2011

Data remanence in New Zealand: 2011

Dax Roberts
University of Otago, New Zealand

H B. Wolfe
University of Otago, New Zealand

Originally published in the Proceedings of the 9th Australian Digital Forensics Conference, Edith Cowan University, Perth Western Australia, 5th -7th December 2011

DATA REMANENCE IN NEW ZEALAND: 2011

D. Roberts, H. B. Wolfe
Security Research Group, University of Otago
New Zealand
droberts@infoscience.otago.ac.nz, hwolfe@infoscience.otago.ac.nz

Abstract

This paper presents findings from a study of computer data remanence in New Zealand and considers three research questions. Those questions are "What is the level of data remanence in New Zealand?", "How does it compare with other countries?", and "Are there industries in New Zealand that are more likely to have data remanence issues?" Computer data remanence is data that remains on a hard disk drive after that hard drive has been prepared for disposal. Typically data remanence research involves purchasing second hand hard drives without knowing the original source and then a variety of tools and techniques are used to determine what if any data remains. That data can range from the mundane such as holiday snapshots, to data of concern such as the credit card details used to book the holiday. This research uses a very similar methodology to the research of an Australian-British led consortium into computer data remanence that has been conducted since 2005 (Jones et al., 2005). For this research, 100 hard drives were sourced from companies based in New Zealand that deal in second hand hard drives.

A total of 24 hard drives were found to have identifying information on them and this is consistent with the results of the consortium. When examining "Are there industries in New Zealand that are more likely to have data remanence issues?" there was an effective sample size of 14 hard drives which was not considered to be a large enough sample size to adequately draw conclusions. The data does suggest that schools are likely to be of concern however.

Keywords

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

INTRODUCTION

An Australian-British led consortium has been conducting research into computer data remanence in Australia and the United Kingdom (UK) since 2005 (Valli and Jones, 2005). The consortium has subsequently extended its research to include other member nations such as France, Germany and the United States (USA). Data remanence is a type of computer forensics that examines second hand hard drives to determine what types of data remains. The New Zealand Government Communications Security Bureau (GCSB) outline in a manual for government agencies (GCSB 2010) suitable procedures for sanitizing or removing data before government agencies dispose of their computer hardware. Ideally when someone sells a hard drive they should scrub or wipe their hard drive as this would permanently (in theory) remove the contents of the hard drive. This is to avoid the possible problems of stalking, espionage, identify theft and fraud as suggested by Medlin and Crazier (Medlin and Crazier, 2011) and others. Formatting a hard drive gives the appearance of the data being removed however with tools such as "Forensics Tool Kit" (FTK, 2011), "Encase" (Encase, 2011) and others it is possible to restore or find references to the previous data.

Data Remanence is different from the computer forensics that law enforcement agencies usually conduct as law enforcement are typically looking for specific criminal activity, must keep a "chain of custody" which clearly shows who had control of the hard drive when and show that the tools they use have not altered the suspect/source hard drive in any way. Law enforcement may also perform live capture to acquire a hard drive at a suspect's location for later analysis in a forensics lab and most likely have access to specialised law enforcement tools that academic and general researchers may not have access to. It was noted that Data remanence research produced by the consortium has not been published with regards to New Zealand. In general there is a lack of information regarding data remanence in New Zealand. This is likely because "*Used equipment is awash with confidential information, but nobody is looking for it—or else there are people looking, but they are not publicizing that fact*" as suggested by Garfinkel and Shelat (Garfinkel and Shelat, 2003).

The goal of this research therefore is to measure the level of data remanence in New Zealand by examining second hand hard drives using a methodology similar to those previously published by the consortium and Garfinkel and Shelat. Once the level of data remanence has been measured in New Zealand it will be compared

with the relevant consortium findings and if possible an attempt will be made to determine which New Zealand industries are more at risk of the issues data remanence presents.

METHODOLOGY

The methodology for this research is to:

1. Purchase second hand hard drives
2. Make a bitwise copy of the hard drive
3. Process the bitwise copy to generate files
4. Manually analyse the files
5. Look for "identifying information" either personally identifying or company identifying.
6. Report the data found

This methodology follows from that used by the consortium in their publications (Jones et al., 2009) with some minor modifications noted below.

The main criteria for the hard drives used in this research

Drives must be legitimately available on the second hand market in New Zealand. This meant the vendor must be selling the drives and not giving them away for free instead of disposing of them at the tip/garbage dump/recycling centre and the vendor must not know the state of the data on the drives themselves. This meant the researchers were not knowingly purchasing drives that were either scrubbed or drives that knowingly had data on them. It was also to ensure that the vendors did not purposefully scrub the hard drives before the researchers purchased the drives. USB drives and Solid state drives were excluded from this project due to cost and lack of availability. It should also be noted that research conducted in 2010 (Bell and Boddington, 2010) shows interesting technical issues surrounding forensic analysis of some Solid-state drives.

Purchasing the second hand hard drives

100 drives in total were purchased for this project. The second hand hard drives for this research were from two types of source. Firstly a total of 69 drives were sourced from land based companies that purchased second hand computer systems from all across New Zealand, reconditioned those systems and sold them nationally or internationally, and the second source was New Zealand based online auction sites.

The source of drives from online sources was "trademe.co.nz" which is currently New Zealand's largest online auction site. Hard drives were purchased in bulk lots of 6-10 at a time and this accounted for the other 31 hard drives in this research.

The standard process once a hard drive arrived was to catalogue each hard drive. This process involved recording the serial number, hard drive size [if printed on the drive], date purchased, and the source of the drive. It was not always possible to determine from the hard drive label the data of manufacture and therefore this was not collected.

Make a bitwise copy

The main reason to make a bitwise copy is so that the researchers do not analyse the original hard drive and potentially alter its contents in case legal matters do arise. Whilst it is very unlikely that illegal material (such as child pornography) would be found, the methodology stated that the police would be contacted immediately. The hard drive would still be considered part of the research project but analysis would stop on the bitwise copy and it would be noted that the analysis could not be completed. Ultimately this was not an issue in this research project but it should always be a consideration for similar work.

A Tableau TD1 (Tableau LLC, 2009), which allows for a fast (up to 6GB/minute) bitwise copy of a hard drive creation, was used for the capturing of the 58 IDE drives and 3 SATA drives.

The 39 SCSI drives were captured using a regular PC Tower with a SCSI card put into it. The "Forensic Tool Kit" (FTK) software created by AccessData was used for analysis of the data, and as part of that suite, the FTK Acquirer software was used to capture the SCSI hard drive data. It should be noted that the PC Tower was never connected to the Internet, had minimal software and services on it, but did have a reasonably up to date anti-virus program running on it.

Processing the bitwise copies

The researchers used a legally licensed version of FTK (Version 1.42) to process the bitwise copies of the hard drives. Whilst this tool is over five years old, it was still fit for purpose as it allowed the researchers to quickly see the types of files that were in the bitwise copy, and if the drive had been formatted or not. FTK also has a feature known as "data carving" where FTK will attempt to reconstitute deleted files. FTK also attempts to classify files into emails, spreadsheets, documents, html and other types which make processing the files found reasonably straight forward. Where the hard drive has not been formatted it may also be possible to see an overall directory structure for the hard drive and therefore it may be possible to see home directories such as "John Doe" and "Jane Doe" which may make finding identifying information easier. AccessData also provide free of charge a library of hash files called "Known File Filter". This was added to the tower to make the processing easier as it allows the user to "filter out" known files that are not likely to be of interest such as common operating system files that do not store any identifying information.

Analyzing the files

A hard drive was classified as having identifying information on it once the researchers were reasonably convinced that they knew who had previously owned the hard drive or had used the hard drive. The goal was not to find every single piece of identifying information on a hard drive and in many cases this could have taken several days of manual processing time per hard drive. The approach therefore was to examine files that were more likely to yield identifying information and the design decision was to examine files in the following order: emails, documents, spreadsheets and databases. Occasionally image files were examined, as FTK classified html files as being "document files" and would regenerate those html pages with images in them. Other times the directory structures and folder names would suggest there might be images of interest and those were then examined. Where a hard drive had been formatted, or no files of an identifying nature were found on a non-formatted drive, data carving was then used to attempt to reconstitute files to find identifying information.

Reporting the data

This research has used a slightly modified version of the reporting style used by the consortium. Below are the definitions used in common by both reporting styles and the differences then follow.

"Total": As the name implies the total number of drives analysed.

"Unreadable": Drives that were unreadable by the researchers. Typically that means drives that are physically broken which can include not powering up, data not being able to read off the hard drive or too many bad sectors, but may include drives in a format not understood by the computer system being used.

"Scrubbed" or "Wiped": Drives where no data (or virtually no data) at all remains, typically this is because a user has overwritten the data with all 1s or all 0s, or a specific or random pattern. This research project considered any drive that had less than 1 megabyte of data remaining on the hard drive as being scrubbed. This is because some scrubbing programs leave artifacts on the drive outlining which scrubbing program was used.

"Remaining": This is the number of readable drives with data on them that remain after unreadable and scrubbed drives are removed.

"Company Identifying" and "Individual Identifying": Drives that are deemed to have some form of data on them that identifies a company or individual.

"Formatted": This is where an attempt has been made to remove the data by formatting the hard drive. As stated earlier this would give non-savvy users the appearance that the hard drive had no remaining data on it.

The consortium also reports a count of hard drives for "Illicit" material. In the 2009 consortium results three of the countries did not report a value for "illicit" material, and the two remaining countries reported 8 hard drives as having illicit material between them (Jones et al, 2010)This represented less than 3% of all drives examined by the consortium for 2009 (they examined 346 in total). As the consortium covers different jurisdictions it was also unclear how exactly New Zealand research would cover "illicit" material. This is because some violent computer games that may be banned or censored in Australia or Germany may be legally purchased in New Zealand in an uncensored format such as "Grand Theft Auto" or "Castle Wolfenstein" (Banned video games, 2011). The laws regarding pornography versus erotica are also different in the various member countries. As described in "making a bitwise copy" above, the decision was made that if anything likely to be deemed illegal in New Zealand was found the police would be contacted, and therefore a specific field "illicit" was not needed and not tracked.

"Total Identifying": This refers to the total number of unique hard drives that had company identifying, individual identifying or both types of data on them. This is an additional field to give clarity not presented by the consortium reporting.

Reporting the results

The first number given is the count of hard drives, and the percentage is that number of hard drives remaining in the sample. For example in Table 1, there are 100 hard drives in total. The 21 unreadable hard drives are then removed leaving 79 readable drives. 34 of those drives are deemed to be scrubbed which gives of 34/79 or 43% of readable drives were scrubbed. The scrubbed drives are then removed from the sample giving the "remaining" value of 45. The result of analysing 100 New Zealand hard drives for data remanence is given in Table 1.

Table 1. Breakdown of New Zealand Results

Total	Unreadable	Scrubbed	Remaining	Company Identifying	Individual Identifying	Formatted	Total Identifying
100	21 21%	34 43%	45	19 42%	19 42%	22 48%	24 53%

Table 2. Breakdown of Identifying Hard Drives

Descriptor of Identifying Data	Occurrence	Percentage
Total Identifying drives	24	100%
Non-formatted Individual only	2	8.3%
Formatted Individual only	3	12.5%
Non-formatted Company only	1	4.2%
Formatted Company only	4	16.7%
Non-formatted Company and Individual	10	41.7%
Formatted Company and Individual	4	16.7%

Table 2 shows the breakdown of identifying information found on the hard drives. Hard drives that are non-formatted are considered to be more concerning than drives that are formatted as less effort is required to find identifying information. Hard drives that contain both Company identifying and Individual identifying are also considered to be more concerning than drives that only contain 1 type of information. From Table 2 the hard drives that are of most concern "Non-formatted Company and Individual" hard drives are also the most common. Of the entire project of 100 hard drives, 10 of those drives could reasonably be put into a new computer, turned on and data that was about a person and a company could be found without additional tools.

Comparing New Zealand 2011 results with Consortium 2009 results

After determining the level of data remanence in New Zealand it makes sense to compare the results with those from the consortium. The consortium was chosen instead of individual member countries as this dealt with individual outliers more effectively and therefore gave a better overall comparison of the results. The results published in 2010 for the 2009 calendar year (Jones et al, 2010) were used for this comparison. The results for New Zealand are from analysing 100 hard drives, whereas the consortium analysed a total of 346 hard drives over 5 nation members.

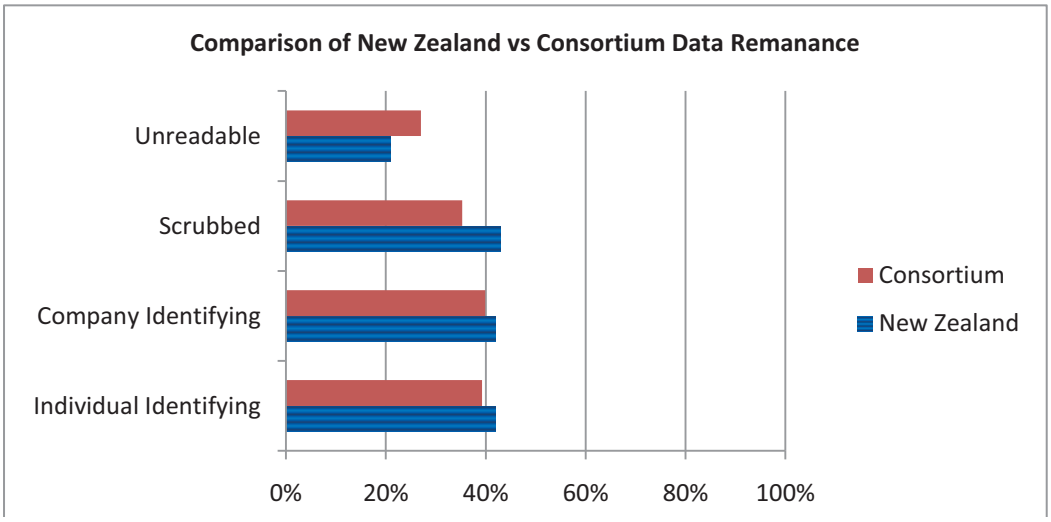

Figure 1. Graph comparing New Zealand and Consortium data remanence

The results in Figure 1 show that the level of data remanence in New Zealand is generally similar to that of the consortium members. In the first instance, a lower value for "unreadable" is desirable as it means more drives are still fit for purpose as working hard drives, and a higher value in scrubbed drives is desirable as it means more drives have had their data fully removed.

Specific hard drives of interest

One hard drive of note was non-formatted and found to have a folder called "Suicide Girls". Whilst that might initially raise concern it should be noted that *"**SuicideGirls** is a website that features softcore pornography and text profiles of goth, punk and indie-styled young women (although styles reminiscent of the 1940s and '50s pin-up models are also incorporated) who are known as the "Suicide Girls"."*(Suicide Girls, 2011) Associated books sold under the Suicide Girls label are legally available in New Zealand. Examining the images using the thumbnail viewer in FTK confirmed that the images were from the "Suicide Girls" website. The researchers did not consider the images to be illegal or illicit and would not have classified them as such.

One hard drive in particular had a directory structure with a number of user names in it. Examining the emails found those user names and then the company domain name, such as johndoe@companynamed.co.nz. Furthermore a webpage found on the hard drive showed that "John Doe" had signed up for a legal "fantasy sports betting" service in New Zealand. The web form on the webpage had kept all the details John Doe provided, such as user name, password (this was obscured with **********), real name, real home address, phone number, cell phone number and email address. The email address matched the one found for him elsewhere on the hard drive, and the home address and phone numbers were consistent with phone numbers for that area. The information was from 2009. Other information on the hard drive such as invoices, spreadsheets and documents made it clear which company it was and other information about the users, including "John Doe" was also available.

Other hard drives of note include four that were from different high schools. Typically those hard drives contained emails from parents to the school explaining why their children were absent (typically for medical reasons). Other data found on the drives included spreadsheets that detailed budgets for hostels or other accommodation for boarders, teacher created assessments and in some cases image files of students attending school activities.

As stated one of the research goals was to determine which industries may have issues with data remanence if possible. Whilst there were 19 instances of company identifying information being found, 4 of those instances were from a very large company in New Zealand and all came from the same source. When the results were re-examined for unique companies, the total count was 14 hard drives, of which 4 were the high schools. The sample size was deemed too small to be able to answer that research question adequately and it has been considered that as sample size of 100 hard drives yielded 14 companies then a sample size of 300 would likely provide a suitable sample size of at least 40.

CONCLUSION

This paper has examined the levels of data remanence in New Zealand, how it compares with other countries and attempted to determine if certain industries are more likely to have issues with data remanence than others.

24 of the 100 New Zealand hard drives analysed had definite identifying information on them of which 10 of those drives could be reasonably put into a computer and that information likely found without use of specialised tools.

When compared with the consortium results, the New Zealand results are consistent with those of the consortium.

When attempting to determine if any industries were likely to have issues with data remanence the sample size of drives with company identifying information was deemed too small at 14 hard drives to be adequate. Given 4 of the 14 hard drives contained information from high schools including budgets, data about students and emails outlining student absence from school (typically for medical reasons) this does highlight an area of further investigation.

REFERENCES

AccessData Group, LLC. (2011). *FTK*. Retrieved 13/10/11 from http://accessdata.com/products/computer-forensics/ftk

Bell, G. B., & Boddington, R. (2010). Solid State Drives: the beginning of the end for current practice in digital forensic recovery? *JDFSL The Journal of Digital Forensics, Security and Law, 5*(3), 1-20. Retrieved 13/10/11 from http://researchrepository.murdoch.edu.au/3714/1/solid_state_drives.pdf

Garfinkle S, & Shelat A. (2003). Remembrance of data passed: A study of disk sanitization practices. *IEEE Security & Privacy, 1*(1), 17-27.

Government Communications Security Bureau. (2011). *New Zealand Information Security Manual*. (v.1.01|June 2011) Wellington, New Zealand: Author. Retrieved 13/10/11 from http://www.gcsb.govt.nz/newsroom/nzism/NZISM_2011_Version_1.01.pdf

Guidance Software, Inc. (2011). *Encase.* Retrieved 13/10/11 from http://www.guidancesoftware.com/forensic.htm

Jones, A., Valli, C., Dardick, G.S., Sutherland, I., Dabibi, G., Davies, G. (2010). The 2009 analysis of information remaining on disks offered for sale on the second hand market. In A. Woodward (Ed.), *Proceedings of the 8th Australian Digital Forensics Conference* (pp. 92-105). Perth, Western Australia: SECAU – Security Research Centre, Edith Cowan University.

List of banned video games. In *Wikipedia, the free encyclopedia*. Retrieved 13/10/11 from http://en.wikipedia.org/w/index.php?title=List_of_banned_video_games&oldid=455126846

Medlin, B., & Cazier, J. (2011). A Study of Hard Drive Forensics on Consumers' PCs: Data Recovery and Exploitation. *Journal of Management Policy and Practice, 12*(1), 27-35.

Suicide girls. In *Wikipedia, the free encyclopedia*. Retrieved 13/10/11 from http://en.wikipedia.org/w/index.php?title=SuicideGirls&oldid=454167575

Tableau, LLC. (2009). *TD1 Forensic Duplicator User Guide.* Waukesha, WI: Author.

Valli, C., & Jones, A. (2005). A UK and Australian Study of Hard Disk Disposal. Paper presented at the 3rd Australian Computer, Information and Network Forensics Conference, Edith Cowan University, Perth, Western Australia. Retrieved 13/10/11 from http://igneous.scis.ecu.edu.au/proceedings/2005/forensics/valli2.pdf