Can compression reduce forensic image time?

Brian Cusack  
Auckland University of Technology, New Zealand  
brian.cusack@aut.ac.nz

Jon Pearse  
Auckland University of Technology, New Zealand  
brian.cusack@aut.ac.nz

Journal of Applied Computing and Information Technology, 15(1). Retrieved June 2, 2015 from  

Abstract

Creating a forensic copy (image) of a hard disk drive is one of the fundamental tasks a computer forensic analyst must perform. Time is often critical, and there is a need to consider a trade-off between a number of factors to achieve best results. This paper reports the results from an exploratory study into the impact of using disk drive compression on the time needed to image (and verify) a hard disk drive. It was found that time reduction may be achieved once the trade-off of contributing variables was properly estimated. The findings led the investigators to suggest a step-by-step decision making process for analysts when considering disk compression as a means for reducing total image processing time.

Keywords

Compression, Time Trade-off, Forensic, Imaging

1. Introduction

How is disk imaging processing time affected when using compression while creating a forensic copy of a computer hard drive? Does compression reduce the image processing time including the time taken to verify the forensic copy?

It is a general accepted within the computing domain, that using compression on a particular computerization process will increase the length of time taken to complete that process. According to (Zirkind, 2007), almost all compression schemes usually involve computational time. If a significant amount of time is required to compress the image data, the reduction in file size should warrant the expense of the computing time. Furthermore, the type of compression algorithm being used for a given process will influence the time needed to complete that process. According to (Yang, Dick, Lekatsas & Chakradhar, 2010), in general, algorithms which achieve higher compression ratios take more time to complete the compression process.

This paper is structured as follows. First, a brief review into past research using
compression with forensic tools is presented. Next, the approach undertaken in this study is specified. This is followed by the results of the research where three variables: amount of data on a drive, data type, and compression level, were tested in eighty one unique tests. A discussion of the research results is then given. Finally, the conclusion identifies areas for further research and discusses the potential trade-offs an analyst can calculate in order to optimize hard drive imaging time.

2. Past Research

Investigation into using compression with forensic tools shows that little research has been conducted on the topic. The literature shows that most people believe that using compression will increase the total processing time. According to (Bunting, 2007), the fastest way to preserve a computer hard drive is not to use compression. Bunting explains that using compression saves disk space (two to three times less space), but costs more in terms of time required to process the compression algorithm (up to five times longer). However, other 'bottle necks' are also noted as factors that slow down the overall image processing time when working with network acquisitions. Faster acquisition speeds can be achieved by using compression if a slow network connection is used. By taking advantage of compression and therefore pushing less data through the network, as opposed to pushing larger amounts of uncompressed data through the network, a faster processing time can be achieved. The slow network connection is the 'bottle neck' in this case. The end result is that less time is taken to complete the overall process. According to Carrier (2005), compression of the evidence file can be useful for transmission to reduce the amount of data sent over a slow network. An example would be when acquiring across a 100MB, or slower, network infrastructure. A decrease in the total acquisition and verification time would be achieved because less data is sent through the slow network, which again acts as the 'bottle neck'.

In an article by Garfinkel, Malan, Dubec, Stevens and Pham (2006), comparison information is given about compression between two forensic imaging tools. One of the tools is EnCase (Guidance Software, 2010), which is commonly used by law enforcement agencies and other professionals. However, the article does not give any figures for the time taken to run the compression tests, details about the parameters of the equipment, or methods used for testing. The article simply points out that one forensic software tool compressed data more than the other.

Online technical readings at the Forensic Focus website are rich with information provided by analysts around the world. A posting by 'jelle' points out two factors that can influence both speed and compression. The first factor is the CPU speed of the imaging workstation which influences compression speeds. The second factor is the type of data on the source drive which can impact the 'compressibility' of the image. A drive which has been wiped with zeroes before installing the operating system (OS) will result in a smaller image than a drive that has been wiped with random data (Encase Compression, 2008).

Another posting by 'cymru100' on the Forensic Focus site says that it really depends on the amount of data on the disk and the type of data on the disk. 'He' has seen completed images the same size as the source drive. This will happen when the source drive has been encrypted which creates a situation where little or no compression occurs. Images of 500 gigabyte (GB) drives can compress down to only several GB in size (Calculate disk size for EnCase or DD image, 2008). The reduction would occur when a drive is mostly empty and the sectors are zeroed out.

It may be concluded that there is some understanding of how compression works and what types of data will compress while using computer forensic tools. Both 'cymru100' and 'jelle' address the 'compressibility' aspect of the data on the drive.

According to Volonino and Anzaldua (2008), network acquisitions are slower than connecting the source directly to the examiner computer via a write blocking device. However, at times the analyst will encounter a server system which has an unknown Redundant Array of Independent Disks (RAID) configuration. In this situation a network acquisition may be the best option. If the drives are removed and acquired separately, the examiner may not be able to recreate the RAID and therefore will
have no accessible evidence.

3. The Approach

The setup of the testing equipment, particularly the setup of how the suspect (source) and destination drives were connected to the analyst's workstation was the same for all testing. USB connectivity for the suspect and destination drives was used because in most live server examinations the analyst would use a USB connected destination drive for the image files as the ways available to extract data from the server may be limited. Normally, when working with a running server there are only two connection options the analyst can extract the data through, USB and network connectivity. The analyst may use a forensic boot CD on the client computer and start a forensic copy. The image files are then saved to an attached USB destination drive. This is also the process used for this study.

The sample data used in this study consisted of 'operating system' and 'program files' data from a Windows XP computer system as this would typically be the type of data which an analyst would encounter in the majority of data collection activities in the field. Ninety forensic copies of a hard drive were created, this included nine pilot test datasets and eighty one unique test datasets (using nine different compression levels and nine different levels of disk drive data load). For consistency the same equipment was used throughout all of the testing. The tests showed that the average compressibility of the data used was around 50%.

All forensic copies were created using FTK Imager by Access Data (AccessData, 2010) and were all bit-for-bit images of the physical disk. A bit-for-bit image of the physical disk contains all logical files (used space), all unallocated data (unreferenced space) and all unused disk space. Evidence, such as deleted files, can reside in the unreferenced and unused disk space which is why the analyst would always create a bit-for-bit image of the physical disk (when possible), as opposed to copying out logical files only. Depending on the situation the analyst may create a bit-for-bit image of a logical volume (partition) which will contain its own unreferenced space. Figure 1 shows a map of a physical disk broken down into its volumes with their associated unreferenced space and the unused space on the disk.

![Figure 1. A physical disk map](image)

If the analyst copied only logical files or created a bit-for-bit image of the two partitions, it would not capture the unused disk space at the end of the disk map. This area may contain evidence. For example, initially this disk drive may have had only one partition using the entire physical space. If data had been allocated in the space at the end of the disk, as shown in Figure 1, some evidence may still reside in this currently unused area. Consequently, bit-for-bit images of the entire physical disk were used in this study.

4. The Results

Initially, a pilot testing phase was run. The pilot included testing with no compression (level 0), mid compression (level 5), and full compression (level 9) settings. The test without compression gave a time of 26 minutes and 16 seconds, this set the maximum time required to acquire the suspect hard drive and verify the forensic image. Any tests which used a level of compression and completed in less time than the above maximum would be of interest for the purposes in this research.

The forensic software was capable of nine different compression levels. All nine levels were used and the drive was filled with nine different datasets, ranging from the disk drive being 10% to 100% filled with data.

Table 1 contains the captured pilot phase data. The results of tests 1, 4 and 7 (no
Tests 2 and 3 (with compression, zeroed data) show that the total processing time is considerably reduced and the size of the image file is significantly smaller as compared to Test 1. These three tests indicate that zeroed data compresses considerably. The output files for Tests 2 and 3 were 18 megabytes (MB) in size. This meant that the computer only had to pass 18 MB of data (image file) through the USB connection to the destination drive when creating the image. In addition, only 18 MB of data from the destination had to be read and passed through the USB connection to the computer to verify the acquisition. This finding clearly shows that the data is decompressed and verified on the computer faster than pulling the entire non-compressed image of 9.54GB's through the USB connection as was done in Test 1. In Test 2 and 3 there is a slight reduction in the image processing time however, the main reduction can be seen in the verify processing time.

Figure 2 shows a chart based on the data in Table 1. Tests 2 and 3 are highlighted by the short vertical time bars for image and verification processing times and the very low data size line of the imaged data.
Comparing Test 1 with Test 9 (Table 2) it can be seen that Test 1’s image is about 55MB larger in size however, the verification process time is shorter. This is because the imaged data are similar in size. Image data with a higher compression level require more computer processing time to decompress and verify the image as opposed to a larger size image with a lower compression level. This is demonstrated by comparing Test 9 (highest compression level) with Test 1 (lowest compression level). The results show a shorter verification time for the larger image. In contrast, the opposite trend can be observed with Test 1 and Test 2 in Table 1 (pilot testing). This is explained by the image data for these two tests being measurably different in size.

![Figure 3. Comparison of the results for Dataset 1 (1 GB)](chart.png)

The remaining test results show characteristics similar to the one exhibited in Figure 3. However, there is a consistent trend showing an increase in the image processing time, particularly from Test 6 to Test 9, in each dataset. Table 3 and Figure 4 show the results generated by testing with Dataset 9 (an almost full disk drive) which demonstrates the differences in process time.

### Table 3. Imaging 9GB of data

<table>
<thead>
<tr>
<th>Test number</th>
<th>Compression Level</th>
<th>Data Size (MB)</th>
<th>Image processing time (min)</th>
<th>Verify processing time (min)</th>
<th>Total processing time (min)</th>
<th>Size of imaged data (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>1</td>
<td>9000.000</td>
<td>15:43</td>
<td>06:39</td>
<td>22:21</td>
<td>4926.627</td>
</tr>
<tr>
<td>Test 2</td>
<td>2</td>
<td>9000.000</td>
<td>17:55</td>
<td>09:39</td>
<td>27:35</td>
<td>4883.565</td>
</tr>
<tr>
<td>Test 3</td>
<td>3</td>
<td>9000.000</td>
<td>18:22</td>
<td>09:31</td>
<td>27:53</td>
<td>4848.657</td>
</tr>
<tr>
<td>Test 4</td>
<td>4</td>
<td>9000.000</td>
<td>20:35</td>
<td>09:31</td>
<td>30:06</td>
<td>4789.994</td>
</tr>
<tr>
<td>Test 5</td>
<td>5</td>
<td>9000.000</td>
<td>20:44</td>
<td>09:20</td>
<td>30:04</td>
<td>4716.499</td>
</tr>
<tr>
<td>Test 6</td>
<td>6</td>
<td>9000.000</td>
<td>23:46</td>
<td>09:32</td>
<td>32:38</td>
<td>4742.122</td>
</tr>
<tr>
<td>Test 7</td>
<td>7</td>
<td>9000.000</td>
<td>25:38</td>
<td>09:09</td>
<td>34:27</td>
<td>4738.847</td>
</tr>
<tr>
<td>Test 8</td>
<td>8</td>
<td>9000.000</td>
<td>30:55</td>
<td>09:15</td>
<td>40:10</td>
<td>4735.289</td>
</tr>
<tr>
<td>Test 9</td>
<td>9</td>
<td>9000.000</td>
<td>36:43</td>
<td>09:22</td>
<td>46:05</td>
<td>4743.508</td>
</tr>
</tbody>
</table>

A finding which stands out in Figure 4 is the difference in the image processing time between Test 1 and Test 9. Test 1 was completed in 16:42 minutes whereas Test 9 took 36:45 minutes to complete. This is well over twice the total processing time of Test 1. In this case, using the highest compression increased the image processing time dramatically. One benefit of using the highest compression is a reduction in the size of the image file. However, in this example the reduction was a mere 192MB or
11.6% between the two tests for the cost of over twice the time to create the image. This is a situation the analyst would want to avoid, particularly when working in professional practice where a trade-off of time and drive space needs to be considered. Since drive space is so much cheaper than time, the analyst should opt for the process which uses the least time. In this case, Test 1 would be the best option.

![Figure 4. Comparison of the results for Dataset 9 (9 MB)](chart_3)

5. Discussion

As shown in Figure 3 all total processing times for dataset 1 were shorter when compression was used as opposed to not being used. In fact the total processing time when compression was not used took 26:16 minutes, so any use of compression which results in a shorter time would be beneficial to the analyst. The total processing time for all tests using Level 1 compression, apart from Dataset 9, are shorter than not using compression. Level 1 compression on Dataset 9 only took five seconds longer than not using compression, an insignificant difference.

Verification time results provided another interesting finding. The verification times were consistent and there was only a maximum of 30 seconds between all verification times in each dataset. This 30 second difference is trivial in relation to the overall image processing time.

Why were the verification times consistent throughout this testing? The size of the imaged data was similar size in each dataset. For example Figure 3 shows a difference of 55MB across all image data sizes and a difference of 12 seconds in verification times across all nine tests for Dataset 1. At the other end of the scale, Figure 4 shows a difference of 192MB between all image data sizes and a difference of 30 seconds in verification time across all nine tests of Dataset 9. The size difference for image data in each dataset is relatively small compared to the total size of the image data for each dataset. This means that the computer will read and pull similar volumes of data from the destination drive through the USB connection while verifying each image in the dataset.

In the environment used to run the tests the decompression processing time is comparatively insignificant when compared to the time needed to read the image data from the destination drive through the 'slow' USB connection. It is because This makes the decompression process time, while verifying, insignificant in the particular testing environment. Therefore, in the example, the verification times are directly related to the data throughput speed of the destination disk drive.

As shown in section 3, the equipment setup for this research was purposely configured with the intention of creating forensic images from online servers and user computer systems via a USB attached destination hard drive. If the suspect and source drives were connected via faster means, such as SATA or USB3, the process timings would be different. The impact of other variables such as these, are opportunities for future research.

6. Conclusion and Recommendations
The results indicate that using compression does reduce the processing time when creating a forensic copy of a hard drive. However, the following three variables need always to be considered: compression level, type of data and amount of data. The research shows that up until a hard drive is 80% full, using level 1 compression will reduce the processing time providing the data on the drive is of similar type to what was used in this research.

If the majority of the data is of non-compressible type, such as MP3 or video files and the drive is filled with this data, then the processing time is unlikely to be shorter when using compression. Two aspects would slow the process down. The first is that almost all of the data from the source drive will need to be pushed across the USB connection to the destination. This is because the data is non-compressible and will have little or no reduction in size. The second aspect is the size of the imaged data that will need to be read back across the USB connection to the computer for verification.

If a drive is barely full with non-compressible data and the remaining sectors on the drive are zeroed, depending on the size of the drive, a significant reduction in processing time can be achieved. As the drive is filled with more data, the processing time will increase. Further research would need to be conducted to identify at which point the processing time would increase to a level where using compression would take more time as opposed to not using compression.

Finally an analyst may use the process outlined below to reduce time when creating a forensic copy of a computer hard drive:

1. Using a software acquisition tool such as FTK Imager, the analyst first previews the suspect hard drive to ascertain how much of the drive is used.
2. Then the analyst establishes whether or not the drive contained zeroed data - to help determine if compression should be used to reduce the processing time.

For example, if an analyst previewed a large hard drive (500GB) and noted that it was only 5% full of data, and the rest of the drive had zeroed sectors, then the analyst could obtain a huge reduction in total processing time by using level 1 compression.

Acknowledgements

The use of the resources at the Auckland University of Technology Digital Forensic Research Laboratories in completing this research is acknowledged.

References


Copyright © 2011 Cusack, B. & Pearse, J.

*Journal of Applied Computing and Information Technology (JACIT): ISSN 2230-4398*

(Incorporating the Bulletin of Applied Computing and Information Technology,