The 2016 Analysis of Information Remaining on Computer Hard Disks Offered for Sale on the Second Hand Market in the UAE

Thomas Martin
Manchester Metropolitan University

Andy Jones
Cyber Security Centre, University of Hertfordshire and Security Research Institute, Edith Cowan University

Mohammed Alzaabi
Khalifa University of Science, Technology and Research, United Arab Emirates

Follow this and additional works at: https://commons.erau.edu/jdfsl

Part of the Computer Engineering Commons, Computer Law Commons, Electrical and Computer Engineering Commons, Forensic Science and Technology Commons, and the Information Security Commons

Recommended Citation
DOI: https://doi.org/10.15394/jdfsl.2016.1428
Available at: https://commons.erau.edu/jdfsl/vol11/iss4/6

This Article is brought to you for free and open access by the Journals at Scholarly Commons. It has been accepted for inclusion in Journal of Digital Forensics, Security and Law by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu, wolfe309@erau.edu.
THE 2016 ANALYSIS OF INFORMATION REMAINING ON COMPUTER HARD DISKS OFFERED FOR SALE ON THE SECOND HAND MARKET IN THE UAE

Thomas Anthony Martin¹, Andy Jones², and Mohammed Alzaabi³

¹Manchester Metropolitan University,
²Cyber Security Centre, University of Hertfordshire and Security Research Institute, Edith Cowan University
³Khalifa University of Science, Technology and Research, United Arab Emirates

T.Martin@mmu.ac.uk, andy1.jones@btinternet.com, msz1621@gmail.com

ABSTRACT

This research describes our survey of data remaining on computer hard disks sold on the second hand market in the United Arab Emirates (UAE). This is a repetition of the first survey conducted in 2012 (Jones, Martin, & Alzaabi, 2012). Similar studies have been carried over the last ten years in the United Kingdom, Australia, USA, Germany and France: (Jones, Mee, Meyler, & Gooch, 2005), (Jones, Valli, Sutherland, & Thomas, 2006), (Jones, Valli, Dardick, & Sutherland, 2008), (Jones, Valli, Dardick, & Sutherland, 2009). This research was undertaken to gain insight into the volumes of data found on second-hand disks purchased in the UAE, as well as any changes that have occurred since the previous survey. We will also compare these results to those produced in other regions of the world to gain an understanding of the relative level of the problem of residual data in the UAE.

The core methodology of the research adopted for this study was the same as has been used for the other global studies. The methodology included the acquisition of a number of second hand computer disks from a range of sources and their subsequent analysis. The goal of the analysis was to determine whether any data could be recovered from the disk and if so, whether the data that it contained could be used to determine the previous owner or user. If information was found on the disks and the previous user or owner could be identified, the research examined whether the information was of a sensitive nature or in a sufficient volume to represent a risk.

Keywords: Computer forensics, disk analysis, data recovery, data disposal, data destruction, data leakage, privacy.

1. INTRODUCTION

As technology advances, computers occupy an ever-more important role in our lives. We rely on them for everything, including communication, education, transport, finance, as well as entertainment. As their importance increases, so does the amount of data...
about us that we allow them to store. Mobile and personal devices, with their wide array of sensors to gather data, as well as close ties to social networks, further exacerbate the problem. While new features and improved convenience are often at the forefront of the consumer’s mind, and security often takes a back seat (although awareness of security issues is slowly improving). The lack of caution is especially a problem when it comes to disposing of devices. The typical lifespan of computing devices is shrinking, and proper sanitization of the confidential data they contain is not always performed before disposal. The potential impact of such lapses could vary from minor embarrassment of individuals to organizations suffering major data breaches. The information recovered in (Jones et al., 2012) ranged from holiday photos to confidential interviews collected as part of a University study.

The research question of this survey is: what are the amounts and types of information that remain on computer disks that have been offered for sale on the second hand market in the UAE. This is the second such survey to be conducted in this region (Jones et al., 2012). We are repeating the survey to determine what, if any, changes have occurred in the practices of data sanitization and disposal in the interim.

The overall finding of our survey was that there is still a dangerously high amount of sensitive data that remains on second-hand disks sold in the UAE. The research revealed that the number of disks that contain personally identifiable information has greatly increased, far above the global figure. The number of disks containing information related to organizations has also increased compared to the previous survey, but is still below the global figure.

The aim of the research was to determine whether any data remained on second hand computer disks and if it did, the potential sensitivity of the information to the previous owner. The research was conducted under the same conditions that had been used during previous second hand disk and hand-held mobile devices studies, carried out by the University of Glamorgan, Edith Cowan University, Longwood University, and British Telecommunication: (Jones et al., 2005), (Jones et al., 2006), (Jones, Valli, Dardick, & Sutherland, 2008), (Jones, Valli, Dardick, & Sutherland, 2009). The tools used were common and easily available, and equated to the undelete and unformat commands in older versions of Windows (Fisher, 2016) and a hexadecimal editor. We have made a new addition to this survey: all files recovered were also scanned for malware. The results of the research are that a number of observations can be made with regard to the level and type of information remaining on second hand computer disks and a number of conclusions and recommendations have been made on ways to improve the situation with regard to data remaining on second hand computer disks.

In Section 2 we will give an overview of similar studies. We describe the methodology of our survey in Section 3. Section 4 consists of the results of the survey, first presented on their own, then compared to the related works. Finally, we give our conclusions and recommendations in Sections 5 and 6 respectively.

2. RELATED WORKS

The first survey of data remaining on second-hard hard disks was performed in (Jones et al., 2005). This set out the general methodology that later surveys followed (including this one). Disks were purchased in small batches from different sources around the UK (92 in total). The disks were imaged and searched for identifying information. This initial survey revealed that sensitive data re-
maining on disks appeared to be widespread. 13% were faulty, and 16% of the remainder were wiped. Of the remaining disks, 70% had commercial data and 59% had individual data present.

In subsequent years, the survey was repeated, both to determine what, if any, changes occurred in the practice of data sanitization before re-sale, and to also include data from other locations. In (Jones et al., 2006) and (Jones, Valli, & G. S. Dardick, 2008), the survey included disks from the UK, Australia, North America, and Germany. France was added to the list of countries in (Jones, Valli, Dardick, & Sutherland, 2008) and (Jones, Valli, Dardick, & Sutherland, 2009). The observed trend shows a slow but steady decline in the number of disks with identifiable data present, but still remaining at a level to raise concern.

The first survey of second hand disks in the UAE was conducted in (Jones et al., 2012). The fraction of devices with individual or organisational data was less than the global rates. Specific results will be presented in comparison to this present survey in Section 4.2.

Other surveys have been conducted on other storage media. Instead of hard drives, USB drives were examined in (Jones, Valli, & Dabibi, 2009). Mobile phones were purchased and studied in (Valli & Jones, 2008). Both surveys revealed similar lapses in proper sanitization of devices before re-sale, but the specific types of personal data varied.

The subject of the ethics of research involving residual data was discussed in (Glisson, Storer, Blyth, Grispos, & Campbell, 2016). As well as reviewing the risks and privacy concerns of past studies, several recommendations were made: Identifying clear research objectives, suitable training for investigators, anonymization of information disclosed, complete lifecycle security practices, liaising with relevant ethics committees, and particular care when data is crossing national boundaries.

3. RESEARCH METHODOLOGY

To ensure that the results of the research provide a realistic and scientifically sound view of the situation, 40 second hand computer disks were obtained. All of the computer disks used in the research were purchased in computer or mobile phone shops in four of the Emirates (Abu Dhabi, Dubai, Sharjah, and Fujairah). The computer disks were purchased discretely, either separately or in small lots by a number of purchasers. This procedure was adopted in order to minimise the possibility of the sellers becoming aware of the purpose for which the computer disks were being obtained and to ensure that the actions of one seller did not have a disproportionate effect on the results of the study.

The computer disks were supplied “blind” to the researchers. This means that the researchers had no indication of the source of the computer disks that they were analysing. The only identifying mark on the computer disks that were provided to the researchers was a unique sequential serial number so that there was no indication of where or how they had been obtained (other than specifying the Emirate). The research methodology used was the same as that used in the previous surveys: (Jones et al., 2005), (Jones et al., 2006), (Jones, Valli, Dardick, & Sutherland, 2008), (Jones, Valli, Dardick, & Sutherland, 2009). The steps taken to examine each disk were as follows:

1. The disk was forensically imaged
2. The image was imported into FTK
3. File carving was performed on the image, seeking to recover Office documents, image files, PDFs, and html files.

4. The filesystem was manually examined for identifiable information, in particular the commonly used areas.

5. Samples of the various filetypes were manually examined for identifiable information.

6. On discovery of any potential identifiable information, searches were performed to find the locations of other matches.

7. Hashes of all recovered files were uploaded to Virustotal.

The forensic imaging was performed with readily available software (Raptor 3.0 Live CD\(^1\) or AccessData Forensic Tool Kit (FTK) Imager 3.1.0.1514\(^2\)). The analysis was then undertaken on the forensic images of the computer disks. There were two main reasons for the adoption of this time consuming step. The first was to preserve the original media in its original state and store it in a secure area in case criminal activity was discovered and there was a requirement to pass the disks on to law enforcement. By adopting this procedure, the chain of custody was preserved for any investigation by law enforcement. This allowed the research to be carried out in a non-intrusive manner that did not affect or change the original data. The second was that if any anomalies were detected with the image, it would be possible to validate the data against a second image created from the original.

The tools used in this study to analyse the images of computer disks were fundamentally the same as those used in previous disk studies (although the versions of the tools have changed). The primary tool used was FTK 4.0.0.35120. The tools performed similar functions to the Windows unformat and undelete commands. File carving of common file types was also performed as it is a reliable method for recovering files from unallocated space (e.g. either deleted files or files on a disk that has been formatted). Where it was necessary to manually examine information that existed in the unallocated portion of the computer disk, hexadecimal editors were used. Tools that perform this type of functionality are free and readily available: examples include forensic analysis tools such as Autopsy\(^3\) and the Linux Kali distribution\(^4\). Freeware Hexadecimal editors include XVI32\(^5\) and HxD\(^6\). These tools can be used effectively by anyone with a moderate level of skill investing some time and effort. I.e. it is reasonable to assume these actions could be performed by a malicious actor.

For this year’s survey, we also introduced the use of malware scans. All executables recovered from the disks were checked with the VirusTotal database\(^7\). This is intended to give an indication of the prevalence of malware on devices in the region, as well as also highlighting the risks posed by the use of second hand disks. All recovered executable files from the disks were hashed.

---

\(^1\)https://www.forensicsandediscovery.com/Pages/Raptor.aspx Accessed 10\(^{th}\) August 2016
\(^3\)http://www.sleuthkit.org/autopsy/ Accessed 12\(^{th}\) August 2016
\(^4\)https://www.kali.org/ Accessed 12\(^{th}\) August 2016
\(^5\)http://www.chmaas.handshake.de/delphi/freeware/xvi32/xvi32.htm Accessed 12\(^{th}\) August 2016
\(^7\)https://www.virustotal.com/ Accessed 15\(^{th}\) August 2016
and the hashes uploaded to the VirusTotal site using a script. The hashes are then checked against a number of anti-malware scans (generally 56 or 57 in our experiments). Since only the hashes were uploaded, this necessarily means that the scans were purely signature based, as opposed to heuristic. In order to reduce the numbers of false positives, we only consider a file to be malware if three or more of the scans flag the file as malicious.

In order to protect the privacy of the previous owners of the disks, appropriate measures were taken. The disks were stored in a physically secure location and all analysis was done on an isolated network. Only those with suitable training took part in the analysis and no identifying details have been included in the publication of the research.

The objectives of the research were the same as those defined for the previous disk and hand held mobile device studies: firstly to determine whether the computer disks that were obtained in the UAE had been effectively cleansed of data or whether they still contained information that was either visible or easily recoverable with the tools identified above. The second objective of the research was to identify whether there was information that could be used to identify the organisation or individual(s) that had used the computer disks.

### 4. RESULTS

#### 4.1 2016 Results

The initial results indicate that there was a significant amount of information found on computer disks in the UAE compared to the previous study, considerably more for personal information. There was also an increase in the number of disks obtained that contained highly sensitive/high risk information.

This section details the results for the study for the 40 computer disks obtained in the UAE:

- **7 disks (17.5%)** were damaged and as a result, unreadable.
- **7 (17.5%)** had been effectively cleaned and contained no recoverable data
- **Of the remaining 26 (65%) disks**
  - **15 (37.5%)** had been deleted or formatted, but still contained recoverable data.
  - **11 (27.5%)** contained data that could be easily recovered,
  - **9 (22.5%)** contained sufficient information for the organisation that they had come from to be identified.
  - **20 (50%)** contained sufficient information for individuals to be identified.
  - **17 (42.5%)** contained malware.

The results are presented as graphs in Figure 1. The significance of the first category of the remaining 26 disks is that the previous owner has made a naive attempt to sanitize the disk, typically through formatting and re-installation of the OS. This is not enough to prevent file carving of many of the user’s files. The second category, where data could be easily recovered, suggests that with more effort expended more files could be recovered. This is likely the case. Since file carving does not work with file fragments for example, one could use known techniques to try to find and match fragments. This was not done for a number of reasons. First, the time involved in such analysis would greatly

\[\text{Either the disk was entirely } \backslash ^00 \text{ bytes or was entirely random bytes.}\]
reduce the size of the sample that could be managed, and the larger the sample size the more reflective the results. Second, it seems unrealistic for a malicious actor to go to such lengths unless they knew there was something of value to be found. Thirdly, our recommendations at the end of the paper do not merely have the aim of making the user’s data “not easy” to recover, but rather impossible to recover.

4.2 Comparison to Previous Results

In Figure 2 we compare these results with those of the previous survey for the UAE, as well as the previous global survey. There is a noticeable increase in the amount of disks obtained that stored sensitive data relating to individuals. CVs, passport scans, and ID card scans were obtained from many of the disks. The fact that such scans are often requested when registering for services means they are more likely to be stored and kept by the users, but also increases the potential damage they could do in terms of fraud. The number of disks storing sensitive data relating to organizations has only slightly increased, but is still much lower than the value from the previous global survey.

Figure 3 shows an assessment of the level of risk that the data recovered would pose to the individual or the organisation. Whereas the previous survey had relatively few disks with medium and high risk data present, this time it appears to have significantly increased. This suggests a worrying trend of the increased use of technology for convenience, without a corresponding use of protection and disposal mechanisms that are necessary to ensure the security and privacy of the data. There also seems to be a blurring of the distinction between devices used for work and for personal use. We often came across disks that had both types of use. The question of whether they were personal devices where the owner often did their work, or work devices with a significant amount of personal use was difficult to decide.

As mentioned in Section 3, we included a malware scan of the disks for the first time in this survey. Of the 26 disks that had recoverable data, 65.4% (17/26) were flagged as having malware by our VirusTotal scan (meaning each of those 17 disks had at least
The 2016 Analysis of Information Remaining on Computer...

Figure 2. Comparison of results from the 2009 global survey, the previous 2012 UAE survey and the current 2016 survey (figures for data present are percentages of the number of disks that had not been wiped/faulty)

Figure 3. Assessment of level of Risk of the data recovered

one file that was flagged as malicious by at least three anti-malware scanners. This is a significant portion of the sample, and indicates an absence of suitable protection mechanisms in the region. It also suggests a risk to using second hand disks or devices.

Details for the top five are given in Table 1. The staggering number of malware files on these disks is surprising. If we take an even more conservative approach and only consider a file malicious if 30 or more anti-virus scanners flag it as malware (over half), then we see the numbers are reduced; dramatically in some cases ($L_1$ and $L_5$), but only slightly in others ($L_4$). This would seem to confirm that although there may be some false positives, there are certainly many cases of malware on the disks. The differences do suggest a wide variance in performance of the anti-virus scanners. On their website, they do disclaim that “VirusTotal is not responsible for false positives generated by any of the resources it uses” and that some of the scans are configured to be more aggressive than the public commercial versions (Unspecified, 2016). This would increase the chances of false positives.

The previous UAE survey noted the pres-
ence of Passport scans and other identity documents in many of the disks. This trend has continued, and even increased. Sensitive documents are required for many services in the UAE, and there are few precautions taken to control how these documents are spread. Other specific instances of valuable data, originating from commercial and private individuals, are detailed below:

- One disk had Funds Transfer Form (including account numbers) for an amount of over 680,000 AED.

- One disk was likely from a bank, and had records of a number of audits of security controls at a specific branch. It also had customer call sheets, with customers’ account and mobile numbers. There was even a document titled “Security Plan & Security Personnel Job Descriptions.”

- One disk contained a file named “Job Sites Userids.txt,” which seemed to have login usernames (and possibly passwords) for various sites and services. The passwords (if that was what they were) were not particularly strong.

- One disk was from a jewelry store and has several invoices for 6 figure sums (in AED).

- One disk was connected to local shipping. Recovered files included an agreement document between a shipping company and travel center, with signatures and seals, a crew list of 50 for a vessel with personal details of each, invoices from the shipping company, and “Sharjah Port Authority Master’s Report.” Also of note was a fax to the Security Office at Ajman Port regarding which people were allowed to go on-board vessels to carry out repairs.

- One disk had a comprehensive collection of financial details of a single individual, with bank account opening forms, credit card statements (with complete numbers), details of investments, and scanned checks for dividends. A summary slide listed account numbers, balances, and PINs. Some statements were encrypted, but using a password based on the individual’s date of birth. This date was readily available from other documents on the disk and the statements were decrypted.

---

9 All financial transactions and balances are stated in UAE dirham (AED). At the time of writing, the exchange rate is 1 AED equals 0.24 Euro or 0.27 USD.

10 We also simulated what would have been necessary if the password policy was known but the user’s
• One disk had several highly sensitive documents from a company in the Oil & Gas industry. Files included geological data, satellite imagery, “Annual Reservoir Performance Report January 2006,” map of “Most Likely Oil-In-Place Potential (Yet to Find),” “Planned Development Wells,” “SCADA Systems for *Region* Wells,” “Review and Audit of key technical processes.” Documents described projects costing tens of millions of dollars. A “Longer Term View” forecast was present with projections up to 2034.

Technology, for the most part, has kept pace with the need to protect information to an appropriate level while it is in use and to destroy it effectively when it is no longer required. The observed failures from commercial organisations seem to be a result of a lack of corporate policies and procedures for the protection and subsequent disposal of obsolete computer disks or a failure to effectively implement these policies where they do exist. For private users, the cause is more likely to be a lack of awareness of the risk arising from failing to erase such information. Equally, they may not be aware of the availability of the tools and techniques to ensure that data is properly erased when it is no longer required.

The subject of the effective erasure of data was first highlighted in a 1996 conference paper (Gutmann, 1996) that reported research on the secure deletion of data. A second paper (Gutmann, 2001) examined Data Remanence in Semiconductor Devices and a further paper (Garfinkel & Shelat, 2003) looked at the measures that could be taken to “sanitize” computer disks. In the period since the issue was first raised, there has been an ongoing series of news reports on high profile cases where personal or organisational information has been found in the public domain. There has also been increasing publicity and awareness of the issue of identity theft and the guidance on how to protect both personal and organisational information.

There are now a good range of tools that can be used to effectively erase data from computer disks; however, there appears to be an ongoing issue with making the users aware of the problem. With the ever increasing levels of theft of both corporate and personal information that are being reported ((Arthur, 2014), (Nakashima, 2015), (Perlroth, 2015), (Harding, 2016)) and the availability of suitable tools to erase data, it is difficult to explain the amount of information that has been found in this survey. It is clear that further investigation needs to be carried out to understand the underlying cause.

For those computer disks that were for personal use, the results indicate that there is an issue with awareness and education of the users as to what information is sensitive and how they can ensure that it is destroyed. For the most part the disks are used to store music, videos, photographs and other, often small items of information that the user will not necessarily consider to be significant. The items of information that were consistently found on the disks that are considered to be of high risk were copies of passports, visas, and bank account details.

For the disks that contained commercial information, there was a range of information on the companies’ operations and customers. This would be embarrassing and potentially damaging to the business if the knowledge that they had released it into the public domain became known.
5. CONCLUSIONS

The issue of the use and subsequent disposal of computer hard disks is a significant problem, and is bound to grow as such threats as identity theft and data disclosures becomes more prevalent. The results of the survey in the UAE show that the volume of relevant information recovered has grown since the previous survey, and also that there is even more extremely sensitive information that is not being properly disposed. The number of disks wiped has even slightly reduced. To compound the matter, because of local customs and requirements, where personal information was recovered it was potentially highly damaging.

Comparing these results to surveys done in other regions shows two significant departures. The disks in the UAE had significantly greater numbers of disks with individual data, but significantly fewer disks with organizational data. The reasons behind the possible differences are hard to precisely determine. It may be that “recycling” of devices is more common a practice among individuals in the UAE than with organizations. If users were aware of, and were following, proper data sanitization procedures, we would have expected to see a greater number of wiped disks. The numbers for this survey were actually less than both the previous UAE survey and those from other regions.

A number of the disks examined were found to contain sensitive corporate or personal information and the consequence of failing to remove or erase this data is that it continues to be available to anyone that might seek to exploit it. There is ongoing clear requirement that staff within organisations be given the appropriate awareness, education and training programmes. They need to be able to ensure that computer disks and other media is properly managed and that the data that they contain is protected properly while in use and then erased before the disks are disposed. For the disks from personal computers, there is a need for the owner to be educated with regard to the potential risks caused by the information and the steps that they can take to effectively erase the data when they dispose of the computer.

6. RECOMMENDATIONS

There are a number of measures that can be taken by both individuals and organisations to reduce the volume of sensitive information that is inadvertently given away when they dispose of computer disks. These include:

1. Education - A public awareness campaign by Government, the media, commerce and/or academia.

2. Risk Assessment The organisation carries out risk assessments to determine sensitivity of the information that may be stored on the computer disks that are being disposed.

3. Roles and Responsibilities The assigning of the roles and responsibilities to those individuals responsible for issuing, managing and disposing of computer disks.

4. Best Practices - The introduction of procedures within organisations to ensure that computer disks are disposed of in an appropriate manner.

5. Physical Destruction - Where appropriate, the physical destruction of the computer disks.

6. Encryption - The encryption of the computer disks to ensure that information
cannot be easily recovered in the event of their loss, theft or disposal.

7. Data Erasure Provide access to the tools and facilities to enable individuals to effectively remove the information from their computer disks.

8. Regulation - Require all manufacturers to include built-in means for the total erasure of all data within devices.

These simple measures could play a significant part in reducing the risks posed by the inadequate disposal of devices.

ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their constructive and insightful comments.
REFERENCES


breach data of 4 million federal workers. ([online]
https://www.washingtonpost.com/
world/national-security/
chinese-hackers-breach-federal-
governments-personnel-office/
2015/06/04/889c0e52-0af7-11e5
-95fd-d580f1c5d44e_story.html)
Perlroth, N. (2015, August). Ashley
Madison Chief Steps Down After
Data Breach. ([online]
http://www.nytimes.com/2015/08/
29/technology/
ashley-madison-ceo-steps-down
-after-data-hack.html?_r=0)
Unspecified. (2016, June). Virus Total:
Frequently Asked Questions. ([online]
faq/)