on the drives bought second-hand, the researchers used the commercial software program "Recover my files". Results showed that data could be retrieved from 67% of the disks among which 26 disks contained the address of their owner. Concerning their domain of interest, they found out that 18% of the disks contained private health information.

For the first part of their study on data leakage, Pointsec Mobile Technologies (Ahlberg, 2004) bought a hundred hard drives and laptops from auction websites and public auctions. They conducted a study similar to the two previously detailed, leading to the finding of data on seven disks out of ten. The second part of their study was a slightly different approach of the data leakage problem and concerned the lifecycle of a lost laptop. They followed the steps through which a lost laptop is going when forgotten in public places such as London airports. Unclaimed laptops

are sold in auctions where potential buyers have the possibility to have a look at it prior to buying it. Their study highlights that this way it is easy for ill-intentioned people to evaluate how they could use the remnant data of those devices.

The German security company O&O Software conducted several times a study (Kehrer, 2007) on second-hand hard disk drives leakages. However, the new part of the last edition of this study was that they also considered memory cards, cameras and USB sticks. 115 storage media were bought via online auctions coming both from Germany and from the USA. If 32 devices were securely deleted, 72.2% of the other disks presented recoverable data. Among the data found, a lot of pictures were available and other sensitive data that used to be backed up on the storage devices that were examined.

3 Methodology

At the beginning of this project it has been decided that a clear methodology was mandatory prior to taking any attempt at analysing the keys. Effectively, the data handled during the investigation are electronic evidence which require an appropriate care in order not to compromise them. Therefore, the ACPO Guidelines (Association of Chief Police Officers, 2003) describing data collection and evidence recovery have been used to build the following methodology. The process starts with the collection procedure. During this step, it has been decided to buy the second hand devices via the auction website eBay and to examine the keys with the help of the forensic software EnCase. Once the keys have been ordered and received, a number has been assigned to each of them and they were gathered with the packages they came with. The second phase is the examination procedure during which forensic images of each of the keys were produced. Electronic evidences can be easily corrupted; as a result there is a need not to work on the initial data but on an exact copy of it (Feldman, 2005). The images are named after the number assigned to the key it comes from and MD5 hashes are produced to ensure the integrity of the data (Scientific Working Group on Digital Evidence, 2006). The third step is the analysis procedure. During this step, several techniques have been used. Firstly, the information given by the key and its package were collected. Then a time analysis was performed to determine when the key was used. Given the relatively small size of a USB storage disk, a file-by-file analysis could be performed, opening and reading all the files that were not too damaged. Finally, keyword searches were used. This powerful technique allowed by EnCase generally lead quickly to an owner name. The last step of the evidence recovery is the reporting procedure during which reports on the keys were produced (Ashcroft et al., 2004).

4 Results

A set of ten USB storage disks has been bought for this experiment via the auction website eBay whose storage capacity distribution can be seen on Table 1. Given the storage capacity of the key bought for this project, it can be seen that the investigation has been conducted on a total capacity of 8,192 Mb of data.

Storage Capacity	128 Mb	256 Mb	512 Mb	1 Gb	2 Gb
Number of keys	1	2	1	5	1

Table 1 - Storage Capacity Distribution

However, all the keys did not contain data. As a result, the useful devices for the study were quickly narrowed to eight because two of them did not present any data. Likewise, two other keys have been considered useless for the purpose of the project which is to evaluate the amount of sensitive data presented by the bought devices. The reason why those two other keys could not be exploited is that they were filled with impersonal data such as movie and music files preventing the investigator to retrieve sensitive data on them. The remaining six USB storage disks all presented either personal or private data. The presence or absence of data on the keys and its sensitivity degree is given in Table 2.

	Private Data	*Personal Data*	*Not sensitive*	*No Data*
KEY-01		✔	✔	
KEY-02	✔	✔	✔	
KEY-03	✔	✔	✔	
KEY-04				✔
KEY-05				✔
KEY-06	✔	✔	✔	
KEY-07	✔	✔	✔	
KEY-08	✔	✔	✔	
KEY-09			✔	
KEY-10			✔	

Table 2 - Data presence on the USB storage disks

It has been decided that sensitive data would be split in two different categories. The first one is “personal” data; in this category, the data allowing identifying the owner of a key will be considered. However, this type of data is often given away by the user and do not constitute the most sensitive data. Along the investigation several examples of personal data were found such as:

- Names
- Addresses
- Email addresses
- Phone numbers
- Links to personal websites
- Education and work experiences

The second category of data is more likely to cause harm in case of a disclosure: it is “private” data. During the investigation, this type of data has been including:

- Photos
- Date of birth
- Invoices
- Nationality
- Credentials

- Network configurations
- Software serial numbers
- Health records
- Bank and credit card details
- Client and account numbers

Considering the eight keys returning data after the investigation, more than four thousand files could be retrieved by the forensic software. However, due to the damaged nature of many of those files, only a few of them could be entirely restored. Nevertheless, the possibility of retrieving the files allowed the investigator to consider what sort of files the owners of the USB storage disks were used to store. Thus, it has been found that pictures were the most commonly stored files on those disks, followed by multimedia files (music and video files for instance). The following categories, nearly equivalent in number, are software and document files. The rest of the results on the nature of the files retrieved on the keys can be seen on Figure 1.

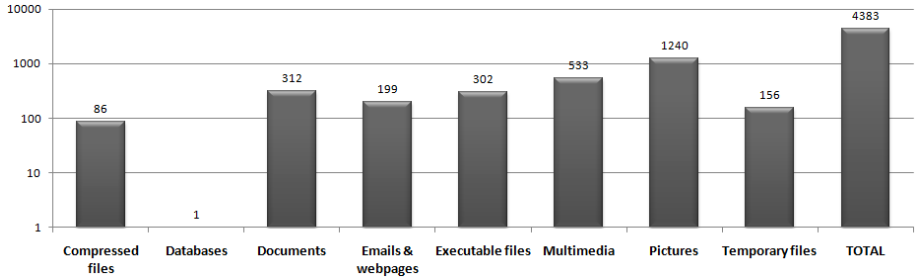

Figure 1 - Nature of the files retrieved on the keys

During the examination of the keys, three of them happened to contain a very large amount of sensitive data. Details on those three keys can be found in this section of this paper.

Data retrieved on Key-02 included the owner's contact details: full name, an email and two postal addresses, his phone number. In addition, what the owner looks like could be determined thanks to photos stored on the keys and entitled with "me" or with his name. Private documents have been restored from the key such as a couple of invoices and a request to obtain a domain name on the Internet. Apart from finding his contact details, those documents allowed the investigator to obtain the owner's signature and serial numbers of software bought on the Internet. Credentials to access a YouTube account were also stored on the key, unencrypted. The most worrying finding was to discover that the owner of this key kept a lot of files giving details on his network. Not only was the kind of network hardware used by the

owner described on the key but also the configurations of this network were available. Thus, unencrypted IP and MAC addresses, credentials to access the router and a WEP key were found. To finish, the key conducted to two personal website owned by the owner, one of them describing his work activity and the other, a weblog giving a lot of details about his life.

The majority of the personal details found on the owner of Key-06 came from remnant parts of his Curriculum Vitae stored on the device. Effectively, his full name, a permanent postal address, four other postal addresses with the corresponding phone numbers, an email address and education and work information were found. In addition more private information was also revealed such as his date of birth, his marital status and his nationality. Unprotected documents also contained credit card numbers and details on a bank account belonging to the owner of this key. If this owner did keep information on him, the analysis revealed that he possessed a large amount of sensitive information on other people. As a result, databases of students applying for the College where the owner worked were found on the key. Among the information contained in this database, names, addresses, phone numbers and sometimes even passport numbers belonging to the students were available.

Used as a backup device by its user, key-08 revealed the owner's most important files. As a result personal information found included a full name, a current postal address and ten previous ones, mobile and phone numbers, an email address and information on work and education. The owner's medical history could be found on the storage disk including his doctor's contact details, his daily medication and his latest hospital admissions. The owner of the key also stored letters he wrote to a number of creditors on which account numbers and client references are specified. Backup files from the web browser Firefox allowed the investigator to access stored credentials of about forty online accounts such as Paypal, eBay and The Carphone Warehouse. To finish, the key conducted to two personal websites created by the owner, the credentials to access the ftp hosting one of the website were found on the device.

5 Discussion

Compared to other research about data leakage, this study has been focusing on the involvement of USB storage drives alone. Studying USB keys often means that no operating system will be available. As a matter of fact, there is less chances to find traces of information automatically recorded by the operating system without the knowledge of the user. For instance, temporary internet files could lead to information a user would not want to disclose but which is automatically kept by the OS. On the contrary, on a USB stick, usually only files and directories are available and in addition the user of the key is likely to have chosen to write the files on his/her device. As a result, if the key is used to transfer music files it will not contain sensitive data. However, when it is used to carry work files or as a backup tape, it is possible that sensitive data can be retrieved.

Threats the data could have lead to have been investigated. Depending on the nature and amount of the data retrieved on each device, the previous owners of the second-

hand USB keys could have faced three different threats: identity theft, fraud and hacking attacks.

Four of the ten keys, presenting an important amount of sensitive data, could have lead to an identity theft scenario. Among them, the three of the keys detailed in the previous section could have been a target for ID theft. In addition to their owner's contact details; Key-02 revealed the signature of its owner, information that could have been useful to falsify documents more easily while impersonating the identity of the owner of this key; Key-06 provided its owner's full profile from his identity to his hobbies passing by his educational and work experience due to stored CVs; Key-08 contained its owner's National Insurance Number and his date of birth information that criminals look for to perform an Identity theft. The large amount of data on four of the keys among the ten bought could also have lead to fraud scenario. Effectively, Key-06 and Key-08 contained banking details that could have allowed a criminal to perform a banking fraud. The owner of Key-06 stored credit cards numbers and a bank account details. The owner of Key-08 made a back up of a software program he used to access his online banking account that would have provided a criminal the necessary credentials to access the accounts. Finally, the owners of two of the keys could have faced a hacking attack. The details on both the hardware and the configurations the owner of key-02 used would have facilitated a hacking attack of his home network; moreover, the details given on his modem router could have incited somebody to misuse his Internet access. The owner of key-08 would have allowed a criminal to access forty of his online accounts by storing files containing his credentials that could be easily decrypted.

To finish, it is of importance to highlight that, according to the results of this investigation, too few people seems aware that formatting their device or erasing the files will not result in a secure wiping of all their data. If the data found during the investigation had been bought by ill-intentioned people they could have been misused and the repercussions could have been serious for the previous owners of those devices. If the Operating System often gives to understand that formatting a USB storage disk will erase ALL data contained by the device, it should be stressed that a lot of the data the owner thought he had wiped are still available when using forensic tools to recover it as it has been proven in this investigation.

6 Conclusion and future work

Evidences of the problem of data leakage have been found in the news, surveys and research. Given some recent data breach incidents, even the UK government decided to take countermeasures about this phenomenon (House of Commons, 2007). In this problem, some sources seem to indicate that mobile devices are a major vector of information leakage. Among those mobile devices, this project has been focusing on handy, cheap and immensely popular USB storage disks. From the ten USB keys bought second-hand on an auction website for this study, six of them revealed who their previous owners were, providing at least their first and last names. As a result, the risk of leakage appeared to be real and its importance needed to be investigated. Therefore the risks, the owners of the USB disks could have faced, have been considered, depending on the amount and nature of both personal and private data provided by the keys. Alarmingly, the data could have been misused in a number of

fraud, hacking or ID theft scenarios. Such results seem to point out that it is of importance to carry on educating the users of computer hardware, raise their awareness on how important their sensitive data can be and how to protect them.

Among the possibilities of future work concerning this research, it has been thought that the set of keys could be extended. Effectively, the study has been considering ten keys which can be considered as a limited number. Conducting this study again or buying additional devices could allow the investigators to confirm the results or to extend the findings. For instance, instead of having USB keys belonging to individuals, it can be expected that USB storage disks belonging to companies could be found. USB storage disks being not the only flash devices, it could be envisaged to extend the study to other mobile devices such as flash cards and mp3 players.

7 References

Ahlberg, M. (2004) 'The lifecycle of a lost laptop' *info4security* [Online] Available: http://www.info4security.com/story.asp?storyCode=3047857&sectioncode=10 [Accessed 25 January 2008]

Ashcroft, J., Daniels, D. J. and Hart, S. V. (2004) 'Forensic Examination of Digital Evidence: A Guide for Law Enforcement' *National Institute Justice*

Association of Chief Police Officers (ACPO) and National Hi-Tech Crime Unit (NHTCU) (2003) 'Good Practice Guide for Computer based Electronic Evidence' Version 3.0 [Online] Available: www.acpo.police.uk/asp/policies/Data/gpg_computer_based_evidence_v3.pdf [Accessed 25 March 2008]

BBC (2008) 'MoD admits loss of secret files' [Online] Available: http://news.bbc.co.uk/1/hi/uk/7514281.stm [Accessed 30 July 2008]

El Emam, K., Neri, E. and Jonker, E. (2007) 'An Evaluation of Personal Health Information Remnants in Second-Hand Personal Computer Disk Drives' *Journal of Medical Internet Research* 9(3):e24 [Online] Available: http://www.jmir.org/2007/3/e24 [Accessed 25 January 2008]

Feldman, J.E. (2005) 'Ten Steps to Successful Computer-Based Discovery' *Computer Forensics Inc.*

House of Commons (2007) 'Justice – First Report' *Justice Committee Publications* [Online] Available: www.publications.parliament.uk/pa/cm200708/cmselect/cmjust/154/15402.htm [Accessed 3 May 2008]

Jones, A., Mee, V., Meyler, C. and Gooch, J. (2005) 'Analysis of Data Recovered from Computer Disks released for Resale by Organisations' *Journal of Information Warfare* 4(2):45-53

Kehrer, O. (2007) 'Data Data Everywhere' *O&O Software GmbH*, Berlin

PGP (2007) 'Annual Study: Cost of a Data Breach' [Online] Available: http://download.pgp.com/pdfs/Ponemon_COB-2007_US_071127_F.pdf [Accessed 25 January 2008]

Scientific Working Group on Digital Evidence (2006) 'Data Integrity within Computer Forensics'