Situational Tool and Method Selection for Digital Forensic Data Collection: Performance Issues

JON PEARSE
CFE, EnCE

a thesis submitted to the graduate faculty of design and creative technologies
AUT University
in partial fulfilment of the
requirements for the degree of
Masters of Forensic Information Technology

School of Computing and Mathematical Sciences

Auckland, New Zealand
2011

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a University or other institution of higher learning, except where due acknowledgement is made in the acknowledgements.

....

Signature

Acknowledgements

The thesis was completed at the Faculty of Design and Creative Technologies in the School of Computing and Mathematical Sciences at Auckland University of Technology in New Zealand. During my research I have received support from many people who have unselfishly given me a great deal of their time. Without their support this thesis would not have been completed in its present form. I would like to take this opportunity to give recognition to those people and to thank them for their support.

I would like to thank my supervisor Dr Brian Cusack for his valued guidance throughout this thesis research. Although Brian supervised a large number of students, he was always able to provide time for discussion sessions at short notice. I appreciate the encouragement, guidance and direction you have provided throughout this thesis.

I would like to express my gratitude to Robbie and Christine Poharama who are my unofficial proof-readers. They have given many hours of their time to proof read many sections of my thesis for proper spelling, grammar and continuity. In addition, thanks goes to Diana Kassabova who conducted the final proofreading for the thesis.

I must mention Peter Mercer from Vound Software. Peter is the developer of great acquisition, analysis and review products such as On Scene Investigator and Intella. Peter actually planted the seed in my head regarding forensic data acquisition techniques several years ago when I was attending a forensic course in Australia. The seed has since grown and developed into this research project.

A great deal of thanks goes to Mark Simms, a Digital Forensic Analyst at the New Zealand Police Electronic Crime Laboratory. For the last two years Mark and I have studied together on the Master of Forensic Information Technology programme. Although Mark is conducting his own thesis, he has always shown an interest in my research and has provided motivational influence and positive discussion which has helped me complete this thesis.

Furthermore, I would also like to acknowledge the support I have received by my employer and the management from the Forensic service line at Deloitte. Deloitte invests large recourses into their employees to continuously grow and improve and have done so throughout my thesis. A special thanks goes to my

work colleague Barry Foster who has been supportive and has provided constructive discussions while I have undertaken this project.

Lastly, I would like to show my gratitude to my wife Clair and my two boys Ryan and Jacob. Clair has been instrumental in looking after the boys, keeping them busy and making sure I’m not disturbed while I have been locked away in the computer room during the two years of this research. The time that I have committed to the research, consisting of full weekends and when I get home after work until late at night, has severely limited my attention to my family. I am truly lucky to have such an understanding family. Now that I have finally concluded this chapter of my life, I will be able to repay the attention that has been absent during this research.

Abstract

Over the last ten years there has been rapid growth in the digital forensics field. Forensically sound computer analysis and testimony is becoming a requirement during investigations related to frauds, missing persons, homicides etc. One of the phases of the digital forensic process is data preservation, where a copy of data from an original electronic storage device is collected in a verifiable manner, producing a forensic copy of the data. A best practice for digital forensics is to capture a bit for bit or *physical* copy of the source device. However, the sizes of hard drive volumes have been increasing exponentially and in 2011, volume sizes for a single hard drive have reached the three terabyte threshold. The increase in volume size equates to an increase in processing time to collect the data and an increase in media capacity to store the acquired data.

The purpose of this research is to explore new tools and methods that will allow an examiner to collect data from a source device in a time-efficient manner. Prior research has been conducted by the author, who concluded that data collection processing times can be reduced by the use of compression algorithms during data collection activities. However, the amount of time reduction depends on the type of data that is resident on the storage device. A reduction in processing time is observed when collecting highly compressible data. Conversely, an increase in processing time can occur when attempting to compress data that does not compress well, during a collection process.

The focus of the research was to develop a means that would be able to determine and report the type of data residing on a storage device. A fast and easy to use scanning tool is developed during the research. The scanning tool is capable of processing a storage device in four minutes and provides a report that accurately details the type of stored data in terms of its compressibility. The information in the report regarding the data's compressibility can assist the examiner when making decisions concerning the use of compression to reduce processing time during data collection activities.

Table of Contents

Chapter 1: Introduction

Chapter 2: Literature Review

Chapter 3: Methodology

Chapter 4: Report of Research Findings

Chapter 5: Discussion

Chapter 6: Conclusion

List of Tables

List of Figures

List of Abbreviations

ATA	Advanced Technology Attachment
ATM	Automated Teller Machine
BIOS	Basic Input Output System
BIT	Built-in Test
CART	Computer Analysis and Response Team
CD	Compact Disk
COEM	Case-Oriented Evidence Mining
CRC	Cyclic Redundancy Check
DAS	Direct Attached Storage
DVD	Digital Video Disk
ED	Electronic Discovery
GB	Gigabytes
Gbits/in^2	Gigabits per square inch
GMR	Giant Magneto Resistive
HEX	Hexadecimal
I/O	Input Output
LE	Law-Enforcement
LMR	Longitudinal Magnetic Recording
LPP	Legal Professional Privilege
MB	Megabyte
MB/s	Megabytes Per Second
Mbit/s	Megabits Per Second
MR	Magneto Resistive
MRx	Magneto Resistive Head
NAS	Network Attached Storage
NIST	National Institute of Standards and Technology
OS	Operating System
PMR	Perpendicular Magnetic Recording
RAID	Redundant Array of Independent Disks
RAM	Random Access Memory
SAN	Storage Area Network

SATA	Serial ATA
SCPS	Scan Cycles Per Second
SMART	Self-Monitoring Analysis and Reporting Technology
SSD	Solid State Drive
SWGDE	Scientific Working Group on Digital Evidence
TB	Terabyte
TCP	Transmission Control Protocol
USB	Universal Serial Bus
VB6.0	Visual Basic Version 6.0
VM	Virtual Machine

Chapter 1

INTRODUCTION

1.0 BACKGROUND

Forensic data collection is one of the most important tasks a forensic examiner must perform. In a Law Enforcement environment, data collection activities are typically triggered by criminal activity. Within the private sector, data collection activities may have been triggered for any one of a number of reasons. However, for both the Law Enforcement and the private sector, the forensic data collection process should produce the same result, which is a verified copy of the source device. If for any reason the acquired data, known as an image, does not verify correctly, the validity of the image may become subject to legal arguments in a court of law. If the image is found to be unreliable, the image, along with any information derived from the image, may be declared inadmissible by the Judge and therefore not able to be used as evidence in court. For many years the examiner has used forensic hardware and software to collect data from the hard drives of computer systems. Forensic hardware typically has built in write blocking functionality whereas forensic software has the ability to activate write blocking functionality via software. The write blocking functionality disables the ability for the source system to be written to. Write blocking methods eliminate any risk of changes being made to the original data on the source device during forensic data collection and therefore provides a forensic process.

The forensic examiner faces challenges with current technology. Over the last ten years, data acquisition transfer rates have not advanced at a rate comparable with the increase of volume sizes of computer hard drives. Although forensic hardware and software collection tools have developed over time, the tools can only acquire data at the same rate as the slowest device in the chain. In many cases, the hard drive controller is where the lowest data transfer rate occurs.

Recently, Seagate released their three terabyte (TB) hard drive to the market. If an examiner were to image the drive using a hardware imaging device which is capable of acquiring data at a rate of five gigabytes (GB) per minute, the time taken to create the image would be in the vicinity of ten hours. However, additional processing time would be also required to verify the image once the acquisition has completed.

Discussions regarding which data to collect and how the data should be collected have been ongoing. According to Spence (2010), increasing digital device capacities are leading to a change in the currently accepted practice of forensically imaging an entire physical storage device to one of imaging only selected data from the device. Some examiners will argue that full forensic copies should always be collected regardless of the size of the data. Collecting a full forensic copy ensures that all possible data is collected. Having a full forensic copy of a system is useful if circumstances change during an investigation and a different avenue of investigation is undertaken. For example, if an investigation was focusing on e-mail activity and for some reason a change in the investigation occurs where the examination of the internet activity is required, the examiner will have access to that data within the forensic image.

On the other hand, other examiners believe that an approach should be used which focuses only on data which is directly relevant to the investigation. One may ask, "Why should a full forensic copy of a 1TB drive be made when the investigation is focused on the creation of fraudulent documents?" Particularly if the 1TB drive is mostly full of movie files. Selective data imaging currently occurs particularly in the corporate sector where often only the data from a custodian who is of interest to an investigation is collected. The advantage is that typically, only the custodian's mail, home drive and shared documents are collected as opposed to creating a full forensic image of the server(s) or storage system. Several advantages and disadvantages exist for both methods of data collection. However, regardless which method is used to collect the data, the integrity of the collected data is of upmost importance.

1.1 PROBLEM AREA

Forensic examiners use several different tools and techniques to create forensic copies of electronic storage devices. The method and tool used to capture the data from a storage device depends greatly on the situation and where the source data resides. For example, if an examiner is required to create a forensic copy of one stand alone computer system that contains an 80GB hard drive, the preferred method would be for the examiner to remove the hard drive from the system and image the source drive with a hardware imaging device. The hardware imaging device provides the highest data transfer rate for data collection and therefore would give the shortest processing time. However, if an examiner is faced with imaging a server system which cannot be turned off, the examiner would most likely opt for an approach where the system is imaged live using a forensic imaging utility. The forensic imaging utility is either run on the system itself or run through the network to which the server system is connected.

Although a full *bit-for-bit* or *physical* forensic copy of a storage device has been the standard practice for numerous years, the substantial development in technology relating to electronic storage media over recent years has started to hinder the examiner when carrying out bit-for-bit data collection activities. Completing a data collection process in a reasonable timeframe is becoming more difficult on large storage volumes. The problem lies with the rate at which data can be transferred out of modern storage devices. While data volume sizes have increased at an astronomical rate, the data transfer rate of the storage devices has not. The result is that full physical data collection processes on large storage devices are taking much longer firstly to create the forensic image and secondly, to verify the image.

1.2 MOTIVATION

Typically a large amount of time is spent on the analysis phase of the computer forensic process. During the analysis phase of an investigation, evidence which supports the case is often discovered. The discovery of relevant evidence is fundamental to winning a case, however, one must understand where the evidence originated from. Before data can be analysed and put before a court, the source of the data needs to be firstly identified as being relevant to an investigation and secondly, preserved to ensure that workable copies of the original data are available for analysis. For example, in a fraud case where evidence supporting the allegation is suspected to be located on an employee's computer system, the identification of the computer system and the preservation of the data that is stored within the computer system are vital. Best practices should be followed to ensure that the integrity of the evidence is maintained. Failure to do so may find the evidence being inadmissible in court, despite how incriminating it may be.

A forensic data collection process has two key functions. The first key function is to create a copy of the data that is stored on the source device. The second key function is to verify that the newly created copy of data is identical to the original data. The data verification process can take up to the same amount of time as the data collection process to complete. Collecting data from large storage volumes is a time-consuming process which needs to be explored. The motivation for conducting this research is to discover innovative methods and/or tools which will assist the examiner to reduce the time taken to forensically acquire data from a source device while continuing to maintain the integrity of the captured data.

The research focuses on developing and testing a data scanning tool that will report the type of data which resides on an electronic storage device. The information gained from using the tool can assist an examiner with data collection activities. The research question is: *"Can a drive scanning tool process a hard drive in a measured period of time and produce a reliable result?"* Section 3.2.2 explains what a measured period of time and a reliable result are.

1.3 RESEARCH FINDINGS

In the course of this research a software tool has been produced that reads a sample of sectors on an electronic storage device and reports what percentage of the read sectors are empty and what percentage of the read sectors contain data. The tool was developed to aid an examiner by providing information about the type of data that resides on a storage device. Knowledge of the type of data on a storage device will allow the examiner to make informed decisions regarding which collection tools and/or methods are best suited when performing data collection activities on that particular storage device. Forty eight tests were carried out over a total of eight different hard drive test subjects. The hard drive test subjects were all operating system (OS) drives where the original data which resides on the test subjects was left intact and was not manually altered in any way.

Several findings were discovered during the research. However, the most astounding finding is that, in almost every test that was performed, the results of a scan which is run for four minutes (standard scan) and the results of a scan which is run for two minutes (two minute scan) produced the same reported result as a full sector scan. A full sector scan is where every sector on the drive is read and processed. Some of the full sector scans for the test subjects, particularly for the larger sized drives, ran for several days. What makes the finding remarkable is that for the 1TB drive which was tested, a sample of only ten megabytes (MB) was collected, yet the reported results were identical to the results from the full sector scan performed on the same drive.

The author hypothesised that the accuracy of a two or four minute scan results would deteriorate as the test subjects capacities increased. The author based the hypothesis on the fact that, as hard drive volume sizes increase, the ratio of sample data which is read versus total data stored on the drive decreases. The author believed that the results would deteriorate due to a relatively smaller sample of data being collected for a larger drive. The author's hypothesis was incorrect. All reported results for the full sector scans and the corresponding two and four minute tests were identical apart from two tests (see Tables 4.4 and 4.5). However, the results of the two tests were within ±1% of their full sector scans and therefore the difference is insignificant.

Chapter 3 outlines the methodology of the research which includes the scanning tool's development. The tool's scanning process was designed to provide high performance. The author believed that an increase in performance would be achieved if only the first eight bytes of each scanned sector were processed as opposed to processing all 512 bytes of data that reside in a sector. Therefore, the tool was designed to read and process only the first eight bytes. However, additional tests which read and process every byte in each scanned sector were also conducted for each corresponding eight byte test. The additional tests were necessary to answer one of the sub-questions. Comparison analysis of the results from the 512 and eight byte tests showed that there were, in some cases, vast differences in the processing times. However, the reported results in regards to the type of data were identical. The time differences were not consistent across all of the test subjects. The author believes the difference in time relates to the type of data which resides on the drive. The author has expressed a theory in Section 5.2.3 as to why the times are different.

1.4 STRUCTURE OF THE THESIS

The thesis consists of six chapters, however, the bulk of the research is detailed in Chapters 2 to 5. Chapter 1 is an introduction to the study. Chapter 1 consists of the background to the research including a discussion regarding forensic data collection methods and tools. The motivation for the research is also discussed. The motivational factors provide the reasons why the research was performed. Chapter 1 concludes with the layout and structure of the thesis.

Chapter 2 provides a review of literature for the topic area of forensic data collection. A focus is placed on the challenges that digital forensic examiners are facing while conducting data collection activities in today's technological age. The challenges include the collection of data from huge storage devices and the limitations of time enforced on the examiner during collection procedures. A selection of innovative ideas regarding increasing data collection performance is discussed. One of the innovative ideas, a data scanning tool, is chosen to form the foundation of the research.

Chapter 3 begins with a review of five previous research papers which are related to software development and testing. Elements of the research question, which is derived from studying the literature discussed in Chapter 2, are defined. The research question is: *"Can a drive scanning tool process a hard drive in a measured period of time and produce a reliable result?"* A measured period of time is defined as a four minute time period and a reliable result is a result which is within ±10% of an accurate result. Two research sub-questions are generated to facilitate further research into two areas. The first sub-question asks: *"Will a two minute scan produce a similar result to a four minute scan?"* The second sub-question asks: *"Will processing all of the bytes in each scanned sector produce a similar result to processing eight bytes in each sector?"* Additional testing was conducted to answer the sub-questions.

Using the information from the literature review in Chapter 2 and the five previous research papers, a methodology for the research is proposed and developed in Chapter 3. The review of the five similar studies and the literature in Chapter 2 assists to identify any potential issues which may occur with the proposed methodology and to ensure the methodology is robust. The methodology includes the design of a software scanning tool which will determine and report which types of data are stored on a source device. A software design methodology is also discussed which includes a software development life cycle detailing the software requirements, testing methods, verification techniques and maintenance of the tool. Chapter 3 also discusses the data collected and placed into the multiple fields in the reports produced by the tool. The chapter closes with the limitations to the proposed tool including the accuracy of the produced results.

Chapter 4 is a report of the testing process and the analysis of the results produced by the tool during the testing phase. An overview of the development of the scanning tool, as set out in Chapter 3, is reported. The overview includes the operational aspects of the tool and the output produced during a test. Alterations to the methodology set out in Chapter 3 are reported. Factors that were not anticipated during the design of the tool, but became evident during the testing phase of the research are reported. The factors have forced alterations to be made to the predefined methodology set out in Chapter 3. Findings from the field testing are analysed and compared to the expected outcomes. The information produced by the tool's function and data verification system, which is built into the tool, is reported and analysed.

Chapter 5 discusses the research findings. The research question and the sub-questions are discussed based on the results from the testing phase in Chapter 4. The hypothesis is tested and is shown to be true for all eight tested systems. A discussion follows regarding the testing process, accuracy of the reported results and issues identified during the research. The data sampling technique used during the testing phase is discussed and the author formulates a theory to explain why the results are extremely accurate. Finally, the results produced by the scanning tool are discussed. The author makes recommendations in regards to forensic data collection techniques based on the reported results.

Chapter 6 concludes the thesis. A summary of the findings and a review of the research question and sub-questions are discussed. Limitations of the research including the scanning tool are discussed and certain constraints are identified. Areas for potential future research involving data collection methods and tools are discussed. References and appendices follow Chapter 6.

Chapter 2

LITERATURE REVIEW

2.0. INTRODUCTION

The use of forensic computing to obtain evidence that substantiates civil or criminal offending has been practiced for over 20 years. According to Cummings (n.d.), the field of computer forensics began in the 1980s, shortly after personal computers became a viable option for consumers. In 1984 an FBI *magnetic media* program was created which was later known as Computer Analysis and Response Team (CART). Although there are several varying descriptions of the forensic process, many of the descriptions have four key phases. According to Kent, Chevalier, Grance and Dang (2006), the forensic process has the following four basic phases: Collection, Examination, Analysis and Reporting.

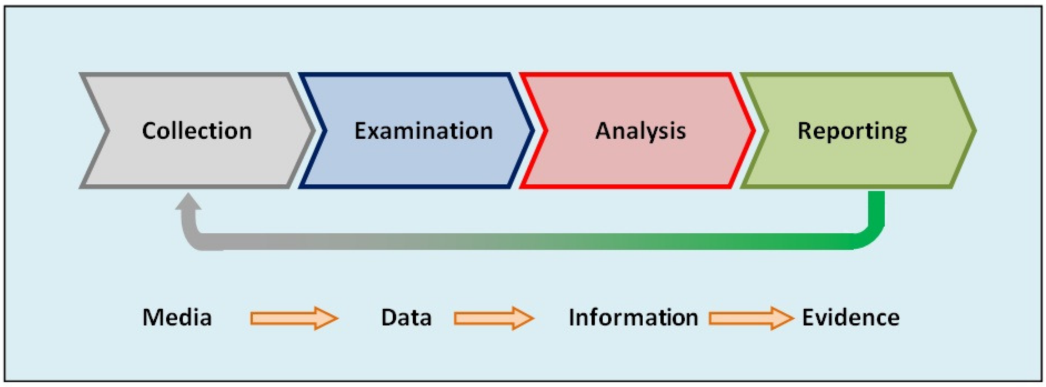


Figure 2.1. The forensic process. Adapted from “Guide to Integrating Forensic Techniques into Incident Response,” by K. Kent, S. Chevalier, T. Grance & H. Dang, 2006, *Recommendations of the National Institute of Standards and Technology,* 800-86, p. 25.

Kent, Chevalier, Grance and Dang (2006) explain that the forensic process transforms media into evidence as shown at the bottom of Figure 2.1. The first transformation occurs when the collected data is examined or when the data is extracted from the media. The process transforms the extracted data into a format that can be processed by forensic tools. Next, the data is transformed into information through analysis. Finally, the information transformation into evidence is similar to transferring knowledge into action by using the information produced by the analysis in one or more ways during the reporting phase. While the four phases work together to make up the complete forensic process necessary for use in a given computer forensic investigation, the research in this thesis is focused on the ‘Collection’ phase of the forensic process.

The collection of computer data is commonly referred to as ‘preserving the data’ or ‘Imaging’. Typically an examiner will preserve the data stored on an electronic storage device by creating a forensic copy of the data. According to Cohen and Schroader (2007), the digital forensic best practice is to obtain a bit-for-bit or *physical* copy of the source device to another media. The forensic copy is then mathematically verified, by a process known as hashing, against the read source data to guarantee that an exact copy of the source data has been captured. According to Bunting (2007a), the original calculated hash value is called the *acquisition hash* and the hash value calculated during verification is called the *verification hash*. To achieve a successful verification, the acquisition hash must match the verification hash. The examiner will use hardware write blocking devices such as the Tableau T35es (Tableau, 2011) or software write blocking technology to ensure that the data stored on the source device is not altered in any way during the data collection process. Chapter 2 summarises research and findings related to the forensic data collection process performed by computer forensic examiners and other researchers.

Chapter 2 contains seven sections. Section 2.1 defines ‘Imaging’ specifically related to computer systems, electronic storage devices and network storage. Section 2.2 reviews the challenges an examiner may face when working in the professional practice. The challenges include collecting data from operating systems that cannot be turned off. Section 2.3 looks at data collection from complex storage systems including encrypted systems. Issues associated with collecting data from large volumes and the risks related to data collection tasks are discussed in section 2.4. Section 2.5 introduces the reader to performance challenges that an examiner may face when conducting data collection activities. Data transfer rates are a critical factor to the overall time required for data collection and methods that may help improve performance and are discussed. In Section 2.6, some innovative methods and tools are considered. These potential innovations can help improve the efficiency of collection processes for a wide range of systems an examiner may encounter. Finally, Section 2.7 addresses issues and problems associated with the forensic process such as privileged information and forensic tools.

2.1. IMAGING

Imaging is the process in which data from an electronic storage device, namely the *source*, is read and a copy of that data is written to separate storage media. According to Sammes and Jenkinson (2007), to obtain an image, it is normal to use specialist software that reads the source disk from its very beginning to its very end and creates an image file. The image file contains all of the data read from the source disk in exactly the same order. The image is then used for the analysis phase of the forensic process. In most cases the image is made up of several large files. For example, EnCase, by Guidance Software creates proprietary image files named E01 files. According to Bunting (2007b), E01 image files contain the imaged data along with case specific information and cyclic redundancy check (CRC) error checking information. Specialised software must be used to mount the image files and view the original file structures within the image. An examiner may also mount the image files onto a virtual computer system that allows the examiner to view the system as the system was viewed by the original user before being imaged.

Hashing algorithms are used to verify that the data read from the source system matches the data written to the destination media. According to Shinder (2002), general procedures that help protect the legal integrity of the evidence include checking the integrity of the image to confirm that the image is an exact duplicate. CRC checking or other programs that use a checksum or hashing algorithm are used to verify that the image is accurate and reliable. The hashing process calculates a hash value over the data as the data is being read during the imaging process. A secondary hash value is generated from the acquired data that is written to the destination drive. The two hash values are compared and if they match, the examiner is assured that there are no differences between the source data and the destination or imaged data.

Section 2.1.1 examines the forensic imaging process for standalone computer systems and various types of removable media. Section 2.1.2 explains imaging methods using hardware and software tools and how an examiner must ensure there are no changes to the original evidence by using write blocking technology. Section 2.1.3 discusses the collection of data from network servers and network devices using specialised tools.

2.1.1. Computer Systems

Computer systems are commonly forensically copied and analysed during investigations. In most cases, the computer or workstation is where the perpetrator committed the crime or carried out the misconduct. Although there are many different types of computer operating systems available, the Microsoft Windows (Microsoft, 2011a) OS is the most commonly used OS worldwide. According to w3schools (OS Platform Statistics, 2011), who have collected statistical data relating to OS usage over the last seven years, the Windows family of OS accounts for almost 90% of all OSs. According to Cardwell et al. (2007), the Windows OS contains a number of locations that can act as a rich source of evidence. An experienced examiner can extract the evidence that may prove or disprove an allegation in an investigation.

While a computer workstation is usually configured with one hard drive, the system may contain several hard drives that require imaging. In most cases, the workstation can be turned off to enable removal of the drive(s) for ease of imaging. Once removed, the hard drive can be connected to a hardware imaging device for imaging or the hard drive can be connected to a hardware write blocking device and imaged with forensic imaging software installed on the examiner's workstation. A hardware imaging device, such as the Tableau TD1 (Tableau, 2003-2011), has write blocking functionality built into the imager. The write blocking functionality ensures that the source drive is not altered during the imaging process. Imaging on a workstation, through a write blocking device and with forensic software is a slower process compared to imaging with a hardware imager. The imaging process is slower because of the added overheads associated with imaging using software on a workstation and using ports with slower throughput rates such as universal serial bus (USB). Overheads such as the workstation's OS and the forensic imaging software can slow down the imaging process significantly. For example, adapting a hard drive's interface connection from serial ATA (SATA) to USB version 2 will reduce the data transfer rate resulting in longer imaging durations.

2.1.2. Electronic Storage Devices

It is not only computer systems or workstations that need to be imaged during investigations. Other electronic storage devices, known as removable media, are frequently encountered onsite or at a scene. These devices include floppy disks, compact disks (CD), digital video disks (DVD), flash drives and a huge number of different types of memory cards that are typically used in digital cameras and cell phones. It's vital that removable media devices are imaged. According to Casey (2008), during an investigation, an examiner may encounter a system that is fully encrypted. The associated keys to decrypt access to the system may be on a removable USB flash drive. Removable media devices have differences to the common hard drive typically located in a computer system. Firstly, apart from floppy disks, all the above mentioned media devices are solid state devices. Unlike hard drives, solid state devices have no moving parts. Secondly, most of the above mentioned media devices require an adapter or external drive unit to gain access to the media and read the data stored on the media. Generally, data on removable media is collected by connecting the media or adapter to a write blocking device. The examiner is then able to connect the write blocking device to a workstation and create an image by using forensic software.

2.1.3. Network Storage

Network storage is quite different to computer system storage and removable media. Network storage is a data storage system that is not physically connected to a user's workstation. It sits on an internal network or on the internet and is usually presented to the user in the form of a mapped network drive. Depending on the type of investigation, the examiner may be required to collect data from a network storage system. Imaging data from network storage is considerably more difficult compared to imaging data from a workstation. Often the examiner does not have physical access to the storage system and some network storage systems have poor data transfer performance that makes the imaging process slow. To a computer user, a network storage location may be represented as a mapped drive, where the user can access and save files to. A mapped drive is commonly known as a logical drive or logical volume. A logical drive may be constructed from many physical hard drives. The physical hard drives may be located in several different geographical locations making a physical imaging process more complicated for the examiner. It may be impossible for the examiner to image all physical hard drives associated with a particular network storage volume when the drives are located in several different geographical locations. In these cases, the examiner will make use of enterprise forensic tools such as Encase Enterprise (Guidance Software Inc, 2011) or F-Response Enterprise Edition (F-Response, 2011). These tools will allow the examiner to access the physical or logical volumes for imaging purposes. Poor data throughput is a limitation of imaging data over a network connection and requires much more processing time compared to imaging when the source drive is directly connected to a hardware imager.

2.2. THE PROFESSIONAL CHALLENGE

Some aspects of forensic data collection in the professional practice are vastly different to those within a Law-Enforcement (LE) environment. According to Burgers (2011), although many of the same tools are used, computer forensic professionals in private practice and those in law enforcement are held to different standards. They have access to different resources and their work results in substantially different outcomes between the criminal and civil cases to which they contribute. According to Crowley (2007), Law Enforcement forensics focuses on gathering evidence for use in a criminal court whilst corporate forensics focuses on maintaining enterprise operations. Enterprise operations include monitoring digital assets to provide evidence that employee actions are in accordance with acceptable use policies.

In an LE environment, computers or electronic storage devices are usually seized under a search warrant and the imaging of those devices is performed in a forensic laboratory. However, in the professional practice, the imaging of computer systems and data storage devices is performed while on site more often than in the security of a forensic laboratory.

Section 2.2.1 reviews on-site data collection and the issues related to collecting data from large data storage volumes. On-site collection involves the initial collection, verification of the data and creating a second copy of the data.

These processes can take considerable time, which is discussed in section 2.2.2. Section 2.2.3 reviews several imaging tools that an examiner may use for data collection activities. The types of tools include hardware imagers, software programs and bootable CD's.

2.2.1. On-Site Collection

On site data collection can be more difficult to perform as opposed to collecting data from a seized standalone computer or storage device in a forensic laboratory. When on site, the examiner will generally need to process server systems, workstations, removable media and on occasion items such as printers, fax machines and photocopiers which all can store electronic information. Collecting data from removable media and workstations is generally straightforward and, in most cases, causes little disruption to the running of the business. For example, the examiner can ask an employee to logon to and use another workstation while the examiner images the employee's workstation.

It's common for an examiner to collect data from running server systems while on site. Data collection from server systems can be a challenging task especially when the server possesses large data storage volumes. According to Adelstein (2006), as systems continue to increase in size, terabytes of disk data are no longer uncommon and imaging can take many hours. Imaging is extremely difficult on network attached storage (NAS), storage area network (SAN), and large redundant array of independent disks (RAID) systems. Often the examiner does not have the option to shut down the system and remove the hard drives for faster imaging on dedicated hardware imaging devices. According to Casey (2004), the examiner faces the risk of being sued (for loss of business) if he/she shuts down and seizes all computer equipment when the files of interest are allocated files and can be extracted with minimal disruption to the business. The risk would also apply if the examiner shuts down and seizes a system when the examiner possesses the tools that have the ability to create an image of a complete running server system with little disruption to business.

2.2.2. Time Constraints

One of the main problems with on-site data collection is the time required to complete all data collection and verification processes. The time required is governed by the speed at which the source data is accessed and the amount of data that needs to be processed. The three processes that have the highest time cost are:

- Extracting the data from the source device (initial acquisition).
- Verifying the acquired data.
- Creating a second verified copy of the acquired data.

Forensic best practices advise that a second verified copy of the imaged data is created so the examiner has two copies of the original data. The original data is what is referred to as the *best evidence*. The first copy is known as the *master copy* and the second copy is known as the *working copy*. According to Carroll, Brannon, Song and Schwarz (2008), there is one cardinal rule for electronic evidence and that is to always work on a copy. Original evidence or the best evidence should be duplicated and kept safe. Only working copies of evidence should be used for review and analysis. Workstations and removable media take the least time and are the least difficult storage devices to image compared to live servers and network storage. Workstations can be imaged in less time because they can be turned off allowing the hard drive to be removed and directly connected to a hardware imaging device for better performance.

On the other hand, server systems can prove to be more difficult for the examiner to process. As mentioned in section 2.2.1, in most cases the examiner does not have the luxury of shutting down a server system to allow for a faster imaging process on a dedicated hardware imaging device because of business requirements. If the server must remain running to fulfil normal business activities, the examiner must image the system live. According to Casey (2008), a live forensic image can be acquired either from the console using tools such as FTK Imager Lite running from removable media or remotely using network tools such as EnCase Enterprise or ProDiscover IR. Data collection from live servers will lengthen the processing time dramatically. The performance issue arises because of poor data transfer rates. In order to extract the data from a running system, the examiner must use the system's own USB or network connection ports for data transfer. Data transfer rates via a USB or network connection are many times slower than that of a current hard drive's SATA interface, which one would use when connecting the drive to a hardware imager.

Another issue encountered by the examiner is the ever increasing size of hard drive data volumes. Not only has the data size of a single hard drive broken through the 3TB threshold, but these drives can also be configured in such a way that several drives are represented as one extremely large data volume. Collecting data from such a large volume over a relatively slow medium, such as a network connection, could take days or weeks to complete. Further issues relating to system volumes and data transfer rates are detailed in sections 2.4 and 2.5 respectively.

2.2.3. Tools Used for Data Collection

The forensic tools used to image storage media come in many forms. However, imaging tools can be split into two main categories. The first category includes hardware imaging devices. The second category includes software imaging applications which are executed on the source system or the examiner's computer system.

Hardware imaging devices are physically small box-type devices with hard drive connection interfaces. The device usually has a *read only* connection interface for the source device. The read only connection ensures that data on the source device cannot be written to and changed by the imaging process. The device also has a *read/write* interface for the destination storage drive. The read/write interface enables the data from the source device to be written and stored to the destination drive. Some commonly used hardware imaging devices are the TD1 from Tableau (Tableau, 2003-2011), the Voom iii from Voom Technologies Inc. (Voom Technologies Inc., 2011) and the Disk Jockey Pro from Diskology (Diskology, 2010). These hardware imagers possess interfaces for fast throughput of data. Therefore, these devices will image a source drive in less time compared to connecting the source drive to a workstation and using forensic software to create an image. A drawback with hardware imaging devices is that the source drive has to be connected directly to the imager. Hardware imaging devices cannot be used for imaging live systems or live network storage.

Software imaging applications such as FTK Imager (AccessData, 2010) and EnCase (Guidance Software Inc, 2011) are commonly used by examiners in the field. The software is typically installed on the examiner's forensic workstation. Some imaging applications are compiled onto a forensic boot CD.

Forensic boot CDs such as Paladin (Sumuri-Forensics Simplified, 2011) or Helix 3 (e-fense, 2011) enable the examiner to boot the source system into a Linux OS. The source system's hard drive is set to *read only* mode to prevent changes and a software imaging tool is used to collect data from the system drive.

The examiner may choose to use a software imaging tool over a hardware imaging tool for imaging. One such situation is when data from a live server system is required for collection. From a CD or USB flash drive, the examiner can run a software imaging tool, such as FTK Imager, on the server system itself. The imaged data is usually sent to a destination hard drive connected to a USB port on the server system. A disadvantage of connecting a destination drive or USB flash drive to a server is that changes are made to data on the server system. According to Carvey and Altheide (2005), when storage devices are connected to USB ports of a Windows system, software on the system is run. Files and registry data are accessed and modified during the detection and mounting processes of the USB device. Small changes to files and registry data can be acceptable providing the process is well documented.

Network forensic tools are also used by practitioners to collect data from live server systems. According to Casey (2008), an image of a live system can be acquired remotely using tools such as EnCase Enterprise or ProDiscover IR. These types of tools have the capability to create a write blocked connection to the source hard drive(s) of the system. A write blocked connection will eliminate any changes being made to the source system during the imaging process. Another network forensic tool is F-Response Enterprise Edition (F-Response, 2011). F-Response is not an acquisition or analysis tool but rather a tool that creates a write blocked network connection to the drive of the source system. An advantage of using F-Response is that once the write blocked connection has been established, the examiner can use any forensic imaging tool for the data collection process.

2.3. SOURCE SYSTEM COMPLEXITY

The examiner may not know what to expect in terms of information technology infrastructure before arriving at a site. The examiner may encounter any of the numerous types of system infrastructure configurations in use today or a configuration that may be out of date. The examiner is at the mercy of the organisation's infrastructure planning, implementation and information technology management personnel. The examiner should be prepared and have the necessary equipment to access those systems. According to Kent, Chevalier, Grance, and Dang (2006), the examiner should receive training and education on forensic related policies, guidelines and procedures. Regular training will ensure the examiner has the knowledge to deal with a wide range of system infrastructures.

The following section reviews the complexity of current system configurations and how the trend is now moving to virtual server infrastructures. Section 2.3.2 looks at storage systems and how they have evolved from storage disks in a computer to large network SAN and NAS systems. Section 2.3.3 discusses the complexities surrounding encryption and the collection of memory from running systems that may contain the keys to decode encrypted data.

2.3.1. System Configurations

This section addresses information system configurations with an emphasis on server configurations. Server system configurations come in many different forms and are processed by the examiner based on their configuration and complexity. In the past, organisations implemented individual server boxes for each business role. For example, an organisation may have separate servers dedicated for file storage, printing services, e-mail and accounting packages. According to Jungl et al. (2009), the one application per one server configuration is still largely common practice. Typically each server system will have its own data storage system that is generally independent form the other server systems. Although data storage on these systems is usually internal, in cases when storage space becomes low, additional storage is usually added to the system by connecting an external USB hard drive.

Due to advancements in technology and the increase in speed at which computer components now operate, there has been an increase in virtualised server configurations. A commonly used virtualised server system is VMware's ESX server, (VMware, 2011). According to Ahmad, Anderson, Holler, Kambo and Makhija (2003), VMware ESX server is a software platform that enables multiple virtual machines to share the hardware resources of a single physical server. Each virtual machine (VM) or virtual server is securely isolated from all other VMs and is given the illusion that it is running directly on dedicated server hardware. An advantage of a virtual system, from a data collection perspective, is that typically several virtualised servers are located logically on the same physical computer system or *host* as large files. Therefore, the examiner can focus imaging tasks on the host systems that are running the virtual servers that are relevant to an investigation. However, a disadvantage with virtual systems is that the physical hardware of the host cannot be shut down without causing disruption to all operating virtual servers. According to Nance, Hay and Bishop (2009), a VM presents the appearance of hardware to those processes or virtual servers that run on it. Virtual servers rely on the VM hardware or host being up and running to operate. Therefore, the host cannot be shut down without causing disruption to the system and potentially to the business. In contrast, single role stand alone server systems may be able to be shut down without disruption to the business that will allow the hard drive to be removed for faster imaging.

2.3.2. Storage Systems

Internal and add-on data storage systems were briefly mentioned in Section 2.3.1. They are commonly used for standalone server systems. The storage system that is accessed by a VM is different to the storage system of a standalone server. Virtual servers typically make use of either NAS or SAN configurations. These types of storage configurations can contain hundreds of hard drives, spread across several physical locations that would make it almost impossible for an examiner to image each drive. The nature of a network storage system adds additional complexities to data collection activities for the examiner. According to Casey (2003), networks present investigators with a number of challenges. Evidence can be distributed on many computer systems when networks are involved. Collection of all storage from such a network can be impractical. In addition, some evidence on a network can be volatile and may be available for only a short period of time. The window of opportunity for collecting such volatile evidence can be very small.

2.3.3. Encrypted Systems and Access Issues

Systems that have encrypted data should be approached with caution. Valuable data can be stored in encrypted containers. According to Hargreaves and Chivers

(2008), the decrypted contents are no longer accessible if the power is removed. With encryption products bundled into operating systems, access to free encryption tools on the internet and encryption technologies becoming more popular, people are using these encryption technologies more often. According to Jo and Hong (2000), people are using anti-forensic methods to obstruct computer forensic examinations. Typically they are using encryption technologies such as BitLocker Drive Encryption (BitLocker, 2011) to encrypt data stored on disks.

Vital data relating to the recovery of encryption passwords can be stored in the memory of a running computer system. Numerous books and writings warn that the potentially valuable data stored in a computer system's memory will be lost when power to the running system is abruptly cut. According to Sutherland, Evans, Tryfonas and Blyth (2008), system memory can store valuable information such as passwords used for encryption and evidence of network intrusion or memory resident root kits. These snippets of evidence would not be available to the examiner on a post mortem, computer turned off, examination. A disadvantage surrounding the capture of the data is that data on the system may be changed when imaging the memory of a live system. According to Bunting and Anson (2007a), it is important to capture the volatile memory for live incident response. The fact is that while capturing the evidence, the examiner will have to access the system which may make changes to data. The changes may be to date and time stamps of system files. However, in these situations disconnecting the power to shut down the system may actually lose more evidence than it preserves (Bunting & Anson, 2007b).

Some hard drives require a password to gain access to the data stored on the drive. The password is setup in the basic input/output system (BIOS) on the computer system. The BIOS of the computer interacts with the firmware of the hard drive and sets the same password in the firmware. According to Xu, He, Zhang and Zhang (2007), the ATA8 protocol defines an optional security feature set. The feature is a password system that restricts access to user data stored on a device. The *Security Set Password* command transfers 512 bytes of data (containing the user's password) from the computer to the firmware of the drive. Setting the password causes the system to prompt for a password on start-up and makes the data on the drive inaccessible if an incorrect password is entered.

Hard disk password locks can create problems for the examiner. Because the password is unknown, the examiner will not be able to gain access to the drive and therefore will not be able to image the data stored on the drive. In some cases the password can be recovered from the BIOS chip on the computer system that created the password. The recovered password can then be used to gain access to the hard drive. However, recovery of a password may be more difficult. For example, the password may not be recovered from the BIOS of the computer or the BIOS password may have been changed and the original password may be unknown. In these situations the examiner will not be able to access the data stored on the drive and may be forced to seek advice from a professional in the area or the hard drive manufacturer.

2.4. SYSTEM VOLUMES

There can be some confusion around the term *volume* in relation to computer and network storage systems. There are two distinctly different meanings for the term volume when data collection is concerned. The first is a *physical volume*, which is commonly referred to as a *physical drive* by forensic examiners. A layperson may understand a physical drive as being the physical hard disk which is installed in a computer system. However, when forensic examiners refer to physical volumes, drives or disks, they refer to the data area of an electronic storage device. The data area begins from the first available sector and continues to the last sector. According to Savoldi and Gubian (2008), a physical data collection is a full copy of the observable content of a storage device, which is all available sectors. In a Windows system, a physical volume is referred to as disk *X*, where *X* represents a number starting from 0 (Microsoft, 2011b).

The second type of volume is known as a *logical volume*. A logical volume is an area or range of sectors within a physical volume. A logical volume can be compiled from several physical volumes spread across several systems. Logical volumes are administered by volume managers within the OS of the system. For example, on a Windows system, the volume manager accesses the physical volume and displays the predefined logical volumes to the user. Logical volumes are typically shown as a drive letter, such as C:\, D:\, E:\ and so on, in a Windows environment. According to Kim, Kim and Shin (2001), logical volume managers have long been key components of storage systems. Their key functions are creation of logical or virtual views of physical storage devices. Volume managers make it possible to overcome the limits to capacity, availability and performance of a physical storage device.

Figure 2.2 illustrates how physical volumes and logical volumes interact within a storage device. The physical volume takes up the entire area of the storage device, whereas the two logical volumes are smaller sections within the physical volume. In addition, Figure 2.2 shows an unused area at the end of the physical volume known as unallocated space. Evidence can reside in the unused area and would be missed if an examiner only collected the data from the two logical volumes, referred to as the C:\ and D:\ drives.

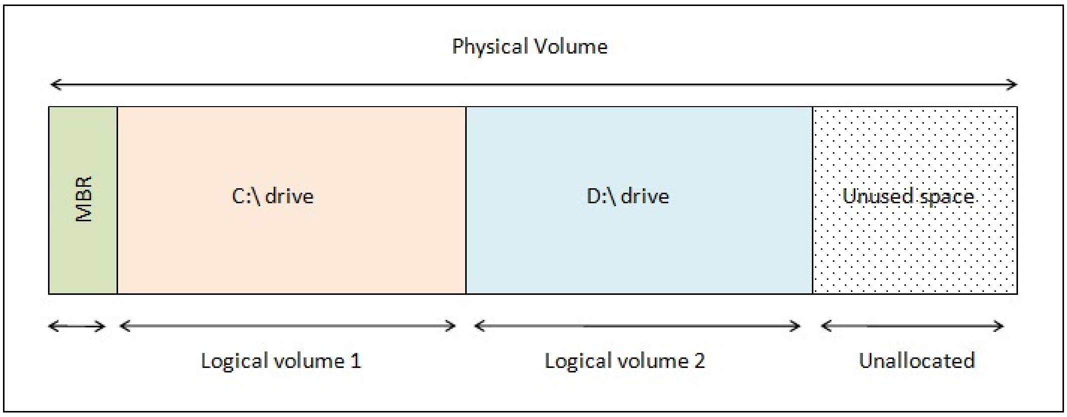


Figure 2.2. An example of a physical and logical volume layout

Logical volumes are not always located within one physical volume. As mentioned in this section, a logical volume can be created from several physical volumes. For example, an organisation may be running low on storage space on a particular system. To remedy the storage space shortage promptly, the organisation may add additional hard drives to the system and merge the space from the new drives with the space from the existing drives to create a larger volume. Merging space together from several hard drives is known as spanning (Using Spanned Volumes, 2011). To the end user it appears that their logical drive has increased in size, however, behind the scenes, the user’s data is now spanned across several physical volumes.

Section 2.4.1 discusses hard drive volume sizes and how the physical aspects of a hard drive have had little change over time, although the data storage area is able to hold many more times the amount of data held by its predecessors.

Volume and storage configurations, particularly network storage, are described in section 2.4.2. Section 2.4.3 looks at the risks associated with data collection and focuses on the risk to business systems if affected by imaging processes. Risks that involve examiners missing relevant evidence when using selective data collection methods are also discussed.

2.4.1. Hard Drives and Logical Sizes

In the 1990's hard drive capacities were breaking through the 500MB threshold. Today hard drives can hold up to 3TB of data, approximately 6,000 times more capacity than a 500MB drive. The increased capacity comes from a number of developments and technology changes over the years. The first was the introduction of the magneto resistive (MR) head by IBM in 1991 (Belleson & Grochowski, 1998). The MR head enabled the development of hard drives with a capacity of 1GB which provided the highest *areal density*, gigabits per square inch (Gbits/in²), available at the time. As technology developed, the introduction of the giant magneto resistive (GMR) head in 1997 aided the production of 16.8GB hard drives for the high performance desktop market. According to Belleson and Grochowski, the most significant advancements in technology have been the MR head, the extended magneto resistive (MRx) head and the follow-on GMR head.

Another advance in technology, that practically doubled the storage capacities of hard drives overnight, was the change from conventional longitudinal magnetic recording (LMR) to perpendicular magnetic recording (PMR). According to Tanaka (2008), PMR hard disk technology offers very high recording density, wide thermal stability, and insensitivity to external fields.

Manufacturers have also made the transition from the older aluminium disk platters to platters made from glass and ceramic substrate. Glass platters can be made smoother and flatter than aluminium platters. They are also much more rigid which reduces noise and vibration and enables the platter to be spun at higher speeds promoting better performance. According to Kozierok (2004), glass platters expand much less than aluminium platters when heated. The reduction in platter expantion gives the glass platter better thermal stability and improved performance.

With the introduction of the GMR head, perpendicular magnetic recording, glass platters and other technological improvements over the years, hard drives have reached capacities of 3TB. These large data volumes can cause issues for the examiner when it comes to the imaging process. Conducting a full forensic copy of a large hard drive is time consuming. In a basic comparison, imaging a 3TB drive could take thirty times longer than imaging a 100GB drive. The vast increase in processing time that is required to obtain a full forensic copy of a large drive has forced the forensic computing industry to re-think the way it collects data. According to Valli (2010), it is time consuming for the forensic examiner to create full forensic copies of large hard drives. Valli suggests than instead of collecting the entire physical drive, the examiner should use an approach where only the data that is relevant to the case or investigation is collected.

2.4.2. Volumes and Storage Configurations

Stand alone computer systems are considerably easier to image as opposed to live server systems. They are easier to image mainly because of their simplistic storage systems. Typically a standard computer system would enclose one physical hard drive. The hard drive may contain several logical volumes but when the physical drive is imaged, all of the logical volumes will be included in the image. Another aspect that makes these systems easier to image is the fact that most systems can be turned off. Once turned off, the examiner can remove the hard drive and connect it directly to a hardware imaging device for fast data collection.

Online systems are more difficult to image, largely for two reasons. Firstly, when a system is online and cannot be turned off, the examiner cannot simply remove the system hard drive(s) and use fast hardware imaging techniques. In these situations the examiner will have to perform *live data collection*. More will be discussed on live data collection in Section 2.5. Secondly, many modern online server systems don't have built in data storage like the earlier standalone server systems. Current server systems have a boot drive that contains little more than the OS and configuration settings. The main data storage for these types of configurations is usually in the form of an externally connected NAS or SAN storage system. NAS and SAN systems were mentioned in Section 2.2.1. These storage systems, particularly SAN systems, have become more popular over recent years with business requirements demanding high performance and large scale storage solutions. According to Sindi, Liu and Al-Shaikh (2011), SAN solutions are becoming very common especially in environments requiring high *input output* (I/O) bandwidth. Compared to direct attached storage (DAS), SAN solutions provide a major improvement in storage performance when shared among many clients and also offers superior scalability, availability, and flexibility.

A SAN is a centralised data storage system with inbuilt redundancy for data reliability. According to Milanovic and Petrovic (2001), SAN is a separate, centrally managed networked environment. It provides a scalable, reliable IT infrastructure to meet the high-availability and high-performance requirements of today's most demanding e-business applications. For the examiner, high-availability and high-performance means that other business systems are likely to rely on the SAN system and data collection from the SAN system will have to be performed live. The sheer size of SAN storage systems and the amount of time that will be required for the collection process could make it impractical for the examiner to collect all of the data retained on the system. Instead of imaging the entire storage system, it may be more practical for the examiner to image a predetermined selection of files relevant to an investigation (Valli, 2010).

2.4.3. Data Collection Risks

Risks associated with data collection fall into two areas: firstly, the risk that the company's business systems are shut down or fail due to the collection process and secondly, the risk that the examiner does not collect all relevant evidence. According to Mandia, Prosise and Pepe (2003), live response is when a live system is being examined, usually following an actual or attempted attack on the system. Evidence may still be resident on the system so data collection has to be performed while the system is live. There is a risk that a collection process may consume a considerable amount of the available network bandwidth and leave the business with degraded or no access to networked data resources for an extended period of time. Kenneally and Brown (2005) point out the problems associated with traditional bit-stream imaging methodologies, particularly with volumes of data over 200GB. They also emphasize the risks to business that are due to poor or interrupted network performance caused by imaging large data volumes over a network. The examiner may choose to collect only relevant data in these situations to avoid liability for a system outage that inhibits business operations.

Because storage volumes have grown so large that full data collection from such volumes is becoming impractical, the trend is to collect only data that is relevant to a given investigation. According to Zhang and Wang (2009), many crime investigation models and processes for data collection have been presented. There are situations where examiners, when faced with large numbers of seized servers and laptops, are overwhelmed due to the lack of an obvious start point from which to begin data collection and analysis processes. Zhang and Wang propose a new forensic investigation model named Case-Oriented Evidence Mining model (COEM) where the examiner has a thorough understanding of the case and is able to extract only the significant data from the computer systems. The COEM collection method has risks associated with it. For example, the examiner may miss relevant evidence when reviewing data and as a consequence, fail to collect all relevant data. On the other hand, if all logical volumes or physical volumes are imaged, then the examiner is assured that all the data on the system will be available for analysis if required at a later date.

An image of a physical volume consists of reading all of the sectors on the physical device. An examiner will, almost always, seek to collect a physical image of a hard drive (Bunting, 2007b). However, if only data that the examiner considers relevant to the investigation is collected, then the examiner will be restricted to analysing only that particular data set. For example, if initially only relevant data is collected and there is a change in the investigation direction that requires analysis of additional data, the examiner will not have access to the additional data.

The collection of only relevant data when processing large datasets is currently occurring. When e-discovery activities are in progress, the examiner may collect only specific types of files. For example, the examiner may collect only e-mail or spreadsheets or a mixture of different file types. The selective data collection methodology, if used in computer forensics, may provide a partial solution to the issues associated with data collection activities from large data volumes. However, the selective data collection method will only collect logical files from the allocated area of the drive. Logical files are files that are intact and readily available. Deleted files do not reside in the allocated area of the drive and will not be collected using a selective collection method.

2.5. PERFORMANCE CHALLENGES

Performance is a major factor in the forensic data collection process. Poor performance increases processing time which in turn increases costs, particularly in the professional practice. There are many factors which will have an influence on imaging speeds and times related to the data collection process. As mentioned in section 2.1.1, using a hardware imaging device to image a hard drive removed from its system, is one of the fastest methods of imaging. Data throughput performance in this case is at its maximum because the drive is connected to the imager directly to its data transfer interface. The limiting factor is the transfer speed of the drive's data interface. However, when imaging a live server or storage system, the examiner does not have direct access to the hard drive's data interface and imaging is performed through a secondary interface such as USB, firewire or ethernet connection. The data transfer rates for these types of connections are much slower than a direct connection to a hardware imaging device. The slower data transfer rate causes an increase in the image processing time.

Section 2.5.1 examines data collection from live server systems. Data throughput and processing time issues accompanied with imaging large amounts of data through slow ports are discussed in section 2.5.2. Section 2.5.3 reviews research related to the use of compression for reducing image processing time.

2.5.1. Collection From a 'Live' System

Data collection from live systems takes significantly longer as opposed to data collection from a standalone system. The increase in imaging time is primarily because the examiner cannot connect directly to the data interface of the operating hard drive and is forced to use a much slower interface or port for data transfer during imaging. In a situation involving an operating server system, connection to a USB or a network port is the preferred method used by examiners to access and extract data from an operating server system.

2.5.2. Data Transfer Rates

USB is one of the most commonly used methods to connect to a computer system. According to McFarland (2007), USB was the most common serial peripheral bus in existence in 2007. It allowed all the most common devices to connect to the computer and to each other through hubs. McFarland states that even wireless USB has become the dominant method of low bandwidth communications between devices and their peripherals. The USB 1.0 specification was introduced in 1996 (USB specification, 2011). It has a data transfer rate of 12Mbit/s. Server systems older than ten years may have USB 1.0 architecture. The USB 2.0 specification was released in April 2000 and has a theoretical data transfer rate of 480Mbit/s. However, according to McFarland, one will never get 480Mbit/s transfers in the real world, only around 240 to 360Mbit/s. USB 3.0 has recently been released onto the market place. While boasting a data transfer rate of five gigabits per second (5Gbit/s), it would most likely be years before the examiner would regularly encounter USB 3.0 hardware in server systems.

Ethernet has been around a lot longer than USB. Ethernet was developed in the mid seventies and was capable of data transfer rates of up to 10Mbit/s. Over time and with continued development, data transfer rates transitioned from the initial 10Mbit/s through to a current rate of 10Gbit/s. However, the examiner currently does not have the luxury of running the imaging process at 10Gbit/s speeds because of three factors. Firstly, 10Gbit/s ethernet hardware is currently expensive and the majority of organisations do not implement it within their network infrastructure. Therefore the examiner will mostly encounter the more commonly used 1Gbit/s network infrastructures. Secondly, a laptop is typically the examiner's forensic workstation and 10Gbit/s ethernet is nonexistent on laptop systems at the moment. Thirdly, hard drives do not operate at 10Gbit/s speed. Current SATA hard drives operate at 6Gbit/s. The best transfer rate that could be achieved is 6Gbit/s minus various data transfer overheads. In contrast, a worst case scenario is where the examiner is required to image a server with a large data volume that cannot be turned off, has 100 or 10Mbit/s ethernet capability and has only USB 1.0 or 1.1 architecture. In this situation the imaging process will be extremely time-consuming because of the considerably low data transfer rates of the slow USB or ethernet ports.

2.5.3. The Use of Compression to Reduce Time

According to Christner and Grevers (2008), data compression is defined as the process of encoding data using fewer bits than an un-encoded representation would use. The encoded data takes up less storage space and uses less bandwidth for transmission. Data compression is typically achieved by the use of encoding techniques known as compression algorithms. There are two types of compression techniques, *lossless* and *lossy*. With a *lossless* compression algorithm, no data is lost during the compression and decompression process. For example, an e-mail message is compressed with a lossless algorithm. The reason why e-mail uses lossless algorithms is because it is important for the recipient to receive the message exactly how it was before it was compressed. According to Mehboob, Khan and Ahmed (2006), a lossless data compression system assures that the data at the decoder output will be exactly identical to the data at the encoder input.

A by-product of a *Lossy* compression algorithm is that some of the original data is lost when data is compressed. Therefore, when decompressed, the data is not the same as the original. Video and audio data is usually compressed with a lossy compression algorithm. Lossy algorithms generally achieve better compression ratios, which can dramatically reduce the output file size. However, when the video or audio file is played back, it can be difficult for the human eyes and/or ears to detect the loss in quality. The two most widely used forensic imaging tools, EnCase and FTK Imager, have compression capability built into the software. A forensic image of a storage device is compressed using a lossless compression algorithm to ensure that none of the original data is lost.

Data compression can be useful for minimising storage and increasing data transfer across slow transmission media. However, using compression algorithms to compress data comes at a cost in the form of time. There are certain situations, particularly when imaging live systems, where the use of a compression algorithm while imaging may reduce the total processing time (Pearse, 2010). Hardware imaging devices cannot be used when a live system cannot be shut down. In most cases, the examiner must image the system through a relatively slow USB or network connection in order to capture the data. A reduction in image processing time can be gained provided there is abundant highly compressible data, such as zero byte sectors, and the compression algorithm can process the data at a faster rate than the transfer rate of data through a network or USB connection. The reduction in processing time comes from the speed of the compression algorithm to reduce the data size. If the speed of the compression algorithm is higher than the data transfer rate then less data is transferred through the slower network or USB connection.

Using compression to firstly reduce the size of data is a faster process as opposed to pushing larger amounts of uncompressed data through the same network or USB connection, which incidentally is where the data flow restriction lies in this case. The end result is that the complete compressed image is collected in less time than an uncompressed image. 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 data across a 100Mbit/s network infrastructure. A decrease in the acquisition time would be achieved because overall less data is transferred across the network. However, in other situations where an examiner has the luxury of shutting down a system and connecting the hard drive directly to a hardware imaging device, the data transfer rate between the drive and imager will be at maximum. Therefore, imaging using compression in this case would almost certainly increase the process time. Bunting (2007a) makes comprehensive mention of compression usage. Bunting explains that using compression saves disk space as it uses two to three times less space. However, it costs more time to process the compression algorithm, up to five times as long. Bunting's assertion of compression suggests that using compression will increase processing time.

Research into the effect of compression while imaging data has been conducted. According to Cusack and Pearse (2011), it is shown that using compression does reduce the processing time when creating a forensic copy of a hard drive in certain circumstances. The testing conducted by Cusack and Pearse involved imaging a hard drive, through a Tableau write blocking device, across a USB2 connection. According to Cusack and Pearse, the amount of time reduction is subject to three variables which are:

- The compression level used;
- The total amount of data on the drive (includes allocated & deleted data);
- The compressibility of the data on the drive.

The research by Cusack and Pearse showed that up until a hard drive is 80% full, using FTK Imager with a compression level of 1 would achieve a reduction in processing time. Cusack and Pearse used test data consisting of operating system and program files from a Windows XP computer system. The test data is a typical workstation dataset and according to Cusack and Pearse, it had an average compressibility of around 50% over the all tests performed.

2.6. POTENTIAL INNOVATIONS

Different types of technical issues can be encountered when imaging computer and data storage systems. Given the complexities of some systems, the examiner must possess the knowledge, the tools and the ability to deal with any situation that is presented. In most cases the difficulties lie with imaging data from online server and network storage systems. These types of systems usually cannot be shut down and data extraction is typically through relatively slow interfaces as indicated in Section 2.5.

Another problem area is when the examiner is faced with imaging large numbers of computer workstations. Typically an examiner will remove the hard drive from a computer system and image it with a hardware imaging device because this process is one of the fastest methods of creating the image. However, some situations call for the examiner to image over fifty workstations in a short time frame. Removing the hard drives from fifty workstation computers for imaging will be a tedious and time-consuming activity, particularly if the majority of these systems are laptop type computer systems. Innovative tools and procedures may assist the examiner to collect data from multiple workstations, server systems and network storage systems.

Section 2.6.1 looks at a method of collecting data from many user workstations by using a forensic boot CD. The method could save the examiner a large amount of time when tasked with numerous workstations to image. A data scanning tool is discussed in section 2.6.2. The tool will allow the examiner to scan a source drive to determine whether processing time can be saved by imaging with a compression algorithm. Section 2.6.3 looks at other methods to speed up the imaging process on older systems by adding modern hardware with faster data transfer capability.

2.6.1. Forensic Boot CD

On occasion the examiner may be required to image a large number of computer workstations. For example, the circumstances may be where an organisation has gone into receivership and the receivers want to collect data from all computer systems on the day the receivers were appointed. The method of removing the hard drives manually will be time-consuming, especially if the hard drives are buried deep within laptop computer systems and many parts have to be removed in order to remove the drives. Additionally, the examiner is restricted by how many write blocking devices, workstations and hardware imaging devices are available for use at the time. The examiner has to use one device for each imaging process. For example, if the examiner's kit contained two hardware imaging devices and one hardware write blocking device, then only three hard drives can be imaged at any one time.

Instead of removing the hard drives of many systems, an innovative approach would be to use a specialised forensic boot CD. The boot CD would boot the workstation into a trusted operating system that does not allow writing to the system's hard drive. Raptor 2 (Forward Discovery, 2011) is such a boot disk. According to Forward Discovery, Raptor has been modified to write-protect all attached media upon booting, thereby preventing accidental writes or having to use expensive physical write-blockers. Once booted, a USB storage drive is attached to the system and the system's hard drive is imaged to the storage drive. The examiner will be able to image more systems in a given day using the boot CD method as opposed to removing the hard drives from each system. The amount of computers an examiner can image simultaneously is determined by how many boot CDs and USB storage drives the examiner has available.

2.6.2. Data Scanning Tool

Although a boot CD will assist the examiner with imaging large numbers of workstation computer systems, there are still issues when imaging server and network storage systems. As mentioned in section 2.5.3, the use of compression while creating an image can reduce the processing time, particularly when imaging is restricted through relatively slow USB or network connectivity. As mentioned in Section 2.5.3, Cusack and Pearse (2011) claimed that the time for imaging depends on three variables.

The amount and type of data on the drive contributes mostly to the processing time. A drive that has been wiped with zeroes before installing the OS will produce a smaller image than a drive that has been wiped with random data (Encase Compression, 2008). For example, if a blank 1TB drive was imaged with no compression, the image file would be 1TB in size. If the same drive was imaged with the lowest compression level set, the image file may be less than 200MBs in size. Since the drive is blank and is practically full of highly compressible zero byte data, the original data will be extremely compressed into a very small size. If the examiner has to pass the image through a USB or network port, like in a live server image situation, the 200MB image would be passed through the USB or network connection in a much shorter time than the larger 1TB image.

Information about the type of data on a particular server would be valuable to the examiner. The development of a data scanning tool that can be used to retrieve information regarding the type of data resident on a hard drive would be beneficial to the examiner when making decisions regarding compression to reduce image processing time.

2.6.3. Adding Hardware Components

The thesis has discussed imaging stand alone computer workstations and live server systems and the issue of not being able to shut down a running system. However, although the examiner may encounter a running system which an organisation relies on for business operations, it may not be necessary mission-critical and may be a system that can be turned off for a short period of time. The offline time may only be for a few minutes which would not be long enough to image the entire system drive. However, the offline time may be useful particularly when an examiner encounters older types of systems where the USB and network ports are very old and data transfer rates through those ports are extremely slow. As mentioned in section 2.5.2, imaging a large hard drive through USB version 1.0 (USB specification, 2011) or 10 megabits per second (Mbit/s) on ethernet connectivity (Ethernet Specification, 1980) would be incredibly time consuming. An approach would be for the examiner to shutdown the system and install a late model USB card or a 1Gbit/s network card. Installing a USB or network expansion card would enable a vast improvement in data transfer rates and would reduce the image processing time substantially. Once imaging is complete the system can be shut down for the removal of the expansion card and restarted for general business.

2.7. ISSUES AND PROBLEMS

The topic of the research focuses on technical issues with forensic data collection processes. However, the examiner has other issues, albeit not so prevalent, that also need to be considered and worked through. Issues, such as changing data on a system while imaging, cannot be left without an explanation. Even an inexperienced defence lawyer would jump at the chance of making evidence potentially inadmissible in court because the evidence had been altered without justification. The investigator must know exactly what to do when arriving on scene, particularly when encountering live systems. Unintentionally changing data on the target system could invalidate the acquired evidence and also cause it to be inadmissible in a court of law (Davis, 2009).

Private or privileged data is another area of concern that needs to be addressed. If privileged data resides on a system, should an examiner create an image of the physical volume? In most cases the answer to the question is no; however, there may be occasions where a party gives permission for a physical image to be created provided the image is sealed and not accessed until all parties involved give consent to access the data. Relevant evidence, including privileged data that is collected from a system, may be inadmissible in a court of law if the privileged data was collected without consent.

Forensic imaging tools are used by the examiner to collect data from a hard drive or electronic storage device. Tools such as EnCase have been accepted in court to operate correctly. It was confirmed by the Appellate Court that the EnCase software used by law enforcement investigators properly preserved data (Zintel, 2002). However, there are several other tools that are designed for data collection. The examiner must ensure that tools used for the imaging process operate as the specification describes and the tool produces accurate and verifiable results. If overlooked, a situation may develop where evidence may be inadmissible because a specific tool is proven not to be accurate or does not function as indicated by the developer.

Issues and problems are addressed in the next three sections. Section 2.7.1 looks at the forensic collection process and how it has changed and adapted to keep up with technology changes. The process now involves making unavoidable changes to data during collection. Section 2.7.2 discusses the issues related to collecting data that contains legal professional privilege and confidential information. Finally, section 2.7.3 looks at forensic data collection tools and the importance of testing the tool after updates or upgrades have been made.

2.7.1. Is it Really a Forensic Process?

The collection phase of the computer forensic process is the foundation of this research. For years law-enforcement and other computer forensic training have focused on the importance of not modifying the time stamps of files on the target system's storage media. The system's screen was photographed, a record of the hardware connections and placement was noted and the system was immediately disconnected from the electrical supply to ensure that data could not be changed (Bunting & Anson, 2007b).

Computing research has seen numerous advancements in technology. As a consequence, forensic practices and processes have had to change in order to adapt to the advancements. Live system imaging is being performed more often to ensure that volatile data is preserved. In some cases, cutting the power to a system may prevent the examiner from capturing imperative data. According to Bunting and Anson (2007b), while cutting the power to a system does have the desired effect of stopping activity on the system and preserving time stamps from that point forward, it also has the undesirable effect of deleting all data from volatile storage such as the system's random access memory (RAM).

Section 2.5.1 discussed that at times the examiner is forced to use a much slower interface or port for data collection from live systems. The data connection methods are usually via USB or network connectivity. Connecting a USB destination drive to a USB port on an operating Windows system triggers the operating system to detect, integrate and install drivers for the device to make the device accessible for the user. The detection process makes unavoidable and uncontrolled changes to the data on the system. According to Carvey and Altheide (2005), when USB storage devices are attached to a running Windows system, drivers on the system collect information from the device and then use that information to create unique artifacts in the system itself. A question that could be raised is: If data on a system has been changed due to a collection process, is the process still considered to be forensic?

According to McKemmish (1999), computer forensics can be briefly described as the process of identifying, preserving, analyzing and presenting digital evidence in a manner that is legally acceptable. McKemmish's section on preserving data illustrates that there are circumstances where changes to data are unavoidable. Where change is inevitable, it is essential that the nature of, and reason for, the change can be explained. Alteration to data that is of evidentiary value must be accounted for and justified. Because of technological advancements, live data collection is becoming standard practice and forensic practices and processes have had to change in order to adapt. Examiners in the industry have accepted that in certain situations, data on a system will be changed by the data collection process.

2.7.2. Confidential Information and Legal Professional Privilege Issues

Some companies work out of the same building and in some cases, two companies may share server and storage infrastructure to cut down on costs. The result is that two companies, company A and company B, are storing data on one server system. If a physical image of the server is created to collect data from company A, the image will contain confidential and possibly privileged information from company B. The situation can be dealt with by using one of the two following methods. Firstly, the examiner may be instructed to create a logical copy of only company A's data. The process should be straightforward providing that both companies' data are segregated. Secondly, company B may give permission for the examiner to create a physical image of the entire system and collect its data provided that the examiner gives legal undertakings that their data is not to be reviewed in any way.

In another scenario, a company may have its own server systems and may only store its own data on those systems. However, an examiner may still be governed by a demand for legal professional privilege (LPP) and may have to filter privileged data out of what has been collected. The filtering process is a common procedure for examiners when conducting electronic discovery (ED) activities. There is no standardized, forensically sound procedure to protect privileged information. However, there are two common approaches used to process privileged information, *selective cloning* and *clone and erase*. Selective cloning tries to conduct a selective data copy instead of cloning the whole storage medium whereas the clone and erase method creates a clone of the system and the privileged information is extracted out of the clone (Law et al., 2008). The term *clone,* used in *clone and erase,* is referring to a bit-for-bit physical image.

2.7.3. Tools Fit for the Job

Examiners rely on forensic imaging tools to collect data from electronic storage systems. The examiner trusts that the tool will perform the way it was designed to perform. However, research of the tools used for data collection should be conducted to ensure that the examiner has sufficient knowledge of the tools in case the examiner is asked about them in legal proceedings. For example, in court, the examiner may be asked if EnCase has been tested and validated by an independent software testing organisation. According to Soe, Manson and Wright (2004), EnCase is the industry standard in computer forensic software. EnCase is used by practitioners worldwide and has been independently tested by the National Institute of Standards and Technology (NIST) (NIST, 2003). In a court of law, the reputation of a tool could be jeopardised if the examiner cannot present evidence supporting the tool's credibility to a jury.

The examiner should use tools that are well recognised in the forensic community for data collection. According to Zintel (2002), in the State v. Cook decision released September 13, 2002 (Ohio App. 2 Dist., 2002-Ohio-4812), the Ohio Appellate Court confirmed that the EnCase software used by law enforcement investigators properly preserved data on a suspect's home computer, thus preventing any alteration of vital digital evidence. These types of findings add integrity to the tool and to the evidence created by the tool.

Another question one may be asked in court is "When did you last test your tools?" NIST (National Institute of Standards and Technology, 2011), runs a computer forensic tool testing program. Reports on most of the common types of tools, write blocking devices and imaging devices that have been tested are available. However, this is not to say that tools and equipment do not need to be tested. To ensure a tool is working correctly, it should be tested using a reputable tool testing process. According to the Scientific Working Group on Digital

Evidence (SWGDE), validation testing is critical to the outcome of the entire examination process. Validation, based on sound scientific principles, is required to demonstrate that examination tools (hardware and software), techniques and procedures are suitable for their intended purpose. Tools, techniques and procedures should be validated prior to their initial use in digital forensic processes. Failure to implement a validation program can have detrimental effects (SWGDE, 2009). The testing of tools and equipment when they have been initially purchased and after applying patches or updates may identify issues before they are used in the field.

2.8. CONCLUSION

Creating a forensically sound image of electronic storage devices is a necessary process, particularly if there is a possibility that discovered evidence will be relied upon in legal proceedings. Section 2.0 begins with a breakdown of the forensic process related to electronic evidence. The forensic process consists of four key areas: Collection, Examination, Analysis and Reporting. This research focuses on the collection phase of the forensic process. The collection phase involves capturing data stored on an electronic device, in a forensically sound manner. To verify the captured data is an exact copy, hashing algorithms are used on both the source data and the collected data. There have been considerable advancements in technology relating to computer systems and electronic storage media. The advancements have caused data collection methods to change from a time where under no circumstances changes were to be made to the source data, to a time where examiners are now required to image live systems and doing so can change data on the system. The change in data collection methods can partially be put down to the demand for availability and reliability of systems for business.

Electronic storage capacities double around every two years. The increase in storage sizes forces examiners to collect large amounts of data from live systems over slow data connections. Some examiners address the issue by collecting only the data that is relevant to an investigation. However, the technique carries the risk that the examiner may unintentionally fail to collect vital evidence. Full disk or device images are preferred as all data is captured as opposed to a selective collection of relevant files. There is a need for new tools and processes to assist the examiner when collecting data from large storage volumes across relatively slow data connections. Section 2.6 suggests an innovative approach to tools which can facilitate in reducing processing time while conducting imaging activities on live systems and multiple workstations. These tools include boot CDs that are used to image multiple workstations simultaneously and a tool that will analyse data on a hard drive to indicate whether the use of compression will reduce image processing time.

Chapter 3 will define the methodology for the testing of performance issues related to data collection from electronic storage devices that are reviewed in this chapter. In Chapter 3, research into the types of data stored on a hard drive is conducted. Research and development of a software tool that can assist an examiner to determine what type of data is stored on a hard drive is carried out. Knowing the type of data stored on a drive can assist the examiner in reducing image processing times.

Chapter 3

METHODOLOGY

3.0. INTRODUCTION

Chapter 2 introduced literature on the thesis topic. The topic is about research into performance issues with forensic data collection activities from electronic storage devices. With considerable advancements in technology relating to computer systems and electronic storage media over recent years, creating full bit-for-bit or *physical* forensic copies of large media devices or storage volumes in a timely manner is becoming more difficult to achieve. Chapter 3 studies pre-data collection testing using tools and methods that can assist an examiner to reduce processing time when creating full forensic copies of media and storage volumes.

Chapter 3 begins with a study into research methods developed by five other researchers. The papers were chosen because the material is based on software development and testing, which is related to this research. Once the methods have been analysed, relevant information from the five papers and ideas from the author will be compiled to develop methods for the research. The research question is defined along with a hypothesis in Section 3.2. Research methods and design of a scanning tool are discussed in Section 3.3. Section 3.4 discusses data collection and analysis. Research limitations and accuracy along with future work are discussed in Section 3.5. Finally, the chapter concludes with a summary of the main points in the chapter.

3.1. REVIEW OF PREVIOUS RESEARCH

The research being undertaken is unique in that the author has not located any material from other researchers in the field. The goal of the research is to develop a software tool that can assist forensic examiners in their data collection activities. Software development and testing methods used by five groups of researchers in their studies are reviewed and presented in Sections 3.1.1 to 3.1.5.

3.1.1. A Framework for Testing Distributed Software Components

Yao and Wang (2005) conducted research into developing a framework for testing distributed software components. They used a technology called Built-in Test

(BIT) which is a technique that embeds BIT software into source code to enhance self-testing and run-time testing. According to Yao and Wang, BIT is useful for automating tests, building run-time testable software, software integration and maintenance. However, there are disadvantages with embedding BIT software into source code. BIT software can become large and occupy a lot of space which increases the overall software component size. In addition, some of the BITs may not be required in the environment where the component is finally deployed.

A distributed testing framework for software component testing is proposed by Yao and Wang (2005). Unlike traditional testing models, which conduct testing after software development, Yao and Wang's incremental testing strategy allows tests to be built and run incrementally. Building and running tests incrementally helps to identify side effects, errors and defects prior to the component being integrated into the rest of the system. The people conducting the component testing build their own tests which match the context of the component through all stages of the software life cycle. The incremental testing framework reduces the time required in component testing and ensures that changes will not break existing code.

Yao and Wang (2005) use a software tool that supports the proposed incremental testing framework for distributed software components. According to Yao and Wang, the testing system can help clients to invoke the operations of a server component under test and thus compare the results with the expectations. The technique has the benefits of reduced time and cost associated with traditional component testing.

Yao and Wang (2005) give an example of component testing for the proposed remote testing framework. An automated teller machine (ATM) bean that simulates a banking system is developed. The ATM bean is capable of the following functions: creating an account, depositing money, withdrawing money and querying a balance. Some of the functions are limited. For example, the deposit function would let any amount be deposited as long as the amount was greater than zero. Similarly, the withdrawal function would allow a user to withdraw money providing the withdrawal was greater than zero and the account balance did not fall below five dollars after the withdrawal. The above mentioned method, *BIT_ATM(),* is shipped with the component. According to Yao and

Wang, by setting up a BIT account, the deposit and withdrawal methods accomplish built-in tests of the ATM bean and can also be used to verify the component's functionality in any application context.

3.1.2. What is Software Testing? And Why is It So Hard?

Whittaker (2000) approaches software testing in four phases: modelling the software's environment, selecting test scenarios, running and evaluating test scenarios and measuring testing progress. The four phases provide software testers with a structure in which to group related problems that they must solve before moving on to the next phase.

When modelling the software's environment, a software tester's task is to simulate interaction between software and the software's environment. According to Whittaker (2000), the software tester must identify and simulate the interfaces that a software system uses and enumerate the inputs that can cross each interface.

The selection of test scenarios can be difficult because there are literally thousands to choose from. Software testers typically seek the test set that will find the most *bugs*. In addition, testers typically execute each source line at least once as minimum criteria to judge the completeness of their work.

Running and evaluating test scenarios are managed by converting the scenarios into executable form to simulate typical user action. Software testers try to automate the test scenarios as much as possible because manually applying test scenarios is a labour intensive and error prone task.

According to Whittaker (2000), measuring testing progress is seldom done. Counting measures give very little insight about the progress of testing. Measuring testing progress is not achieved by counting how many bugs have been found and corrected or counting the number of times the application has been terminated successfully. According to Whittaker, interpreting such counts is difficult and many software testers supplement counted data by answering questions designed to ascertain structural and functional testing completeness. A list of such questions is shown in Table 3.1.

Table 3.1

Questions Designed to Ascertain Structural and Functional Testing Completeness (Adapted from Whittaker, 2000, p. 78)

To check for structural completeness, testers might ask these questions:	Have I tested for common programming errors?
	Have I exercised all of the source code?
	Have I forced all the internal data to be initialized and used?
	Have I found all seeded errors?
To check for functional completeness, testers might ask these questions:	Have I thought through the ways in which the software can fail and selected tests that show it doesn't?
	Have I applied all the inputs?
	Have I completely explored the state space of the software?
	Have I run all the scenarios that I expect a user to execute?

3.1.3. Evaluating Testing Methods by Delivered Reliability

Frankl, Hamlet, Littlewood and Strigini (1998) conducted research into software testing methods that achieve higher program reliability. They defined two main goals in testing software. The first is to use debug testing techniques to identify and remove defects in order to achieve adequate quality. The second goal is to perform operational testing to gain confidence that the software is reliable for its intended purpose.

The researchers describe the *debug testing* method as systematic testing methods used to locate as many bugs as possible either by sampling all situations likely to produce failures or by concentrating on situations that are considered most likely to do so, such as *stress testing*. On the other hand, *operational testing* is described as when software is subjected to the same statistical distribution of inputs that is expected in operation. Instead of actively looking for failures, the tester, in an operational testing case, waits for failures to surface spontaneously. Frankl et al. (1998) research studies the testing effectiveness based on the reliability of a program after the program is tested. A measure is used to compare debug testing to operational testing by exploring circumstances under which each technique is likely to yield towards superior reliability.

The authors found that reliability improves under either testing scheme. When failures are found, the software is successfully changed, and the operational failure probability decreases. The researchers believe that analytic, probabilistic methods are the best tools for studying software reliability. The methods indicate which empirical measurements could provide indirect evidence that, in a particular project and phase of development, a certain test method is best. Trusting the ability of the debuggers can be risky and Frankl et al. (1998) states that comparing the effectiveness of their testing profiles with that of operational profiles is possible and can be accomplished at a low cost.

3.1.4. Classification of Software Testing Tools Based on the Software Testing Methods

Mustafa, Al-Qutaish and Muhairat (2009) conducted research into classifying a set of testing tools and distributing them over a series of testing methods to determine if the software testing process can be assisted by software tools to make the software testing process automated. To begin with, software products were classified into software groups based on their intended usage, complexity and development technology. The researchers then collected 135 testing tools and distributed them into each of the software groups, shown in Table 3.2, based on which software product could be applied.

Table 3.2

Number of Testing Tools Associated With Each Software Group (Adapted from Mustafa, Al-Qutaish & Muhairat, 2009, p. 230)

Software Group	Number of tools
Web Application	63
Network Protocol (TCP)	27
Application Software	18
Java Software	16
Open Source Software	10
Database	7
System Software	2
Embedded Software	2

The researchers focused their research by selecting the three software groups that had the highest number of testing tools associated to them. These groups were *web application, network protocol* and *application software*. To discover which testing type is most used for each of the three software groups the researchers categorised the testing tools by using a selection of nine commonly used testing types. These are:

- Stress Testing
- Load Testing
- Regressions Testing
- Functional Testing
- Unit Testing
- Performance Testing
- Acceptance Testing
- Security Testing
- Open Source Testing

After analysis of the three main software groups, the researchers concluded that the functional testing type was mostly used for *web applications* and *application software*. However, for transmission control protocol (TCP), the performance testing type was the most used. The classifications given by the researchers will assist practitioners in software testing to know which testing methods have automation tools and which types of tests have a limited number of automated tools.

3.1.5. Software Testing

Freeman (2002) discusses several different types of software testing methods and provides information from his own personal experiences in regards to software testing.

Freeman (2002) begins with *glass box testing*, also known as white box or clear box testing, where testing is performed on the internal components of the software. Glass box testing relies on knowledge of the actual source code. Glass box testing consists of static and dynamic analysis. The dynamic analysis technique is performed by running the system and examines the behaviour of a software system before, during and after the software execution. The dynamic analysis technique is used for testing software products such as a database.

According to Freeman (2002), there are several different types of static techniques. These techniques include *statement coverage testing* where every statement is executed at least once, *branch coverage testing* where a series of tests are performed to ensure that all branches of a test requirement or software component are tested at least once and *path coverage testing* where testing all the various paths of each test requirement or software are performed.

Another software testing method is black box testing or closed box testing. The black box testing method tests the functionality of the system using the functional requirements. Black box testing can be used when testing an order tracking system. For example, when purchasing a product over the internet, the user will be required to create an account, select items to purchase, enter personal information and enter payment information. All the functionality of the software must be met before the software product is ready for production. Knowledge of the code is not necessary during black box testing, knowledge of the requirements is.

3.2. THE RESEARCH QUESTION AND HYPOTHISES

The literature review in Chapter 2 has provided abundant knowledge in the areas of electronic data storage systems and forensic data collection methods that examiners use during digital forensic investigations. In Section 2.6, the author detailed three potential areas of research which could assist an examiner to reduce processing time while creating a forensic copy of a data storage device. Section 2.6.2 discussed a data scanning tool that allowed the examiner to scan a source drive to identify the types of data that reside on the drive. A review of five previous works exclusively relating to software design and application testing was conducted in Section 3.1. The knowledge gained from both the literature review and the previous works is used to develop the research question and will also assist with defining a sound methodology for the project. The topic of the research is '*Situational Tool and Method Selection for Digital Forensic Data Collection: Performance Issues*'. In essence the study conducts research into discovering how data collection performance can be improved using tools and procedures.

3.2.1. The Research Question

Fundamentally the data scanning tool proposed in Section 2.6.2 reads data within a selection of sectors that are scanned across the range of the storage media. The scanning tool determines whether the read sectors are *empty* or *used*. Empty sectors are sectors where all of the byte values in the sector are identical. Used sectors are sectors where the bytes within the sector contain different values. The tool calculates a percentage for both empty and used sectors and produces a report with the results.

Section 2.5.3 discusses how the use of compression algorithms, while creating forensic copies of data storage devices, can reduce the image processing time, particularly when data transfer rates are restricted through relatively slow USB or network connectivity. Imaging through such connectivity is typically encountered when imaging live systems. If the examiner can determine that a particular storage device contains a high percentage of highly compressible data, such as empty sectors, then the examiner may choose to use a compression algorithm while imaging the device to reduce the processing time.

In order for a scanning tool to complete a scan of a drive in a reasonably short period of time, it will be designed in such a way that the tool will read a sample of sectors and not every sector on the drive. Reading all of the sectors on a drive would take a considerable amount of time and would defeat the purpose of the scanning tool. To enable tool performance, only a small selection of all available sectors is read, although there remains a requirement for the scanning tool to produce an acceptable assessment of the type of data that resides on a drive, i.e. a reliable result. The research question for the thesis is hence: *"Can a drive scanning tool process a hard drive in a measured period of time and produce a reliable result?"*

3.2.2. Defining the Research Question

There are two key factors within the research question that need to be defined. Firstly, *a measured period of time* provides no description as to how long the time period actually is. Secondly, apart from being an acceptable assessment of the type of data that resides on a drive, there is no clear indication as to what a *reliable result* is.

Because the examiner is usually under a high level of pressure to complete all data collection activities in a limited time frame, extra processes should be kept to a minimum in terms of time. To keep extra processes, such as using a scanning tool, to a minimum amount of time, the author has defined *a measured period of time* to be four minutes. The four minute time period is also chosen for providing a starting point or reference for testing processes. In addition, the four minute time period is also short enough to allow for ample tests to be conducted during the research. The four minute time period will be referred to as the *standard test time* or a *standard scan* throughout the thesis.

In terms of reliability of results, an accurate scan result would only be possible if every sector on the entire drive were scanned during a scan process. Therefore, the calculation of an accurate result would be based on the ratio between occupied and empty sectors from all available sectors on the disk. However, scanning an entire drive out in the field would most probably take a longer amount of time as opposed to actually imaging the entire disk. If a *standard scan* could be performed where the result produced from a scan is within ±10% of an accurate result, then it would be considered as a reliable result. The result is reliable because it is 90% accurate and an examiner would benefit from such a result.

3.2.3. Hypothesis

The author has stated that a four minute time period would represent the standard test time as a starting reference for the research. If, on the other hand, the scanning process took a considerable amount of time to process a drive, then the scanning tool would not be beneficial to an examiner. For example, if an examiner was using the drive scanning tool to scan an 80GB hard drive and the scanning process was going to take one hour to complete, the information from the scan would be of no benefit to the examiner because the examiner could have imaged the entire drive in a similar amount of time. In order for the scanning tool to be beneficial, it must process a drive quickly and also produce a result that is reliable as explained in Section 3.2.2. However, the issue with running the tool in such a short timeframe is that only a fraction of all of the data on the drive would be read and processed. Not processing all of the available data could decrease the accuracy of the results to a degree where the result becomes unreliable.

The proposed hypothesis is based on a theoretical 20GB hard drive that contains 5GB of data and 15GB of wiped space, i.e. the drive is one quarter full of data and three quarters empty. The scanning tool would be designed to scan a disk from the first usable sector to the last usable sector with a calculated number of sectors in between the scanned sectors. The scanning tool will be set by the examiner to process a number of sectors or scan cycles for a particular hard drive. The tool will calculate the offset between each scan cycle which will enable even reading of sectors across the range of the drive within a four minute period. Therefore, the even scanning process reads both the area where data resides and the remaining wiped area or unallocated space on the disk.

Based on the theoretical example above, the author hypothesises that a scanning tool can process a hard drive in a measured period of time, namely four minutes, and the tool will produce a reliable result that is within ±10% of an accurate result. If the example above were to be tested, theoretically the result would indicate that the drive is one quarter full and three quarters empty. The author believes that evenly scanning a drive from the beginning to the end will produce a comprehensive sample of data which will provide a result that is within ±10% of an accurate result.

Theoretically, if the scanning time frame was doubled to eight minutes, twice as many scan cycles would occur resulting in a more defined or concentrated scan of the data on the drive. A more concentrated scan may produce a variance in the result compared to a standard four minute test. However, any variance would most probably be small because the tool is scanning a disk that remains one quarter full of data and three quarters empty.

3.2.4. Research Sub-Questions

Two research sub-questions are generated to facilitate further research into two areas of the experimental design of the testing. Firstly, as mentioned in Section 3.2.2, the author has chosen the four minute time period to provide a starting point for the test processing times during the research. The first sub-question asks: *"Will a two minute scan produce a similar result to a four minute scan?"* The question is designed to expand the research to include an additional series of tests run over a two minute time period. The objective is to determine whether similar results would be produced from a two minute and a four minute scan. If similar results are produced by the two minute tests and the test results remain reliable, then the standard test time would become redundant as a two minute test runs in half the time.

The second sub-question has to do with the amount of data that is processed during a scan. An explanation of the sub-question and more details surrounding the author's reasoning behind the sub-question is explained under the third software requirement in Section 3.3.1.1.

3.3. RESEARCH METHODS AND DESIGN

The goal of the study is to research and develop a tool that will assist with reducing processing time when creating forensic copies of hard drives or electronic storage devices. If the examiner has knowledge of the type of data on a system, he or she will be able to make informed decisions as to whether the processing time can be reduced by using compression algorithms while imaging. The knowledge gained from a scanned device can assist an examiner in evaluating the type of data on the storage device and to better determine a course of action in terms of imaging processes and technologies. A *quantitative* research approach has been chosen for the research. *Descriptive* methods will be used to study the output results in relation to input variables within laboratory experiments.

3.3.1. Software Development Life Cycle

According to Lewallen (2005), software life cycle models describe phases of the software cycle and the order in which those phases are executed. The use of a software development life cycle model assists the software development project team members to ensure the accuracy of the software development efforts and that deadlines are met.

There are several types of software development life cycle models in existence. A common type of software development life cycle is the *waterfall model* by Royce (1970). Figure 3.1 illustrates a reconstructed waterfall model. The waterfall approach for software development is separated into five process phases comprising the requirement phase, software design, implementation, verification/testing and maintenance. The process starts from the top level. Each process phase cascades to the next phase below when the phase has been completed. The cascading effect means that the phase below is not started until the phase above has been completed, hence the name waterfall model.

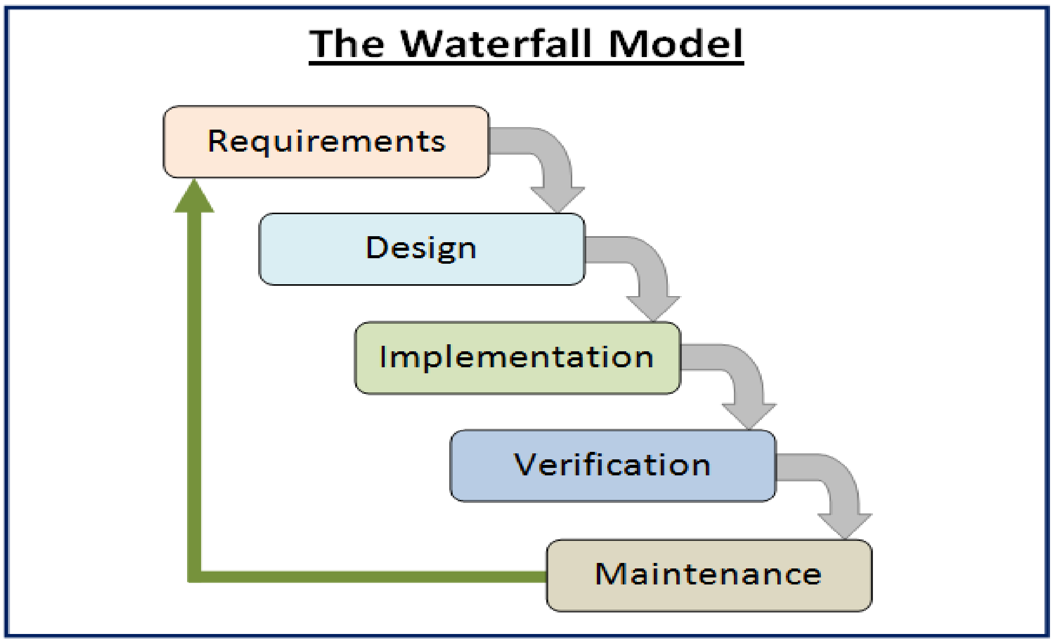


Figure 3.1. Software development life cycle – The waterfall model

According to Parekh (2011), a disadvantage with the waterfall model is that often the requirements of the customer continue to be added to the list after the requirements phase has been completed. Adding more requirements when development is beyond the requirements phase causes the requirements to not be fulfilled which results in the development of an almost unusable system.

However, the waterfall model will be used for the research. The waterfall model has been chosen because of the model's chronological flow design where phases are completed before work starts on the next phase. The disadvantages noted by Parekh (2011) should not be an issue with the research providing the author, prior to commencing the design phase, has exhausted the list of requirements to a point where no new requirements will be added after the requirement phase. If a situation arises where new requirements need to be added to the software life cycle to enable better tool functionality, the requirements will be added after the current version has been completed. Further amendments of the requirements can appear in the future versions of the software.

The waterfall method is an excellent model for the project and has the following advantages:

- Any design errors will be captured before the software is written.
- Capturing design errors early could save a considerable amount of time during the implementation phase.
- The approach is very structured and focuses on one layer at a time.
- Testing is easier as testing can be performed by reference to the scenarios defined in the functional specification.
- There are no issues with functionality when the application is finally completed as all of the application's features have been fully specified.

3.3.1.1 Software Requirements

The scanning tool is a piece of software that will scan a hard drive to determine what type of data resides on the drive. The results from processing a hard drive will show the percentage of the processed sectors that are empty and the percentage of sectors that contain data. An empty sector is defined by the author as a sector where every byte, within the sector, has the same value. Sectors that contain the same value for each byte will compress extremely well. According to Cusack and Pearse (2011), image processing time can be considerably reduced when imaging highly compressible data. Knowledge of type of data that reside on a hard drive will enable an examiner to make informed choices of whether to use a compression algorithm while imaging to reduce processing time.

Four operating requirements for the software tool will be discussed. The first requirement is time. In section 3.2.2 the author stated that a measured period of time will be defined as four minutes. The four minute time period is referred to as the *standard test time.* The standard test time will be used for the majority of the scanning tests.

The second requirement is to collect a comprehensive data sample. The scanning tool will be designed to read individual sectors on a hard drive. During a scanning process the drive is read from the first sector of the disk and will finish on the last sector of the disk with sectors being read at even spaced intervals between the first and last sectors. The proposed method, scanning across the entire disk, ensures that a comprehensive data sample will be collected.

The third requirement is to process as little data from each sector as practicable without compromising the accuracy of the results. If accuracy can be maintained when processing a portion of a read sector, the author predicts that the processing will take less time to complete as opposed to processing all of the data stored within the sector. Less time taken to process each sector will allow for more scan cycles to take place within a standard test. The author believes that the accuracy of the overall result will be increased if more scan cycles are processed within the same timeframe.

The third requirement, which addresses processing a sample of data from each scanned sector in order to reduce processing time, is purely theoretical. If processing time can be reduced by processing a fraction of the data contained in each sector, then that amount of reduced time would not be measureable without additional processes. Therefore, the second research sub-question, *"Will processing all of the bytes in each scanned sector produce a similar result to processing eight bytes in each sector?"* is designed to conduct additional research and testing to establish two points of interest. The first point is to identify how much additional time is required to process all of the bytes in each scanned sector during a given scan. A benefit of processing all of the data in a sector is that the result is definitive and is 100% accurate. There are only two possible outcomes when testing all of the data in a sector: either the values of all of the bytes in the sector are the same or there are different values for one or more of the bytes. Nonetheless, if all of the data in a sector is read and stored in memory during a scanning process, then there is a possibility that processing all of the bytes directly from the memory may not require a substantial amount of additional time. The second point of interest is to compare and confirm the accuracy of processing only a portion of the data in each sector to reduce the scan processing time. Processing all of the bytes in each scanned sector will produce an accurate result and will provide a benchmark to which other results can be compared.

The fourth requirement is reporting. When a scan process has finished, the results are printed on the screen in a pie chart. The pie chart displays the ratio of sectors that are empty and the sectors that contain data. The examiner can use the information from the pie chart to determine the best course of action in regards to data collection activities. However, to uphold best forensic practices, there is a requirement to save the results and associated test information into a report file so that the information is available in the future if required. The information included in a full text report is presented in Table 3.3. The report can be opened by the examiner upon completion of a scan and can be saved to a desired location by the examiner.

Table 3.3

Information Included in a Full Text Report

Category	Property	Data Types
Case Information	Case name or number	String
	Examiner	String
	Location	String
Source drive details	Hard drive make	String
	Hard drive model	String
	Hard drive serial number	String
	Hard drive size	Integer
	Total sector count	Double
Test Information	Scan start date	String
	Scan start time	String
	Type of test	String
	First sector scanned	Integer
	Last sector scanned	Double
	Sector offset	Double
	Total sectors scanned	Double
	Bytes scanned per sector	Integer
	Scan finish date	String
	Scan finish time	String
	Test duration	String
Results	Empty sectors	Double
	Used sectors	Double
Scan Status	Completed successfully or with errors	String

3.3.1.2 Analysis and Design

Software application architecture is the process of defining a structured solution that meets all of the requirements, while optimizing common quality attributes such as performance, security and manageability (Software architecture, 2011). In the analysis/design phase, the four operating requirements discussed in Section 3.3.1.1 are analysed and the software tool design is established.

While a four minute time period for a scan process seems simple to achieve, some testing will be necessary to ascertain timing data that will be used to calculate an accurate time period for running the tests. The length of time a test will run for is subject to several variables that influence the processing time. The primary variable is the number of scan cycles that are processed during a scan process. However, the number of scan cycles can be modified for a particular hard drive to produce a four minute time block for a given scan, on a given drive. For example, a faster hard drive will be able to process more scan cycles in four minutes as opposed to a slower drive. The number of scan cycles can be increased for the faster drive and decreased for the slower drive so each scan finishes in the four minute timeframe.

Two other variables that influence processing time are hard drive read rates and data throughput rates. There is a significant contrast in performance between two hard drives that have a difference of ten years of technology between them. To assist in answering the first sub-question, additional testing will be conducted using a two minute time period. The results from both the two minute and four minute tests will be analysed to draw answers to the first sub-question.

To eliminate inconsistencies produced by the variables associated with the hard drives that will be tested and the computer workstations used to conduct the testing, one source drive will be used for each batch of tests and each batch of tests will be performed on the same workstation. Calculating how many scan cycles fall into a four minute time period for a particular drive will be defined once the tool is operational. The scanning tool will have the ability to adjust the number of scan cycles which will alter the processing time.

Section 3.3.1.1 discusses how the sectors on the disk are processed to collect a comprehensive data sample. Scan cycles start from the first sector and process across the range of the disk at evenly spaced intervals and then finish on the last sector. The method of scanning across the range of the disk will return a comprehensive data sample because no particular area of the disk has a higher concentration of scanning compared to another area on the disk. Because there are far less scan cycles compared to the total number of sectors on the drive, the scanning tool will automatically calculate the distance (in number of sectors) between the scan cycles in relation to the standard test time. The calculated distance between the scan cycles is referred to as the *sector offset.* A smaller capacity disk will have a smaller sector offset, which means the scan cycles will be closer together as opposed to those for a larger capacity disk. If the scan cycles are close together the *granularity* of the scanning is higher which the author believes will increase the overall accuracy of the results.

The theory behind scanning across the entire range of the drive is to ensure that a comprehensive sample of data is collected. For example, if a scanning process was run on a drive and it only read the first quarter of the drive as shown in Figure 3.2, the results most probably would not be as accurate as the results produced by scanning across the full range of the drive. The reason is that the scanning tool would not collect a comprehensive data sample. Typically, most operating systems are written to the beginning of the disk. If the sectors were empty beyond the operating system, as shown in the example given in Figure 3.2, most of the empty sectors would not be processed. In addition, because the test time is the same, the granularity or density of scan cycles would be four times higher for the first quarter of the disk compared to scanning across the entire drive. The example shown in Figure 3.2 would produce a result that would overstate the used sectors and understate the empty sectors for the entire drive.

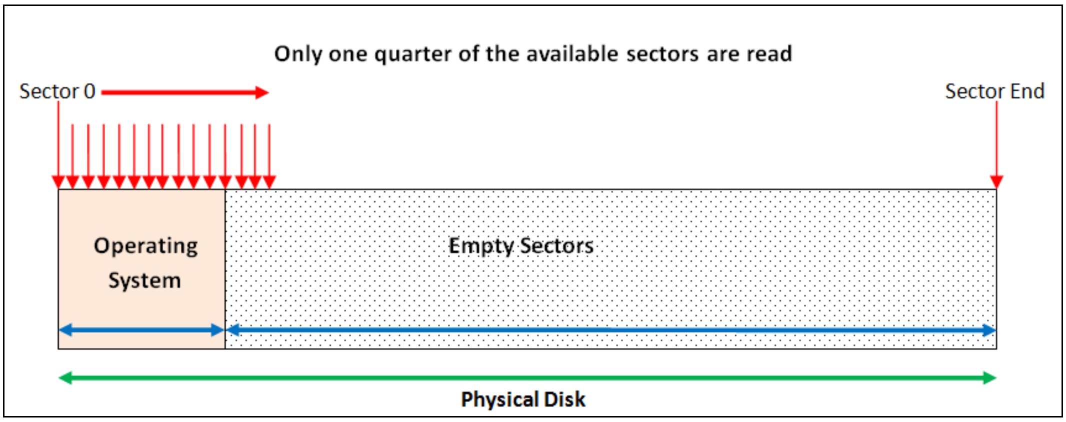
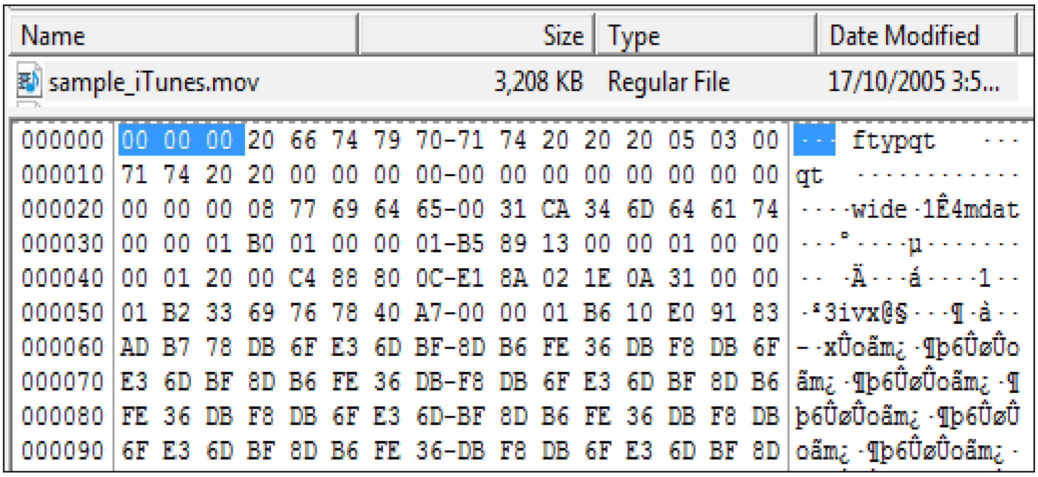


Figure 3.2. An illustration of sectors not being read across an entire disk

The third software requirement is to process as little data for each scan cycle as practicable without compromising the accuracy of the results. A sector contains 512 bytes of data. When a scan cycle reads a sector, the entire sector is read into memory. However, only the first eight bytes, the *byte sample*, of the sector are processed by the tool. Like the duration of the *standard test time*, a byte sample of eight bytes is chosen initially as a starting reference for the testing. The size of the byte sample may be modified if discovered that the byte sample is not suitable during the testing phase of the research. However, eight bytes will ensure a significant amount of data is processed to avoid false positives or inaccurate results that can occur if the byte sample is too small. For example, if the byte sample is two bytes in length and a sector with two leading zeros is processed, the software would flag the sector as empty. However, a file may be occupying the sector and the file may have a file header, or *signature,* containing two leading zeros. The byte sample in this example is too small and an incorrect result is produced. An example of a file having leading zeros in its header is shown in Figure 3.3. The particular file is a Quick Time movie file.



Name	Size	Type	Date Modified
sample_iTunes.mov	3,208 KB	Regular File	17/10/2005 3:5...

```
000000 00 00 00 20 66 74 79 70-71 74 20 20 20 05 03 00 ··· ftypqt   ···
000010 71 74 20 20 00 00 00 00-00 00 00 00 00 00 00 00 qt  ············
000020 00 00 00 08 77 69 64 65-00 31 CA 34 6D 64 61 74 ····wide·1Ê4mdat
000030 00 00 01 B0 01 00 00 01-B5 89 13 00 00 01 00 00 ···°····µ·······
000040 00 01 20 00 C4 88 80 0C-E1 8A 02 1E 0A 31 00 00 ·· ·Ä···á····1··
000050 01 B2 33 69 76 78 40 A7-00 00 01 B6 10 E0 91 83 ·²3ivx@§···¶·à··
000060 AD B7 78 DB 6F E3 6D BF-8D B6 FE 36 DB F8 DB 6F -·xÛoãm¿·¶þ6ÛøÛo
000070 E3 6D BF 8D B6 FE 36 DB-F8 DB 6F E3 6D BF 8D B6 ãm¿·¶þ6ÛøÛoãm¿·¶
000080 FE 36 DB F8 DB 6F E3 6D-BF 8D B6 FE 36 DB F8 DB þ6ÛøÛoãm¿·¶þ6ÛøÛ
000090 6F E3 6D BF 8D B6 FE 36-DB F8 DB 6F E3 6D BF 8D oãm¿·¶þ6ÛøÛoãm¿·
```

Figure 3.3. Signature of a quick time file.

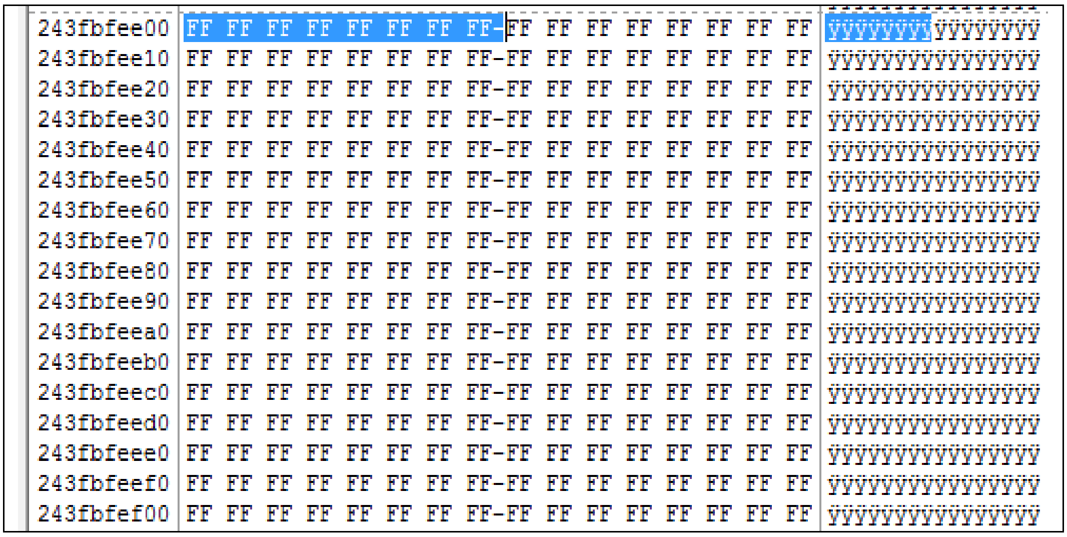
When processing the first eight bytes of data, the bytes are compared with each other to determine if all eight bytes are the same. The comparison process begins with comparing the first and second byte. Subsequently the first and third bytes are compared and the process continues until the first and eighth bytes are compared. If the bytes are equal for the last comparison test, the sector is flagged as empty by advancing an incremental counter, the *empty sector counter*, by one. However, if the comparison process detects two bytes that have different values in any of the comparison tests, the comparison process will cease for that particular scan cycle and the *occupied sector counter* will be advanced by one.

The first eight bytes are compared to establish if all eight bytes are the same. If the first eight bytes are the same, the author hypothesises that the remaining bytes in the sector are also the same and the sector is empty. The hypothesis is backed by research into file signatures by the author. A review of the file signature table in EnCase version 6.14 reveals that only one file type has a file header where the first eight bytes are the same. The file is an *Oracle parameter file* and has a signature containing ten consecutive hexadecimal (HEX) 23 bytes. The file that has the second highest number of identical bytes in the signature is a *DOS system driver.* The DOS system driver has a signature containing four consecutive HEX FF bytes. However, a search on a Windows XP system did not locate any files containing these signatures.

Processing only the first eight bytes of a sector is designed to increase the efficiency of the tool. However, the author believes that processing only an eight-byte sample of a sector may produce false positive results. As stated in Section 3.3.1.1, and in accordance with answering the second research sub-question, additional testing will be undertaken where a change is made to the size of the byte sample. The software will be modified so that all 512 bytes for each scanned sector are read and compared with each other. Changing the byte sample from eight bytes to 512 bytes for the extra set of tests will be the only modification made to the tool for the additional tests.

Typically new hard drives have all of their bytes set to zero. Having zeroed byte sectors means that when the drive is put to use, all sectors on the drive, excluding sectors occupied by the operating system and other data, will be empty. Some disk wiping utilities have options to wipe or sterilise a drive with any single HEX character or with random HEX characters. Sterilising a hard drive is a process whereby the data on the drive is destructively cleansed before data from an imaging process is written to the drive. According to Craiger (2011), a *forensic wipe* removes any vestiges of previous contents on the drive. The process ensures that a defence attorney cannot claim that any evidence recovered from the subject's hard drive was from the previous contents on the disk, caused by the co-mingling of evidence.

A benefit of the scanning tool design is that the tool will operate normally with a drive that has been wiped with any single HEX character and does not rely on a drive being wiped with the HEX zero character. If all of the bytes possessed in a particular sector have the same value, the actual value of each byte would be of no consequence. During the comparison process the first byte is compared with each other byte in the sector and not with a predefined value. The comparison method provides the scan tool with an advantage where a drive can be wiped with any single HEX character and the drive will be scanned accurately. Figure 3.4 shows a portion of a sector where the bytes have been wiped with the HEX character FF. The scanning tool will flag this sector as empty because the first eight bytes are identical.



```
243fbfee00 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfee10 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfee20 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfee30 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfee40 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfee50 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfee60 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfee70 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfee80 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfee90 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfeea0 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfeeb0 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfeec0 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfeed0 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfeee0 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfeef0 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
243fbfef00 FF FF FF FF FF FF FF FF-FF FF FF FF FF FF FF FF ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
```

Figure 3.4. Sectors which are wiped with the HEX character FF.

If a drive has been wiped with random characters and for every sector there are no two consecutive bytes with the same value, the scanning tool's compare function will detect that the first and second bytes are different and the sector being processed will be flagged as containing data. The process is repeated throughout each scan cycle and the final result would indicate that every scanned sector contained data. Using compression in such a scenario would most likely increase the processing time and size of the overall image. The reason is that the time overhead of the compression algorithm and the extra data produced by the algorithm will be added to the processing time and the size of the image file respectively.

Information supporting the fourth software requirement is captured following a scan process. Once the scanning process has completed, the data collected by the empty sector counter and the occupied sector counter are passed into a pie chart which is displayed on the screen for the examiner. The results will show the percentage of the read sectors that contain the same byte values and are flagged as empty sectors and the percentage of the read sectors that contain different byte values. In addition, the examiner can view and save a comprehensive report once the scanning process has completed.

The operation of the scanning tool will be based on the flow chart shown in Figure 3.5. The examiner starts the scanning tool and the tool automatically detects and displays the storage devices that are connected to the system. The examiner selects the drive to be tested and selects the case and scan details button. The examiner's name, the case name or number and the scan location are entered into the form. One of the following three tests is selected: a standard scan, a full sector scan or a user defined scan where the examiner can enter the number of sectors to be scanned. The scan disk button is selected and the disk is scanned accordingly to the selected type of test. Once the test is complete, a summary of the results is shown on the screen in the form of a pie chart. If desired, the examiner can view a full report that includes all the information detailed in Table 3.3, and can save a copy of the report to a location on the system for future use.

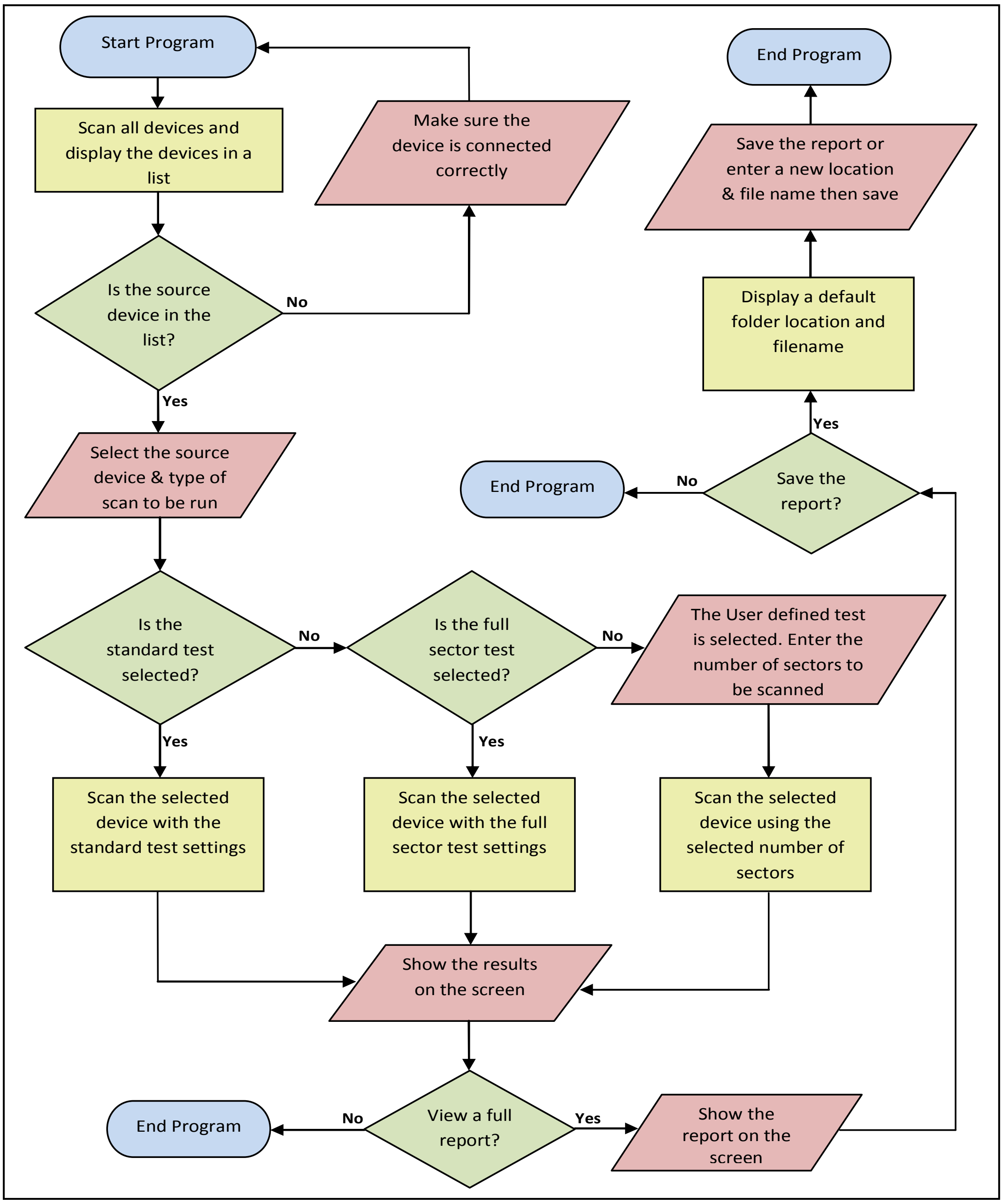

Start Program
End Program
Scan all devices and display the devices in a list
Make sure the device is connected correctly
Save the report or enter a new location & file name then save
Is the source device in the list?
No
Yes
Display a default folder location and filename
Select the source device & type of scan to be run
End Program
No
Save the report?
Yes
Is the standard test selected?
No
Is the full sector test selected?
No
The User defined test is selected. Enter the number of sectors to be scanned
Yes
Yes
Scan the selected device with the standard test settings
Scan the selected device with the full sector test settings
Scan the selected device using the selected number of sectors
Show the results on the screen
End Program
No
View a full report?
Yes
Show the report on the screen

Figure 3.5. An operational flow chart for the scanning tool

3.3.1.3 Code Development and Implementation

Code development is the phase where the design ideas are converted into actual code and the author can see the tool in operation. The scanning tool will be written in Visual Basic version 6 and will be initially available for the Windows XP system. The tool will be installed on the examiner’s workstation and will be run to scan locally attached hard drives. The intention is that source drives will be connected through a write blocking device to ensure that data on the source drive will not be changed during the scan process.

3.3.1.4 Testing and Verification

The testing and verification phase is a vital part of software development. Testing is a process that will determine whether the software is functioning correctly or there are issues that need to be addressed. Questions identified by Whittaker (2000), and designed to ascertain structural and functional testing completeness will be used to test the structural and functional aspects of the scanning tool. The questions are presented in Table 3.1.

Testing and verification of the output from a scanning process is addressed in Section 3.4.1. The scan tool requires three input variables to run a test. All three values will be tested and verified to ensure the accuracy and test for correct functionality of the tool. The properties for the three variables are shown in Table 3.4.

Table 3.4

Variables Required by the Tool for Scanning a Drive

Variable Name	Type	Static/Dynamic	Description
Standard Test Time	Time	Static (4 minutes)	Four minute time period
Sector Offset	Numeral	Dynamic	A number which is automatically calculated by the tool based on the total number of sectors divided by the number of scan cycles
Byte Sample	Numeral	Static (8 bytes)	The first eight bytes of each sector

To verify that the process runs for the predetermined time period, the author will time each of the scanning processes based on the standard test time of four minutes. However, the author predicts that a standard test time of four minutes may be difficult to achieve because of factors relating to varying hard drive technologies, particularly data transfer rates. The standard test time is calculated by timing a scanning process with a predefined number of scan cycles set. By using the results, the number of scan cycles is calculated and adjusted accordingly to modify the processing time to equal a four minute period.

The sector offset is a numeral variable calculated by dividing the total number of sectors by the number of scan cycles set for a test. The sector offset will be a unique value depending on the physical size of the hard drive. For example, a larger drive will have a higher value for the sector offset as opposed to a smaller drive that will have a smaller value. To verify that the correct sectors are being read and processed, the author will include built-in *verification code* during the development of the scan tool. Data from seven sectors is extracted by the tool's verification code. The sectors include the first and last sectors, and five randomly selected sectors. Verification of the data involves opening the drive that has been scanned by a 3rd party analysis tool such as EnCase or FTK Imager and comparing the data within the sector with the sector data which has been extracted by the scan tool's verification code during the scan process. If the sector numbers are the same and the content of both datasets are identical, then the scan tool has read and processed the correct sector.

In Section 3.3.1.2, the byte sample was defined as the first eight bytes of a read sector. The testing of the first, last and randomly selected sectors will include verifying the byte sample of each of the seven sectors. Byte sample verification ensures that when the scanning tool is running, the correct sector is being read and the correct bytes in the sector are being processed. If the byte samples are different under comparison analysis, then the scan tool has read and processed the wrong sector and amendments to the software need to be made before full testing takes place. Verification testing will be conducted during the development of the software to ensure the finished product operates as proposed.

3.3.1.5 Maintenance

The final phase of the software development life cycle is the maintenance phase. The maintenance phase includes making small changes in the software to meet changing requirements. For example, the amount of scan cycles will need to be adjusted, depending on the source drive, to ensure that a scan runs for four minutes. Alternatively, the byte sample may need to be modified to assist with improving the performance and/or accuracy of the tool. Errors that have been identified would be corrected in the maintenance phase. The maintenance phase also allows for new requirements to be submitted, although the new requirements will be included into the next revision of the software.

3.3.2. Testing Requirements

Section 3.3.1.4 discusses how the three required input variables integrate with the software to produce a result when a scan is run. More importantly, Section

3.3.1.4 explains how the three values, standard test time, sector offset and byte sample are tested and verified. Testing is performed by using two analysis tools. The first tool is the scanning tool itself with its built in verification function. Yao and Wang (2005) conducted research into testing distributed software components and used a technique that embeds BIT software into source code to enhance self-testing and run time testing. Similar to Yao and Wang's testing method the author will include extra code or modules within the scanning tool during development to enable extraction of data during a scan process. The extracted data will be used to verify the functionality and accuracy of the scanning tool. The data that has been extracted during a scan will be verified using a secondary analysis tool.

3.3.2.1 Testing Environment

The testing environment will comprise standalone computer systems and hard drive test subjects. A variety of hard drives will be tested during the research. Each of the hard drives being tested will be different in regards to technological age, capacity and data throughput rates. Multiple standalone computer systems will be used for the testing phase. Using multiple computer systems will allow the tests to be spread out and run on several systems simultaneously, resulting in reduced overall time to run all of the tests for the testing phase of the research. However, hard drive test subjects are tested several times and therefore the test subject will be tested on the same computer system to ensure that a consistent testing environment is maintained during testing of that particular drive.

The requirements for the standalone computer systems include multiple connection methods for the hard drive test subjects. The connection methods include USB, IDE and SATA interfaces. The computer systems will have Windows XP installed as the scanning tool will be initially developed to operate on Windows XP. The hard drive test subjects will comprise IDE and SATA hard drives that will be connected directly to the system or through a USB connection if mounted in a drive caddy. The research will conduct all testing on mechanical hard drives that do not possess any solid state drive (SSD) memory. The size of the hard drive test subjects will range from 20GB to 2TB.

3.3.2.2 Testing Process

The testing is in accordance with the main research question where the results of the two scans, standard scan and full sector scan, will determine whether a hard drive can be scanned in a measured period of time and produce a reliable result. Hard drive test subjects will not be compared against each other because variables, such as data throughput rates and faster interface connection types, will influence the results. All tests will be documented and the test data will be collected in a report file by the scanning tool. The test data will be analysed to draw conclusions to answer the research question for each of the hard drive test subjects.

Additional testing in relation to the two sub-questions will be carried out. Firstly, the testing will involve running standard tests over the hard drive test subjects. However, on this occasion, the test will be run for half the amount of time. The two minute test results will be compared with the results from the standard tests to determine whether there is a significant difference in the reported ratio of empty and used sectors. Secondly, additional tests will be conducted whereby the byte sample will be modified to scan and process all 512 bytes in each sector as opposed to only eight bytes. In accordance with the second sub-question the results of the 512 byte tests will be compared with the results of the eight byte tests for the full sector scans to determine the extent, if any, of increased processing time. In regards to the fixed length tests, i.e. standard test and two minute tests, the number of processed sectors will be analysed to estimate processing speed differences.

3.4. DATA COLLECTION AND ANALYSIS

When designing a hard drive scanning tool, a great deal of testing is required to ensure that the tool will function exactly the way the developer intends the tool to function. Before the outputs and results are collected and analysed, the developer will ensure that the input variables are accurate and the functional processing of those variables is correct. The proposed input variables will be either statically hard-coded into the scanning tool or dynamically calculated whilst the software operates. Analysis of the input variables is discussed in Section 3.3.1.4.

3.4.1. Analysis of the Results

The output will be analysed to satisfy the author that the process built into the scanning tool that produces the output will function correctly. However, manually examining many millions of sectors to determine the integrity of the output would take an extremely long time and is therefore outside the scope of the research. Verifying the output from a scanning process will be facilitated in the same fashion as the verification process of the input variables detailed in Section 3.3.1.4 by utilising built-in verification code. Data from five randomly selected sectors along with the first and last sectors will be extracted by the verification code. The data will be analysed using a 3rd party analysis tool. In addition, the output verification code will also check the data in the sector and produce an indication regarding the sector's *state* (i.e. whether the sector is used or empty) for each of the seven selected sectors. When the data in the sector is manually verified using the 3rd party analysis tool, the usage *state* will also be checked and verified.

3.4.1.1. Testing for a Reliable Result

Achieving a reliable result was not part of the software requirements. However, discovering the reliability of the results is part of the research and can only be accomplished once the scanning program has been developed. Only when data from the tests have been collected and analysed, can the author determine whether the results were reliable or not.

Once the data processing and output have been examined and verified to be correct, as detailed in Section 3.4.1, the examiner will run tests to determine whether the results for a standard test are reliable. Section 3.2.2 generated some clarity to the research question in terms of a reliable result. A reliable scan result was defined as a result from a standard scan that has achieved 90% or more accuracy of an accurate result. An accurate result is produced during a *full sector scan* where all available sectors on a disk are read and processed. In other words, an accurate result is produced from sampling 100% of the data contained on the disk.

To establish whether the result from a standard scan is reliable, a full sector scan of the same device is required. The results can then be compared to each other to determine whether the result from the standard scan falls within

10% of the result produced by a full sector scan and can be considered reliable. The scanning tool will have an option that will allow all of the sectors on the drive to be read and processed. The option to scan all of the sectors on a hard drive is purely for testing and verification purposes. The author's hypothesis states that there may be some variance when comparing a full sector scan with a standard scan, although the variance is predicted to be small. In addition, the accuracy of the result from a standard scan can have up to 10% variance of the result from a full sector scan before the result from the standard scan becomes unreliable.

Figure 3.6 shows hypothetical results from a standard scan and a full sector scan. To determine whether the standard scan is within 10% of the full sector scan, the author divides the smaller percentage of *highly compressible data* (the standard scan in this case) into the larger percentage of highly compressible data. The calculated result from the hypothetical example in Figure 3.6 shows that the results from the standard scan is over 93% accurate and is therefore considered to be a reliable result.

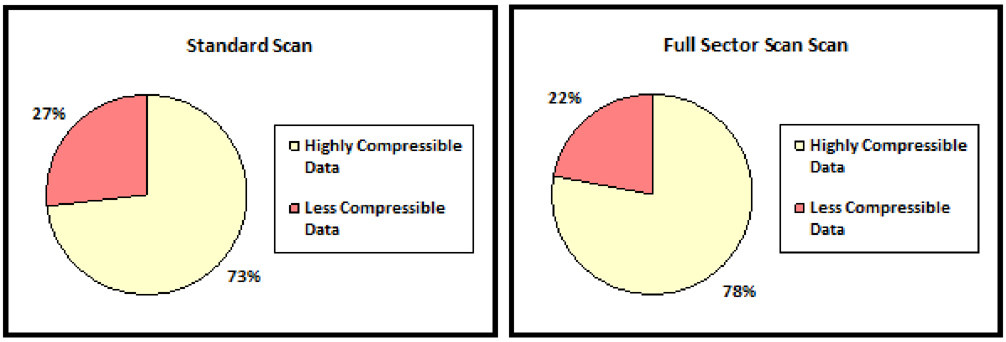


Figure 3.6. Hypothetical results from a standard and a full sector scan

3.4.1.2. Expected Outcomes

The proposed data scanning tool used to determine the type of data on a hard drive is believed to be unique. The author's research has failed to identify such a tool that has the capability to determine the type of data that resides on a storage device. Because no information, studies or trials have been discovered, determining the expected outcomes from the testing phase of the research will be difficult.

Section 3.2.3 discusses a hypothetical situation where a 20GB hard drive that contains 5GB of standard system data and the rest of the space on the drive is completely wiped of data. Physically the drive is one quarter full and three quarters empty. Determining the expected outcome from a test in this case would be straightforward. In this scenario one could logically assume that the result from a scan would indicate that the drive is one quarter full and three quarters empty. However, not all hard drives will have their data stored in this way. Furthermore, predicting the outcome of a scan process on a drive where the examiner has no knowledge of the type of data would be difficult.

The size of the hard drive test subjects will provide more certainty in the expected outcomes. Given that the standard test is four minutes, the results of a scan on a smaller capacity drive is believed to be more accurate as opposed to a scan on a larger capacity drive. Therefore, the author predicts that the accuracy of the results for a standard test will deteriorate as tests are run on drives with larger capacities.

3.5. LIMITATIONS

Section 3.5 draws attention to the limitations with the research and proposed disk scanning tool. Section 3.5.1 discusses the tool's standard scan and full sector scan settings and how they should be used. Section 3.5.2 examines the accuracy of the tool. Given that a standard scan of a hard drive does not read all of the sectors, the author believes that the tool will not be one hundred percent accurate and should be used solely to gain an idea of what type of data is stored on a hard drive. Section 3.5.3 discusses future work and suggestions that would improve the accuracy of the tool are put forward.

3.5.1. Completeness

The final version of the disk scanning tool will have options where an examiner can override the standard four minute test. For example, if an examiner wanted to run an eight minute test, the examiner could enter the desired parameters for the test. Alternatively, if an examiner wanted to scan every sector on a particular drive then the full sector scan option could be selected. There will always be a trade-off between time and the number of sectors scanned. For example, an eight minute test would obviously take twice as much time to process compared to a four minute test. However, the size of the sector offset will be halved which increases the granularity and therefore will produce a more accurate result. How much more accurate the result would be is unknown as tests would have been conducted and results will have to be analysed to discover the answer. In terms of completeness, the scanning tool has the capability to scan every sector of a drive. Scanning every sector of a drive is critical for the author's testing methodology; however, scanning every sector is not how the tool is intended to be used in the field.

The disk scanning tool should be used as a tool for triaging hard drives in a timely manner and therefore the author recommends that an examiner uses the standard test time when conducting scan processes in the field. The term *triage* has many meanings. A commonly known meaning is used in medical situations and is defined as, *a process for sorting injured people into groups based on their need for or likely benefit from immediate medical treatment* (The Free Dictionary, 2011). However, the term triage used by the author in this text is more to do with the process of determining how digital evidence items are ranked, in terms of importance or priority, for collection. In order for the examiner to work out how electronic storage devices should be imaged in the most efficient manner, the examiner needs to know two variables. The first variable is the capacity of the drive and the second variable is the type of data on the drive. The term triage, used by the author in reference to the tool, refers to a simplistic method of gathering and processing storage device information to produce a summary of sizes and data types for all devices. By analysing the summary, the examiner can make an informed decision regarding the best order and approach to image the storage devices.

3.5.2. The Accuracy of the Results

The author predicts that the results returned by the scanning tool will vary according to the size of the hard drive being scanned. The tool is designed to scan a sample of sectors from the beginning to the end of the hard drive. The sampled sectors are evenly spaced apart so the range covers an area from the first sector, *sector 0,* to the last sector, *sector end,* as shown in Figure 3.7.

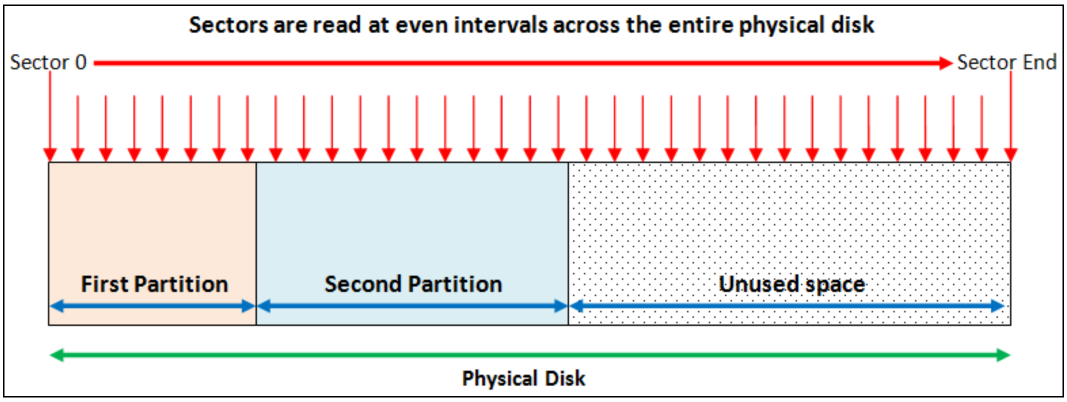


Figure 3.7. A representation of sectors being read across an entire drive

Although Figure 3.7 shows only thirty-five sectors being read, in reality, tens of thousands of sectors will be read during a standard scan process. Briefly discussed in Section 3.4.1.2, the author believes that the size of the hard drive will have an effect on the accuracy of the results. To complete a standard test in four minutes, a larger drive will have a larger sector offset, which in turn creates a lower ratio of total scan cycles to total sectors on the drive. The lower ratio of scan cycles will produce a less accurate result because more data is being missed as opposed to a smaller drive with a higher ratio of scan cycles to total sectors. In-depth discussion regarding the accuracy of the testing is detailed in the Discussion section of Chapter 5.

As mentioned in Section 3.3.1.2, when a scanning process is completed, a summary of the overall results is displayed on the screen in the form of a pie chart. The results are separated into two categories named *highly compressible data* and *less compressible data,* as shown in Figure 3.6. Fundamentally the categories show the percentage of scanned sectors that contain different valued bytes and the percentage of scanned sectors that contain potentially empty sectors or sectors where all bytes have potentially the same value based on a scan using a byte sample of eight bytes in size. What is not mentioned is that the sectors that contain different types of data, labelled as *less compressible data*, may well be able to be compressed. Therefore, data from a drive half-filled with highly compressible Microsoft Word documents for example, would compress a great deal more than the same drive half filled with mp3 music files, which are less compressible. A great deal of compressible files, such as Word documents, would be beneficial to the examiner. The files would compress to a smaller size and could be transferred across a slow data connection faster than if they were uncompressed. However, the research is focused on identifying empty, highly compressible sectors and therefore determining the rate of compression for particular file types is outside the scope of the research.

3.6. CONCLUSION

The time required to create a forensic bit-for-bit copy of an electronic storage device is dramatically increasing due to the advancements of hard drive technology that enable drives to be built with huge storage capacities. Chapter 3 defines a methodology to build and test a software scanning tool that can assist an examiner to determine what type of data is stored on a hard drive. Knowing the type of data stored on a drive can assist the examiner in reducing image processing times. A review of previous work related to software development and testing by five researchers was conducted in Section 3.1 to gain software development and testing knowledge. The knowledge gained will be used in the development phase of the scanning tool. The research question, *"Can a drive scanning tool process a hard drive in a measured period of time and produce a reliable result?"* is defined and the associated hypothesis is discussed. Additional sub-questions are discussed. The design of the scanning tool is developed incorporating a software development life cycle to ensure that the scanning tool has minimal development problems. The testing requirements have been defined and are discussed in Section 3.3.2. The testing requirements create a boundary where the testing is contained within. When the tool is used, data is collected during the scanning process. The collection of data is part of a verification mechanism where the data is verified to ensure the input and the output of the tool are correct and can be relied on. Finally, limitations of the proposed software tool and future work are discussed.

The specification for the drive scanning tool is outlined in Chapter 3 and is used to develop the proposed drive scanning tool. Once the scanning tool has been developed and is fully operational, the testing phase will begin. Data from the tests will be captured and the results of the testing will be analysed and reported in Chapter 4.

Chapter 4

REPORT OF RESEARCH FINDINGS

4.0. INTRODUCTION

Several innovative ideas that could assist an examiner in the data collection phase of the forensic process were discussed in Section 2.6. One of the ideas, a data scanning tool, was identified in Chapter 3 as being a viable topic for further research and was chosen for the subject of the thesis. The research question was formulated based around the attributes of accuracy and efficiency of the proposed scanning tool. A set of logical hypothesis were developed. The subsequent research methodology was defined with expected theoretical outcomes forecasted and limitations to the specified research explained.

Chapter 4 reports the research findings in regards to the results from the testing phase. Section 4.1 discusses the scanning tool including the inbuilt functions. Section 4.2 reports alterations to the original methodology specified in Chapter 3. Section 4.3 presents the results from the testing phase of the tool and provides an analysis of the results. The chapter finishes with a summary in Section 4.4.

4.1. TOOL DEVELOPMENT AND OPERATION

Section 4.1 provides an overview of the scanning tool. Section 4.1.1 reports on the development of the tool. Section 4.1.2 provides a step by step process detailing the operation of the tool. Section 4.1.3 provides detailed examination of a report that is automatically generated once the scanning process has completed.

4.1.1. Tool Development

The author developed a scanning tool which is used to scan a storage device to determine the type of data that is stored on the device. The scanning tool complies with the four *operating requirements* set out in Section 3.3.1.1. The tool is written and compiled in Microsoft Visual Basic version 6 (VB6.0). VB6.0 was used because VB6.0 is an object-based programming language and it enables the rapid build of applications. Other software compilers were evaluated but VB6.0 provides sufficient flexibility and support to achieve the software build.

The structure and functionality of the scanning tool was drawn up on a whiteboard and incorporated the operating requirements listed in Section 3.3.1.1. The *operational flow chart,* shown in Figure 3.5, was produced from analysing the information drawn on the whiteboard. The scanning tool was developed using the operational flow chart as a guide and satisfies the four operating requirements.

The first operating requirement for the tool is to run a scanning process for a period of four minutes. The scanning tool allows the examiner to specify the number of sectors to scan for a test. For a particular hard drive, entering the correct number of sectors will allow the examiner to perform a four minute scan.

The second operating requirement is to scan a comprehensive sample of data during a standard scan. The tool is designed to start scanning from the first available sector and scan at evenly spaced intervals (the sector offset) up until the last sector of a storage device. An evenly distributed scan across the device will provide a comprehensive sample of data during a scanning process. The research included identifying a method to programmatically access an electronic storage device at sector level.

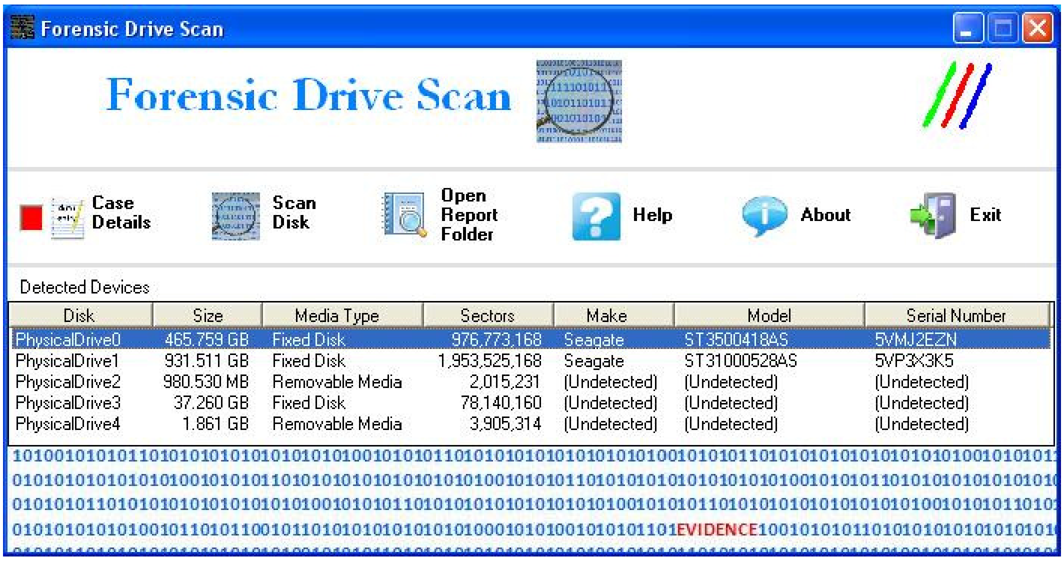
The third operating requirement is to process as little data as possible while maintaining a level of accuracy. The tool is designed to process the first eight bytes of each scanned sector. Processing only eight of the 512 bytes for each sector is believed to reduce the processing time of a scan and therefore will allow more sectors to be scanned within a four minute period.

The forth operating requirement is reporting. Forensic best practices include creating notes of items and processes so that the information is available at a later date if required. The tool automatically saves the test results, test parameters and drive information into a report file on the examiner's workstation when a scan is complete. Although the automatic saving of the data is not mentioned in the methodology, the automation of reports is discussed in Section 4.2.3 which reports alterations to the methodology. The information saved in the report file meets the *reporting requirements* set out in Section 3.3.1.1. The information written to the report is from several sources. The sources include data that is manually entered by the examiner, data that is generated by a scanning process and data that is read directly from the drive.

4.1.2. Tool Operation

The scanning tool is compiled into a distribution package and is installed onto the examiner’s forensic workstation by using the executable setup file. Once the program is installed, the examiner connects the source hard drive to a write blocking device and then connects the write blocking device to the workstation. Although the scanning tool does not have code that allows writing to a source drive, computer forensic best practices need to be followed. The author recommends that the examiner makes use of a write blocking device when using the scanning tool. Using a write blocking device will eliminate the data on the source drive being altered. Once the source drive has been powered on, the scanning tool can be operated and the source drive can be scanned.

Figure 4.1 shows a screen capture of the scanning tool when the tool is executed. On execution, the tool automatically scans for storage devices that are connected to the system. In this example, Figure 4.1 shows that five devices have been detected and the devices are displayed in the *detected devices* list. The detected devices comprise three hard drives, shown as *fixed disk* in the *media type* column and two *removable media* devices.

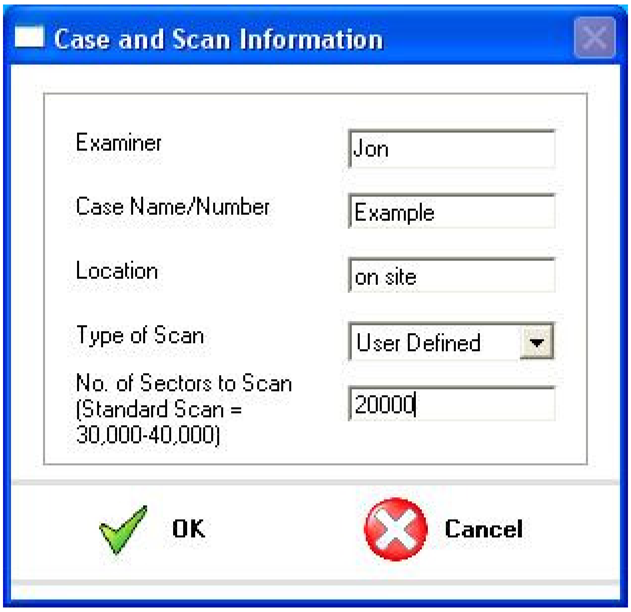


Detected Devices

Disk	Size	Media Type	Sectors	Make	Model	Serial Number
PhysicalDrive0	465.759 GB	Fixed Disk	976,773,168	Seagate	ST3500418AS	5VMJ2EZN
PhysicalDrive1	931.511 GB	Fixed Disk	1,953,525,168	Seagate	ST31000528AS	5VP3X3K5
PhysicalDrive2	980.530 MB	Removable Media	2,015,231	(Undetected)	(Undetected)	(Undetected)
PhysicalDrive3	37.260 GB	Fixed Disk	78,140,160	(Undetected)	(Undetected)	(Undetected)
PhysicalDrive4	1.861 GB	Removable Media	3,905,314	(Undetected)	(Undetected)	(Undetected)

Figure 4.1. The scanning tool displaying all detected devices

Operational simplicity has been built into the tool’s graphical interface. To scan a drive, the examiner selects a device from the detected devices list. Once a device has been selected, the examiner clicks the *case details* button. A case details window opens, shown in Figure 4.2, which enables the examiner to enter information regarding the case and type of scan to be performed. Information such as the examiner’s name, case name or number and test location are entered into the case details window. Figure 4.2 shows an example of the case details window open with the fields populated. By default the *type of scan* is set to *standard test;* however, the examiner has the option to select the *user defined* option from the dropdown menu, which the examiner has done in this particular case. Because the examiner has selected the *user defined* option, the examiner will be required to enter a numeral for the number of sectors that the examiner wishes to scan. Figure 4.2 shows that the examiner has entered 20,000 sectors into the field.



Examiner: Jon

Case Name/Number: Example

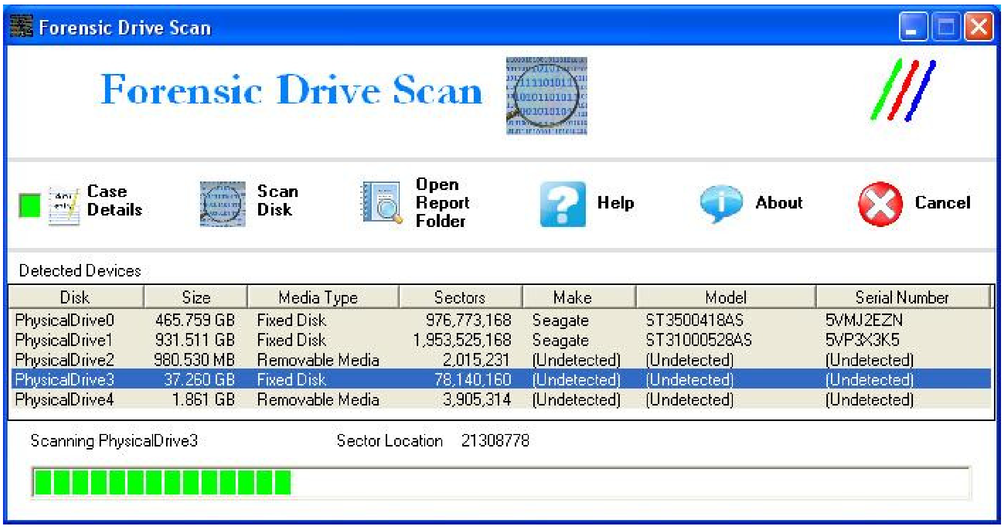
Location: on site

Type of Scan: User Defined

No. of Sectors to Scan (Standard Scan = 30,000-40,000): 20000

Figure 4.2. The case details window with populated fields

Once the case details have been entered, the red indicator on the case details button turns green, as shown in Figure 4.3, to indicate that valid data has been entered into the fields. The *detected devices* list is disabled and the drive is ready to be processed. During a scanning task, a progress bar, located at the bottom of the main window, provides the examiner with information regarding the progress of the scanning task. Information concerning the sector location is shown in Figure 4.3 just above the progress bar. The sector location information is purely for software testing and troubleshooting purposes and will be removed in future development of the scanning tool.



Detected Devices

Disk	Size	Media Type	Sectors	Make	Model	Serial Number
PhysicalDrive0	465.759 GB	Fixed Disk	976,773,168	Seagate	ST3500418AS	5VMJ2EZN
PhysicalDrive1	931.511 GB	Fixed Disk	1,953,525,168	Seagate	ST31000528AS	5VP3X3K5
PhysicalDrive2	980.530 MB	Removable Media	2,015,231	(Undetected)	(Undetected)	(Undetected)
PhysicalDrive3	37.260 GB	Fixed Disk	78,140,160	(Undetected)	(Undetected)	(Undetected)
PhysicalDrive4	1.861 GB	Removable Media	3,905,314	(Undetected)	(Undetected)	(Undetected)

Scanning PhysicalDrive3 Sector Location 21308778

Figure 4.3. A screen capture of a hard drive being scanned

When the scanning process is complete, the *scan results* window is automatically displayed. The scan results window shows a summary of the results in a pie chart form. The percentages of highly compressible and less compressible data are shown for the quantity of empty and used sectors respectively. In this example, Figure 4.4 shows that 1% of the scanned sectors on PhysicalDrive3 are empty and 99% of the scanned sectors contain data. Although 99% of the scanned sectors contain data, the amount and the type of data contained within those sectors is not determined by the scanning tool. The scanning tool is designed purely to report whether the scanned sectors contain data or not. The examiner has the option to open the automatically generated full text report using the *open report* button from the scan results window. An explanation of the automatically generated text report and its contents is detailed in Section 4.1.3.

The tool has been tested with drives connected to internal SATA and PATA ports, e-SATA ports and USB ports. The testing phase of the research was conducted where the tests have been performed without the use of write blocking hardware or write blocking software for ease of testing. However, both the functionality and output of the scanning tool have been verified to operate and report correctly with SATA and PATA drives connected through a *Tableau* T35e forensic bridge in write blocked mode. In addition, the scanning tool has been verified to operate correctly with a drive connected to the USB port and the system running *USB Write Protect Version 1.0* developed by Joz Ong of the New South Wales Police in Australia (USB Write Protect, 2005). USB Write Protect is a software write blocking utility. If the examiner chooses to connect a source drive directly to an e-SATA port or a USB port on a forensic workstation, the author recommends that the examiner maintains forensic standards by using software enabled write blocking technologies, such as the abovementioned USB Write Protect software or Safe Block from ForensicSoft (2010). The use of write blocking technologies will minimise the risk of changes being made to the data on the source drive.

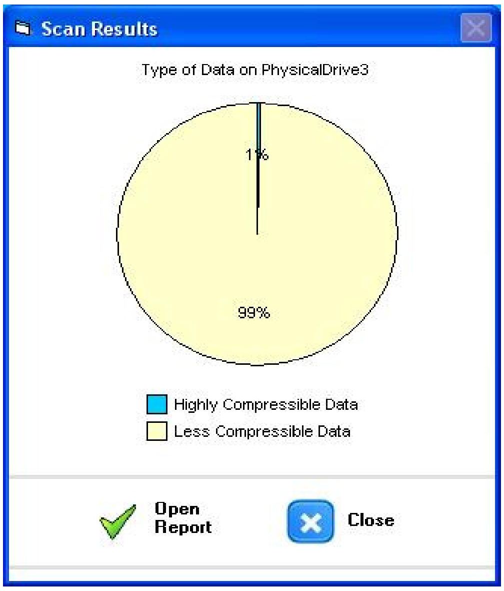
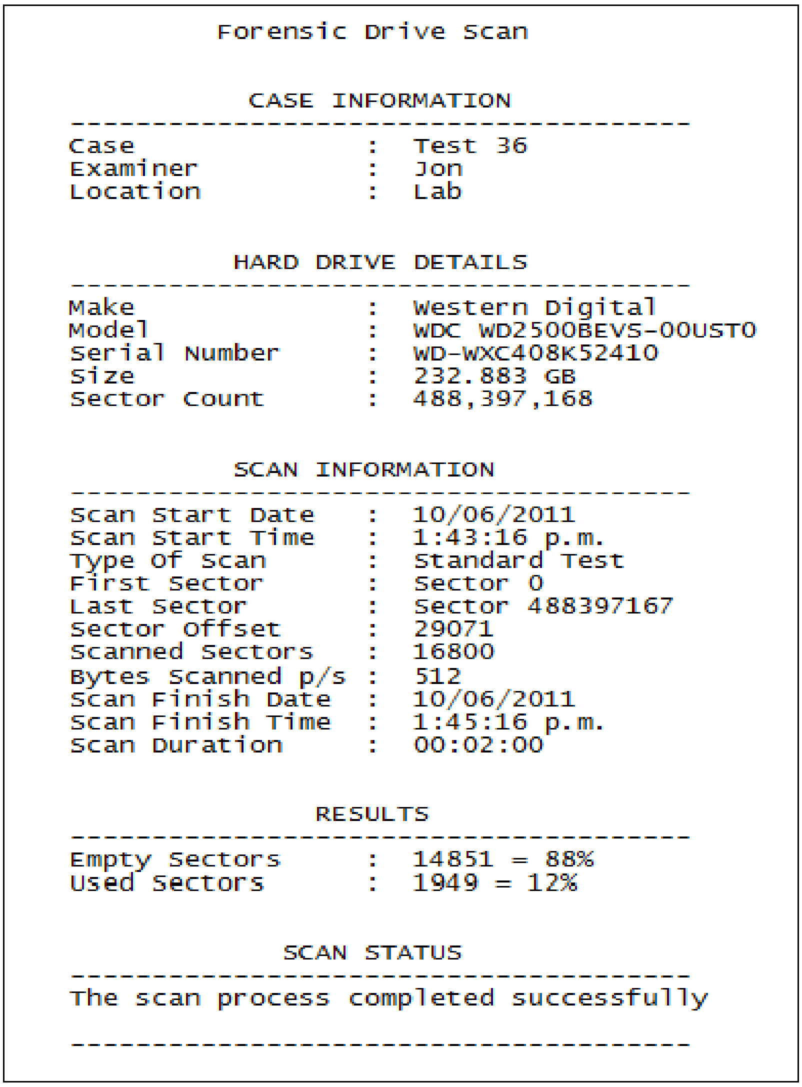


Figure 4.4. The scan results window

4.1.3. Output

Tables 4.4 and 4.5 are populated with data taken from the reports generated by each of the 48 tests that were performed during the testing phase. Figure 4.5 shows the report that was created for Test 36. The report is divided into five sections of relevant information. The first section shows the *case information*. Data for the case information fields are entered before the examiner starts the scanning process. The case information fields comprise a case name or number, the examiner's name and the location where the scan is taking place, i.e. on site or in the laboratory.

The next section, *hard drive details,* shows information about the hard drive. Some of the information, such as the model and serial number, is automatically read from the Self-Monitoring Analysis and Reporting Technology (SMART) monitoring system built into the drive. Reading data from the SMART occurs when the scanning tool is executed and the tool is detecting devices that are connected to the system. If a device is not SMART enabled or the system cannot access SMART, the make, model and serial number fields, in both the detected devices window and the report, will be populated with *Undetected.* Figure 4.3 shows that SMART could not be accessed for PhysicalDrive3 and the three fields are populated with Undetected. Issues have been identified where SMART cannot be read via a USB connection. A discussion regarding the issues with SMART and USB connected devices is provided in Chapter 5. However, Figure 4.5 shows a situation where SMART could be accessed on a drive and the make, model and serial number information is included in the report. Other information such as the *total sector count* and *logical size* are not obtained via SMART and are always available in the report.



```
             Forensic Drive Scan

               CASE INFORMATION
    --------------------------------------
    Case                 :  Test 36
    Examiner             :  Jon
    Location             :  Lab


              HARD DRIVE DETAILS
    --------------------------------------
    Make                 :  Western Digital
    Model                :  WDC WD2500BEVS-00UST0
    Serial Number        :  WD-WXC408K52410
    Size                 :  232.883 GB
    Sector Count         :  488,397,168


              SCAN INFORMATION
    --------------------------------------
    Scan Start Date      :  10/06/2011
    Scan Start Time      :  1:43:16 p.m.
    Type Of Scan         :  Standard Test
    First Sector         :  Sector 0
    Last Sector          :  Sector 488397167
    Sector Offset        :  29071
    Scanned Sectors      :  16800
    Bytes Scanned p/s    :  512
    Scan Finish Date     :  10/06/2011
    Scan Finish Time     :  1:45:16 p.m.
    Scan Duration        :  00:02:00


                   RESULTS
    --------------------------------------
    Empty Sectors        :  14851 = 88%
    Used Sectors         :  1949 = 12%


                 SCAN STATUS
    --------------------------------------
    The scan process completed successfully

    --------------------------------------
```

Figure 4.5. The report generated for Test 36

The section of the report regarding the *scan information* contains data relating to the scan process. The scan information fields include *start date and time* fields that represent when the scan was started. The date and time information is captured from the system clock and is therefore subject to verification by the examiner. The *type of scan* field shows which option has been selected by the examiner. In this example the examiner has selected *standard test* for the scan type. The *sector offset* is calculated by dividing the number of sectors to be scanned into the total number of sectors on the drive. The number of bytes to be processed per sector, *byte sample*, is a fixed number and is set to eight bytes for the initial testing phase. The byte sample was modified and set to 512 bytes to conduct additional testing to answer the second sub-question. Finally the *finish date and time* is shown and the *duration* for the scan is calculated from the start and finish date and time fields.

The results section shows the percentage of scanned sectors that are empty or considered highly compressible, and the percentage of sectors that contain data and are therefore less compressible. The overall result is the key piece of information that the examiner is seeking. The report shown in Figure 4.5 confirms that 12% of the scanned sectors contain data and 88% of the scanned sectors are empty. According to Cusack and Pearse (2011), the overall processing time can be reduced by using a level of compression when imaging a drive that is up to 80% full of data. The testing by Cusack and Pearse was conducted using drives connected to USB ports.

4.2. ALTERATIONS TO THE METHODOLOGY

There are three main areas where the methodology defined in Chapter 3 has been modified. The changes were made to accommodate issues that were encountered while conducting the testing phase of the research. Section 4.2.1 reports the issues met while testing large capacity hard drives. Section 4.2.2 examines the accuracy of the standard and two minute scan time periods and a tolerance is introduced for the two time periods. Changes to the tool's reporting functionality are discussed in Section 4.2.3.

4.2.1. Hard Drive Test Subjects

The proposed testing included running the scanning tool over a selection of different hard drives. Nine conventional, non solid state, hard drives were selected as test subjects for the research. The nine drives provide a broad range of capacities and a variety of performance technologies for the research. The original data that resided on the selected hard drives was left *intact* and was not manually manipulated or altered in any way. The hard drive test subjects are listed in Table 4.1.

Table 4.1

Hard Drive Test Subjects Selected for the Research

HDD Number	Make	Model	Logical Size	Interface
1	IBM	IBM-DJSA-220	20GB	ATA
2	Toshiba	MK3017GAP	30GB	ATA
3	Samsung	SV4002H	40GB	ATA
4	Seagate	ST380021A	80GB	ATA
5	Seagate	ST3120022A	120GB	ATA
6	Western Digital	WD2500BEVS-00UST0	250GB	SATA
7	Seagate	ST3500418AS	500GB	SATA
8	Seagate	ST31000528AS	1000GB	SATA
9	Seagate	ST32000542AS	2000GB	SATA

Table 4.1 lists a total of nine hard drives which were selected for the testing phase of the research. The drives range in capacity from 20GB to 2000GB (2TB) providing an overall size difference of one hundred times between test subject 1 and test subject 9. The test subjects possess a wide range of technologies that have been developed over the last ten years. These technologies include a transition from the Advanced Technology Attachment (ATA) data interface, shown in Table 4.1 for test subjects 1 to 5, to the SATA interface shown for test subjects 6 to 9. The maximum data transfer rate for the ATA standard is 133MB/s. The maximum data transfer rate for the SATA test subjects listed in Table 4.1 is 300MB/s, which is over double the transfer rate of ATA drives. However, data transfer rates are much higher for current devices. According to Seagate (2011), the *Seagate Barracuda XT 3TB* is capable of data transfers at a rate of 600MB/s.

There were a total of six tests, comprising of two sets of three tests, performed on each test subject. Each set of three tests includes a two minute test, a standard (four minute) test and a full sector scan. Every sector on the hard drive is read and processed during a full sector scan. The first set of tests is run where only the first eight bytes in each scanned sector are processed. The second set of tests is similar to the first set of tests, apart from one modification that is made to the scanning tool. The tool is modified so that all of the data in each scanned sector is processed as opposed to only processing the first eight bytes. On the larger capacity drives, the duration of the full sector scans have extended for several days.

Table 4.5 shows that the 512 byte, full sector scan for test subject 8 took over 97 hours to process. The full sector scan was run on a new computer system with quad core processing power and 24GB of RAM. Using the same rate of time for test subject 8 as a guide, the expected processing time for a 2TB hard drive would be 194 hours. However, the new computer system was not available for the expected test duration so a decision was made not to test the 2TB hard drive and test subject 9 was removed from the list of drives to be tested. Although the 2TB hard drive was not tested during the research, the results from testing the remaining eight drives produced sufficient data for analysis.

4.2.2. Standard Scan Duration

In Section 3.2.2, the standard test time was defined as a four minute time period. The standard test time is achieved by modifying the number of scan cycles until a test runs for four minutes. The method used to calculate the correct number of scan cycles is as follows: A scan is run and the number of scan cycles and the duration of the scan are noted. The *scan cycles per second* (SCPS) are calculated by dividing the number of seconds into the number of scan cycles. Multiplying the SCPS by 240, which is the number of seconds in four minutes, will theoretically provide the correct number of scan cycles required to produce a scan duration of four minutes for that particular drive. However, calculating the correct number of scan cycles using this method was not accurate in practice for any of the test subjects. To achieve the correct duration for a standard scan, several scans were run on a test subject. By increasing or decreasing the number of scan cycles, using a *trial by error* process, a scan duration of four minutes was eventually achieved. The *trial by error* process was replicated for each test subject and is shown in Table 4.2. Scan A represents the initial test scan for test subject 1. Subsequent scans, Scan B onwards, are run with altered scan cycles until a scan duration of four minutes is achieved. Some of the final results, which are shown in the blue rows in Table 4.2, are one second either side of the four minute scan duration. These results are explained later in the section.

Table 4.2

Four Minute Scan Durations for Test Subjects 1 to 8

Hard Drive	Scan	Scan Cycles	Scan Duration (mins)	Seconds	Normal Distribution
Test subject 1 (20GB)	A	26400	00:04:20	260	0.011876437
	B	23400	00:03:32	212	0.004626222
	C	25000	00:03:52	232	0.021967276
	D	25150	00:03:49	229	0.019336175
	E	25300	00:03:50	230	0.020260308
	F	25500	00:03:56	236	0.024567068
	G	25650	00:03:54	234	0.023424948
	H	26000	00:04:16	256	0.015833419
	I	25800	00:04:01	241	0.02572875
Test subject 2 (30GB)	A	24820	00:04:18	258	0.01382752
	B	24100	00:03:59	239	0.025574865
Test subject 3 (40GB)	A	31720	00:04:04	244	0.025163712
	B	31680	00:04:04	244	0.025163712
	C	31580	00:04:07	247	0.023706548
	D	31480	00:04:08	248	0.023047243
	E	31380	00:04:06	246	0.024283468
	F	31280	00:04:09	249	0.022313244
	G	31100	00:04:08	248	0.023047243
	H	30900	00:04:01	241	0.02572875
Test subject 4 (80GB)	A	35400	00:03:07	187	6.35534E-05
	B	39400	00:04:27	267	0.006117363
	C	37700	00:04:00	240	0.025705111
Test subject 5 (120GB)	A	40425	00:04:00	240	0.025705111
Test subject 6 (250GB)	A	28000	00:03:45	225	0.015388607
	B	29000	00:03:33	213	0.005202593
	C	33000	00:04:21	261	0.010938221
	D	31100	00:03:57	237	0.025002317
	E	31200	00:04:02	242	0.025645488
	F	31150	00:04:04	244	0.025163712
	G	31100	00:03:57	237	0.025002317
	H	31125	00:04:01	241	0.02572875
Test subject 7 (500GB)	A	40000	00:04:20	260	0.011876437
	B	38000	00:03:37	217	0.007982225
	C	39000	00:03:47	227	0.017393974
	D	39650	00:04:11	251	0.020655199
	E	39320	00:03:59	239	0.025574865
Test subject 8 (1000GB)	A	39600	00:03:57	237	0.025002317
	B	39650	00:03:56	236	0.024567068
	C	40000	00:04:36	276	0.001932068
	D	39800	00:04:11	251	0.020655199
	E	39750	00:04:06	246	0.024283468
	F	39730	00:04:04	244	0.025163712
	G	39715	00:03:59	239	0.025574865

Analysis of the duration data in Table 4.2 was carried out to identify the normal distribution. The scan durations were converted into seconds for each of the tests and added to the *Seconds* column in Table 4.2. The average processing time, or mean, was calculated over all of the tests by using the data in the Seconds column.

The average for the processing time is 240.72 seconds. The standard deviation was calculated from the duration data to measure the variability in the dataset. The standard deviation is ±15.50 seconds. The normal distribution was calculated for each test by using the duration data and standard deviation parameters and is included in Table 4.2. The normal distribution is shown as a bell curve along with two standard deviations in Figure 4.6.

The standard deviation segments shown in Figure 4.6 show how much variation there is from the mean. Within a normal distribution, 68.27% of the data will fall within 1 Sigma (σ) which is shown in Figure 4.6 as one standard deviation line either side of the mean. 2σ accounts for 95.45% of the data and is shown as two standard deviation lines either side of the mean. From a total of 43 tests shown in Table 4.2, 31 tests fell within 1σ and 41 tests fell within 2σ.

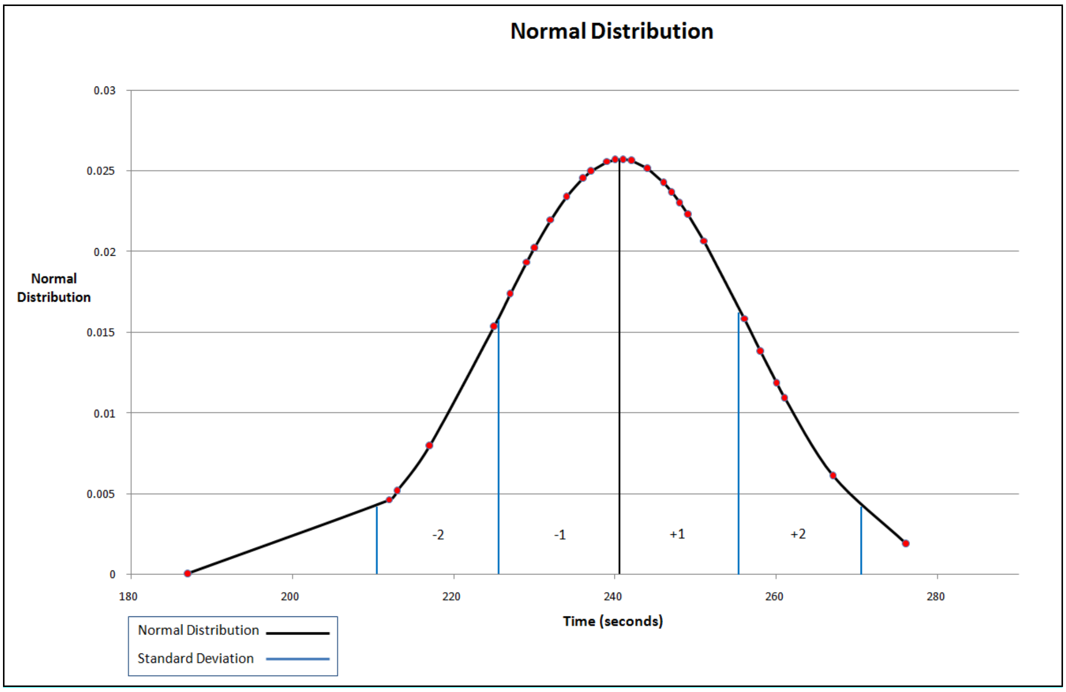


Figure 4.6. Normal distribution showing two standard deviations

Anomalies were identified in the duration data collected from several hard drive test subjects. The anomalies are in regards to the relationship between the number of scan cycles and the overall duration of the scan. As mentioned in Section 4.2.2, calculating the exact number of scan cycles to produce a scan duration of four minutes was not accurate in practice and manual adjustments to the number of scan cycles were required in order to produce a four minute duration. For example, Table 4.2 shows that 28,000 scan cycles were set for Test A on test subject 6. The scan duration for the test is 225 seconds and the SCPS for the test is 124.44. However, multiplying the SCPS by 240 seconds produced 1,258 fewer scan cycles than what was required to produce a scan duration of four minutes.

Some of the results from manually adjusting the scan cycles were inconsistent. In some cases, a reduction in the number of scan cycles actually increased the duration of the scan. For example, Table 4.2 shows that 25000 scan cycles were set for Scan C of test subject 1. Scan C produced a scan duration of 3:52 mins. Although the scan cycles were increased to 25150 for Scan D, the resulting duration for Scan D decreased to 3:49 mins. However, despite the anomalies, Figure 4.8 shows evidence of a linear relationship between the number of scan cycles and the scan duration for each test subject.

The anomalies were also evident on several different computer systems used for the testing and not confined to a single computer system. To achieve a level of consistency when testing, efforts were undertaken to ensure that the state of the testing platforms, which are shown in Table 4.3, were set to the same baseline before testing each test subject. The following tasks were performed on the testing systems to create the baseline.

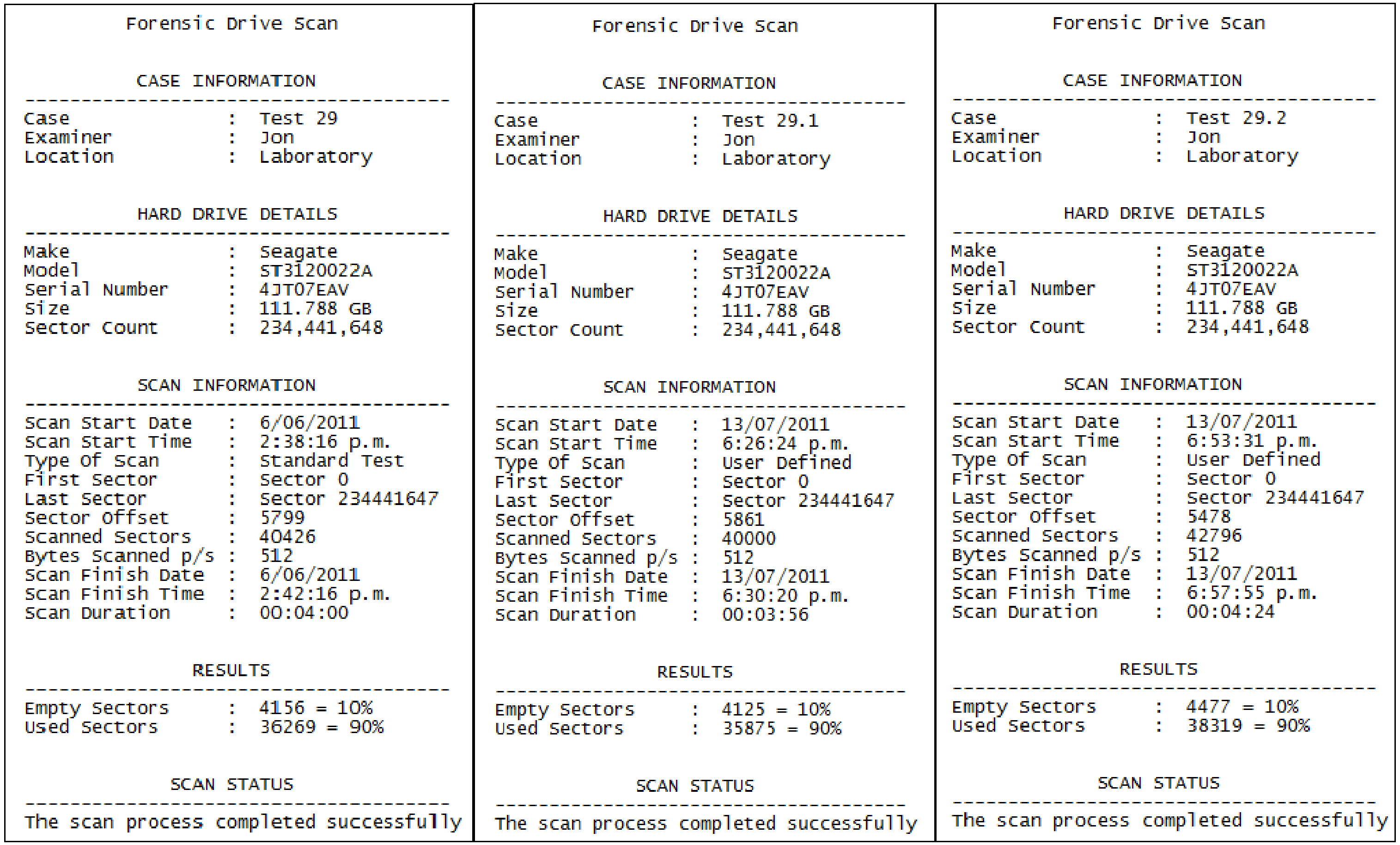
- Disable programs from running in the background.
- Cease usage of the test systems during the tests.
- Reboot the systems between each of the tests.
- After rebooting the systems, give the systems ample time to fully load the operating system and other services prior to testing.

Table 4.3

Testing Platforms

Test System	Operating System	CPU	RAM	Interface	Test Subject	Comments
1	Window XP Pro, SP3	P4 3.0GHz	2GB	SATA/PATA	1, 2, 3, (6), 7	Tests 31, 32, 33, 35 and 36 for test subject 6 were run on this test system
2	Window XP Pro, SP3	P4 3.0GHz	1GB	SATA/PATA	5	
3	Window XP Pro, SP3	P4 2.0GHz	1GB	SATA/PATA	4	
4	Window XP Pro, SP3	6 Core i5	24GB	SATA	8 (6)	Test 34, the 512 byte full sector scan for test subject 6 was run on this test system

Due to the abovementioned timing anomalies, difficulties were experienced during the testing phase to achieve a four minute time period for several of the standard scan tests. Figure 4.6 shows that although there is not a high concentration of points around the mean, most of the tests fall within one standard deviation of the mean. The final scan durations for most of the test subjects shown in Table 4.2 were either 3:59mins or 4:01mins. When these tests were re-run using the same number of scan cycles, some tests finished with different results. Because of the timing anomalies and the difficulties encountered when creating a four minute scan duration, the author has modified the original *standard test time* requirements. The author has allowed for a tolerance of ± 1 second for the four minute time period being acceptable for the testing phase. The same timing anomalies were present when attempting to achieve a time period for the two minute tests therefore, the ± 1 second tolerance has also been applied to the two minute tests. Table 4.2 shows that only two of the eight test subjects had a *standard test time* of exactly four minutes.



```
           Forensic Drive Scan

            CASE INFORMATION
-------------------------------------
Case                 :  Test 29
Examiner             :  Jon
Location             :  Laboratory

           HARD DRIVE DETAILS
-------------------------------------
Make                 :  Seagate
Model                :  ST3120022A
Serial Number        :  4JT07EAV
Size                 :  111.788 GB
Sector Count         :  234,441,648

            SCAN INFORMATION
-------------------------------------
Scan Start Date      :  6/06/2011
Scan Start Time      :  2:38:16 p.m.
Type Of Scan         :  Standard Test
First Sector         :  Sector 0
Last Sector          :  Sector 234441647
Sector Offset        :  5799
Scanned Sectors      :  40426
Bytes Scanned p/s    :  512
Scan Finish Date     :  6/06/2011
Scan Finish Time     :  2:42:16 p.m.
Scan Duration        :  00:04:00

                RESULTS
-------------------------------------
Empty Sectors        :  4156 = 10%
Used Sectors         :  36269 = 90%

               SCAN STATUS
-------------------------------------
The scan process completed successfully
```

```
           Forensic Drive Scan

            CASE INFORMATION
-------------------------------------
Case                 :  Test 29.1
Examiner             :  Jon
Location             :  Laboratory

           HARD DRIVE DETAILS
-------------------------------------
Make                 :  Seagate
Model                :  ST3120022A
Serial Number        :  4JT07EAV
Size                 :  111.788 GB
Sector Count         :  234,441,648

            SCAN INFORMATION
-------------------------------------
Scan Start Date      :  13/07/2011
Scan Start Time      :  6:26:24 p.m.
Type Of Scan         :  User Defined
First Sector         :  Sector 0
Last Sector          :  Sector 234441647
Sector Offset        :  5861
Scanned Sectors      :  40000
Bytes Scanned p/s    :  512
Scan Finish Date     :  13/07/2011
Scan Finish Time     :  6:30:20 p.m.
Scan Duration        :  00:03:56

                RESULTS
-------------------------------------
Empty Sectors        :  4125 = 10%
Used Sectors         :  35875 = 90%

               SCAN STATUS
-------------------------------------
The scan process completed successfully
```

```
           Forensic Drive Scan

            CASE INFORMATION
-------------------------------------
Case                 :  Test 29.2
Examiner             :  Jon
Location             :  Laboratory

           HARD DRIVE DETAILS
-------------------------------------
Make                 :  Seagate
Model                :  ST3120022A
Serial Number        :  4JT07EAV
Size                 :  111.788 GB
Sector Count         :  234,441,648

            SCAN INFORMATION
-------------------------------------
Scan Start Date      :  13/07/2011
Scan Start Time      :  6:53:31 p.m.
Type Of Scan         :  User Defined
First Sector         :  Sector 0
Last Sector          :  Sector 234441647
Sector Offset        :  5478
Scanned Sectors      :  42796
Bytes Scanned p/s    :  512
Scan Finish Date     :  13/07/2011
Scan Finish Time     :  6:57:55 p.m.
Scan Duration        :  00:04:24

                RESULTS
-------------------------------------
Empty Sectors        :  4477 = 10%
Used Sectors         :  38319 = 90%

               SCAN STATUS
-------------------------------------
The scan process completed successfully
```

Figure 4.7. Three reports from tests run on test subject 5

Although the scan durations for most of the tests were one second either side of the two and four minute time periods, the one second tolerance has had no effect on the final results. To test whether the output is affected by the tolerance, two additional tests, Tests 29.1 and 29.2, were created and manually configured to run well under and well over the four minute time periods respectively. Figure 4.7 shows the reports that were automatically generated for Tests 29, 29.1 and 29.2. Although the duration shown in the report for Test 29.1 is 3:56 and the reported duration for Test 29.2 is 4:24, the reported percentages for empty and used sectors for both, tests 29.1 and 29.2 are identical to Test 29.

The time difference between tests 29.1 and 29.2 is twenty eight seconds, which equates to Test 29.2 processing 7% more scan cycles compared to Test 29.1. Although the number of scan cycles that were processed for Test 29.2 is 7% more compared to Test 29.1, the empty and used sector ratios reported by the scanning tool remain the same. Because the reported results are the same when Test 29.2 has run for twenty eight seconds longer than Test 29.1, the author has determined that a tolerance of one second would most probably not be significant to change the reported percentages of a standard or two minute test.

Figure 4.8 shows a scatter plot chart representing a linear regression of the scan duration vs. scanned sectors for the four minute tests. The data used to generate the linear regression derives from the *scan cycles* and *seconds* columns in Table 4.2. The data was recorded when the author was testing for four minute scan periods for each of the test subjects. The *Y* axis shows a time scale measured in seconds. Figure 4.8 shows that the average or mean is slightly over 240 seconds which is four minutes.

The test subjects are represented as HDD*x* where *x* is the number of the test subject. A linear regression is calculated for each test subject and is shown in Figure 4.8. The linear regression demonstrates the closeness of the linear relationship between the scan duration and the scanned sectors. Outliers for HDD4 and HDD8 are evident well below and well above the mean respectively.

Two standard deviations are shown on Figure 4.8. As expected, most of the results fall within two standard deviations.

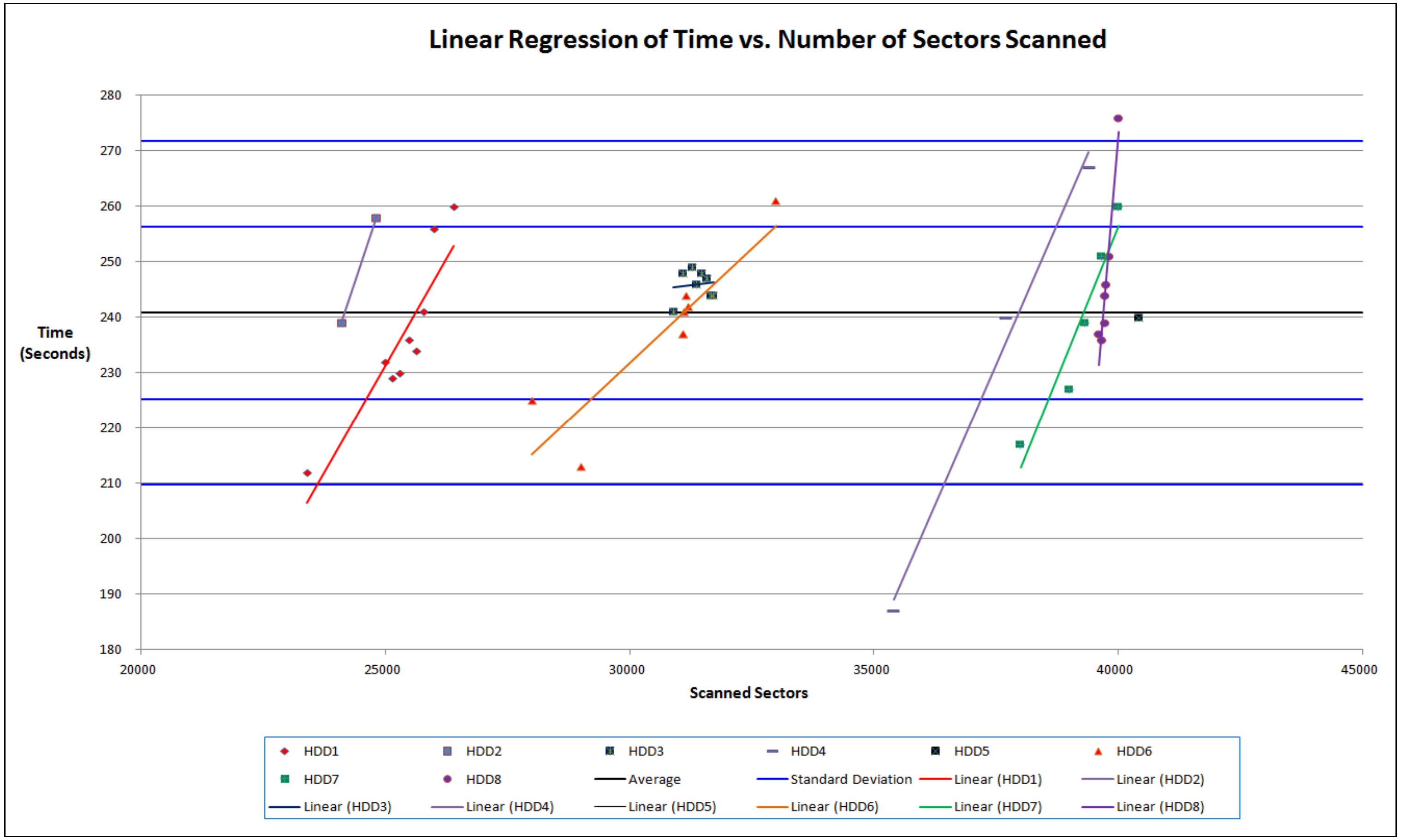


Figure 4.8. Linear regression of time versus number of sectors scanned for test subjects 1 to 8

4.2.3. Reporting

The *Operational Flow Chart* in Section 3.3.1.4 shows that the scanning tool's proposed reporting function was in the form of two questions. Upon completion of a scanning process, the examiner would be asked whether the examiner wishes to view a summary. When the summary is displayed on the screen, the examiner would be asked whether the examiner would like to save the report to disk. If the examiner chooses to save the report to disk, a *save file* dialog would be opened so the examiner can navigate to a folder location and enter a file name for the report.

The reporting function of the scanning tool was modified in two ways. Firstly, the two questions, to view the summary and to save the report, are not asked. Secondly, the summary and report are automatically generated when the test is complete. The changes were made to firstly reduce the tasks an examiner has to perform when using the tool and secondly, automatically writing the report to disk reduces the risk of losing the report information by securing a copy on the hard drive when the scanning process is complete. Securing the reports at the conclusion of the tests was vital during the testing phase of the research, particularly while undertaking scanning activities that stem into many hours or several days as they did during the testing. With the originally proposed reporting method, if a report had not been manually saved by the examiner and a power failure occurs, the report data will be lost resulting in the same test having to be re-processed. This particular scenario occurred during the testing phase.

With the modified reporting function, the summary is automatically displayed on the screen post processing and the report is written to the *Reports* folder which is located in the installation path of the application. The format for the report filename is 'Report_*(device serial number)_(date, yyyy-mm-dd)_(time, hh-mm-ss)*.txt'. For example, Report_5VMJ2EZN _2011-06-08_19-51-48.txt. The reports for the testing phase are included in the appendix. The reports produced by the tests have been renamed to make it easier for the author to identify which test a specific report belongs.

4.3. TEST RESULTS AND FINDINGS

All test results are included in the appendix of the thesis. There were a total of 48 tests run on eight hard drive test subjects. The test subjects range from 20GB to

1TB in size and contain a combination of SATA and PATA interfaces. Section 4.3.1 reports on the expected outcomes from the testing as detailed in Chapter 3. The field testing methodology is discussed in Section 4.3.2. Section 4.3.3 shows the results of the field tests conducted in the testing phase. Data is extracted during a scan process to verify the operation of the tool. Section 4.3.4 reports on both the verification method used during the testing and accuracy of the results.

4.3.1. Expected Outcomes

In Chapter 3 the author expressed some uncertainty surrounding the expected outcomes for the testing phase of the research. The uncertainty is generated because there does not appear to be any previous studies or trials available on the topic. Predicting the outcome of the tests has proved difficult. However, the author expected that the accuracy of the scan results would deteriorate as the hard drive test subject's logical size or capacity increases. The theory behind the prediction is based on the fact that a standard scan for each test subject has a similar number of scan cycles, regardless of the drives capacity. Although the number of scan cycles is similar, the ratio of scanned sectors to total sectors on a drive is lower as the drive capacity increases. For example, if two drives were scanned with 10,000 scan cycles each and one drive has twice the capacity, the ratio of scanned sectors and total sectors would be half for the larger drive.

Both Table 4.4 and Table 4.5 show a column for the *percentage of total sectors* which were scanned during each test. Table 4.4 shows that the percentage of total sectors scanned for Test 2, the 20GB drive, is 0.0676%. Table 4.5 shows that Test 38, the 500GB drive, has scanned and processed 0.004% of the total sectors on that particular drive. Test 2 has processed over fifteen times more of its overall sectors compared to Test 38.

Although reading and processing fifteen times more data on the 20GB drive is an astounding difference compared to the 500GB drive, the reported results for the standard and full sector tests are identical for each drive. The comparison of the results for the full sector and standard tests, over all hard drive test subjects, showed that the author's expectation that the accuracy of a scan result would deteriorate as the drive capacities increase was incorrect for the dataset used for testing.

4.3.2. Results From the Field Testing

The results from the 48 field tests are shown in Tables 4.4 and 4.5. Section 4.1.1 mentioned there were a total of six tests performed on each test subject. Table 4.4 shows information regarding the six tests for hard drive test subjects 1 to 4. Similarly, Table 4.5 shows the information for hard drive test subjects 5 to 8.

Section 3.3.1.1 discusses the software requirements for the scanning tool. The third software requirement is to process as little data for each sector as practicable without compromising the accuracy of the results. To assist with achieving the software requirement, only the first eight bytes of each read sector were processed for the first three tests run on each test subject. The first test which is run is a *full sector scan* where every sector on the drive is processed. The full sector scan provides a benchmark for accuracy and the maximum processing time that would be encountered on each test subject. During the analysis of the results, other test results can be compared with the benchmark to determine the accuracy and the performance of those tests. The second test that is run is a *standard scan.* The standard scan runs for a period of four minutes, regardless of the capacity of the hard drive. The third test is a two minute test. The two minute tests are performed to provide information that will be used to answer the first research sub-question.

The remaining three tests for each test subject are identical to the first three tests. However, they are carried out where all 512 bytes of each read sector are processed as opposed to processing only the first eight bytes. Full sector scans using the 512 byte option create an accurate result. The result is accurate because every sector on the drive is fully read and processed. In addition, every byte in an empty sector is processed to ensure that the sector is, in fact, empty. The full sector, 512 byte scan provides an account for all empty sectors on the drive and presents an accurate report of the *empty sector* percentage.

Table 4.4

Results from Field Tests 1 to 24

Test Subject	Test Information								Results			
HDD Number & Size	Test Number	Scan Type	Duration	Sectors Scanned	Sector Offset Size (MB)	Percentage of Total Sectors Scanned	Data Read During Scan	Bytes Processed/ps	Empty Sectors	Used Sectors	Reported Percentage Empty/Used	Percentage - 2 Decimal Places Empty/Used
1 20GB	1	Full Sector	02:01:59	39,070,080	0.000512	100.0000%	20GB	8	18,256,609	20,813,471	47% / 53%	46.73% / 53.27%
	2	Standard	00:04:01	26,398	0.76	0.0676%	14MB	8	12,335	14,063	47% / 53%	46.72% / 53.28%
	3	2 Minute	00:02:00	12,290	1.63	0.0315%	6MB	8	5,741	6,549	47% / 53%	46.71% / 53.29%
	4	Full Sector	06:30:52	39,070,080	0.000512	100.0000%	20GB	512	18,186,584	20,883,496	47% / 53%	46.55% / 53.45%
	5	Standard	00:04:01	25,801	0.78	0.0660%	13MB	512	12,003	13,797	47% / 53%	46.52% / 53.48%
	6	2 Minute	00:02:01	12,290	1.63	0.0315%	6MB	512	5,720	6,570	47% / 53%	46.54% / 53.46%
2 30GB	7	Full Sector	01:11:25	58,605,120	0.000512	100.0000%	30GB	8	51,486,577	7,118,543	88% / 12%	87.85% / 12.15%
	8	Standard	00:04:01	24,821	1.21	0.0424%	13MB	8	21,821	2,999	88% / 12%	87.91% / 12.09%
	9	2 Minute	00:01:59	9,999	3.00	0.0171%	5MB	8	8,788	1,211	88% / 12%	87.89% / 12.11%
	10	Full Sector	11:00:42	58,605,120	0.000512	100.0000%	30GB	512	51,446,800	7,158,320	88% / 12%	87.79% / 12.21%
	11	Standard	00:03:59	24,097	1.25	0.0411%	12MB	512	21,147	2,950	88% / 12%	87.76% / 12.24%
	12	2 Minute	00:02:01	9,999	3.00	0.0171%	5MB	512	8,783	1,216	88% / 12%	87.84% / 12.16%
3 40GB	13	Full Sector	03:22:41	78,242,976	0.000512	100.0000%	40GB	8	50,670,865	27,572,111	65% / 35%	64.76% / 35.24%
	14	Standard	00:03:59	35,694	1.12	0.0456%	18MB	8	23,124	12,570	65% / 35%	64.78% / 35.22%
	15	2 Minute	00:02:01	18,049	2.22	0.0231%	9MB	8	11,694	6,355	65% / 35%	64.79% / 35.21%
	16	Full Sector	12:45:16	78,242,976	0.000512	100.0000%	40GB	512	50,661,104	27,581,872	65% / 35%	64.75% / 35.25%
	17	Standard	00:04:01	30,901	1.30	0.0395%	16MB	512	20,013	10,887	65% / 35%	64.76% / 35.24%
	18	2 Minute	00:02:01	15,861	2.53	0.0203%	8MB	512	10,268	5,592	65% / 35%	64.74% / 35.26%
4 80GB	19	Full Sector	07:44:52	156,301,488	0.000512	100.0000%	80GB	8	1,829,033	154,472,455	01% / 99%	01.17% / 98.83%
	20	Standard	00:03:59	37,699	2.12	0.0241%	19MB	8	445	37,254	01% / 99%	01.18% / 98.82%
	21	2 Minute	00:02:00	18,360	4.36	0.0117%	9MB	8	224	18,136	01% / 99%	01.22% / 98.78%
	22	Full Sector	08:05:29	156,301,488	0.000512	100.0000%	80GB	512	857,068	155,444,420	01% / 99%	00.55% / 99.45%
	23	Standard	00:04:00	37,699	2.12	0.0241%	19MB	512	229	37,470	01% / 99%	00.61% / 99.39%
	24	2 minute	00:01:59	18,360	4.36	0.0117%	9MB	512	97	18,263	01% / 99%	00.53% / 99.47%

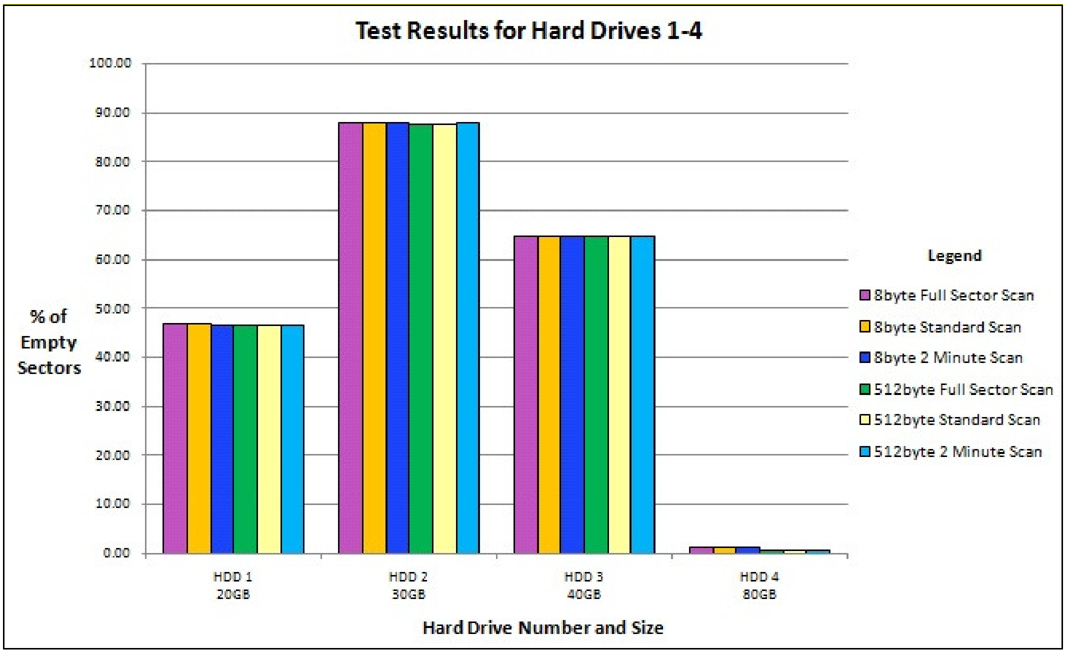


Figure 4.9. Percentage of empty sectors reported for Test Subjects 1-4

The reported percentages for empty sectors from Tables 4.4 and 4.5 have been charted to help make the data easier to interpret. Figure 4.9 shows the charted data for hard drive test subjects 1 to 4 and Figure 4.10 shows the charted data for hard drive test subjects 5 to 8. Both charts show a pattern for each of the hard drive test subjects that were tested. For the six tests processed over each hard drive test subject, the empty/used sector ratio is noticeably similar, regardless of which type of test was performed. For example, Figure 4.9 shows that the percentage of empty sectors for hard drive test subject 3 is the same for all six tests. Table 4.4 shows the percentages in more detail. The column that shows the *percentage calculated to two decimal places* shows that, for test subject 3, all empty sector percentages range from 64.74% to 64.79%. The difference over all six tests is 0.05%. Fundamentally, the standard scan produced the same reported result as the full sector scan. The difference is that the full sector scan ran for over three hours and twenty minutes compared to four minutes of the standard scan.

Table 4.5

Results from Field Tests 25 to 48

Test Subject	Test Information								Results			
HDD Number & Size	Test Number	Scan Type	Duration	Sectors Scanned	Sector Offset Size (MB)	Percentage of Total Sectors Scanned	Data Read During Scan	Bytes Processed/ps	Empty Sectors	Used Sectors	Reported Percentage Empty/Used	Percentage - 2 Decimal Places Empty/Used
5 120GB	25	Full Sector	07:32:07	234,441,648	0.000512	100.0000%	120GB	8	27,088,947	207,352,701	12% / 88%	11.55% / 88.45%
	26	Standard	00:03:59	40,511	2.96	0.0173%	21MB	8	4,631	35,879	11% / 89%	11.43% / 88.57%
	27	2 Minute	00:01:59	20,189	5.95	0.0086%	10MB	8	2,403	17,786	12% / 88%	11.90% / 88.10%
	28	Full Sector	12:59:40	234,441,648	0.000512	100.0000%	120GB	512	24,109,201	210,332,447	10% / 90%	10.28% / 89.72%
	29	Standard	00:04:00	40,425	2.97	0.0172%	21MB	512	4,156	36,269	10% / 90%	10.28% / 89.72%
	30	2 Minute	00:02:01	20,189	5.95	0.0086%	10MB	512	2,136	18,053	11% / 89%	10.58% / 89.42%
6 250GB	31	Full Sector	14:36:06	488,397,168	0.000512	100.0000%	250GB	8	431,836,668	56,560,500	88% / 12%	88.42% / 11.58%
	32	Standard	0:04:01	31,359	7.97	0.0064%	16MB	8	27,724	3,635	88% / 12%	88.41% / 11.59%
	33	2 Minute	0:02:01	16,930	14.77	0.0035%	9MB	8	14,970	1,960	88% / 12%	88.42% / 11.58%
	34	Full Sector	56:04:41	488,397,168	0.000512	100.0000%	250GB	512	431,783,507	56,613,661	88% / 12%	88.41% / 11.59%
	35	Standard	00:04:01	31,125	8.03	0.0064%	16MB	512	27,517	3,608	88% / 12%	88.41% / 11.59%
	36	2 Minute	00:02:00	16,800	14.88	0.0034%	9MB	512	14,851	1,949	88% / 12%	88.40% / 11.60%
7 500GB	37	Full Sector	26:04:11	976,773,168	0.000512	100.0000%	500GB	8	330,618,809	646,154,359	34% / 66%	33.85% / 66.15%
	38	Standard	00:03:59	39,319	12.72	0.0040%	20MB	8	13,306	26,013	34% / 66%	33.84% / 66.16%
	39	2 Minute	00:02:01	19,529	25.61	0.0020%	10MB	8	6,596	12,933	34% / 66%	33.78% / 66.22%
	40	Full Sector	53:36:40	976,773,168	0.000512	100.0000%	500GB	512	303,937,162	672,836,006	31% / 69%	31.12% / 68.88%
	41	Standard	00:03:59	39,599	12.63	0.0041%	20MB	512	12,357	27,242	31% / 69%	31.21% / 68.79%
	42	2 Minute	00:01:59	19,799	25.26	0.0020%	10MB	512	6,178	13,621	31% / 69%	31.20% / 68.80%
8 1TB	43	Full Sector	36:44:23	1,953,525,168	0.000512	100.0000%	1TB	8	778,267,428	1,175,257,740	40% / 60%	39.84% / 60.16%
	44	Standard	00:03:59	39,750	25.16	0.0020%	20MB	8	15,824	23,926	40% / 60%	39.81% / 60.19%
	45	2 Minute	00:01:59	20,000	50.01	0.0010%	10MB	8	7,992	12,008	40% / 60%	39.96% / 60.04%
	46	Full Sector	97:38:58	1,953,525,168	0.000512	100.0000%	1TB	512	774,714,546	1,178,810,622	40% / 60%	39.66% / 60.34%
	47	Standard	00:03:59	39,714	25.19	0.0020%	20MB	512	15,763	23,951	40% / 60%	39.69% / 60.31%
	48	2 Minute	00:02:00	19,859	50.37	0.0010%	10MB	512	7,865	11,994	40% / 60%	39.60% / 60.40%

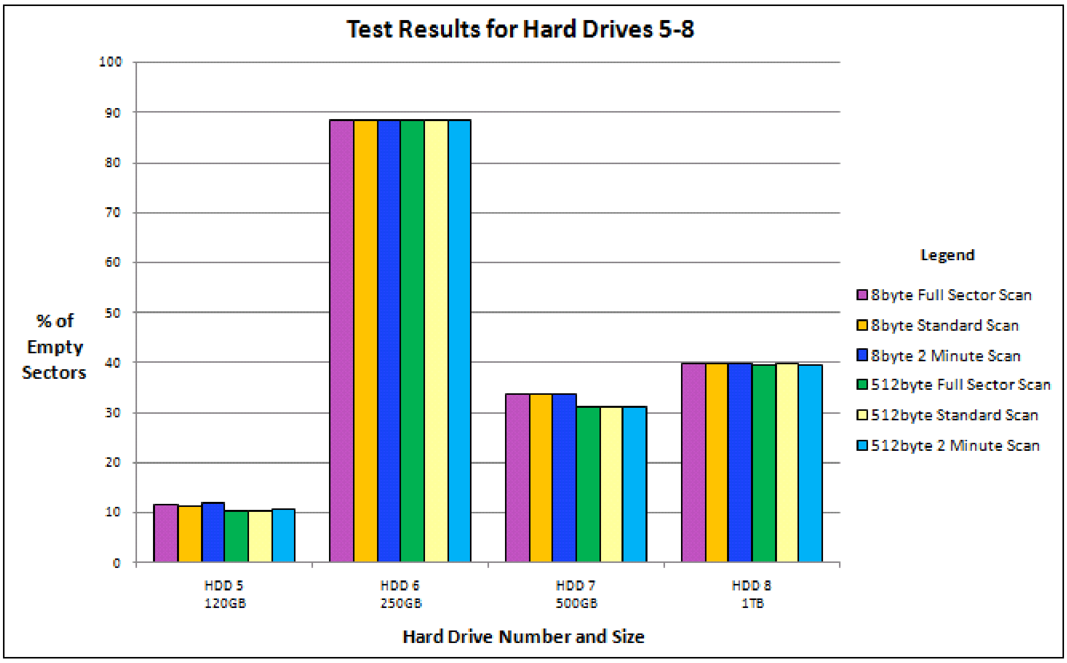


Figure 4.10. Percentage of empty sectors reported for Test Subjects 5-8

The results of empty sector percentages for test subjects 5 to 8, shown in Figure 4.10, are similar to the results shown in Figure 4.9. Test subject 6, in Figure 4.10, shows a similar trend to test subject 3, where the reported empty sector percentages are identical for all six tests. Reviewing the data in Table 4.5 for test subject 6 shows that the empty sector percentages range from 88.40% to 88.42% over the six tests. The difference in this case is 0.02%. Again, the four minute scan produced the same reported result as the full sector scan for test subject 6. However, the full sector scan ran for over fourteen hours and twenty six minutes.

In Figure 4.10, test subject 7 shows a distinctive difference between the eight byte tests and the 512 byte tests. Table 4.5 shows that the empty sector percentages for the eight byte tests range from 33.78% to 33.85%. However, the percentages for the three 512 byte tests range from 31.12% to 31.20%. The average difference between the eight byte and 512 byte tests is 2.64% for test subject 7. Theoretical explanations why these variations exist in the reported percentages, between the eight byte and 512 byte full sector tests, are discussed in Chapter 5.

Figure 4.11 shows the total number of sectors scanned for the full sector and standard scans for each test subject. Logically, the number of sectors scanned for the full sector scans is equal to the number of sectors on each test subject. Figure 4.11 shows that the total number of sectors that were scanned for the test subjects range from forty million to two billion sectors. Tables 4.4 and 4.5 show that the time to process every sector on each test subject ranges from two hours for test subject 2 to almost thirty seven hours for test subject 8.

Figure 4.11 shows that the number of total sectors scanned for the standard tests is significantly smaller than the number of sectors scanned for the full sector tests. The *Y* axis in Figure 4.11 shows a logarithmic scale for the number of sectors scanned. The number of sectors scanned for the standard scans range from 23,000 to 40,000 whereas the full sector scans range from 39 million to 2 billion sectors.

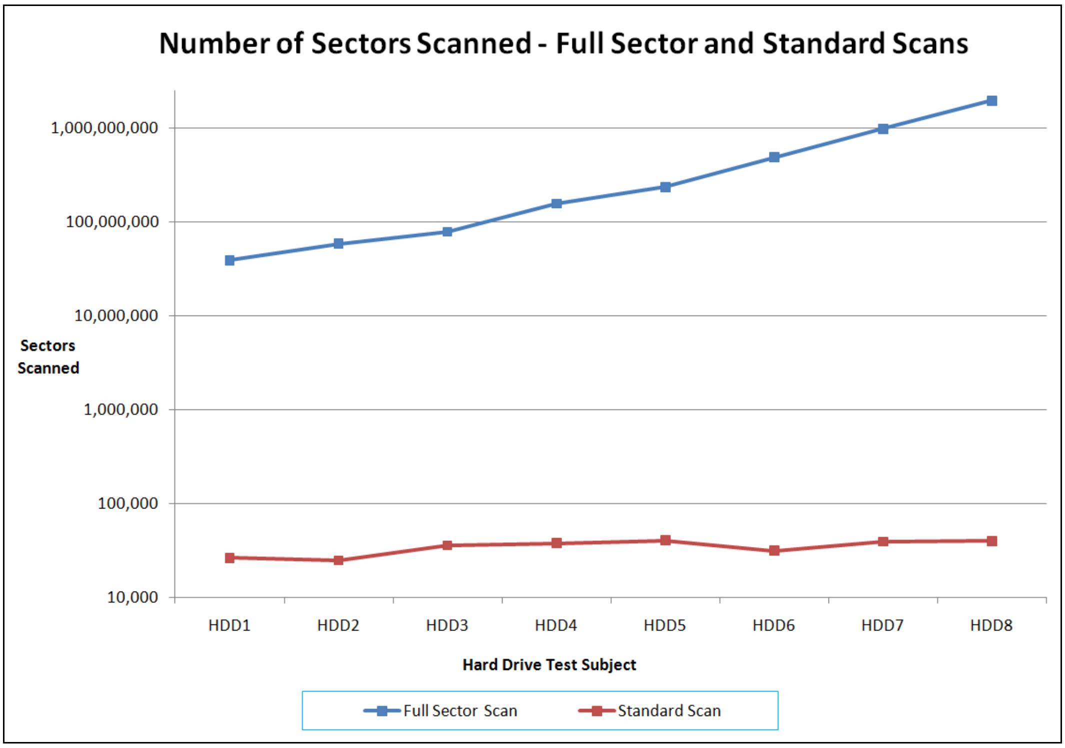


Figure 4.11. Number of sectors scanned in full sector scans and in standard scans

Figure 4.12 shows the error percentage taken from the results of the full sector and standard scans. The error percentage is calculated by subtracting the result of a standard scan from the result of a full sector scan. The result from a full sector scan is used because the result is accurate, as described in Section 4.3.2. However, because a standard scan reads and processes a significantly small number of sectors on a drive, the results are believed to be less accurate compared to the result of a full sector scan. Therefore, the difference between the two results is the *error percentage*.

Figure 4.12 shows that the error percentages for all standard scans are exceptionally low. An exceptionally low error percentage means that the accuracy of a standard scan is extremely high. For example, test subject 8, shown as HDD8 in Figure 4.12, has an error percentage of 0.03%. The low error percentage equates to test subject 8s standard scan being 99.97% accurate. Although test subject 8s standard scan is 99.97% accurate, the duration of the scan is only 03:59 minutes as opposed to the duration of the full sector scan which took almost thirty seven hours to process. In relation to Figure 4.11, Figure 4.12 shows the same metric for the standard scans, however, the metric is shown in the form of a *percentage for the total number of sectors scanned* for each test subject as opposed to the number of sectors scanned during the standard scans.

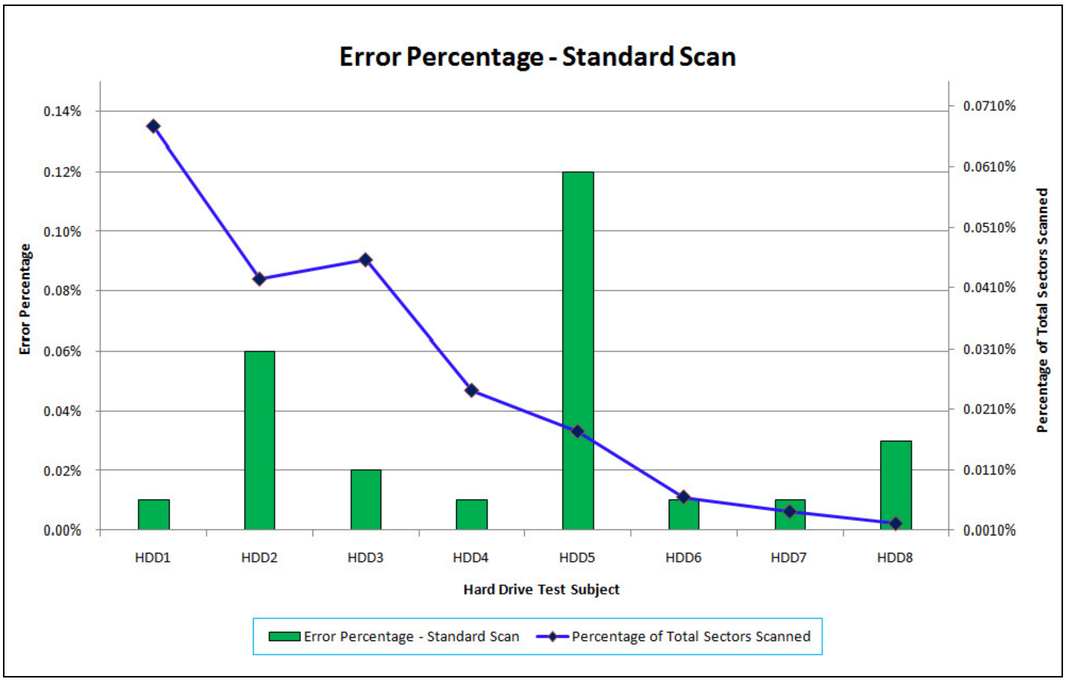


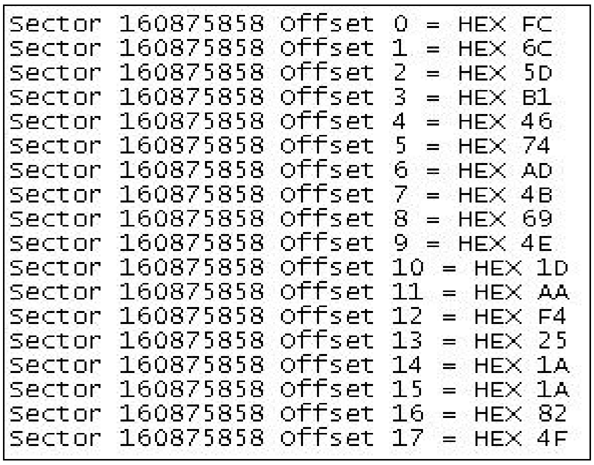
Figure 4.12. Error percentage for all eight byte standard scans

4.3.3. Results That are Invalid for Analysis

Some data and results for test subject 6 cannot be used for comparison analysis. Due to errors encountered when performing Tests 31 and 34, which are the eight byte and 512 byte full sector scans for test subject 6, the tests were performed on two different workstations. In summary, the scanning tool encountered reading errors on the drive. On rebooting the system, the BIOS did not detect the drive. The drive was subsequently connected to another system and Test 34 was run and completed without error. Test 34, the 512 byte full sector scan, was performed on a system having 24GB of RAM and a 6 core i5 processor. Test 31 was performed on a system with 2GB RAM and 1 processor. Time-related information for Tests 31 and 34 will not be used for comparison analysis to eliminate potential skewing of the results. However, the reported empty/used sector results for both tests remain valid for comparison analysis.

4.3.4. Verification of the Test Results

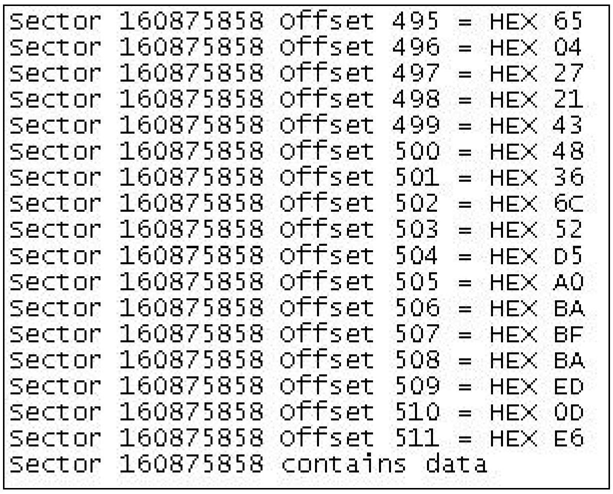
An important element of the research is to ensure that the results provided from the testing process are accurate and can be verified. Section 3.4.1 briefly discusses the output verification code that was built into the scanning software during the tool's development. The output verification code enables byte data from five randomly selected sectors to be captured and writes the data to a log file for analysis. The purpose of capturing the data is so that the data in the log can be verified against the values of the same sector on the original drive. In addition, the verification code determines whether the sector is empty or the sector contains data and sets the *sector state* flag accordingly. The sector's state is written to the same log file so that the programmatic routine, which determines whether the sector is used or empty, can be verified. Figure 4.13 shows the first eighteen bytes of the verification log for sector 160875858. The partial log shown in Figure 4.13 contains the sector number and the HEX value for each byte offset within the sector.



```
Sector 160875858 offset 0 = HEX FC
Sector 160875858 offset 1 = HEX 6C
Sector 160875858 offset 2 = HEX 5D
Sector 160875858 offset 3 = HEX B1
Sector 160875858 offset 4 = HEX 46
Sector 160875858 offset 5 = HEX 74
Sector 160875858 offset 6 = HEX AD
Sector 160875858 offset 7 = HEX 4B
Sector 160875858 offset 8 = HEX 69
Sector 160875858 offset 9 = HEX 4E
Sector 160875858 offset 10 = HEX 1D
Sector 160875858 offset 11 = HEX AA
Sector 160875858 offset 12 = HEX F4
Sector 160875858 offset 13 = HEX 25
Sector 160875858 offset 14 = HEX 1A
Sector 160875858 offset 15 = HEX 1A
Sector 160875858 offset 16 = HEX 82
Sector 160875858 offset 17 = HEX 4F
```

Figure 4.13. First eighteen entries of sector 160875858s verification log

Figure 4.14 shows the last seventeen entries of the log for sector 160875858. The last line of the log shows a human readable interpretation of the sector's state. In this case the scanning tool has correctly determined that the sector contains data and has reported the fact in the log.

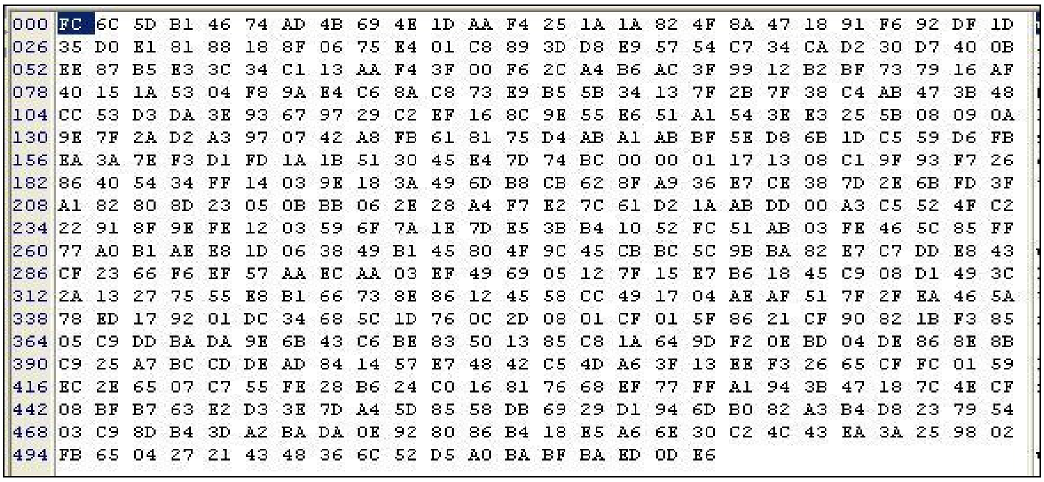


```
Sector 160875858 offset 495 = HEX 65
Sector 160875858 offset 496 = HEX 04
Sector 160875858 offset 497 = HEX 27
Sector 160875858 offset 498 = HEX 21
Sector 160875858 offset 499 = HEX 43
Sector 160875858 offset 500 = HEX 48
Sector 160875858 offset 501 = HEX 36
Sector 160875858 offset 502 = HEX 6C
Sector 160875858 offset 503 = HEX 52
Sector 160875858 offset 504 = HEX D5
Sector 160875858 offset 505 = HEX A0
Sector 160875858 offset 506 = HEX BA
Sector 160875858 offset 507 = HEX BF
Sector 160875858 offset 508 = HEX BA
Sector 160875858 offset 509 = HEX ED
Sector 160875858 offset 510 = HEX 0D
Sector 160875858 offset 511 = HEX E6
Sector 160875858 contains data
```

Figure 4.14. The end of sector 160875858s verification log

Figure 4.15 is a screen capture taken while viewing sector 160875858 in EnCase. All of the byte values within the sector are represented as HEX values. To ensure correct operation of the scanning tool, the data within the sector must be verified. To verify the data for a given sector, the value of each byte reported in EnCase is compared with the value of the corresponding byte in the verification log. In this example, the byte values in the verification log shown in Figures 4.13 and 4.14 match the byte values in the same sector viewed in EnCase and shown in Figure 4.15. Therefore, the tool has been verified to operate correctly by extracting data from the correct sector location and reporting the correct sector state.

During the research, verification of the sector data and verification of the state of the sector were checked on one of the six tests for each of the hard drive test subjects. However, the verification checks were performed over all of the six different types of tests to ensure there were no issues isolated to a particular type of test. None of the hard drive test subjects failed the data verification and sector state verification checks.



```
000 FC 6C 5D B1 46 74 AD 4B 69 4E 1D AA F4 25 1A 1A 82 4F 8A 47 18 91 F6 92 DF 1D
026 35 D0 E1 81 88 18 8F 06 75 E4 01 C8 89 3D D8 E9 57 54 C7 34 CA D2 30 D7 40 0B
052 EE 87 B5 E3 3C 34 C1 13 AA F4 3F 00 F6 2C A4 B6 AC 3F 99 12 B2 BF 73 79 16 AF
078 40 15 1A 53 04 F8 9A E4 C6 8A C8 73 E9 B5 5B 34 13 7F 2B 7F 38 C4 AB 47 3B 48
104 CC 53 D3 DA 3E 93 67 97 29 C2 EF 16 8C 9E 55 E6 51 A1 54 3E E3 25 5B 08 09 0A
130 9E 7F 2A D2 A3 97 07 42 A8 FB 61 81 75 D4 AB A1 AB BF 5E D8 6B 1D C5 59 D6 FB
156 EA 3A 7E F3 D1 FD 1A 1B 51 30 45 E4 7D 74 BC 00 00 01 17 13 08 C1 9F 93 F7 26
182 86 40 54 34 FF 14 03 9E 18 3A 49 6D B8 CB 62 8F A9 36 E7 CE 38 7D 2E 6B FD 3F
208 A1 82 80 8D 23 05 0B BB 06 2E 28 A4 F7 E2 7C 61 D2 1A AB DD 00 A3 C5 52 4F C2
234 22 91 8F 9E FE 12 03 59 6F 7A 1E 7D E5 3B B4 10 52 FC 51 AB 03 FE 46 5C 85 FF
260 77 A0 B1 AE E8 1D 06 38 49 B1 45 80 4F 9C 45 CB BC 5C 9B BA 82 E7 C7 DD E8 43
286 CF 23 66 F6 EF 57 AA EC AA 03 EF 49 69 05 12 7F 15 E7 B6 18 45 C9 08 D1 49 3C
312 2A 13 27 75 55 E8 B1 66 73 8E 86 12 45 58 CC 49 17 04 AE AF 51 7F 2F EA 46 5A
338 78 ED 17 92 01 DC 34 68 5C 1D 76 0C 2D 08 01 CF 01 5F 86 21 CF 90 82 1B F3 85
364 05 C9 DD BA DA 9E 6B 43 C6 BE 83 50 13 85 C8 1A 64 9D F2 0E BD 04 DE 86 8E 8B
390 C9 25 A7 BC CD DE AD 84 14 57 E7 48 42 C5 4D A6 3F 13 EE F3 26 65 CF FC 01 59
416 EC 2E 65 07 C7 55 FE 28 B6 24 C0 16 81 76 68 EF 77 FF A1 94 3B 47 18 7C 4E CF
442 08 BF B7 63 E2 D3 3E 7D A4 5D 85 58 DB 69 29 D1 94 6D B0 82 A3 B4 D8 23 79 54
468 03 C9 8D B4 3D A2 BA DA 0E 92 80 86 B4 18 E5 A6 6E 30 C2 4C 43 EA 3A 25 98 02
494 FB 65 04 27 21 43 48 36 6C 52 D5 A0 BA BF BA ED 0D E6
```

Figure 4.15. The byte values of sector 160875858 viewed in EnCase

4.4. CONCLUSION

Chapter 4 is a report that focuses on three main areas related to the tool development and testing phase of the research. An overview of the scanning tool's operation was presented. The overview included detailed information regarding the tool's development, its operational functions and the output which is produced in the form of a report.

Three alterations to the research methodology, outlined in Chapter 3, were reported. The alterations include a reduction in the number of hard drive test subjects, altering the fixed scan durations for the standard and two minute tests to allow for a one second tolerance either side of the fixed four and two minute time periods and adding the ability for the report to be displayed on the screen and to be written to the examiner's workstation automatically when a scan has completed.

Section 4.3.2 reported the results and findings from the testing phase. Section 4.3.4 reported the verification methodology used to verify the operation and accuracy of the scanning tool and the results that were produced from the testing process. Chapter 5 discusses the findings identified in Chapter 4 in reference to the results from the testing phase. The research question and sub-questions will also be answered.

Chapter 5

DISCUSSION

5.0. INTRODUCTION

A review of forensic data collection literature was presented in Chapter 2. An innovative idea for a hard drive scanning tool was formed from the literature review and was discussed in Section 2.6.2. Chapter 3 defined the research methodology for the development and testing of a hard drive scanning tool along with the research question, the two sub-questions and a hypothesis. The operation and the testing of the hard drive scanning tool were reported in Chapter 4. In addition, Chapter 4 reported the results from the testing phase of the research where the scanning tool was tested on eight different hard drive test subjects.

Chapter 5 presents a discussion regarding the research and findings. Section 5.1 answers the research question based on the results shown in Chapter 4. The two sub-questions are answered and the results from the testing phase shown in Chapter 4 assist with testing the hypothesis. Findings from the research are discussed in Section 5.2. Section 5.3 discusses issues identified during the testing phase of the research. Recommendations regarding the results produced by the scanning tool are discussed in Section 5.4. The chapter is brought to a close with a conclusion in Section 5.5.

5.1. THE RESEARCH QUESTION AND HYPOTHESES

The research question was initially derived from information related to forensic data collection and gathered during the literature review in Chapter 2. The research question was further refined and proposed in Chapter 3. The research question is answered in Section 5.1.1.

The two sub-questions were created because some aspects relating to the research question were unclear. The sub-questions were designed to expand on the scanning tool’s efficiencies by conducting further research and hard drive scanning tests. As a result of the further testing, the two sub-questions can be answered. Furthermore, the results from the additional testing created possibilities for future research in the area of reducing scan durations while maintaining the accuracy and reliability of the test results.

5.1.1. Research Question

The research question asked: "*Can a drive scanning tool process a hard drive in a measured period of time and produce a reliable result?*" Section 3.2.1 defined aspects of the research question to clarify how long a *measured period of time* is and the *reliability* aspect of a result. The data collected during the testing phase of the research was used to answer the research question.

The research question has two elements to it. Firstly, the question asked if a scanning tool can process a hard drive in a measured period of time. Section 3.2.2 defined the measured period of time as four minutes and referenced the four minute time period as the *standard test time* or a *standard scan*. Secondly, the research question asked if the scanning tool can provide a reliable result. A reliable result was defined as a result that is within ±10% of an accurate result. An accurate result was defined in Section 3.2.2 as a result produced when every sector on a drive is processed. The scanning tool's *full sector scan* settings enable an examiner to read and process every sector on the drive.

To provide an answer to the research question, a baseline is required for each test subject. The baseline would allow for comparison analysis to take place between the baseline and standard scan tests. The baseline was created by running full sector scans on all eight hard drive test subjects to produce an accurate result for each drive. Tables 4.4 and 4.5 show the data for each of the test subject's full sector scans.

The results for the standard scans were compared with the results from the full sector scans for each test subject. The comparison analysis shows that every test, apart from one, reported the same results for each drive's full sector scan and standard scan. The test which did not produce the same result, reported a mere 1.0% difference between the full sector scan and the standard scan results. Given that the result for every standard scan was within 10% of the corresponding full sector scan result and all of the standard test durations were between 3:59 minutes and 4:01 minutes, the answer to the research question is: yes, a drive scanning tool can process a hard drive in a measured period of time and produce a reliable result.

5.1.2. Research Sub-Questions

Two research sub-questions were defined in Section 3.2.4. The sub-questions were generated based on the scanning tool's expected performance and accuracy. The first sub-question asked: "*Will a two minute scan produce a similar result to a four minute scan?"* The first sub-question was asked to determine whether the result of a two minute scan is similar to the result of a four minute scan and to measure the differences, if any, between the two scan results. If the results from a two minute scan fall within ±10% of a full sector scan for the same test subject, the four minute scan would become obsolete and the two minute scan would be used in place of the four minute scan out in the field.

Two minute scans were performed over the eight hard drive test subjects. Tables 4.4 and 4.5 show the results for each of the drives two minute scans. An analysis of the results of the two minute and four minute scan results was conducted. Tables 4.4 and 4.5 show that apart from one test subject, the results from the two minute and four minute scans are identical. Although the results from one test subject were not identical, the results were within 1% of each other. Therefore, the answer to the first sub-question is: yes, a two minute scan does produce a similar result to a four minute scan.

Because almost all of the results for the two minute and four minute tests are identical for each test subject, the four minute test will become obsolete and will be replaced by the two minute scan. One may wonder how much a two minute test can be reduced in time before the accuracy of the result is affected and the result become unreliable. Further research in the area may provide the answer.

The second sub-question asked: *"Will processing all of the bytes in each scanned sector produce a similar result to processing eight bytes in each sector?"* The question was asked to determine the accuracy of reading and processing only the first eight bytes, or less than 1.6%, of each scanned sector. The scanning tool was designed to process the first eight bytes of a sector for improved performance. However, the author was uncertain whether the accuracy of the results would be affected by processing only the first eight bytes of each scanned sector. The output from the additional research is able to show whether the results from an eight byte and 512 byte test are similar.

As articulated in Section 4.2.1, an additional 512 byte test was conducted for the two minute, standard and full sector tests for each test subject. The results of the 512 byte tests are shown in Tables 4.4 and 4.5. Analysis of the results revealed that all of the test subjects produced identical empty sector percentages for both the eight and 512 byte tests apart from test subject 5 and test subject 7. The difference between the eight and 512 byte tests for test subjects 5 and 7 is only 2% and 3% respectively. Therefore, the answer to the second sub-question is: yes, processing all of the bytes in each scanned sector does produce a similar result to processing eight bytes in each scanned sector. More importantly, the results show that the accuracy of the output is very high when processing only the first eight bytes of each sector.

An interesting finding was revealed during the analysis of the processing times for the eight and 512 byte scans. The test subjects that produced the largest and the smallest time difference between their eight and 512 byte, full sector tests, were test subjects 2 and 4 respectively. Table 5.2 shows that test subject 2 produced a time difference of over nine and a half hours between the eight and 512 byte scans, while test subject 4 produced a time difference of only twenty minutes. However, the empty sector percentages between the eight and 512 byte scans are identical for test subjects 2 and 4. The author provides a theory as to why there is a vast time difference between the two full sector tests for each drive in Section 5.2.3.

5.1.3. Hypothesis

A theorised hypothesis was developed in Section 3.2.3. The hypothesis is based on a 20GB hard drive where the drive is one quarter full of data and three quarters empty. Testing the hypothesis on one or more of the test subjects would mean that changes to the data on the drive(s) would have to be made. As stated in Section 4.2.1, the original data that resided on the hard drive test subjects was left *intact* and was not manually manipulated or altered in any way during the research. Therefore, to ensure that the data was kept intact on the test subjects, the hypothesis was not tested on any of the test subjects by changing data on the drive.

The author hypothesised that the scanning tool would process a hard drive within four minutes and the tool would produce a result that would be within ±10% of an accurate result. The author based their hypothesis on the fact that the scanning tool collects a sample of data evenly across the range of the hard drive, as opposed to random sampling or sampling from predefined groups of sectors on the disk. Collecting data evenly across a disk eliminates the possibility of inconsistent results being reported caused by excessive sectors being read in either populated or vacant areas of the disk.

Although the hypothesis was not physically tested on the test subjects as outlined in Section 3.2.3, the data collected during the testing phase was used to test the hypothesis over all of the hard drive test subjects by using a comparison technique. The benefit of using the comparison technique is that the data on the test subjects did not have to be changed. Table 5.1 shows the results for the full sector, standard and two minute scans during the testing phase. The data in the *actual percentage empty/used* sector column for the full sector scans was used as a baseline for comparison analysis of the test results for the standard and two minute scans for each test subject. All durations for the standard and two minute scans are shown in Table 5.1. Some of the durations include the acceptable variance of ±1 second as detailed in Section 4.2.2.

Comparison analysis was conducted on the full sector and standard scans as well as the full sector and two minute scans. The analysis reported that the results for each test subject, apart from test subject 5, are identical. Test 26, shown in Table 5.1, revealed an *actual variance* of 1% from the full sector scan. However, the 1% variance is well within the ±10% *allowable tolerance* of an accurate result produced by a full sector scan. Therefore, the hypothesis that *a scanning tool can process a hard drive in a measured period of time and will produce a reliable result* was found to hold true for all eight test subjects used in the research.

Table 5.1

Scan Results to Test Hypothesis

HDD Number	Test Number	Scan Type	Duration	Actual % Empty/Used	Allowable Tolerance	Actual Variance	Hypothesis True/False
1 20GB	1	Full Sector	02:01:59	47% / 53%	±10%	N/A	N/A
	2	Standard	00:04:01	47% / 53%	±10%	0%	TRUE
	3	2 Minute	00:02:00	47% / 53%	±10%	0%	TRUE
2 30GB	7	Full Sector	01:11:25	88% / 12%	±10%	N/A	N/A
	8	Standard	00:04:01	88% / 12%	±10%	0%	TRUE
	9	2 Minute	00:01:59	88% / 12%	±10%	0%	TRUE
3 40GB	13	Full Sector	03:22:41	65% / 35%	±10%	N/A	N/A
	14	Standard	00:03:59	65% / 35%	±10%	0%	TRUE
	15	2 Minute	00:02:01	65% / 35%	±10%	0%	TRUE
4 80GB	19	Full Sector	07:44:52	01% / 99%	±10%	N/A	N/A
	20	Standard	00:03:59	01% / 99%	±10%	0%	TRUE
	21	2 Minute	00:02:00	01% / 99%	±10%	0%	TRUE
5 120GB	25	Full Sector	07:32:07	12% / 88%	±10%	N/A	N/A
	26	Standard	00:03:59	11% / 89%	±10%	1%	TRUE
	27	2 Minute	00:01:59	12% / 88%	±10%	0%	TRUE
6 250GB	31	Full Sector	14:36:06	88% / 12%	±10%	N/A	N/A
	32	Standard	0:04:01	88% / 12%	±10%	0%	TRUE
	33	2 Minute	0:02:01	88% / 12%	±10%	0%	TRUE
7 500GB	37	Full Sector	26:04:11	34% / 66%	±10%	N/A	N/A
	38	Standard	00:03:59	34% / 66%	±10%	0%	TRUE
	39	2 Minute	00:02:01	34% / 66%	±10%	0%	TRUE
8 1TB	43	Full Sector	36:44:23	40% / 60%	±10%	N/A	N/A
	44	Standard	00:03:59	40% / 60%	±10%	0%	TRUE
	45	2 Minute	00:01:59	40% / 60%	±10%	0%	TRUE

5.2. DISCUSSION

Section 5.2 discusses the analysis and findings discovered during the testing phase in Chapter 4. The actual outcomes are compared with the expected outcomes in Section 5.2.1. Inconsistencies between the processing durations for the eight byte and 512 byte full sector scans are discussed in Section 5.2.2 along with a theory by the author as to why the inconsistencies exist. Section 5.2.3 discusses the accuracy of the reported percentages for empty and used sectors and how the tool’s rounding function has an effect on the reported percentages.

5.2.1. The Testing Process

The testing process was discussed in Section 3.3.2.2. It was designed to provide sufficient information to answer the research question and the two sub-questions. The information required to answer the research question was produced by performing full sector and standard scans and comparing the results. The information required to answer the two sub-questions was produced by performing two separate sets of tests. Firstly, a two minute scan was conducted for each test subject. The results from the two minute and standard scans were compared to determine whether a two minute scan would produce a similar result to a standard scan. The output of the comparison allowed the first sub-question to be answered as discussed in Section 5.1.1. Secondly, a corresponding 512 byte test was conducted for every eight byte test that was processed during the testing phase. The results of the 512 and eight byte tests were compared to determine whether the accuracy of the eight byte test results was affected in any way. The comparison of the 512 and eight byte tests allowed the second sub-question to be answered as discussed in Section 5.1.2.

Section 4.2.1 discussed the hard drives that were used for the research. The author wanted to ensure that a diverse sample of test drives were used to provide a rich source of output data for analysis after testing. A variety of differently sized drives, ranging from 20GB to 1TB, were chosen as test subjects. The drives were removed from functioning computer systems and the drives contained Windows operating system data amongst other data. One drive gave problems during the testing phase and could not be read by one of the workstations for one of the tests. However, the issue was overcome by connecting the unreliable drive to another workstation and completing the test. Section 4.3.3 discussed the validity of the results from the unreliable drive and identifies which tests could be used for comparison analysis.

Four different workstations were used for the testing phase of the research. Using several workstations for the testing allowed some of the tests to be processed simultaneously resulting in a reduction of the overall testing time for the research. The ability to process the test subjects simultaneously was advantageous as some of the full sector scans took days to process and tied up testing systems, which would have halted testing if only one system was used.

5.2.2. Expected Outcomes Versus Actual Outcomes

Sections 5.1.1 and 5.1.3 discuss the standard test results and how they fall well within ±10% of their corresponding full sector scans. In Section 4.3.1, the *Expected Outcomes* section, the author hypothesised that the scanning tool would produce standard test results that fall within ±10% of their corresponding full sector scans. The results of the testing phase have been consistent with the author's hypothesis. The author also hypothesised that the accuracy of the results would deteriorate as the hard drive test subject's capacity increases. However, the test results showed that the author's hypothesis was false.

The actual outcomes were far from the expected outcomes outlined in Section 4.3.1. As shown in Tables 5.6 and 5.7, when the hard drive test subject's logical sizes increase, the accuracy of the result does not deteriorate as the author expected. Analysis of the results revealed that, in all apart from one case, a standard four minute scan produces an identical result to a full sector scan. During a standard test, an extremely small amount of the total number of sectors on the drive are read and processed. However, results show that the reported results are the same as the results from their corresponding full sector scans.

5.2.3. Scan Duration for Eight Byte and 512 Byte Tests

The differences in processing time between the eight byte and 512 byte full sector scans are not consistent throughout all of the test subjects. Table 5.2 shows that test subject 4, the 80GB drive, had the smallest time difference between both the eight and 512 byte full sector scans. The difference in time is a mere 20:37 min. In contrast, test subject 2, the 30GB drive, produced a time difference of almost ten hours between the eight and 512 byte full sector scans. To ensure the results from the comparison analysis have validity, comparisons can only be made between the eight byte and 512 byte scans which were run on the same drive and performed on the same computer system. Performing the tests on the same workstation ensures that factors such as system performance and data throughput rates will have little or no effect on the results.

Excluded from comparison analysis are tests 31 and 34, the eight byte and 512 byte full sector scans for test subject 6. The author has chosen not to use these results as the tests were performed on different computer systems. Because the eight byte and 512 byte full sector scans were performed on vastly different systems, time comparison for both of the full sector tests will be unreliable and therefore incomparable.

Table 5.2

Results for Test Subjects 2 and 4

Test Subject	Test Number	Scan Type	Bytes Processed/ps	Duration	Time Difference – 8 & 512 byte Full Sector Tests	Empty/Used Percentage
2 30GB	7	Full Sector	8	01:11:25	09:49:17	88% / 12%
	8	Standard	8	00:04:01		88% / 12%
	9	2 Minute	8	00:01:59		88% / 12%
	10	Full Sector	512	11:00:42		88% / 12%
	11	Standard	512	00:03:59		88% / 12%
	12	2 Minute	512	00:02:01		88% / 12%
4 80GB	19	Full Sector	8	07:44:52	00:20:37	01% / 99%
	20	Standard	8	00:03:59		01% / 99%
	21	2 Minute	8	00:02:00		01% / 99%
	22	Full Sector	512	08:05:29		01% / 99%
	23	Standard	512	00:04:00		01% / 99%
	24	2 minute	512	00:01:59		01% / 99%

One may ask why test subject 2 and test subject 4, shown in Table 5.2, produced vastly different processing times between each drive’s eight byte and 512 byte full sector scans. The author believes that the vastly different scan durations are related to the *types of data* that reside on each drive and the way in which the data is processed by the tool during a scanning process. The *type of data* is determined by how many sectors contain identical byte values and how many sectors contain bytes with different values. If identical byte values are observed throughout a sector, then the type of data for that sector is considered to be empty. On the contrary, if a sector contains different byte values, then the type of data for the sector is considered to be used.

Table 5.2 shows that test subject 2 produced a result showing that 88% of all processed sectors on the drive contain the same byte values in each sector, and 12% of all processed sectors contain different byte values. The results show that the hard drive is 88% empty. In contrast, test subject 4, shown in Table 5.2, produced a result showing that 1% of all processed sectors are empty and 99% of all processed sectors contain data, therefore test subject 4 is almost full of data.

Although the volume size of test subject 2 is under half the size of test subject 4, the difference in processing time between the eight byte and 512 byte test show that test subject 2 took over nine hours longer to run compared to test subject 4. The author has the following theory as to why the time difference between the two drives is substantial. In the case for test subject 4, the scanning tool would have read every sector on the drive. However, because the drive is mostly full of data, the byte comparison would have terminated early during the comparison process because the sector contained data. If any byte values do not match the value of the first byte in a sector during the comparison check, the comparison test is terminated for that particular sector and the next sector is processed. Because the comparison check would have terminated early for the majority of the sectors, the processing time would have been significantly reduced.

In the case of test subject 2, 88% of the sectors on the drive are completely empty. Therefore, while processing each sector, the comparison process would have to compare every byte within each empty sector before the next sector could be read and processed. This equates to fully processing 88% of the sectors on the drive. The processing time will be considerably longer if the comparison check has to process all 512 bytes in each empty sector as opposed to the first few bytes, as the case may have been for test subject 4.

5.2.4. Accuracy of the Reported Results

The reports produced by the scanning tool during the testing phase contain the percentage of the scanned sectors that are empty and the percentage of the scanned sectors that contain data.

Table 5.3

Accuracy of the Standard Scans for Test Subjects 1-4

Test Subject	Test Information			Results		
HDD Number & Size	Test Number	Scan Type	Bytes Processed Per Sector	Reported Percentage Empty/Used	Error Percentage	Accuracy of the Standard Scan
1 20GB	1	Full Sector	8	47% / 53%		
	2	Standard	8	47% / 53%	0.01	99.99%
2 30GB	7	Full Sector	8	88% / 12%		
	8	Standard	8	88% / 12%	0.06	99.94%
3 40GB	13	Full Sector	8	65% / 35%		
	14	Standard	8	65% / 35%	0.02	99.98%
4 80GB	19	Full Sector	8	01% / 99%		
	20	Standard	8	01% / 99%	0.01	99.99%

Section 4.3.2 reported the error percentage for the standard tests for each of the test subjects. The error percentages calculated from all standard and two minute scans performed during the testing are extremely low. As mentioned in Section 4.3.2, an extremely low error percentage for a particular scan means that the accuracy of the scan is extremely high due to the result being similar to the result of the corresponding full sector scan.

Tables 5.3 and 5.4 show the accuracy of the standard tests for all of the test subjects based on the error percentage. The accuracy of the standard scan results is calculated by subtracting the error percentage from 100%.

Table 5.4

Accuracy of the Standard Scans for Test Subjects 5-8

Test Subject	**Test Information**			**Results**		
HDD Number & Size	**Test Number**	**Scan Type**	**Bytes Processed Per Sector**	**Reported Percentage Empty/Used**	**Error Percentage**	**Accuracy of the Standard Scan**
5 120GB	25	Full Sector	8	12% / 88%		
	26	Standard	8	11% / 89%	0.12	99.88%
6 250GB	31	Full Sector	8	88% / 12%		
	32	Standard	8	88% / 12%	0.01	99.99%
7 500GB	37	Full Sector	8	34% / 66%		
	38	Standard	8	34% / 66%	0.01	99.99%
8 1TB	43	Full Sector	8	40% / 60%		
	44	Standard	8	40% / 60%	0.03	99.97%

Tests 25 to 30 shown in Table 5.5 for test subject 5 produced three different results for the percentage of empty sectors on the drive. The percentages range from 10% to 12%. The results were not consistent with those for most of the other test subjects, where the percentage results for empty sectors were identical over all six tests per drive. The fluctuating results for tests 25 to 30 were analysed. The apparently inconsistent results were generated by a rounding function from a calculation that the tool performs. When a scanning process is complete, the scan tool calculates the percentage amounts for the used and empty sectors. Table 5.5 shows the results for tests 25 to 30 that were processed on test subject 5.

Table 5.5

Test Results for Test Subject 5

Test	Scan Type	Duration	Sectors Scanned	Bytes Sector	Empty Sectors	Used Sectors	Percentage Empty/Used	Percentage two Decimal Places
25	Full Sector	07:32:07	234,441,648	8	27,088,947	207,352,701	12% / 88%	11.55% / 88.45%
26	Standard	00:03:59	40,511	8	4,631	35,879	11% / 89%	11.43% / 88.57%
27	2 Minute	00:01:59	20,189	8	2,403	17,786	12% / 88%	11.90% / 88.10%
28	Full Sector	12:59:40	234,441,648	512	24,109,201	210,332,447	10% / 90%	10.28% / 89.72%
29	Standard	00:04:00	40,425	512	4,156	36,269	10% / 90%	10.28% / 89.72%
30	2 Minute	00:02:01	20,189	512	2,136	18,053	11% / 89%	10.58% / 89.42%

When analysing the inconsistencies in regards to the reported percentages for test subject 5, the reported percentages for empty and used sectors were manually calculated to verify that the scanning tool was producing the correct figures for the reported percentages. When the manual calculations were conducted, the author calculated the percentages to two decimal places. The manual calculations are shown in the last column of Table 5.5. Upon analysis of the manually calculated percentages, the author discovered the reason why there are apparent inconsistencies in the reported percentages for test subject 5.

The reason why the inconsistencies appeared is because of an oversight in the programming of the scanning tool that is directly related to the reporting function. The scanning tool produces a report when a scan on a drive has completed. The report includes percentages for both the used and empty sector counts, but the percentages are rounded into whole numbers. In Table 5.5, tests 25 and 26 show an example of how the calculated percentages produce apparently inconsistent results due to rounding. Tests 25 and 26 show that the percentages for used sectors are 12% and 11% respectively. However, calculating the percentages to two decimal places shows that the results are actually similar, 11.55% for Test 25 and 11.43% for Test 26. Although the two figures are within 0.12% of each other, due to the tool's rounding function the percentages are rounded up to the nearest percent for Test 25 and rounded down to the nearest percent for Test 26. The rounding process creates an apparent percentage difference of one percent between the two tests when in reality, the difference is only 0.12%.

5.2.5. Data Sampling

Section 4.3.2 reported that the number of sectors scanned in a four minute test was considerably small compared to the total number of sectors that reside on the hard drive test subjects. The full sector scans presented in Tables 5.6 and 5.7 show the total number of sectors that reside on each test subject. The number of scanned sectors for the standard and two minute tests are also shown for each test subject.

Table 5.6

Scan Information for Test Subjects 1-4

HDD Number & Size	Test Number	Scan Type	Duration	Sectors Scanned	Sector Offset Size (MB)	Percentage of Total Sectors Scanned	Data Read During The Scan	Reported Percentage Empty/Used
1 20GB	1	Full Sector	02:01:59	39,070,080	0.000512	100.0000%	20GB	47% / 53%
	2	Standard	00:04:01	26,398	0.76	0.0676%	14MB	47% / 53%
	3	2 Minute	00:02:00	12,290	1.63	0.0315%	6MB	47% / 53%
	4	Full Sector	06:30:52	39,070,080	0.000512	100.0000%	20GB	47% / 53%
	5	Standard	00:04:01	25,801	0.78	0.0660%	13MB	47% / 53%
	6	2 Minute	00:02:01	12,290	1.63	0.0315%	6MB	47% / 53%
2 30GB	7	Full Sector	01:11:25	58,605,120	0.000512	100.0000%	30GB	88% / 12%
	8	Standard	00:04:01	24,821	1.21	0.0424%	13MB	88% / 12%
	9	2 Minute	00:01:59	9,999	3.00	0.0171%	5MB	88% / 12%
	10	Full Sector	11:00:42	58,605,120	0.000512	100.0000%	30GB	88% / 12%
	11	Standard	00:03:59	24,097	1.25	0.0411%	12MB	88% / 12%
	12	2 Minute	00:02:01	9,999	3.00	0.0171%	5MB	88% / 12%
3 40GB	13	Full Sector	03:22:41	78,242,976	0.000512	100.0000%	40GB	65% / 35%
	14	Standard	00:03:59	35,694	1.12	0.0456%	18MB	65% / 35%
	15	2 Minute	00:02:01	18,049	2.22	0.0231%	9MB	65% / 35%
	16	Full Sector	12:45:16	78,242,976	0.000512	100.0000%	40GB	65% / 35%
	17	Standard	00:04:01	30,901	1.30	0.0395%	16MB	65% / 35%
	18	2 Minute	00:02:01	15,861	2.53	0.0203%	8MB	65% / 35%
4 80GB	19	Full Sector	07:44:52	156,301,488	0.000512	100.0000%	80GB	01% / 99%
	20	Standard	00:03:59	37,699	2.12	0.0241%	19MB	01% / 99%
	21	2 Minute	00:02:00	18,360	4.36	0.0117%	9MB	01% / 99%
	22	Full Sector	08:05:29	156,301,488	0.000512	100.0000%	80GB	01% / 99%
	23	Standard	00:04:00	37,699	2.12	0.0241%	19MB	01% / 99%
	24	2 minute	00:01:59	18,360	4.36	0.0117%	9MB	01% / 99%

The columns labelled *percentage of total sectors scanned* in Tables 5.6 and 5.7 show the percentage of the total number of addressable sectors on the drive that have been scanned during the tests. As you would expect, the full sector scans show 100% for the percentage of processed sectors during a full sector scan. However, the percentages for the standard and two minute tests are so low that it may be difficult to understand how much data was actually processed during the tests. Therefore, the author has calculated the total amount of data that is read and processed during each test. These figures are shown in the *data read during the scan* columns of Tables 5.6 and 5.7. Immediately, one can see the contrast between the full sector scans and the standard and two minute scans. For example, test subject 1 shows that 20GB data is read and processed for the full sector scan, while only 14MB is read for the standard scan and a mere 6MB is read for the two minute scan, yet the reported results regarding the empty and used sectors are identical for all three tests.

Table 5.7

Scan Information for Test Subjects 5-8

HDD Number & Size	Test Number	Scan Type	Duration	Sectors Scanned	Sector Offset Size (MB)	Percentage of Total Sectors Scanned	Data Read During The Scan	Reported Percentage Empty/Used
5 120GB	25	Full Sector	07:32:07	234,441,648	0.000512	100.0000%	120GB	12% / 88%
	26	Standard	00:03:59	40,511	2.96	0.0173%	21MB	11% / 89%
	27	2 Minute	00:01:59	20,189	5.95	0.0086%	10MB	12% / 88%
	28	Full Sector	12:59:40	234,441,648	0.000512	100.0000%	120GB	10% / 90%
	29	Standard	00:04:00	40,425	2.97	0.0172%	21MB	10% / 90%
	30	2 Minute	00:02:01	20,189	5.95	0.0086%	10MB	11% / 89%
6 250GB	31	Full Sector	14:36:06	488,397,168	0.000512	100.0000%	250GB	88% / 12%
	32	Standard	0:04:01	31,359	7.97	0.0064%	16MB	88% / 12%
	33	2 Minute	0:02:01	16,930	14.77	0.0035%	9MB	88% / 12%
	34	Full Sector	56:04:41	488,397,168	0.000512	100.0000%	250GB	88% / 12%
	35	Standard	00:04:01	31,125	8.03	0.0064%	16MB	88% / 12%
	36	2 Minute	00:02:00	16,800	14.88	0.0034%	9MB	88% / 12%
7 500GB	37	Full Sector	26:04:11	976,773,168	0.000512	100.0000%	500GB	34% / 66%
	38	Standard	00:03:59	39,319	12.72	0.0040%	20MB	34% / 66%
	39	2 Minute	00:02:01	19,529	25.61	0.0020%	10MB	34% / 66%
	40	Full Sector	53:36:40	976,773,168	0.000512	100.0000%	500GB	31% / 69%
	41	Standard	00:03:59	39,599	12.63	0.0041%	20MB	31% / 69%
	42	2 Minute	00:01:59	19,799	25.26	0.0020%	10MB	31% / 69%
8 1TB	43	Full Sector	36:44:23	1,953,525,168	0.000512	100.0000%	1TB	40% / 60%
	44	Standard	00:03:59	39,750	25.16	0.0020%	20MB	40% / 60%
	45	2 Minute	00:01:59	20,000	50.01	0.0010%	10MB	40% / 60%
	46	Full Sector	97:38:58	1,953,525,168	0.000512	100.0000%	1TB	40% / 60%
	47	Standard	00:03:59	39,714	25.19	0.0020%	20MB	40% / 60%
	48	2 Minute	00:02:00	19,859	50.37	0.0010%	10MB	40% / 60%

Table 5.7 shows that during the standard scan process for test subject 5, a data sample of only 21MB was read and processed from a total data size of 120GB on the hard drive. The data samples read for the standard tests across all remaining test subjects is below 21MB. The size of the data sample does not necessary increase as the drive capacities increase. For a given test, the size of the data sample is more so based on the speed at which the drive can be accessed by the tool. In some cases the data sample can actually be smaller for a larger drive. The reason is that the tests are processed on a set time basis. Therefore, a drive that has greater data throughput capability would process a larger data sample in four minutes compared to a drive that has smaller data throughput capability. For example, Table 5.7 shows that the standard test for test subject 8 read 20MB of data from a total data size of 1TB,

which is 1MB less than test subject 5 where 21MB of data was read. Test subject 5 is several years older than test subject 8.

Although less data is read on a drive that is not as old and is over eight times larger than test subject 5, the accuracy of the result was not affected. Table 5.7 shows identical results for the standard and full sector scans for test subject 8. In addition, the two minute test for test subject 8, which reads half the amount of sample data of the standard scan, also produced an identical result. The accuracy of the results for the standard and two minute tests for test subject 8 came as a surprise as the author expected a decrease in accuracy as volume sizes increased.

How can the results of the standard and two minute tests be so accurate when an extremely small data sample is collected? The author believes that the reason is to do with the way the author designed the tool's scanning method. In Section 5.1.3, the author hypothesised that the scanning tool would process a hard drive within four minutes and the tool would produce a result that is within ±10% of an accurate result. The author's hypothesis was based on the way the scanning tool processes a hard drive. Firstly, a sample of data is taken from all available data stored on a drive. Secondly, the data sampling or scanning process reads the data evenly across the range of the hard drive.

The data sampling technique used for the research is not random sampling or representative sampling. Random sampling is where data is randomly taken from the entire storage device. Representative sampling is where data is sampled from a subset of data that reflects the data on the storage device. The author believes that collecting data evenly across the entire device will produce a sample of data that is reliable and is a true representation of the actual data stored on a device.

5.2.6. Issues Identified During the Research

Two main issues were identified while conducting the research. Section 5.2.6.1 looks at the issues encountered with one of the hard drive test subjects where the drive may have been starting to fail and how the issues were overcome. Issues related to obtaining information, such as the model and serial numbers, from hard drives that are connected via USB connection are discussed in Section 5.2.6.2.

5.2.6.1. Hard Drive Test Subjects

One of the hard drive test subjects encountered problems during the testing phase of the research that resulted in one of the six tests not being performed on the same computer system where the five previous tests had been completed. Hard drive test subject 6, a Western Digital 250GB drive, returned an error during Test 34, the 512byte full sector scan. Once the error message was closed, the scanning tool exited. Before the error occurred, the test had been running for four full days and was about three quarters of the way through the scanning process.

Although the system was shut down and restarted, the drive was not detected by the BIOS correctly. On a second restart, the drive was detected correctly and Test 34 was restarted. The second scan attempt did not run properly and performance was severely degraded. The rate of processed sectors would slow down dramatically for several seconds then the rate would increase to the normal rate again. The drive appeared to be faulty or was in the process of failing. A decision was made to attempt to complete Test 34 on another computer system.

A new powerful computer system was chosen to process Test 34. The computer system was chosen to minimise the overall processing time that the drive would have to endure by leveraging from the system's high performance capabilities. The author believed that if the processing time could be reduced, the chances of completing the test without an error would be increased. The test completed without any errors being reported.

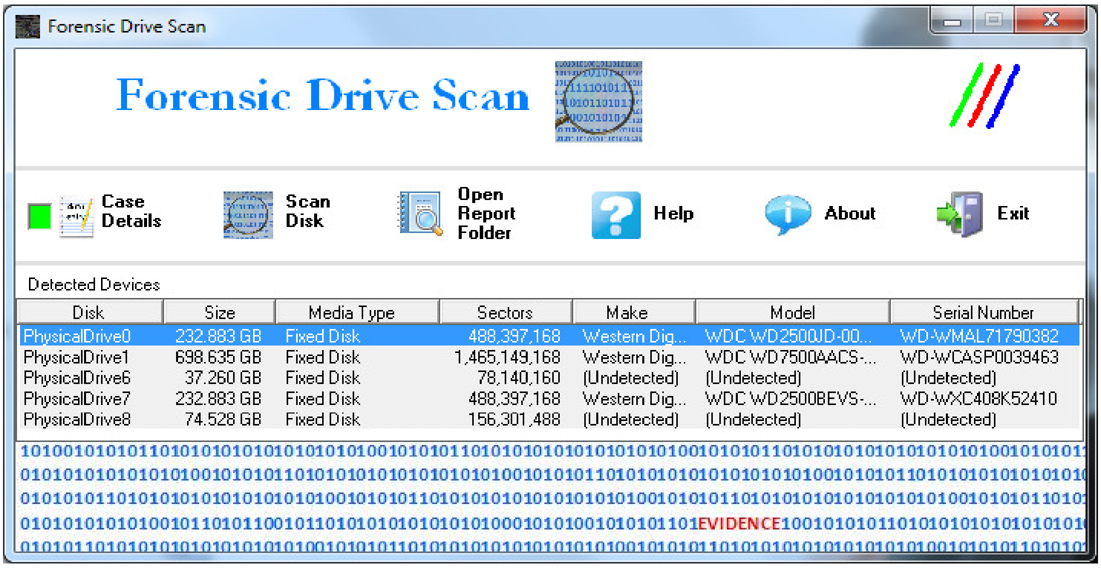
As mentioned in Section 4.3.3, the author will not use the results produced by Test 34 for timing comparison analysis. The reason why the test results will not be used is that the test was performed on a vastly different system and the results would not be comparable for analysis of process timings. However, Test 34s information regarding the empty/used sector ratio will still be valid for comparison analysis.

5.2.6.2. Information From USB Connected Devices

While conducting the testing phase, an issue was identified where devices that were connected via USB connectivity did not show information for the make, model or serial number of the device. Figure 5.1 shows an example where two of the five detected devices are connected to the system via a USB connection. The device list shows information regarding the logical size, media type and sector count for

Physical drives 6, and 8; however, the information related to the make, model and serial number is reported as *undetected*.

According to Allen (2004), SMART enabled disk drives internally monitor their own health and performance. The disk itself can provide advance warning regarding errors and can help to avoid sudden disk failures. When the scanning tool is operated, the tool automatically detects each device connected to the system and reads the model and serial number information directly from the SMART monitoring system on the drive. However, the drives manufacturer information is not available; therefore, the scanning tool uses the drive's model number to determine the manufacturer information which is listed in the *make* column of the detected devices table shown in Figure 5.1. If a drive does not support the SMART monitoring system, information for the model and serial number is unavailable and therefore the associated fields in the tool's detected devices list are populated with *undetected*.



Detected Devices

Disk	Size	Media Type	Sectors	Make	Model	Serial Number
PhysicalDrive0	232.883 GB	Fixed Disk	488,397,168	Western Dig...	WDC WD2500JD-00...	WD-WMAL71790382
PhysicalDrive1	698.635 GB	Fixed Disk	1,465,149,168	Western Dig...	WDC WD7500AACS-...	WD-WCASP0039463
PhysicalDrive6	37.260 GB	Fixed Disk	78,140,160	(Undetected)	(Undetected)	(Undetected)
PhysicalDrive7	232.883 GB	Fixed Disk	488,397,168	Western Dig...	WDC WD2500BEVS-...	WD-WXC408K52410
PhysicalDrive8	74.528 GB	Fixed Disk	156,301,488	(Undetected)	(Undetected)	(Undetected)

Figure 5.1. SMART information is unavailable for physical drives 6 and 8 as they are connected by USB

Access to the SMART monitoring system is available when the device is connected directly to SATA or PATA ports and the device is SMART enabled. However, the SMART monitoring system is unavailable when a device, such as a USB flash drive or a USB hard drive, is connected to a computer system by a USB port. There are mixed opinions regarding the access of SMART information through USB connectivity. According to EASIS (2011), it is impossible to get SMART information from media such as USB drives and external hard drives. However, Ariolic Software (2011) sell a product named *ActiveSmart* which can read SMART information from some USB connected hard drives. Research into the complications and associated issues with accessing SMART information through USB connectivity is outside the scope of the research and may be researched in future work.

5.3. RECOMMENDATIONS

The author has developed a hard drive scanning tool that will enable an examiner to identify what type of data, in terms of used verses empty sectors, resides on a particular hard drive. The research shows that the tool can be operated efficiently, in terms of time, and the results are extremely close to being accurate for the selection of hard drive test subjects which were tested.

The author has mentioned throughout the thesis that having knowledge of the type of data that resides on a hard drive will allow the examiner to make informed decisions regarding collection tools and methods when performing data collection activities. Now that the tool has been developed and the testing shows that the tool is capable of producing accurate results in a time efficient manner, the question may be asked: “How are the results from a scan used to assist the examiner?”

Section 4.3.1 references published research by the author and Dr Brian Cusack regarding the use of compression algorithms while imaging hard drives. Cusack and Pearse (2011) claim that the overall image processing time can be reduced by using a level of compression when imaging a drive which can be up to 80% full of data. The research shows that time can be reduced by imaging a drive with compression when the drive is connected by slow connectivity such as USB2 or ethernet. The sample data used in Cusack and Pearse’s study consisted of *operating system* and *program files* data typically found on Windows XP computer systems.

There are situations when an examiner will be forced to image storage media over a slow connection such as USB2 or ethernet. For example, the examiner may be collecting data from a live system and a USB connected destination drive is attached to the system or the examiner may be using a boot disk on a suspect’s workstation and connects a USB destination drive for the image to be written to. In these situations, an examiner may use the research by Cusack and Pearse (2011) as a guideline regarding the use of compression when conducting imaging activities. Cusack and Pearse claim that a reduction in image processing time is possible with a drive that is 80% full of data. However, the dataset used by Cusack and Pearse is different to the dataset used in the research. Therefore, if a hard drive is 80% filled with data which is different to the dataset used by Cusack and Pearse, a reduction in processing time may not occur.

The results from the research by Cusack and Pearse (2011) showed that the FTK Imager compression level which produced the best results is level 1. Although FTK Imager's *level 1* compression setting provides the least amount of compression, the level 1 setting produces the most efficient compression algorithm in terms of time and compresses empty sectors extremely well. Cusack and Pearse's research did not show results for level 1 compression on a completely blank drive but did show the results when compression level 5 was used on a blank drive. The result showed that the data from a 10GB hard drive can be compress down to 18MB.

Further research is required to discover accurate processing time reductions and the most efficient use of compression when creating forensic copies of data. However, the author recommends using FTK Imager's level 1 compression setting up until a drive is 80% full of data. The recommendation is based on the research of Cusack and Pearse (2011); however, the author suggests that metrics from data collection activities should be collected when undertaking imaging tasks. The metrics can be analysed to establish up to which percentage of used sectors the use of compression produces a reduction in image processing time.

5.4. CONCLUSION

Chapter 5 is a discussion of the research project. The findings which were identified and reported in Chapter 4 were discussed. The research question and sub-questions proposed in Section 3.2 have been answered by using the results which were produced from the testing phase of the research. To ensure that the data which resides on the test subjects was kept intact and unchanged, the theoretical hypothesis was not tested as outlined in Section 3.2.3 on any of the test subjects. However, analysis of the data collected from the testing phase showed that the hypothesis was correct for all of the hard drive test subjects.

Discussion regarding the test results and how the results from the standard tests are, in most cases, identical to an accurate result has identified that the scanning tool does produce a reliable result. The results from both the standard and two minute tests were accurate across all test subjects and therefore showing that the expected outcomes outlined in Section 4.3.1 were incorrect. The expectation was that as the logical sizes of hard drives increase, the accuracy of the test results will deteriorate due to the smaller ratio of scanned sectors and the total number of sectors on the drive. However, the results did not deteriorate as the author expected.

Chapter 6 provides a summary of the findings identified during the thesis and concludes the research project. Limitations which restricted the research will be discussed along with recommendations for further research in the topic area.

Chapter 6

CONCLUSION

6.0. INTRODUCTION

A review of literature related to data storage systems, forensic tools and data collection techniques was conducted in Chapter 2. The review identified the issues surrounding the creation of forensic copies of large storage devices such as computer and server hard disk drives. The issues are becoming more apparent as the technological advancements are constant and storage volumes continue to increase.

Section 2.6 discusses a selection of innovative ideas regarding how data collection performance can be increased and one of the ideas was chosen for the research. The research involves the development and testing of a data scanning tool that will report the type of data residing on an electronic storage device. Chapter 3 sets out the methodology for the research. To assist with developing the methodology a review of five previous research papers that are related to software development and testing was conducted. Together, the literature review and the review of the five previous research papers aid in constructing a sound methodology. In addition, Chapter 3 derives the research question and sub-questions from the review of the literature in Chapter 2.

The scanning tool was developed according to the methodology set out in Chapter 3. However, three areas of the methodology were modified. The modifications were conducted to rectify unforeseen issues which were encountered while conducting the testing phase of the research. The modifications include introducing a ± 1 second tolerance to the standard and two minute scan durations, changing the reporting function to automatically save a full text report upon completion of a scanning task and choosing not to test the 2TB hard drive that was originally selected as one of the test subjects. The 2TB hard drive was not tested due to the anticipated long processing times associated with testing the drive and the unavailability of high-end testing systems.

Section 6.1 is a summary of the findings from the testing phase in Chapter 4 and the discussion in Chapter 5. The answers to the research question and sub-questions are summarised along with a review of the hypothesis in Section 6.2.

Limitations to the research, the testing process and the scanning tool are addressed in Section 6.3. Areas for potential future research involving data collection methods and tools are discussed in Section 6.4. References and an appendix follow Chapter 6.

6.1. SUMMARY OF FINDINGS

The testing process and the analysis of the results from the testing phase of the research are reported in Chapter 4. The key findings identified while conducting the research include the accuracy of the test results from the standard and two minute scans, the author's hypothesis regarding the accuracy of the results deteriorating as drive sizes increase, the amount of data which is read during a standard and two minute scan and the accuracy of the tool's operation and the test results produced by the tool.

Testing was conducted on eight hard drive test subjects where a total of 48 tests were performed. All reports created by the scanning tool during the testing phase were analysed and the results are shown in Tables 4.4 and 4.5. The tables show that the reported used and empty sector percentages for the standard and two minute tests are all within ±1% of their corresponding full sector scans.

The author hypothesised that the accuracy of a two or four minute scan result would deteriorate as the test subjects capacities increased. The hypothesis was based around the relationship of the standard or two minute scans and the increase in data volume sizes. For example, if a 20GB and a 40GB drive were scanned for a period of four minutes, the tool would most likely process a similar number of sectors for both drives. However, the amount of data between the scanned sectors would be doubled for the 40GB drive compared to the 20GB drive. Put another way, the ratio of read sample data versus total data stored on the drive would be half for the 40GB drive. The author believed that the results would deteriorate due to a less concentrated sample of data being collected for the larger drive. However, the author's hypothesis was incorrect. All reported results for the full sector scans and the corresponding two and four minute tests were identical apart from two tests. However, the results of the two tests were within ±1% of the results of their full sector scans and therefore the difference is insignificant.

The results from the standard and two minute tests are remarkably similar to the full sector scan considering that in some cases, processing the full sector scan ran for several days. The amount of data which is read and processed during a scanning procedure is small compared to the overall size of the data that resides on the drive. The largest amount of data that was read during a standard scan was 21MB. For test subject 5, 21MB equates to 0.0173% of the total data that resides on the drive. However, the two minute test on test subject 8, which is the 1TB drive, read 10MB of data. For test subject 8 10MB equates to a mere 0.001% of the total available data on the drive. Although the scanning tool read between 5MB and 21MB of data for each of the standard and two minute tests, the reported results were identical to the results from the full sector scan performed on the same drive.

The author believes that the accuracy of the results from the standard and two minute tests is produced by the data sampling technique engineered into the scanning tool. Firstly, the sampling method collects a sample of data taken from all available data stored on a drive. Secondly, the sample is collected by reading the data evenly across the entire hard drive. The author believes that collecting a sample of data evenly across the entire drive will produce a sample that is reliable and provides a true representation of the actual data stored on a drive.

The accuracy of the tool's operation and the test results produced by the tool are verified by using the tool's built-in data verification system and an external forensic tool. Information gained from the review of the five previous studies was used to develop a data verification module which is hard-coded into the tool. The data verification module is actioned during every scan process. The data verification system performs two functions that confirm that the tool is operating correctly and that the tool is producing accurate data in regards to the empty and used sector percentages. The first function reads five randomly selected sectors and stores the data into a file for analysis. The second function is performed after each randomly selected sector is read and indicates whether the sector is flagged as used or empty by the tool. Manual analysis of the data written to the file is conducted to verify that the correct sector is being read and that the tool is correctly distinguishing between used and empty sectors.

Section 4.3.4 reports the results from analysing the data produced by the tool's data verification system. Verification analysis was conducted on a sample of the tests that were performed over all of the test subjects to ensure that the correct sector was being read and that the tool is correctly distinguishing between used and empty sectors. Verification analysis was performed on only one of the tests conducted for each of the hard drive test subjects; however, the verification analysis was performed over every different type of test to ensure there were no issues isolated to a particular type of test. None of the hard drive test subjects failed the verification process which indicates that the tool is functioning correctly for those tests.

6.2. RESEARCH QUESTION, HYPOTHESIS AND SUB-QUESTIONS

The research question and two sub-questions were derived from the literature review in Chapter 2. The research question is: "*Can a drive scanning tool process a hard drive in a measured period of time and produce a reliable result?*" The author defined a measured period of time as a four minute time period. The four minute time period, ±1 second, was used for all standard tests that were preformed during the testing phase. The author defined a reliable result as a result produced by a scanning task that is within ±10% of an accurate result. An accurate result is produced when a full sector scan is performed. The result is accurate because every sector on the source device is read and processed and therefore, an accurate result is based on processing 100% of the data on the device as opposed to processing a sample of the available data.

The analysis of the results in Chapter 4 provided sufficient information that was used to answer the research question. For all eight hard drive test subjects, the results showed that all standard tests completed within the predefined four minute time period with a tolerance of ±1 second. Likewise, the two minute tests completed within two minutes with a ±1 second tolerance. In addition, all standard and two minute tests results were within ±1% of their corresponding full sector scans. Therefore, the answer to the research question is: yes, a drive scanning tool can process a hard drive in a measured period of time and produce a reliable result. The author hypothesised that the scanning tool would process a hard drive within a four minute time period and the tool would produce a result that is within ±10% of an accurate result. The author's hypothesis is true for all of the test subjects.

Two sub-questions were derived to test whether the four minute time period can be reduced while maintaining a similar result and also to determine whether the accuracy of the result is affected by reading only the first eight bytes of each scanned sector as opposed to reading all 512 bytes. The first sub-question asks: *"Will a two minute scan produce a similar result to a four minute scan?"* The second sub-question asks: *"Will processing all of the bytes in each scanned sector produce a similar result to processing eight bytes in each sector?"* Additional testing was conducted to produce data that will assist with answering the sub-questions. The additional tests consisted of a series of two minute tests and a series of tests where 512 bytes in each read sector were processed. Given that all of the two minute scan results are within ±1% of their corresponding four minute scan results, the answer to the first sub-question is: yes, a two minute scan does produce a similar result to a four minute scan. Similarly, all of the eight byte scan results are within ±1% of their corresponding 512 byte scan results. Therefore, the answer to the second sub-question is: yes, processing all of the bytes in each scanned sector does produce a similar result to processing eight bytes in each scanned sector. The finding regarding the similarity of the results from the eight and 512 byte tests demonstrates that there is minimal effect on the accuracy of the results when processing only eight bytes in each scanned sector as opposed to processing all 512 bytes of the sector.

6.3. LIMITATIONS

During the research, all forty eight tests were performed by the author on four different testing platforms. The testing platforms all had the Microsoft Windows XP operating system installed. The scanning tool has therefore been thoroughly tested on the Window XP operating system.

The author also used the scanning tool on a Windows 7 platform, however, using the tool on Windows 7 was performed to satisfy the author that the tool does operate on a Windows 7 system and was not performed as part of the research testing process for this research. There was no data collected and/or analysed including tool verification while the tool was used on the Windows 7 system. Further testing will be necessary to verify that the tool operates correctly and produces correct information on systems such as Windows Vista and Windows 7.

Currently the tool must be installed onto a computer system before it is used. The intention is that the tool is installed on the examiner's system prior to deployment to a site or scene. The examiner will then have the opportunity to remove the drive from the target system and connect the drive to a write blocking device that is connected to the examiner's workstation. The source drive will then be ready for processing.

Originally nine hard drive test subjects were selected for the testing phase of the research. Modifications were made to the methodology where only eight of the nine test subjects were tested. Test subject 9, which is a 2TB hard drive, was not tested. The decision not to test the 2TB was made because of expected long processing times associated with testing the drive and the unavailability of high-end testing systems.

All eight remaining hard drive test subjects used during the testing phase of the research had a Windows operating system installed on them. The data on the test subjects was not modified in any way to ensure that the results from the testing produced a true representation of the type of data stored on the drives. There were eight test subjects used for testing phase of the research. The author believed that the eight drives would be a sufficient sample for testing. However, because the results of the standard and two minute tests were so accurate and no tests showed any results that were more than ±1% different to the corresponding full sector scan, the author has reservations as to whether eight test subjects were a sufficient data set for testing.

6.4. FUTURE RESEARCH

As mentioned in Section 6.3, the tool must be installed onto an examiners workstation before the tool can be used. Therefore, in order to scan a device such as a hard drive, the examiner would be required to remove the hard drive from the target system and connect the drive to the examiner's workstation through a write blocking device. Currently the tool cannot be used as a standalone executable file. Further research is suggested into developing a scanning tool with a standalone executable file as it would be beneficial to the examiner. A standalone executable file would allow the examiner to process an operating server system by running the scanning tool from a removable storage device such as a USB flash drive.

The research is focused on developing a tool that will determine the percentages of empty and used sectors on a storage device. Although in Section 5.3 the author has provided some recommendations on how the results from the tests could be used in terms of creating forensic copies of storage devices, the recommendations are based on the research conducted by Cusack and Pearse (2011). Further research into testing and discovering the actual amount of reduction in image processing time by using compression algorithms could be carried out. Further study could be undertaken where the researcher would use a scanning tool on a sample of hard drives, similar to what the author has undertaken in this work, to measure the percentage of empty sectors. The drives would then be initially imaged without compression and then reimaged using several different compression algorithms. Analysis of the image processing times for various compression levels and the empty/used sector ratios would provide valuable benchmark data that can be used by an examiner in the future.

The scan duration is another area that requires research. The first sub-question in Section 6.2 asked whether a two minute scan would produce a similar result to the four minute scan. Astonishingly, the results for both the two and four minute scans are identical. The author suggests that further research and testing be conducted where the duration of a test, for a specific hard drive, is continuously shortened to identify at which point the accuracy of the results become affected and unreliable.

PUBLICATIONS

Cusack, B. & Pearse, J. (2011). Can Compression Reduce Forensic Image Time? *Journal of Applied Computing and Information Technology* Vol. 15, Issue 1. ISSN 2230-4398. Retrieved from http://www.citrenz.ac.nz/jacit/JACIT1501/2011Cusack_Compression.html

REFERENCES

AccessData. (2010). http://www.accessdata.com

Adelstein, F. (2006). *Live forensics: diagnosing your system without killing it first.* doi:10.1145/1113034.1113070

Ahmad, I., Anderson, J., Holler, A., Kambo, R. & Makhija, V.(2003). *An Analysis of Disk Performance in VMware ESX Server Virtual Machines.* doi: 10.1109/WWC.2003.1249058

Allen, B. (2004). *Monitoring hard disks with SMART*. Retrieved from http://www.linuxjournal.com/magazine/monitoring-hard-disks-smart

Ariolic Software. (2011). Retrieved from http://www.ariolic.com/activesmart/usb-smart.html

Belleson, J. & Grochowski, E. (1998). *The era of giant magnetoresistive heads.* Retrieved from https://www1.hitachigst.com/hdd/technolo/gmr/gmr.htm

BitLocker. (2011). *BitLocker Drive Encryption Overview.* Retrieved from http://technet2.microsoft.com/WindowsVista/en/library/ce4d5a2e-59a5-4742-89cc-ef9f5908b4731033.mspx

Bunting, S. (2007a). *The official encase certified examiner study guide: Second edition* (pp. 136-183). Indianapolis, IN: Wiley Publishing, Inc.

Bunting, S. (2007b). *The official encase certified examiner study guide: Second edition* (pp. 140-145). Indianapolis, IN: Wiley Publishing, Inc.

Bunting, S., & Anson, S. (2007a). *Mastering windows network forensics and investigations* (pp. 129-159). Indianapolis, IN: Wiley Publishing, Inc.

Bunting, S., & Anson, S. (2007b). *Mastering windows network forensics and investigations* (pp. 129). Indianapolis, IN: Wiley Publishing, Inc.

Burgess, S. (2011). *Computer Forensics - Criminal vs Civil: What's the Difference?* Retrieved from http://www.governmentsecurity.org/articles/computer-forensics-civil-criminal.html

Cardwell, K., Clinton, T., Cross, M., Gregg, M., Varsalone, J. & Wright, C. (2007). *The official CHFI study guide (Exam 312-49) for computer hacking forensic investigators* (pp. 288-289). Burlington, MA: Syngress Publishing, Inc.

Carroll, O. L., Brannon, S. K., Song, T. & Schwarz, J. M. (2008). *Managing Large Amounts of Electronic Evidence.* Retrieved from http://www.justice.gov/usao/eousa/foia_reading_room/usab5601.pdf

Carrier, B. (2005). *File system forensic analysis* (pp. 47-66). Upper Saddle River, NJ: Pearson Education, Inc.

Carvey, H., & Altheide, C. (2005). Tracking USB storage: Analysis of windows artifacts generated by USB storage devices. *Digital Investigation (2005)*, 2, 94-100. Burlington, MA: Elsevier Academic Press. doi:10.1016/j.diin.2005.04.006

Casey, E. (2003). *Handbook of computer crime investigation* (pp. 201-205). London, UK: Academic Press.

Casey, E. (2004). *Digital evidence and computer crime* (pp. 48-57). London, UK: Academic Press.

Casey, E. (2008). *The impact of full disk encryption on digital forensics.* doi:10.1145/1368506.1368519

Christner, J. & Grevers, T. (2008). *Overcoming Transport and Link Capacity Limitations Through WAN Optimization.* Retrieved from http://www.ciscopress.com/articles/article.asp?p=769557&seqNum=4

Cohen, T., & Schroader, A. (2007). *Alternate data storage forensics* (pp. 5-21). Burlington, MA: Syngress Publishing, Inc.

Craiger, J. P. (2011). Computer Forensics Procedures and Methods. Retrieved http://ncfs.ucf.edu/craiger.forensics.methods.procedures.final.pdf

Crowley, E. (2007). Corporate forensics class design with open source tools and live cd's. *Journal of Computing Sciences in Colleges, 22*(4), 1-7. Retrieved from http://portal.acm.org/citation.cfm?id=1229667

Cummings, T. (n.d.). *The History of Computer Forensics.* Retrieved from http://www.ehow.com/about_5813564_history-computer-forensics.html

Cusack, B. & Pearse, J. (2011). Can Compression Reduce Forensic Image Time?. *Journal of Applied Computing and Information Technology* Vol. 15, Issue 1. ISSN 2230-4398. Retrieved from http://www.citrenz.ac.nz/jacit/JACIT1501/2011Cusack_Compression.html

Davis, N. (2007). Live *memory acquisition for windows operating systems.* Retrieved from http://www.linkpdf.com/download/dl/live-memory-acquisition-for-windows-operating-systems--.pdf

Diskology. (2010). http://www.diskology.com/djforensic.html

e-fense. (2011). http://www.e-fense.com

EASIS. (2011). Retrieved from http://www.easis.com/easis-drive-check.html

Encase Compression. (2008). *Forensic Focus*. Retrieved April 25, 2010, from www.forensicfocus.com/index.php?name=Forums&file=viewtopic&t=3091

Ethernet Specification. (1980). Retrieved from http://ethernethistory.typepad.com/papers/ethernetspec.pdf

F-Response. (2011). http://www.f- response.com/index.php?option=com_content &view=article&id=171:f-response-enterprise-edition&catid=36

ForensicSoft. (2010). Retrieved from http://www.forensicsoft.com/safeblock.php

Forward Discovery. (2011). http://forwarddiscovery.com/

Freeman, H. (2002). Software testing. *Instrumentation & Measurement Magazine, IEEE,* vol.5, no.3, pp. 48-50, Sep 2002 doi: 10.1109/MIM.2002.1028373

Frankl, P.G., Hamlet, R.G., Littlewood, B. & Strigini, L. (1998). *Evaluating testing methods by delivered reliability.* doi:10.1109/32.707695

Guidance Software Inc. (2011). http://www.guidancesoftware.com

Hargreaves, C. & Chivers, H. (2008). Recovery of encryption keys from memory using a linear scan. doi:10.1109/ARES.2008.109

HTCIA. (2011). High technology crime investigation association.

Jo, S. & Hong, D. (2008). *Defense technology of anti forensic.* doi: 10.1109/ICCAS.2008.4694617

Jungl, Y., Kiml, J., Bael, S., Kohl, K., Wool, Y. & Kim, S. (2009). Standard-based Virtual Infrastructure Resource Management for Distributed and Heterogeneous Servers. In *Proceedings of the 11th international conference on Advanced communication technology, 2009. ICACT 2009.* ISBN: 978-8-9551-9138-7

Kenneally, E. & Brown, C. (2005). *Risk sensitive digital evidence collection.* doi:10.1016/j.diin.2005.02.001

Kent, K., Chevalier, S., Grance, T. & Dang, H. (2006). *Guide to Integrating Forensic Techniques into Incident Response.* Retrieved from http://csrc.nist.gov/publications/nistpubs/800-86/SP800-86.pdf

Kim, C., Kim, G. & Shin, B. (2001). *Volume management in SAN environment.* Paper presented at the Eighth International Conference on Parallel and Distributed Systems.

Kozierok, C. (2004). Platter Substrate *Materials.* Retrieved from http://www.pcguide.com/ref/hdd/op/mediaMaterials-c.html

Law, F., Lai, P., Jiang, Z., Ieong, R., Kwan, M., Chow, K., Hui, L., Yiu, S., & Chong, C. (2008). *Protecting digital legal professional privilege (LPP) data.* doi: 10.1109/SADFE.2008.19

Lewallen, R. (2005). *Software development life cycle models.* Retrieved from http://codebetter.com/raymondlewallen/2005/07/13/software-development-life-cycle-models

Mandia, K., Prosise, C. & Pepe, M. (2003). *Incident response and computer forensics, second edition* (pp. 12-32). USA: McGraw-Hill Companies, Inc.

Max Visual Basic. (2010). *Resources for visual basic programmers.* Retrieved from http://www.max-visual-basic.com/history-of-visual-basic.html

McFarland, P. (2007). *A brief history of USB (and a look at the future).* Retrieved from http://adterrasperaspera.com/blog/2007/03/29/why-powered-usb-is-needed-part-1-the-short-history-of-usb/

McKemmish, R. (1999). What is forensic computing? Retrieved from http://www.aic.gov.au/documents/9/C/A/%7B9CA41AE8-EADB-4BBF-9894-64E0DF87BDF7%7Dti118.pdf

Mehboob, R., Khan, S. A., Ahmed, Z. (2006) *High speed lossless data compression architecture.* doi:10.1109/INMIC.2006.358141

Microsoft. (2011a). www.microsoft.com

Microsoft. (2011b). A description of the diskpart command-line utility. Retrieved from http://support.microsoft.com/kb/300415

Microsoft Cluster. (2011). Retrieved from http://support.microsoft.com/kb/140365

Milanovic, S. & Petrovic, Z. (2001). *Building the enterprise-wide storage area network.* doi: 10.1109/EURCON.2001.937782

Mustafa, K., Al-Qutaish, R. & Muhairat, M. (2009). *Classification of software testing tools based on the software testing methods.* doi:10.1109/ICCEE.2009.9

Nance, K., Hay, B. & Bishop, M. (2009). *Investigating the implications of virtual machine introspection for digital forensics.* doi:10.1109/ARES.2009.173

National Institute of Standards and Technology. (2011). http://www.nist.gov/index.html

NIST. (2003). *Test results for disk imaging tools: EnCase 3.20.* Retrieved from http://www.ncjrs.gov/pdffiles1/nij/200031.pdf

OS Platform Statistics. (2011). Retrieved February 11, 2011, from http://www.w3schools.com/browsers/browsers_os.asp

Parasuraman, A. (2000). Technology readiness index (Tri): A multiple-item scale to measure readiness to embrace new technologies. Journal of service research, 2(4), 307-320.

Parekh, N. (2011). The Waterfall Model Explained. Retrieved from http://www.buzzle.com/editorials/1-5-2005-63768.asp

Pearse, J. (2010). Using Compression to Reduce Processing Time. In the *proceedings of the 2010 Digital Forensics International Conference,* 6-7 September 2010, (pp. 21-28). Auckland, New Zealand.

Royce, W. (1970). *Managing the development of large software systems.* Proceedings of IEEE WESCON 26 (August): 1–9, http://www.cs.umd.edu/class/spring2003/cmsc838p/Process/waterfall.pdf

Sammes, T. & Jenkinson, B. (2007). *Forensic computing - a practioners guide, second edition* (pp. 291-299). London, UK: Springer-Verlag

Savoldi, A. & Gubian, P. (2008). Logical and Physical Data Collection of Windows CE Based Portable Devices. In *Proceedings of the 2008 ACM symposium on Applied computing.* ACM, New York, NY, USA. doi:10.1145/1363686.1364023

Seagate. (2011). *Barracuda XT Desktop Hard Drives.* Retrieved from http://www.seagate.com/www/en-us/products/desktops/barracuda_xt

Shinder, D. (2002). *Scene of the cybercrime computer forensic handbook* (pp. 548-557). Rockland, MA: Syngress Publishing, Inc.

Sindi, M., Liu, E. and Al-Shaikh, R. (2009). *SAN performance evaluation testbed.* doi: 10.1109/TRIDENTCOM.2009.4976191. Retreived from http://ieeexplore.ieee.org.ezproxy.aut.ac.nz/search/srchabstract.jsp?tp=&arnumber=4976191&openedRefinements%3D*%26

Soe, L. L., Manson, D. & Wright, M. (2004). Establishing network computer forensics classes. In *Proceedings of the 1st annual conference on Information security curriculum development* (InfoSecCD '04). ACM, New York, NY, USA, 76-81. doi:10.1145/1059524.1059540

Software architecture. (2011). *Microsoft software application architecture.* Retrieved from http://msdn.microsoft.com/en-us/library/ee658098.aspx

Spence, M. E. (2010). *Factors influencing digital evidence transfer across international borders: A case study* (Master’s thesis, Auckland University of Technology, Auckland, New Zealand). Retrieved from http://aut.researchgateway.ac.nz/handle/10292/1187

Sumuri - Forensics Simplified. (2011). http://www.sumuri.com/software.html

Sutherland, I., Evans, J., Tryfonas, T., & Blyth, A. (2008). *Acquiring volatile operating system data tools and techniques.* doi: /10.1145/1368506.1368516

SWGDE. (2009). *SWGDE Recommended Guidelines for Validation Testing.* Retrieved from http://www.swgde.org/documents/current-documents/ 2009-01-15%20SWGDE%20Recommendations%20for%20Validation %20Testing%20Version%20v1.1.pdf

Tableau. (2003-2011). http://www.tableau.com/index.php?pageid =products&model=TD1

Tableau. (2011). http://www.tableau.com/index.php?pageid=products&model =T35es

Tanaka, Y. (2008). *Perpendicular recording technology: From research to commercialization.* doi: 10.1109/JPROC.2008.2004309

The Free Dictionary. (2011). Retrieved from http://www.thefreedictionary.com/triage

TechFive Systems. (2011). *Software Development Life Cycle.* Retrieved from http://www.techfivesystems.com/about/offshore-application-development

USB specification. (2011). http://www.usb.org/developers/docs/

USB Write Protect. (2005). *Developed by Joz Ong:* NSW Police, State Electronic Evidence Branch.

Using Spanned Volumes. (2011). http://www.microsoft.com/resources/ documentation/windows/xp/all/proddocs/en-us/sag_diskconcepts_15. mspx?mfr=true

VMware. (2011). http://www.vmware.com/products/vsphere/esxi-and-esx/index.html

Villi, C. (2010). Hardening the business information system with forensic readiness. In the *proceedings of the 2010 Digital Forensics International Conference,* 6-7 September 2010. Auckland, New Zealand.

Voom Technologies Inc. (2011). http://www.voomtech.com/hc3.html

Whittaker, J.A. (2000). What is software testing? And why is it so hard? *Software, IEEE,* vol.17, no.1, pp.70-79, Jan/Feb 2000
doi: 10.1109/52.819971

Xu, M., He, J., Zhang, B. & Zhang, H. (2007). *A new data protecting scheme based on TPM.* doi: 10.1109/SNPD.2007.276

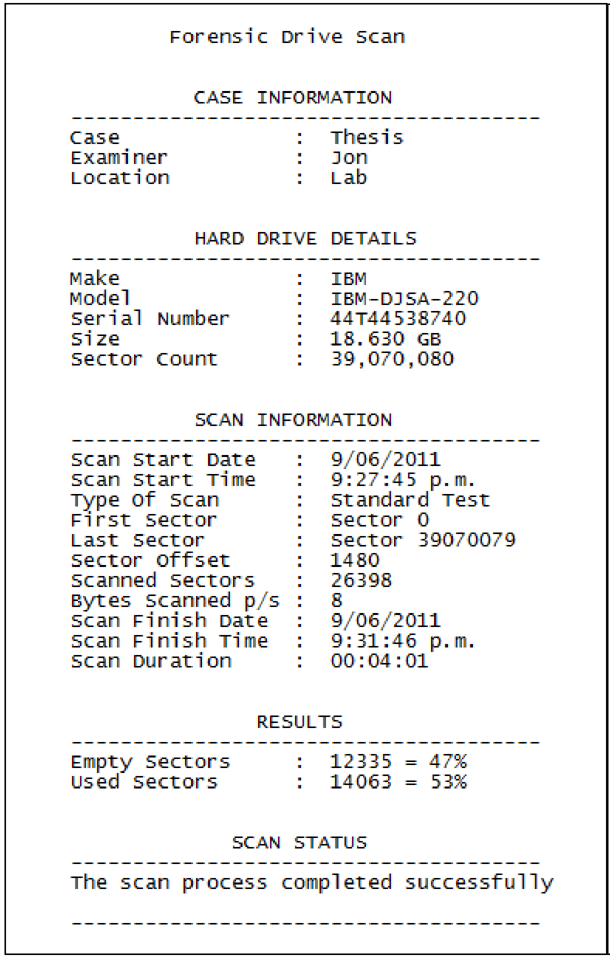
Yao, Y. & Wang, Y. (2005). A framework for testing distributed software components. Canadian Conference on Electrical and Computer Engineering 2005 (2005) Issue: May, Publisher: Ieee, Pages: 1574-1577
doi:10.1109/CCECE.2005.1557280

Zhang, J. & Wang, L. (2009). *Application of Case-oriented Evidence Mining in Forensic Computing.* doi:10.1109/MINES.2009.178

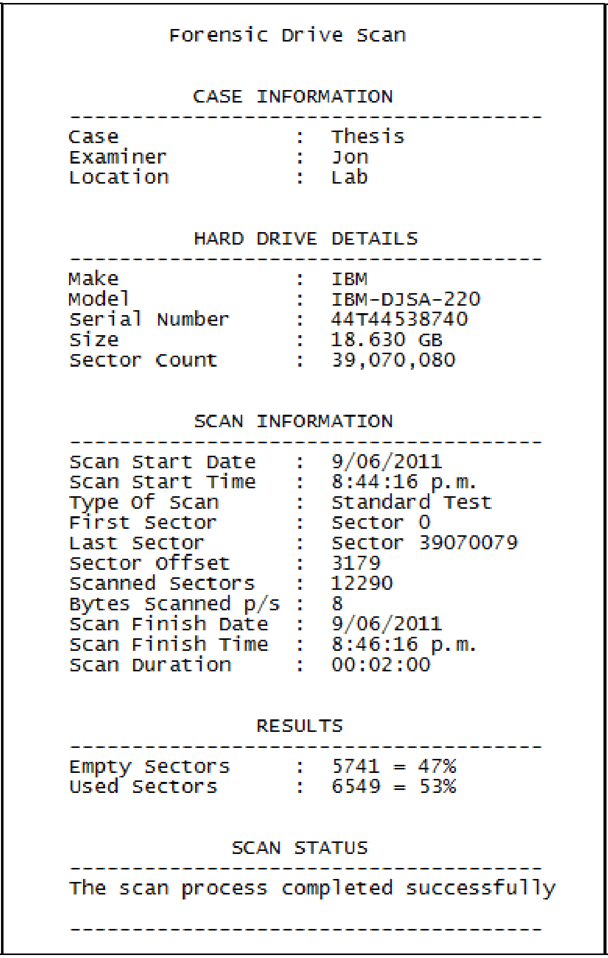
Zintel, M. (2002). Ohio appellate court upholds use of *guidance software's encase computer forensic software.* Retrieved from http://www.thefreelibrary.com
/Ohio+Appellate+Court+Upholds+Use+of+Guidance+Software's+EnCase.
..-a092810999

APPENDIX

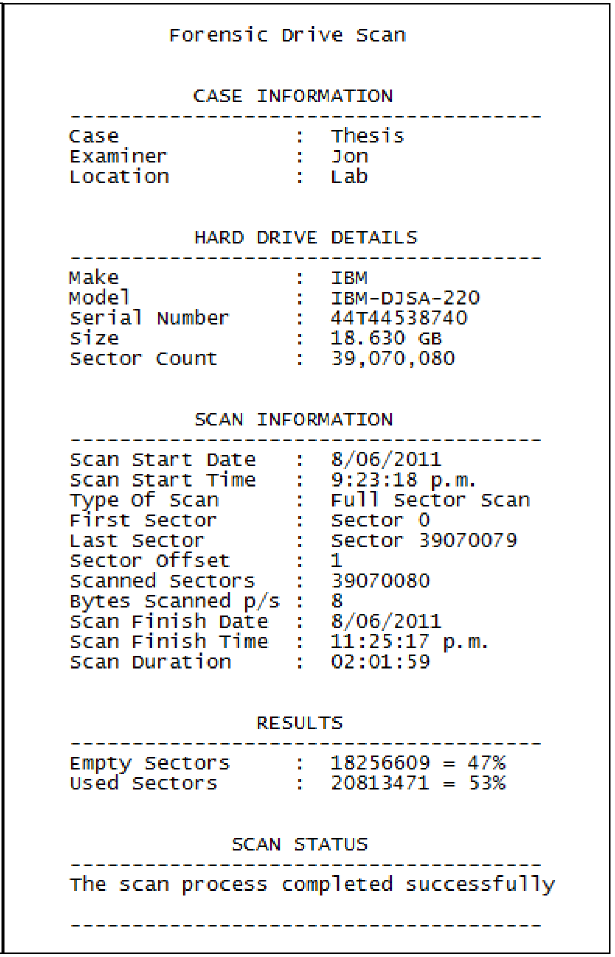
Eight Byte Test Reports for Test Subject 1



```
          Forensic Drive Scan

            CASE INFORMATION
     ----------------------------------------
     Case                :  Thesis
     Examiner            :  Jon
     Location            :  Lab

            HARD DRIVE DETAILS
     ----------------------------------------
     Make                :  IBM
     Model               :  IBM-DJSA-220
     Serial Number       :  44T44538740
     Size                :  18.630 GB
     Sector Count        :  39,070,080

            SCAN INFORMATION
     ----------------------------------------
     Scan Start Date     :  9/06/2011
     Scan Start Time     :  9:27:45 p.m.
     Type Of Scan        :  Standard Test
     First Sector        :  Sector 0
     Last Sector         :  Sector 39070079
     Sector Offset       :  1480
     Scanned Sectors     :  26398
     Bytes Scanned p/s   :  8
     Scan Finish Date    :  9/06/2011
     Scan Finish Time    :  9:31:46 p.m.
     Scan Duration       :  00:04:01

                  RESULTS
     ----------------------------------------
     Empty Sectors       :  12335 = 47%
     Used Sectors        :  14063 = 53%

                SCAN STATUS
     ----------------------------------------
     The scan process completed successfully

     ----------------------------------------
```

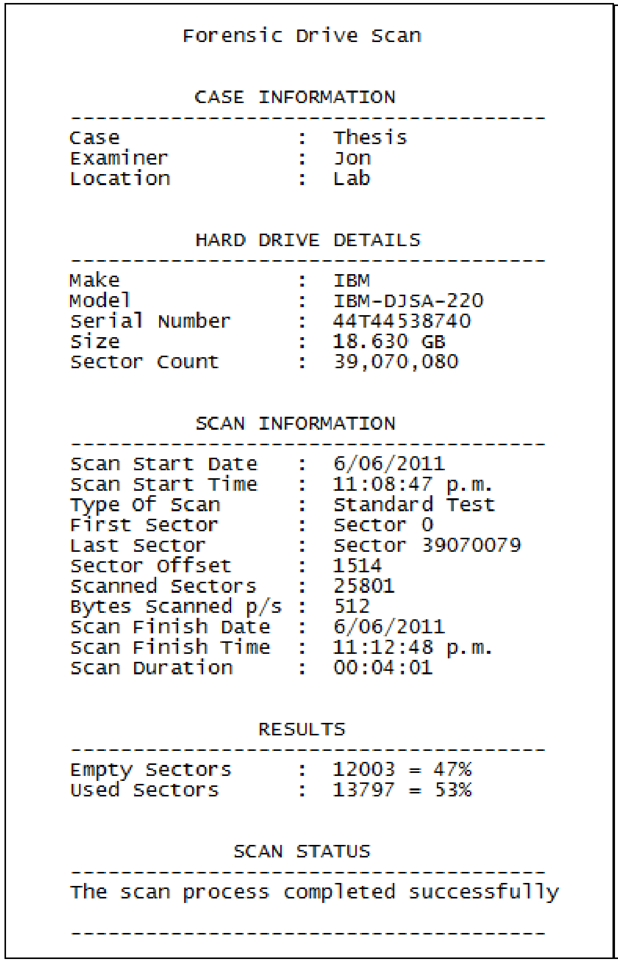


```
          Forensic Drive Scan

            CASE INFORMATION
     ----------------------------------------
     Case                :  Thesis
     Examiner            :  Jon
     Location            :  Lab

            HARD DRIVE DETAILS
     ----------------------------------------
     Make                :  IBM
     Model               :  IBM-DJSA-220
     Serial Number       :  44T44538740
     Size                :  18.630 GB
     Sector Count        :  39,070,080

            SCAN INFORMATION
     ----------------------------------------
     Scan Start Date     :  9/06/2011
     Scan Start Time     :  8:44:16 p.m.
     Type Of Scan        :  Standard Test
     First Sector        :  Sector 0
     Last Sector         :  Sector 39070079
     Sector Offset       :  3179
     Scanned Sectors     :  12290
     Bytes Scanned p/s   :  8
     Scan Finish Date    :  9/06/2011
     Scan Finish Time    :  8:46:16 p.m.
     Scan Duration       :  00:02:00

                  RESULTS
     ----------------------------------------
     Empty Sectors       :  5741 = 47%
     Used Sectors        :  6549 = 53%

                SCAN STATUS
     ----------------------------------------
     The scan process completed successfully

     ----------------------------------------
```

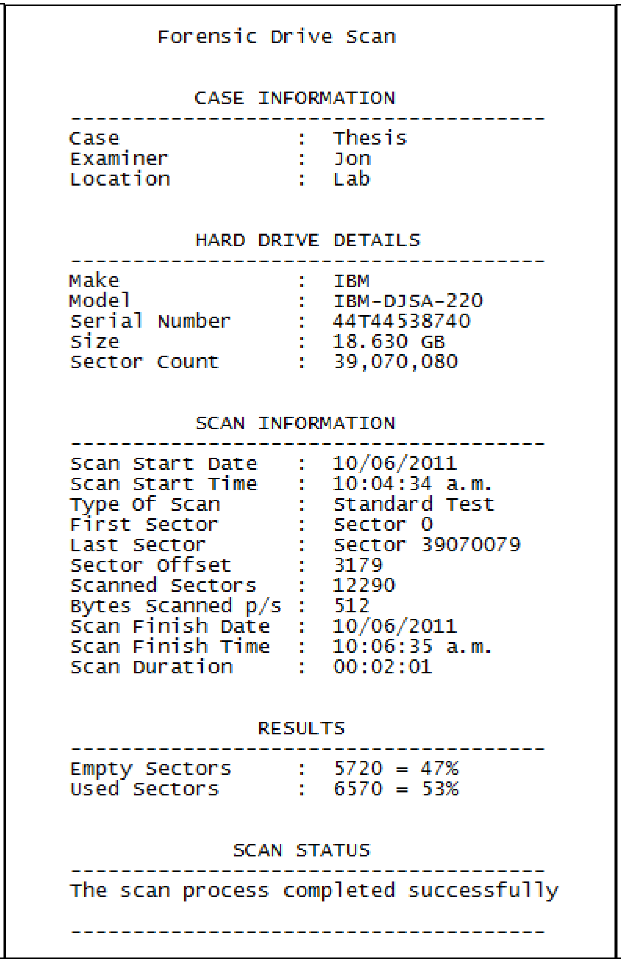


```
          Forensic Drive Scan

            CASE INFORMATION
     ----------------------------------------
     Case                :  Thesis
     Examiner            :  Jon
     Location            :  Lab

            HARD DRIVE DETAILS
     ----------------------------------------
     Make                :  IBM
     Model               :  IBM-DJSA-220
     Serial Number       :  44T44538740
     Size                :  18.630 GB
     Sector Count        :  39,070,080

            SCAN INFORMATION
     ----------------------------------------
     Scan Start Date     :  8/06/2011
     Scan Start Time     :  9:23:18 p.m.
     Type Of Scan        :  Full Sector Scan
     First Sector        :  Sector 0
     Last Sector         :  Sector 39070079
     Sector Offset       :  1
     Scanned Sectors     :  39070080
     Bytes Scanned p/s   :  8
     Scan Finish Date    :  8/06/2011
     Scan Finish Time    :  11:25:17 p.m.
     Scan Duration       :  02:01:59

                  RESULTS
     ----------------------------------------
     Empty Sectors       :  18256609 = 47%
     Used Sectors        :  20813471 = 53%

                SCAN STATUS
     ----------------------------------------
     The scan process completed successfully

     ----------------------------------------
```

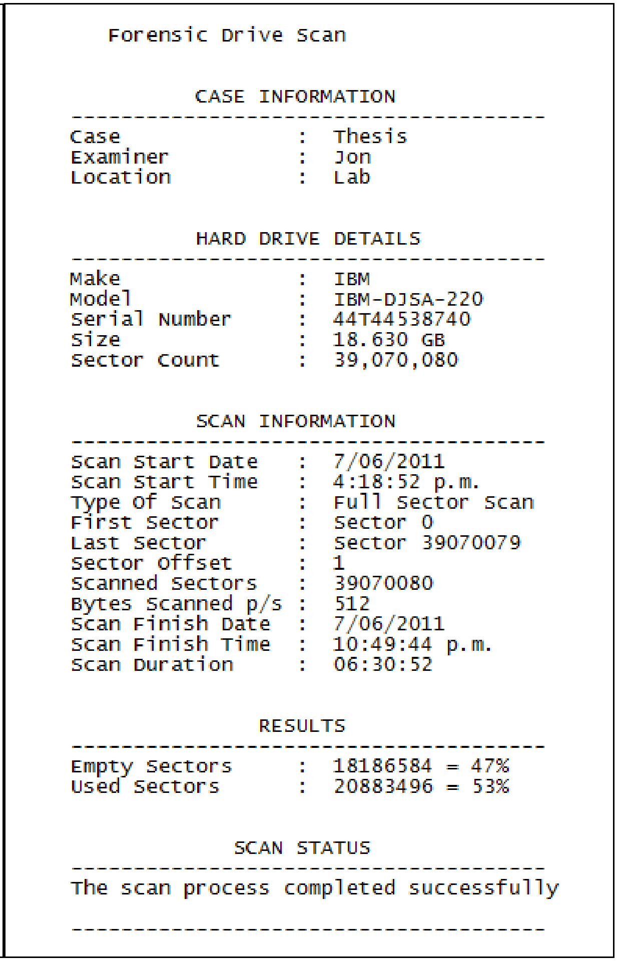
512 Byte Test Reports for Test Subject 1



```
            Forensic Drive Scan

             CASE INFORMATION
    ----------------------------------------
    Case               :  Thesis
    Examiner           :  Jon
    Location           :  Lab

             HARD DRIVE DETAILS
    ----------------------------------------
    Make               :  IBM
    Model              :  IBM-DJSA-220
    Serial Number      :  44T44538740
    Size               :  18.630 GB
    Sector Count       :  39,070,080

             SCAN INFORMATION
    ----------------------------------------
    Scan Start Date    :  6/06/2011
    Scan Start Time    :  11:08:47 p.m.
    Type Of Scan       :  Standard Test
    First Sector       :  Sector 0
    Last Sector        :  Sector 39070079
    Sector Offset      :  1514
    Scanned Sectors    :  25801
    Bytes Scanned p/s  :  512
    Scan Finish Date   :  6/06/2011
    Scan Finish Time   :  11:12:48 p.m.
    Scan Duration      :  00:04:01

                  RESULTS
    ----------------------------------------
    Empty Sectors      :  12003 = 47%
    Used Sectors       :  13797 = 53%

                SCAN STATUS
    ----------------------------------------
    The scan process completed successfully

    ----------------------------------------
```

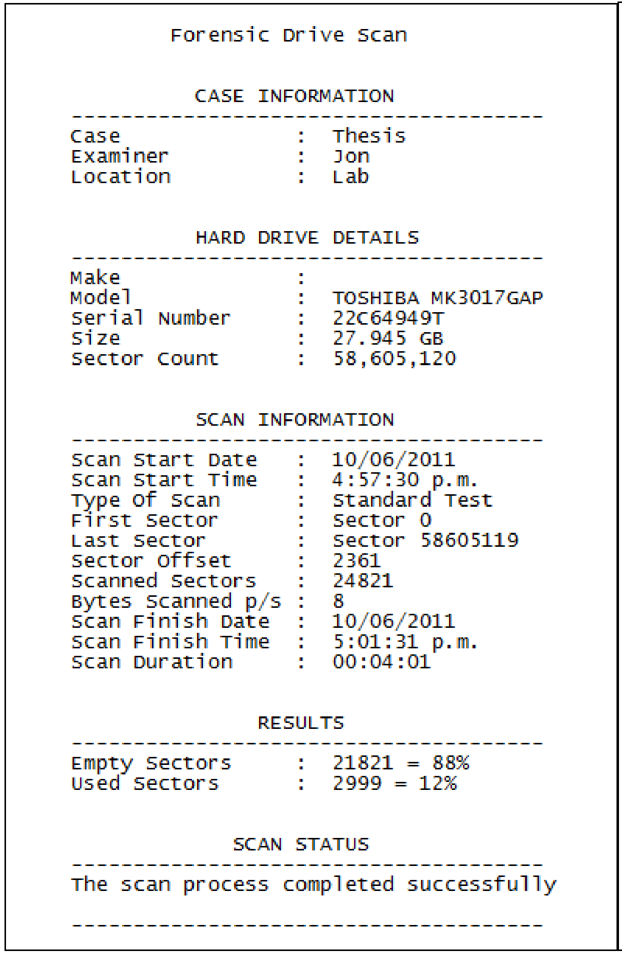


```
            Forensic Drive Scan

             CASE INFORMATION
    ----------------------------------------
    Case               :  Thesis
    Examiner           :  Jon
    Location           :  Lab

             HARD DRIVE DETAILS
    ----------------------------------------
    Make               :  IBM
    Model              :  IBM-DJSA-220
    Serial Number      :  44T44538740
    Size               :  18.630 GB
    Sector Count       :  39,070,080

             SCAN INFORMATION
    ----------------------------------------
    Scan Start Date    :  10/06/2011
    Scan Start Time    :  10:04:34 a.m.
    Type Of Scan       :  Standard Test
    First Sector       :  Sector 0
    Last Sector        :  Sector 39070079
    Sector Offset      :  3179
    Scanned Sectors    :  12290
    Bytes Scanned p/s  :  512
    Scan Finish Date   :  10/06/2011
    Scan Finish Time   :  10:06:35 a.m.
    Scan Duration      :  00:02:01

                  RESULTS
    ----------------------------------------
    Empty Sectors      :  5720 = 47%
    Used Sectors       :  6570 = 53%

                SCAN STATUS
    ----------------------------------------
    The scan process completed successfully

    ----------------------------------------
```

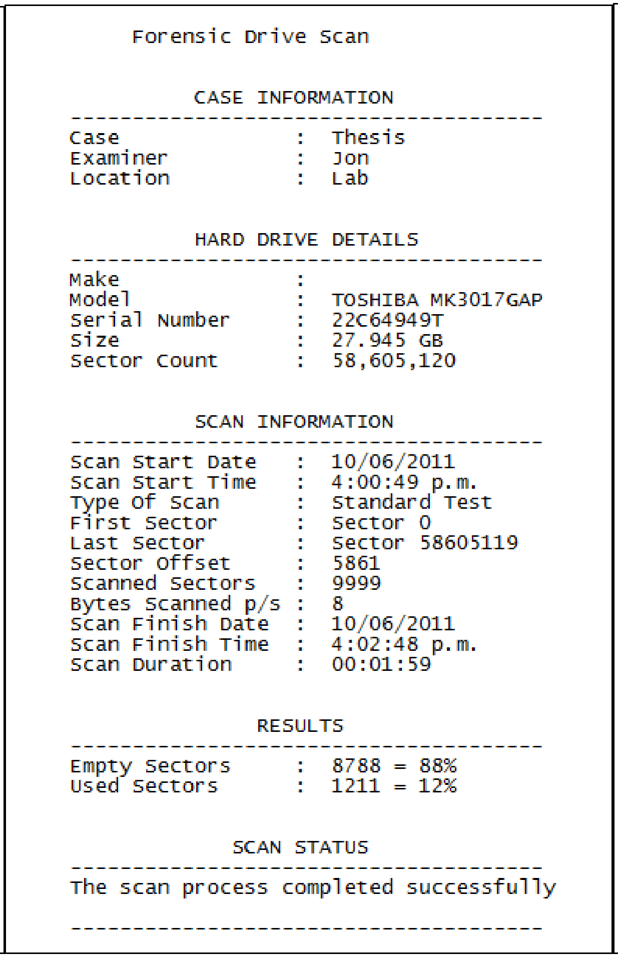


```
            Forensic Drive Scan

             CASE INFORMATION
    ----------------------------------------
    Case               :  Thesis
    Examiner           :  Jon
    Location           :  Lab

             HARD DRIVE DETAILS
    ----------------------------------------
    Make               :  IBM
    Model              :  IBM-DJSA-220
    Serial Number      :  44T44538740
    Size               :  18.630 GB
    Sector Count       :  39,070,080

             SCAN INFORMATION
    ----------------------------------------
    Scan Start Date    :  7/06/2011
    Scan Start Time    :  4:18:52 p.m.
    Type Of Scan       :  Full Sector Scan
    First Sector       :  Sector 0
    Last Sector        :  Sector 39070079
    Sector Offset      :  1
    Scanned Sectors    :  39070080
    Bytes Scanned p/s  :  512
    Scan Finish Date   :  7/06/2011
    Scan Finish Time   :  10:49:44 p.m.
    Scan Duration      :  06:30:52

                  RESULTS
    ----------------------------------------
    Empty Sectors      :  18186584 = 47%
    Used Sectors       :  20883496 = 53%

                SCAN STATUS
    ----------------------------------------
    The scan process completed successfully

    ----------------------------------------
```

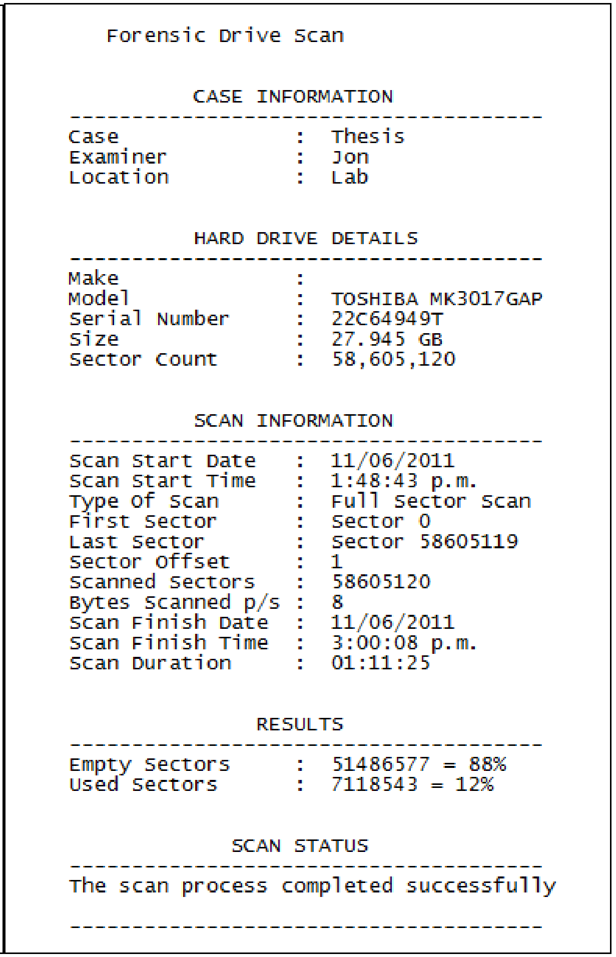
Eight Byte Test Reports for Test Subject 2



```
          Forensic Drive Scan

            CASE INFORMATION
 ---------------------------------------
 Case                 :  Thesis
 Examiner             :  Jon
 Location             :  Lab

          HARD DRIVE DETAILS
 ---------------------------------------
 Make                 :
 Model                :  TOSHIBA MK3017GAP
 Serial Number        :  22C64949T
 Size                 :  27.945 GB
 Sector Count         :  58,605,120

           SCAN INFORMATION
 ---------------------------------------
 Scan Start Date      :  10/06/2011
 Scan Start Time      :  4:57:30 p.m.
 Type Of Scan         :  Standard Test
 First Sector         :  Sector 0
 Last Sector          :  Sector 58605119
 Sector Offset        :  2361
 Scanned Sectors      :  24821
 Bytes Scanned p/s    :  8
 Scan Finish Date     :  10/06/2011
 Scan Finish Time     :  5:01:31 p.m.
 Scan Duration        :  00:04:01

                RESULTS
 ---------------------------------------
 Empty Sectors        :  21821 = 88%
 Used Sectors         :  2999 = 12%

              SCAN STATUS
 ---------------------------------------
 The scan process completed successfully
 ---------------------------------------
```

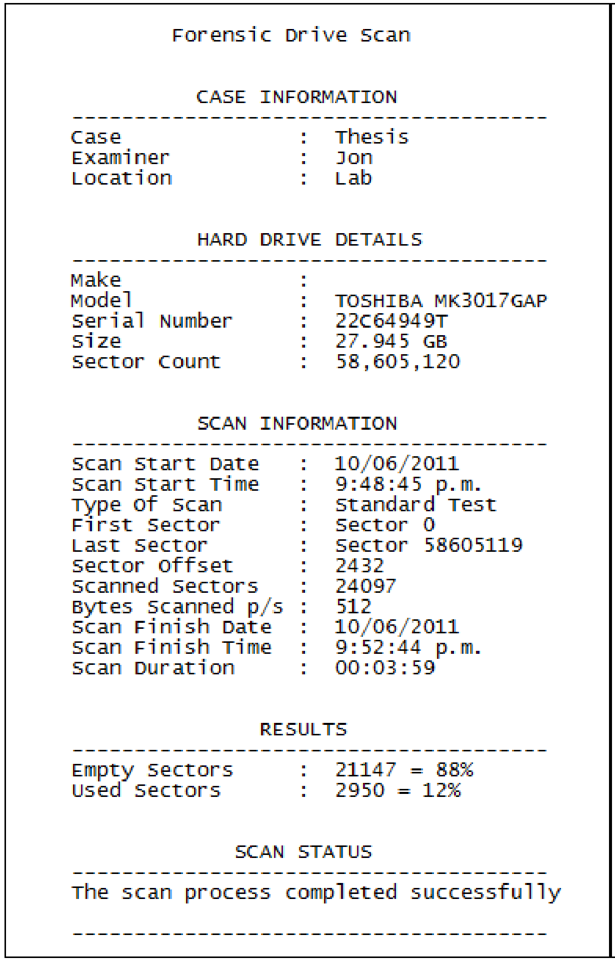


```
          Forensic Drive Scan

            CASE INFORMATION
 ---------------------------------------
 Case                 :  Thesis
 Examiner             :  Jon
 Location             :  Lab

          HARD DRIVE DETAILS
 ---------------------------------------
 Make                 :
 Model                :  TOSHIBA MK3017GAP
 Serial Number        :  22C64949T
 Size                 :  27.945 GB
 Sector Count         :  58,605,120

           SCAN INFORMATION
 ---------------------------------------
 Scan Start Date      :  10/06/2011
 Scan Start Time      :  4:00:49 p.m.
 Type Of Scan         :  Standard Test
 First Sector         :  Sector 0
 Last Sector          :  Sector 58605119
 Sector Offset        :  5861
 Scanned Sectors      :  9999
 Bytes Scanned p/s    :  8
 Scan Finish Date     :  10/06/2011
 Scan Finish Time     :  4:02:48 p.m.
 Scan Duration        :  00:01:59

                RESULTS
 ---------------------------------------
 Empty Sectors        :  8788 = 88%
 Used Sectors         :  1211 = 12%

              SCAN STATUS
 ---------------------------------------
 The scan process completed successfully
 ---------------------------------------
```

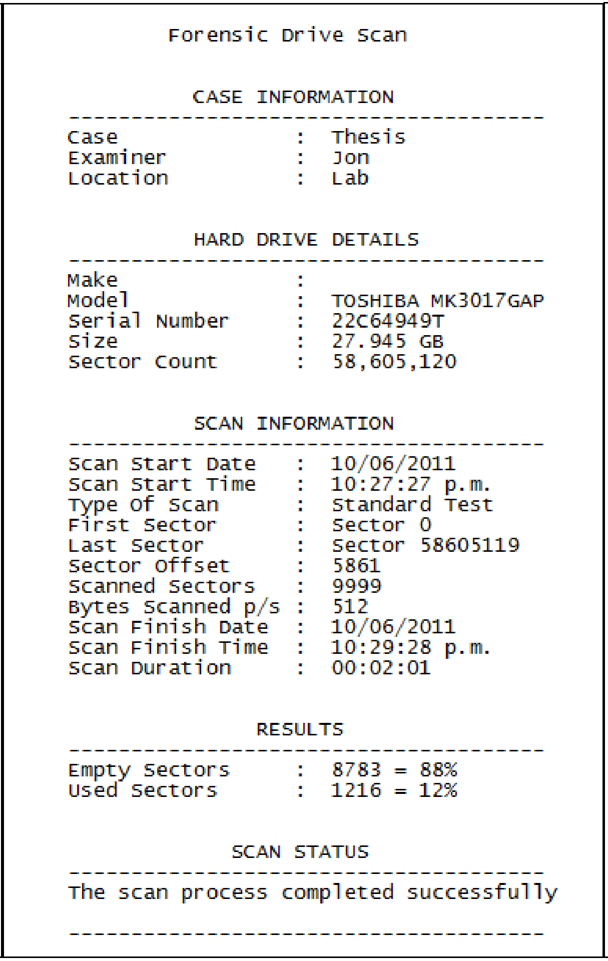


```
          Forensic Drive Scan

            CASE INFORMATION
 ---------------------------------------
 Case                 :  Thesis
 Examiner             :  Jon
 Location             :  Lab

          HARD DRIVE DETAILS
 ---------------------------------------
 Make                 :
 Model                :  TOSHIBA MK3017GAP
 Serial Number        :  22C64949T
 Size                 :  27.945 GB
 Sector Count         :  58,605,120

           SCAN INFORMATION
 ---------------------------------------
 Scan Start Date      :  11/06/2011
 Scan Start Time      :  1:48:43 p.m.
 Type Of Scan         :  Full Sector Scan
 First Sector         :  Sector 0
 Last Sector          :  Sector 58605119
 Sector Offset        :  1
 Scanned Sectors      :  58605120
 Bytes Scanned p/s    :  8
 Scan Finish Date     :  11/06/2011
 Scan Finish Time     :  3:00:08 p.m.
 Scan Duration        :  01:11:25

                RESULTS
 ---------------------------------------
 Empty Sectors        :  51486577 = 88%
 Used Sectors         :  7118543 = 12%

              SCAN STATUS
 ---------------------------------------
 The scan process completed successfully
 ---------------------------------------
```

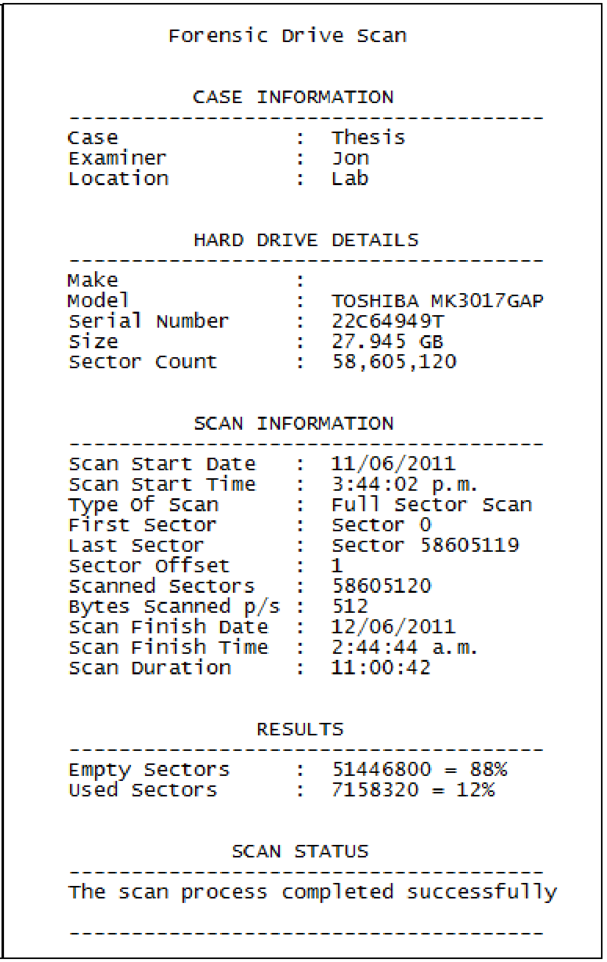
512 Byte Test Reports for Test Subject 2



```
          Forensic Drive Scan

            CASE INFORMATION
--------------------------------------
Case                :  Thesis
Examiner            :  Jon
Location            :  Lab

            HARD DRIVE DETAILS
--------------------------------------
Make                :
Model               :  TOSHIBA MK3017GAP
Serial Number       :  22C64949T
Size                :  27.945 GB
Sector Count        :  58,605,120

            SCAN INFORMATION
--------------------------------------
Scan Start Date     :  10/06/2011
Scan Start Time     :  9:48:45 p.m.
Type Of Scan        :  Standard Test
First Sector        :  Sector 0
Last Sector         :  Sector 58605119
Sector Offset       :  2432
Scanned Sectors     :  24097
Bytes Scanned p/s   :  512
Scan Finish Date    :  10/06/2011
Scan Finish Time    :  9:52:44 p.m.
Scan Duration       :  00:03:59

                 RESULTS
--------------------------------------
Empty Sectors       :  21147 = 88%
Used Sectors        :  2950 = 12%

               SCAN STATUS
--------------------------------------
The scan process completed successfully

--------------------------------------
```

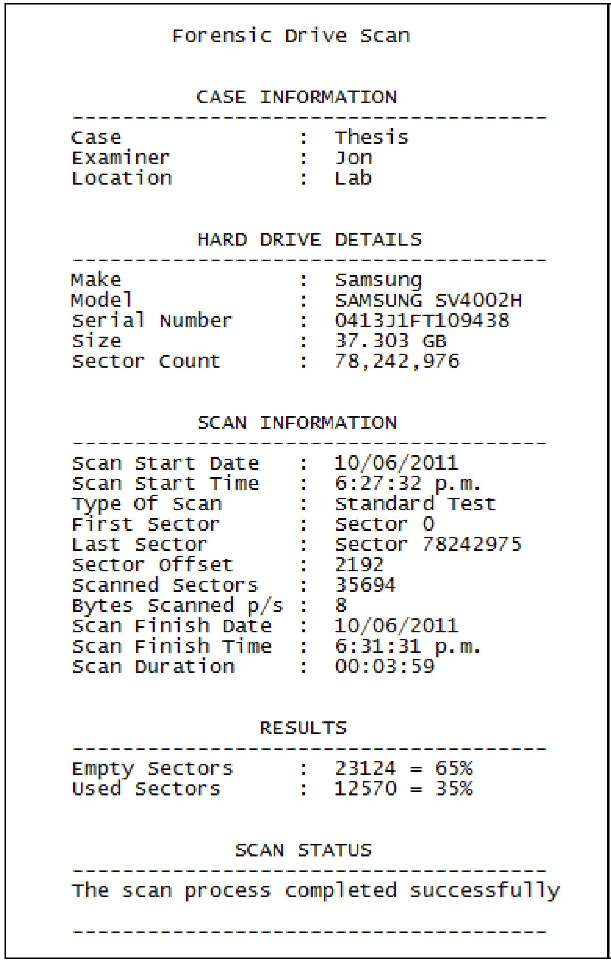


```
          Forensic Drive Scan

            CASE INFORMATION
--------------------------------------
Case                :  Thesis
Examiner            :  Jon
Location            :  Lab

            HARD DRIVE DETAILS
--------------------------------------
Make                :
Model               :  TOSHIBA MK3017GAP
Serial Number       :  22C64949T
Size                :  27.945 GB
Sector Count        :  58,605,120

            SCAN INFORMATION
--------------------------------------
Scan Start Date     :  10/06/2011
Scan Start Time     :  10:27:27 p.m.
Type Of Scan        :  Standard Test
First Sector        :  Sector 0
Last Sector         :  Sector 58605119
Sector Offset       :  5861
Scanned Sectors     :  9999
Bytes Scanned p/s   :  512
Scan Finish Date    :  10/06/2011
Scan Finish Time    :  10:29:28 p.m.
Scan Duration       :  00:02:01

                 RESULTS
--------------------------------------
Empty Sectors       :  8783 = 88%
Used Sectors        :  1216 = 12%

               SCAN STATUS
--------------------------------------
The scan process completed successfully

--------------------------------------
```

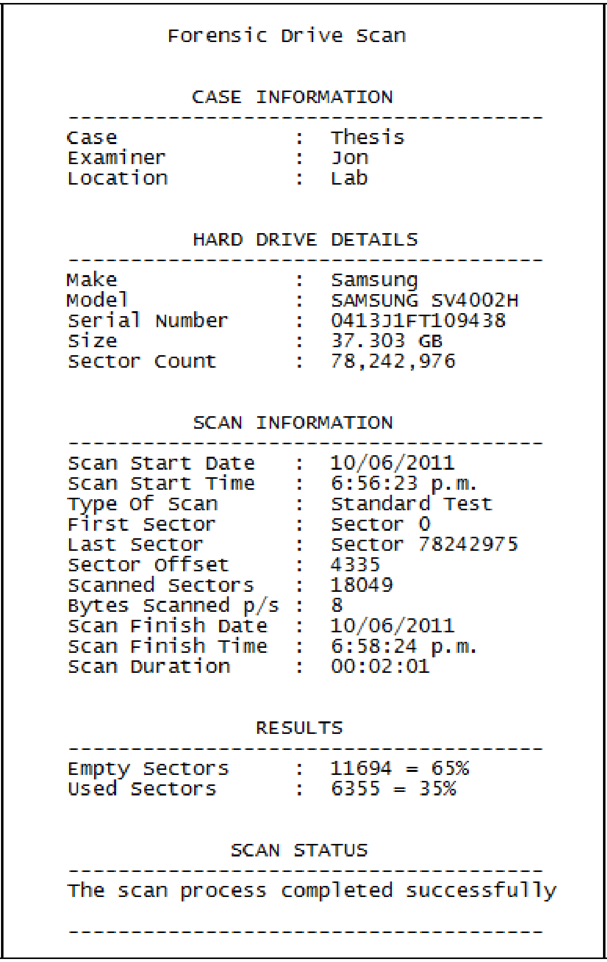


```
          Forensic Drive Scan

            CASE INFORMATION
--------------------------------------
Case                :  Thesis
Examiner            :  Jon
Location            :  Lab

            HARD DRIVE DETAILS
--------------------------------------
Make                :
Model               :  TOSHIBA MK3017GAP
Serial Number       :  22C64949T
Size                :  27.945 GB
Sector Count        :  58,605,120

            SCAN INFORMATION
--------------------------------------
Scan Start Date     :  11/06/2011
Scan Start Time     :  3:44:02 p.m.
Type Of Scan        :  Full Sector Scan
First Sector        :  Sector 0
Last Sector         :  Sector 58605119
Sector Offset       :  1
Scanned Sectors     :  58605120
Bytes Scanned p/s   :  512
Scan Finish Date    :  12/06/2011
Scan Finish Time    :  2:44:44 a.m.
Scan Duration       :  11:00:42

                 RESULTS
--------------------------------------
Empty Sectors       :  51446800 = 88%
Used Sectors        :  7158320 = 12%

               SCAN STATUS
--------------------------------------
The scan process completed successfully

--------------------------------------
```

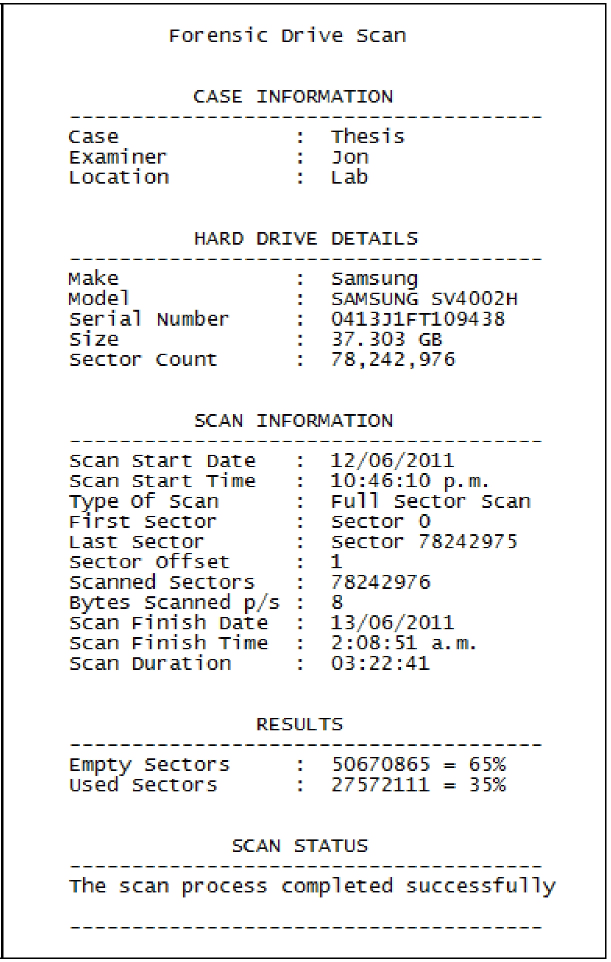
Eight Byte Test Reports for Test Subject 3



```
           Forensic Drive Scan

            CASE INFORMATION
---------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

           HARD DRIVE DETAILS
---------------------------------------
Make                 :  Samsung
Model                :  SAMSUNG SV4002H
Serial Number        :  0413J1FT109438
Size                 :  37.303 GB
Sector Count         :  78,242,976

            SCAN INFORMATION
---------------------------------------
Scan Start Date      :  10/06/2011
Scan Start Time      :  6:27:32 p.m.
Type Of Scan         :  Standard Test
First Sector         :  Sector 0
Last Sector          :  Sector 78242975
Sector Offset        :  2192
Scanned Sectors      :  35694
Bytes Scanned p/s    :  8
Scan Finish Date     :  10/06/2011
Scan Finish Time     :  6:31:31 p.m.
Scan Duration        :  00:03:59

                 RESULTS
---------------------------------------
Empty Sectors        :  23124 = 65%
Used Sectors         :  12570 = 35%

              SCAN STATUS
---------------------------------------
The scan process completed successfully

---------------------------------------
```

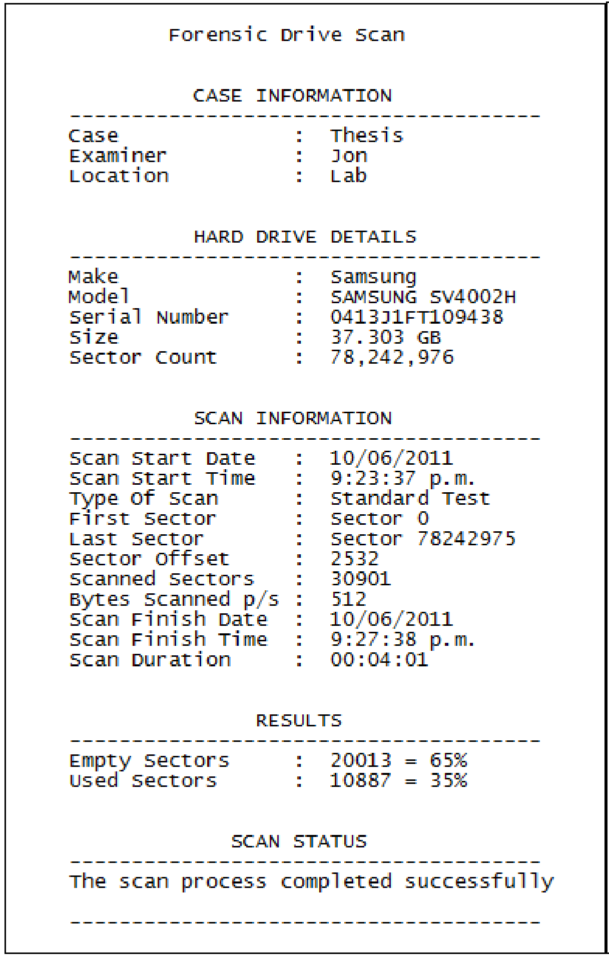


```
           Forensic Drive Scan

            CASE INFORMATION
---------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

           HARD DRIVE DETAILS
---------------------------------------
Make                 :  Samsung
Model                :  SAMSUNG SV4002H
Serial Number        :  0413J1FT109438
Size                 :  37.303 GB
Sector Count         :  78,242,976

            SCAN INFORMATION
---------------------------------------
Scan Start Date      :  10/06/2011
Scan Start Time      :  6:56:23 p.m.
Type Of Scan         :  Standard Test
First Sector         :  Sector 0
Last Sector          :  Sector 78242975
Sector Offset        :  4335
Scanned Sectors      :  18049
Bytes Scanned p/s    :  8
Scan Finish Date     :  10/06/2011
Scan Finish Time     :  6:58:24 p.m.
Scan Duration        :  00:02:01

                 RESULTS
---------------------------------------
Empty Sectors        :  11694 = 65%
Used Sectors         :  6355 = 35%

              SCAN STATUS
---------------------------------------
The scan process completed successfully

---------------------------------------
```

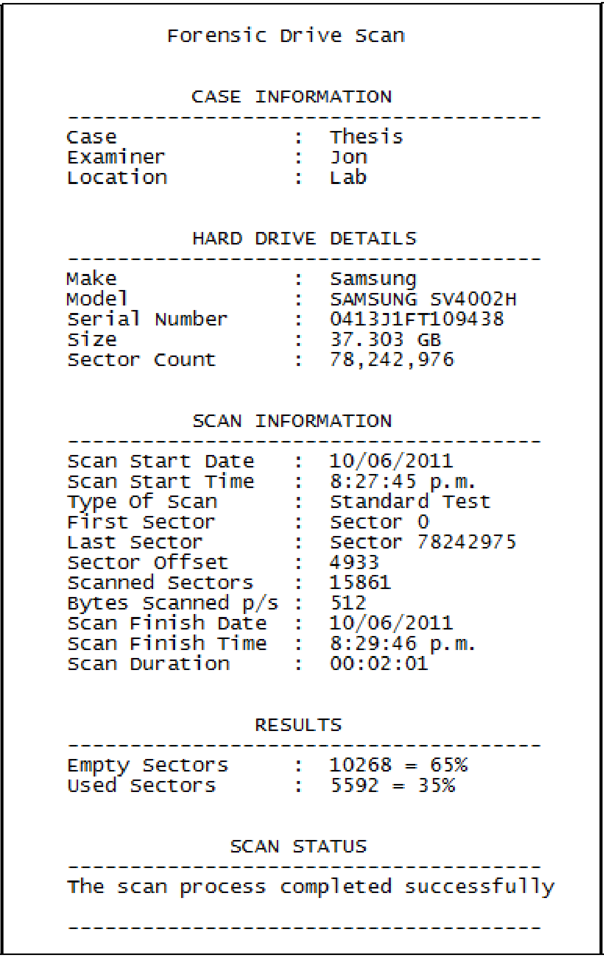


```
           Forensic Drive Scan

            CASE INFORMATION
---------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

           HARD DRIVE DETAILS
---------------------------------------
Make                 :  Samsung
Model                :  SAMSUNG SV4002H
Serial Number        :  0413J1FT109438
Size                 :  37.303 GB
Sector Count         :  78,242,976

            SCAN INFORMATION
---------------------------------------
Scan Start Date      :  12/06/2011
Scan Start Time      :  10:46:10 p.m.
Type Of Scan         :  Full Sector Scan
First Sector         :  Sector 0
Last Sector          :  Sector 78242975
Sector Offset        :  1
Scanned Sectors      :  78242976
Bytes Scanned p/s    :  8
Scan Finish Date     :  13/06/2011
Scan Finish Time     :  2:08:51 a.m.
Scan Duration        :  03:22:41

                 RESULTS
---------------------------------------
Empty Sectors        :  50670865 = 65%
Used Sectors         :  27572111 = 35%

              SCAN STATUS
---------------------------------------
The scan process completed successfully

---------------------------------------
```

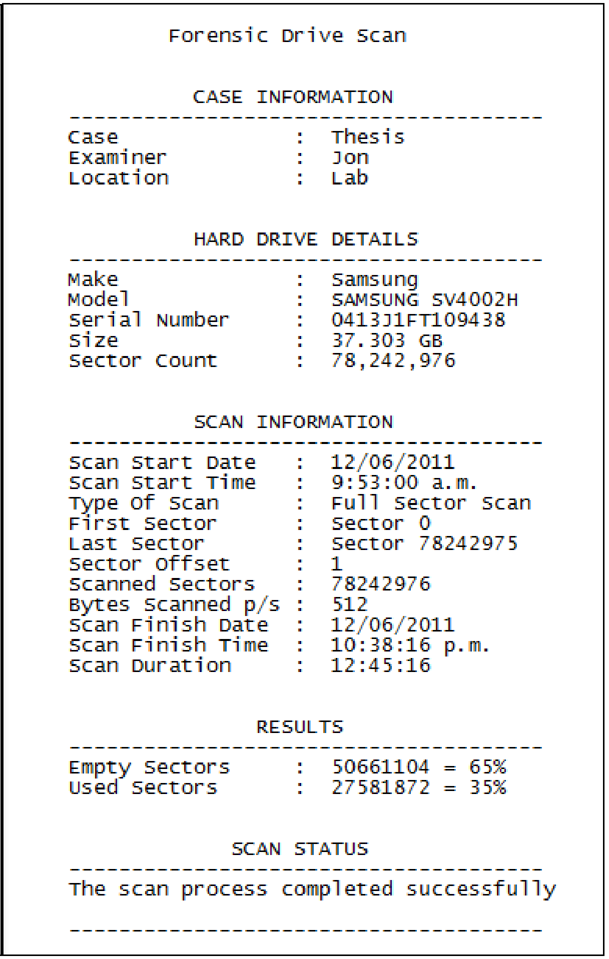
512 Byte Test Reports for Test Subject 3



```
            Forensic Drive Scan

             CASE INFORMATION
--------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

            HARD DRIVE DETAILS
--------------------------------------
Make                 :  Samsung
Model                :  SAMSUNG SV4002H
Serial Number        :  0413J1FT109438
Size                 :  37.303 GB
Sector Count         :  78,242,976

             SCAN INFORMATION
--------------------------------------
Scan Start Date      :  10/06/2011
Scan Start Time      :  9:23:37 p.m.
Type Of Scan         :  Standard Test
First Sector         :  Sector 0
Last Sector          :  Sector 78242975
Sector Offset        :  2532
Scanned Sectors      :  30901
Bytes Scanned p/s    :  512
Scan Finish Date     :  10/06/2011
Scan Finish Time     :  9:27:38 p.m.
Scan Duration        :  00:04:01

                  RESULTS
--------------------------------------
Empty Sectors        :  20013 = 65%
Used Sectors         :  10887 = 35%

                SCAN STATUS
--------------------------------------
The scan process completed successfully

--------------------------------------
```

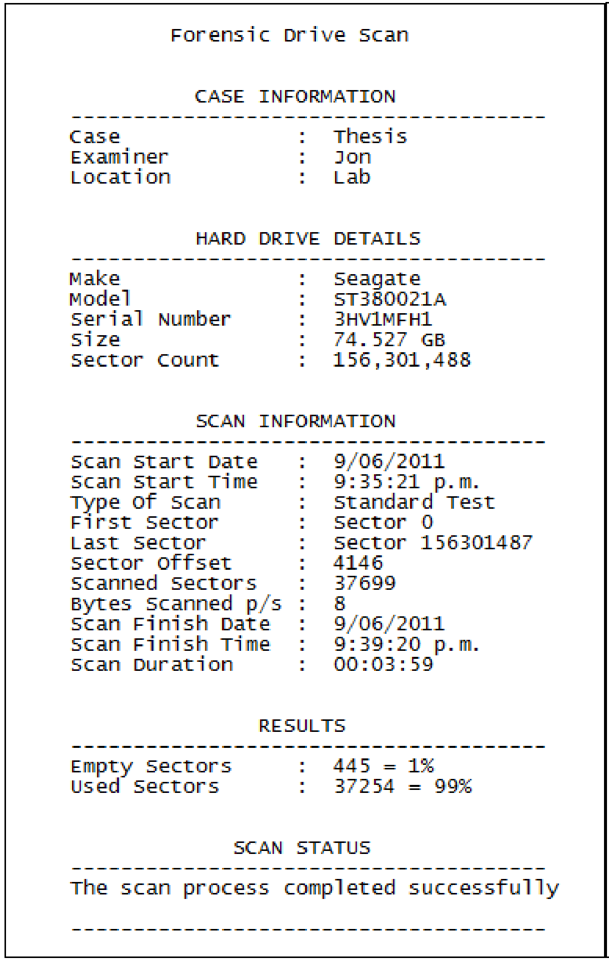


```
            Forensic Drive Scan

             CASE INFORMATION
--------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

            HARD DRIVE DETAILS
--------------------------------------
Make                 :  Samsung
Model                :  SAMSUNG SV4002H
Serial Number        :  0413J1FT109438
Size                 :  37.303 GB
Sector Count         :  78,242,976

             SCAN INFORMATION
--------------------------------------
Scan Start Date      :  10/06/2011
Scan Start Time      :  8:27:45 p.m.
Type Of Scan         :  Standard Test
First Sector         :  Sector 0
Last Sector          :  Sector 78242975
Sector Offset        :  4933
Scanned Sectors      :  15861
Bytes Scanned p/s    :  512
Scan Finish Date     :  10/06/2011
Scan Finish Time     :  8:29:46 p.m.
Scan Duration        :  00:02:01

                  RESULTS
--------------------------------------
Empty Sectors        :  10268 = 65%
Used Sectors         :  5592 = 35%

                SCAN STATUS
--------------------------------------
The scan process completed successfully

--------------------------------------
```

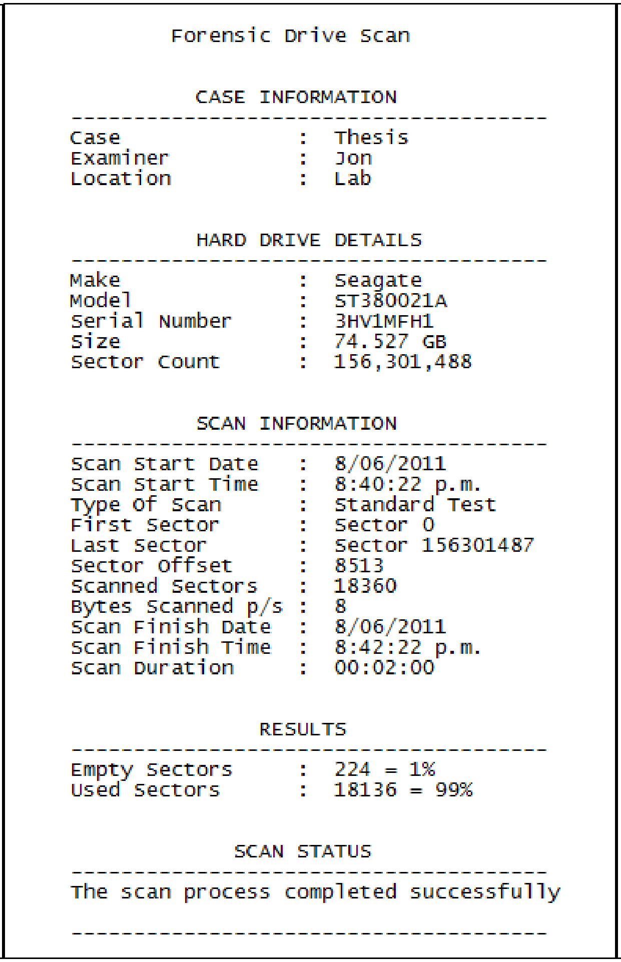


```
            Forensic Drive Scan

             CASE INFORMATION
--------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

            HARD DRIVE DETAILS
--------------------------------------
Make                 :  Samsung
Model                :  SAMSUNG SV4002H
Serial Number        :  0413J1FT109438
Size                 :  37.303 GB
Sector Count         :  78,242,976

             SCAN INFORMATION
--------------------------------------
Scan Start Date      :  12/06/2011
Scan Start Time      :  9:53:00 a.m.
Type Of Scan         :  Full Sector Scan
First Sector         :  Sector 0
Last Sector          :  Sector 78242975
Sector Offset        :  1
Scanned Sectors      :  78242976
Bytes Scanned p/s    :  512
Scan Finish Date     :  12/06/2011
Scan Finish Time     :  10:38:16 p.m.
Scan Duration        :  12:45:16

                  RESULTS
--------------------------------------
Empty Sectors        :  50661104 = 65%
Used Sectors         :  27581872 = 35%

                SCAN STATUS
--------------------------------------
The scan process completed successfully

--------------------------------------
```

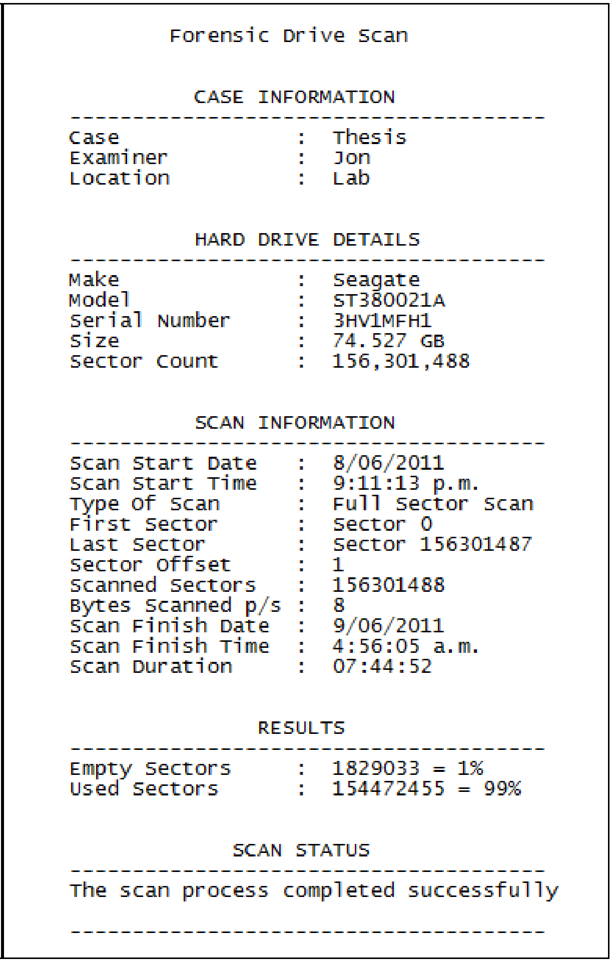
Eight Byte Test Reports for Test Subject 4



```
              Forensic Drive Scan

                CASE INFORMATION
-------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

               HARD DRIVE DETAILS
-------------------------------------
Make                 :  Seagate
Model                :  ST380021A
Serial Number        :  3HV1MFH1
Size                 :  74.527 GB
Sector Count         :  156,301,488

                SCAN INFORMATION
-------------------------------------
Scan Start Date      :  9/06/2011
Scan Start Time      :  9:35:21 p.m.
Type Of Scan         :  Standard Test
First Sector         :  Sector 0
Last Sector          :  Sector 156301487
Sector Offset        :  4146
Scanned Sectors      :  37699
Bytes Scanned p/s    :  8
Scan Finish Date     :  9/06/2011
Scan Finish Time     :  9:39:20 p.m.
Scan Duration        :  00:03:59

                     RESULTS
-------------------------------------
Empty Sectors        :  445 = 1%
Used Sectors         :  37254 = 99%

                   SCAN STATUS
-------------------------------------
The scan process completed successfully

-------------------------------------
```

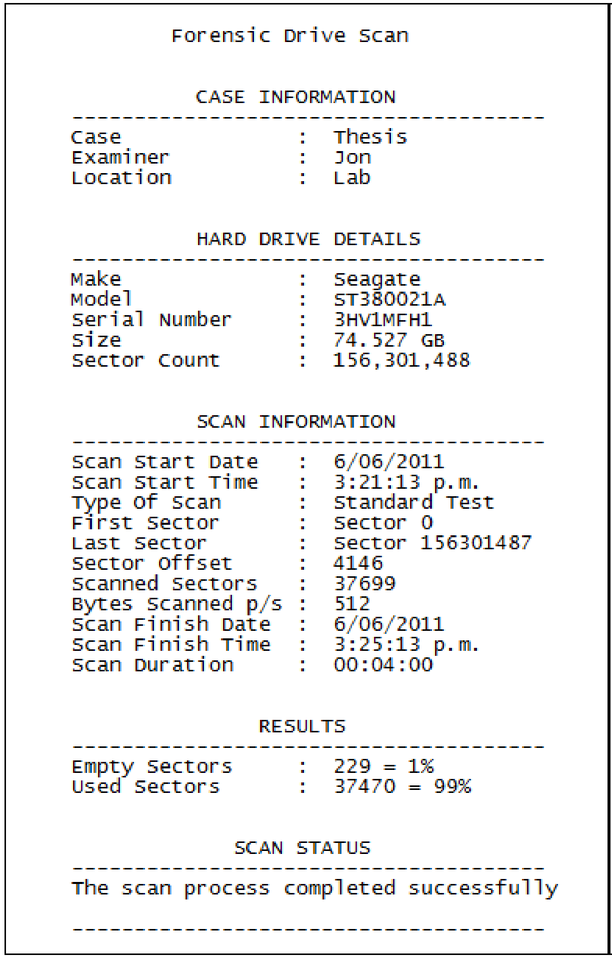


```
              Forensic Drive Scan

                CASE INFORMATION
-------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

               HARD DRIVE DETAILS
-------------------------------------
Make                 :  Seagate
Model                :  ST380021A
Serial Number        :  3HV1MFH1
Size                 :  74.527 GB
Sector Count         :  156,301,488

                SCAN INFORMATION
-------------------------------------
Scan Start Date      :  8/06/2011
Scan Start Time      :  8:40:22 p.m.
Type Of Scan         :  Standard Test
First Sector         :  Sector 0
Last Sector          :  Sector 156301487
Sector Offset        :  8513
Scanned Sectors      :  18360
Bytes Scanned p/s    :  8
Scan Finish Date     :  8/06/2011
Scan Finish Time     :  8:42:22 p.m.
Scan Duration        :  00:02:00

                     RESULTS
-------------------------------------
Empty Sectors        :  224 = 1%
Used Sectors         :  18136 = 99%

                   SCAN STATUS
-------------------------------------
The scan process completed successfully

-------------------------------------
```

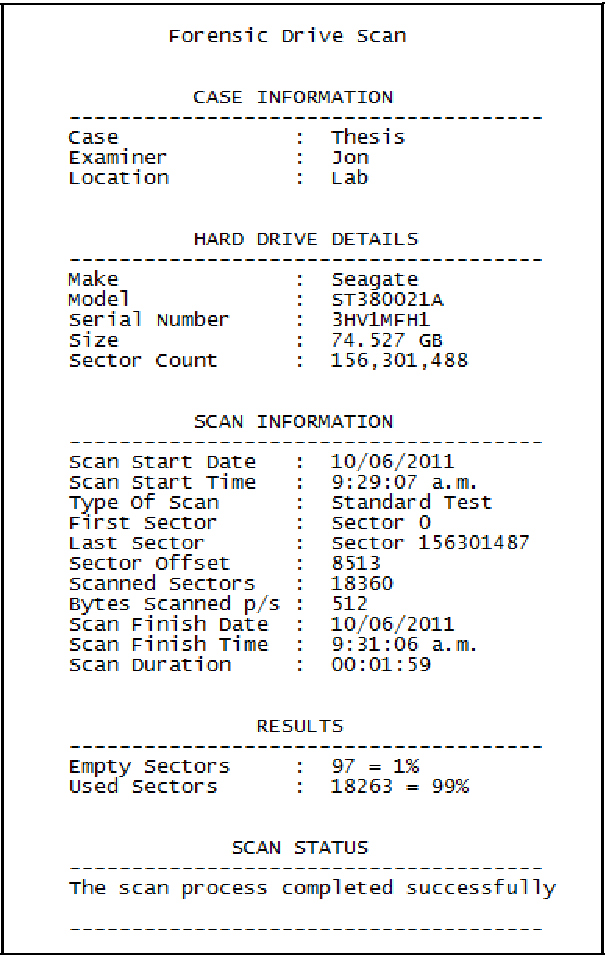


```
              Forensic Drive Scan

                CASE INFORMATION
-------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

               HARD DRIVE DETAILS
-------------------------------------
Make                 :  Seagate
Model                :  ST380021A
Serial Number        :  3HV1MFH1
Size                 :  74.527 GB
Sector Count         :  156,301,488

                SCAN INFORMATION
-------------------------------------
Scan Start Date      :  8/06/2011
Scan Start Time      :  9:11:13 p.m.
Type Of Scan         :  Full Sector Scan
First Sector         :  Sector 0
Last Sector          :  Sector 156301487
Sector Offset        :  1
Scanned Sectors      :  156301488
Bytes Scanned p/s    :  8
Scan Finish Date     :  9/06/2011
Scan Finish Time     :  4:56:05 a.m.
Scan Duration        :  07:44:52

                     RESULTS
-------------------------------------
Empty Sectors        :  1829033 = 1%
Used Sectors         :  154472455 = 99%

                   SCAN STATUS
-------------------------------------
The scan process completed successfully

-------------------------------------
```

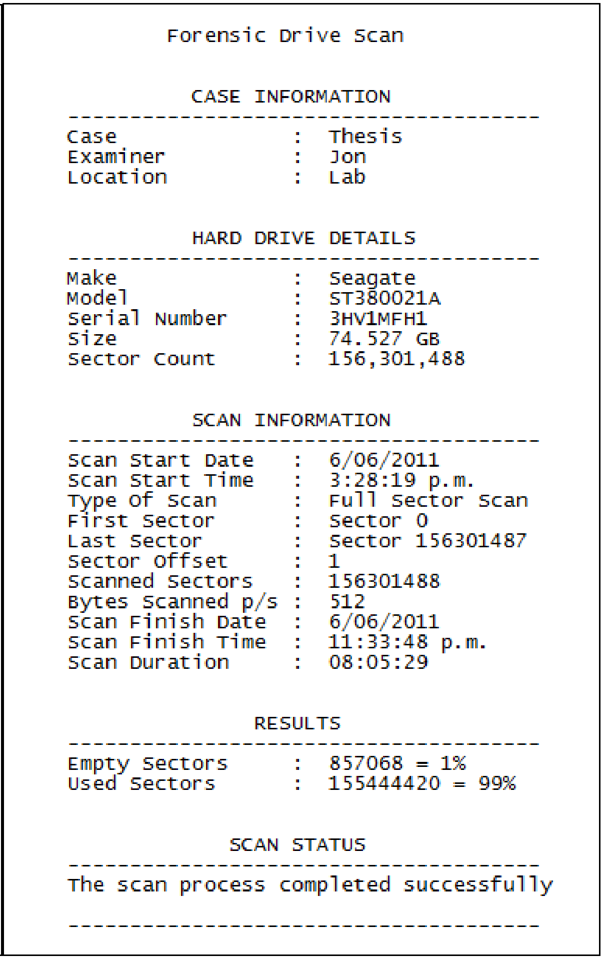
512 Byte Test Reports for Test Subject 4



```
            Forensic Drive Scan

             CASE INFORMATION
 --------------------------------------
 Case               :  Thesis
 Examiner           :  Jon
 Location           :  Lab

            HARD DRIVE DETAILS
 --------------------------------------
 Make               :  Seagate
 Model              :  ST380021A
 Serial Number      :  3HV1MFH1
 Size               :  74.527 GB
 Sector Count       :  156,301,488

             SCAN INFORMATION
 --------------------------------------
 Scan Start Date    :  6/06/2011
 Scan Start Time    :  3:21:13 p.m.
 Type Of Scan       :  Standard Test
 First Sector       :  Sector 0
 Last Sector        :  Sector 156301487
 Sector Offset      :  4146
 Scanned Sectors    :  37699
 Bytes Scanned p/s  :  512
 Scan Finish Date   :  6/06/2011
 Scan Finish Time   :  3:25:13 p.m.
 Scan Duration      :  00:04:00

                 RESULTS
 --------------------------------------
 Empty Sectors      :  229 = 1%
 Used Sectors       :  37470 = 99%

               SCAN STATUS
 --------------------------------------
 The scan process completed successfully

 --------------------------------------
```

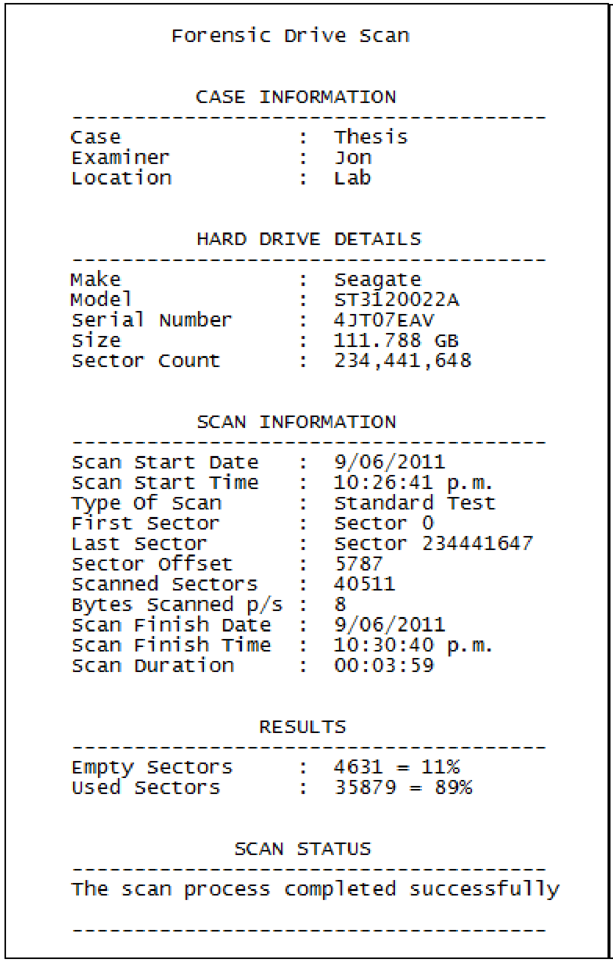


```
            Forensic Drive Scan

             CASE INFORMATION
 --------------------------------------
 Case               :  Thesis
 Examiner           :  Jon
 Location           :  Lab

            HARD DRIVE DETAILS
 --------------------------------------
 Make               :  Seagate
 Model              :  ST380021A
 Serial Number      :  3HV1MFH1
 Size               :  74.527 GB
 Sector Count       :  156,301,488

             SCAN INFORMATION
 --------------------------------------
 Scan Start Date    :  10/06/2011
 Scan Start Time    :  9:29:07 a.m.
 Type Of Scan       :  Standard Test
 First Sector       :  Sector 0
 Last Sector        :  Sector 156301487
 Sector Offset      :  8513
 Scanned Sectors    :  18360
 Bytes Scanned p/s  :  512
 Scan Finish Date   :  10/06/2011
 Scan Finish Time   :  9:31:06 a.m.
 Scan Duration      :  00:01:59

                 RESULTS
 --------------------------------------
 Empty Sectors      :  97 = 1%
 Used Sectors       :  18263 = 99%

               SCAN STATUS
 --------------------------------------
 The scan process completed successfully

 --------------------------------------
```

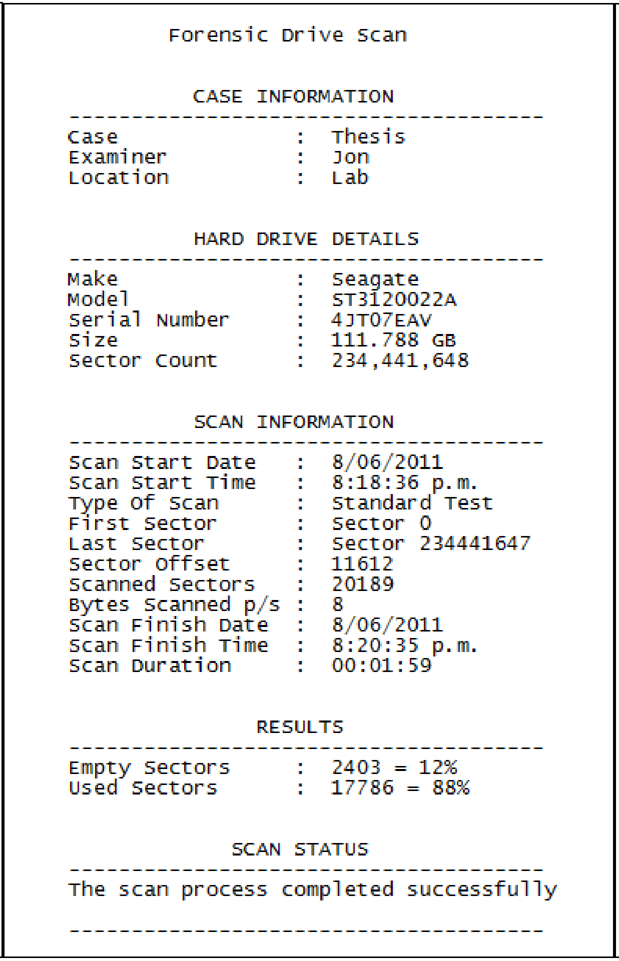


```
            Forensic Drive Scan

             CASE INFORMATION
 --------------------------------------
 Case               :  Thesis
 Examiner           :  Jon
 Location           :  Lab

            HARD DRIVE DETAILS
 --------------------------------------
 Make               :  Seagate
 Model              :  ST380021A
 Serial Number      :  3HV1MFH1
 Size               :  74.527 GB
 Sector Count       :  156,301,488

             SCAN INFORMATION
 --------------------------------------
 Scan Start Date    :  6/06/2011
 Scan Start Time    :  3:28:19 p.m.
 Type Of Scan       :  Full Sector Scan
 First Sector       :  Sector 0
 Last Sector        :  Sector 156301487
 Sector Offset      :  1
 Scanned Sectors    :  156301488
 Bytes Scanned p/s  :  512
 Scan Finish Date   :  6/06/2011
 Scan Finish Time   :  11:33:48 p.m.
 Scan Duration      :  08:05:29

                 RESULTS
 --------------------------------------
 Empty Sectors      :  857068 = 1%
 Used Sectors       :  155444420 = 99%

               SCAN STATUS
 --------------------------------------
 The scan process completed successfully

 --------------------------------------
```

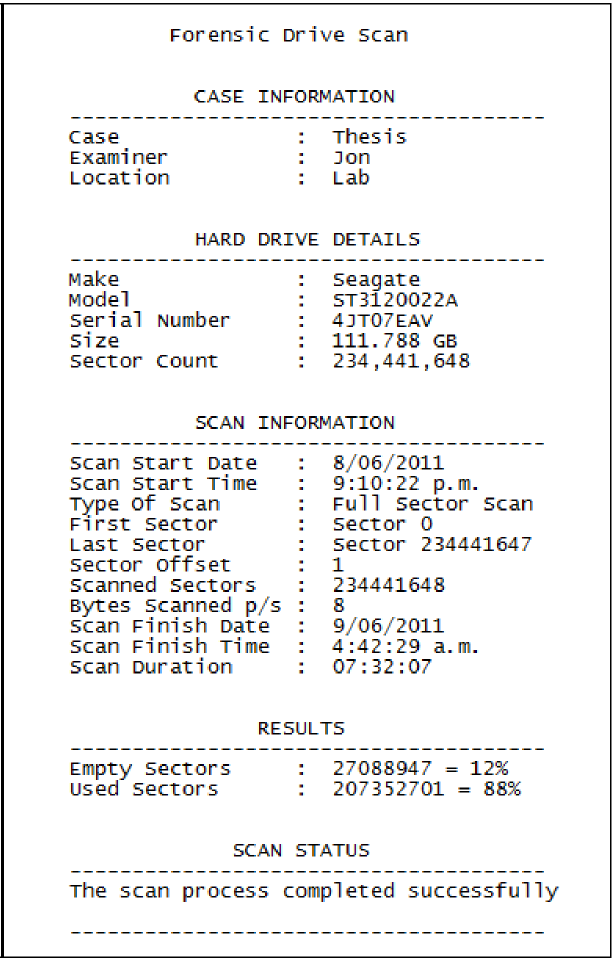
Eight Byte Test Reports for Test Subject 5



```
                Forensic Drive Scan

                 CASE INFORMATION
      --------------------------------------
      Case                :  Thesis
      Examiner            :  Jon
      Location            :  Lab

                 HARD DRIVE DETAILS
      --------------------------------------
      Make                :  Seagate
      Model               :  ST3120022A
      Serial Number       :  4JT07EAV
      Size                :  111.788 GB
      Sector Count        :  234,441,648

                 SCAN INFORMATION
      --------------------------------------
      Scan Start Date     :  9/06/2011
      Scan Start Time     :  10:26:41 p.m.
      Type Of Scan        :  Standard Test
      First Sector        :  Sector 0
      Last Sector         :  Sector 234441647
      Sector Offset       :  5787
      Scanned Sectors     :  40511
      Bytes Scanned p/s   :  8
      Scan Finish Date    :  9/06/2011
      Scan Finish Time    :  10:30:40 p.m.
      Scan Duration       :  00:03:59

                      RESULTS
      --------------------------------------
      Empty Sectors       :  4631 = 11%
      Used Sectors        :  35879 = 89%

                    SCAN STATUS
      --------------------------------------
      The scan process completed successfully

      --------------------------------------
```

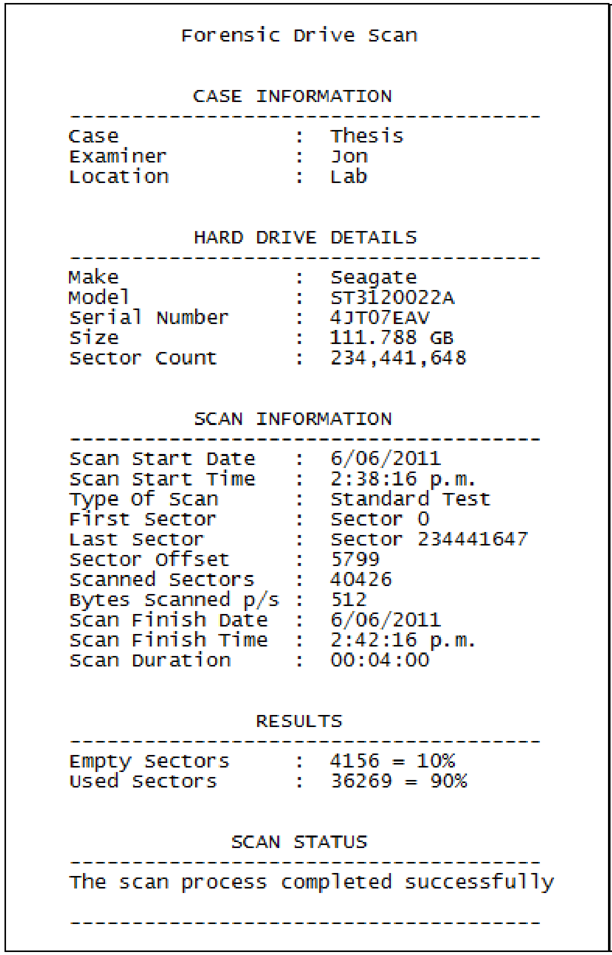


```
                Forensic Drive Scan

                 CASE INFORMATION
      --------------------------------------
      Case                :  Thesis
      Examiner            :  Jon
      Location            :  Lab

                 HARD DRIVE DETAILS
      --------------------------------------
      Make                :  Seagate
      Model               :  ST3120022A
      Serial Number       :  4JT07EAV
      Size                :  111.788 GB
      Sector Count        :  234,441,648

                 SCAN INFORMATION
      --------------------------------------
      Scan Start Date     :  8/06/2011
      Scan Start Time     :  8:18:36 p.m.
      Type Of Scan        :  Standard Test
      First Sector        :  Sector 0
      Last Sector         :  Sector 234441647
      Sector Offset       :  11612
      Scanned Sectors     :  20189
      Bytes Scanned p/s   :  8
      Scan Finish Date    :  8/06/2011
      Scan Finish Time    :  8:20:35 p.m.
      Scan Duration       :  00:01:59

                      RESULTS
      --------------------------------------
      Empty Sectors       :  2403 = 12%
      Used Sectors        :  17786 = 88%

                    SCAN STATUS
      --------------------------------------
      The scan process completed successfully

      --------------------------------------
```

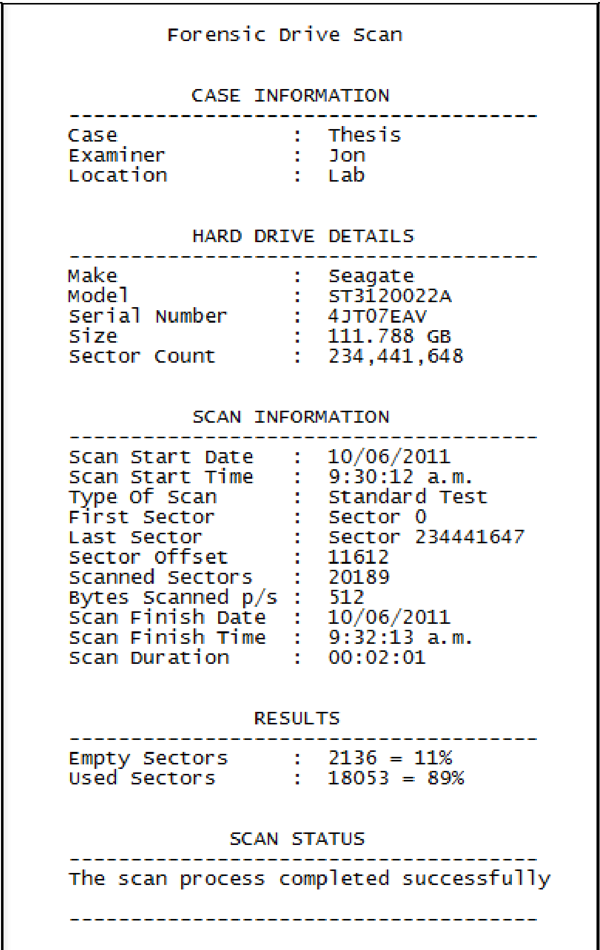


```
                Forensic Drive Scan

                 CASE INFORMATION
      --------------------------------------
      Case                :  Thesis
      Examiner            :  Jon
      Location            :  Lab

                 HARD DRIVE DETAILS
      --------------------------------------
      Make                :  Seagate
      Model               :  ST3120022A
      Serial Number       :  4JT07EAV
      Size                :  111.788 GB
      Sector Count        :  234,441,648

                 SCAN INFORMATION
      --------------------------------------
      Scan Start Date     :  8/06/2011
      Scan Start Time     :  9:10:22 p.m.
      Type Of Scan        :  Full Sector Scan
      First Sector        :  Sector 0
      Last Sector         :  Sector 234441647
      Sector Offset       :  1
      Scanned Sectors     :  234441648
      Bytes Scanned p/s   :  8
      Scan Finish Date    :  9/06/2011
      Scan Finish Time    :  4:42:29 a.m.
      Scan Duration       :  07:32:07

                      RESULTS
      --------------------------------------
      Empty Sectors       :  27088947 = 12%
      Used Sectors        :  207352701 = 88%

                    SCAN STATUS
      --------------------------------------
      The scan process completed successfully

      --------------------------------------
```

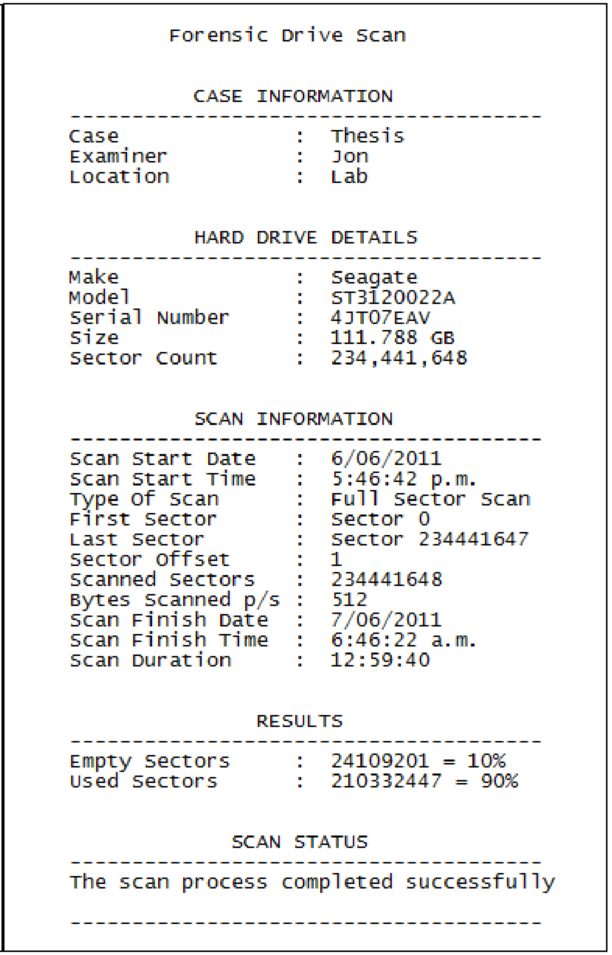
512 Byte Test Reports for Test Subject 5



```
              Forensic Drive Scan

               CASE INFORMATION
      --------------------------------------
      Case                 :  Thesis
      Examiner             :  Jon
      Location             :  Lab

              HARD DRIVE DETAILS
      --------------------------------------
      Make                 :  Seagate
      Model                :  ST3120022A
      Serial Number        :  4JT07EAV
      Size                 :  111.788 GB
      Sector Count         :  234,441,648

               SCAN INFORMATION
      --------------------------------------
      Scan Start Date      :  6/06/2011
      Scan Start Time      :  2:38:16 p.m.
      Type Of Scan         :  Standard Test
      First Sector         :  Sector 0
      Last Sector          :  Sector 234441647
      Sector Offset        :  5799
      Scanned Sectors      :  40426
      Bytes Scanned p/s    :  512
      Scan Finish Date     :  6/06/2011
      Scan Finish Time     :  2:42:16 p.m.
      Scan Duration        :  00:04:00

                    RESULTS
      --------------------------------------
      Empty Sectors        :  4156 = 10%
      Used Sectors         :  36269 = 90%

                  SCAN STATUS
      --------------------------------------
      The scan process completed successfully

      --------------------------------------
```

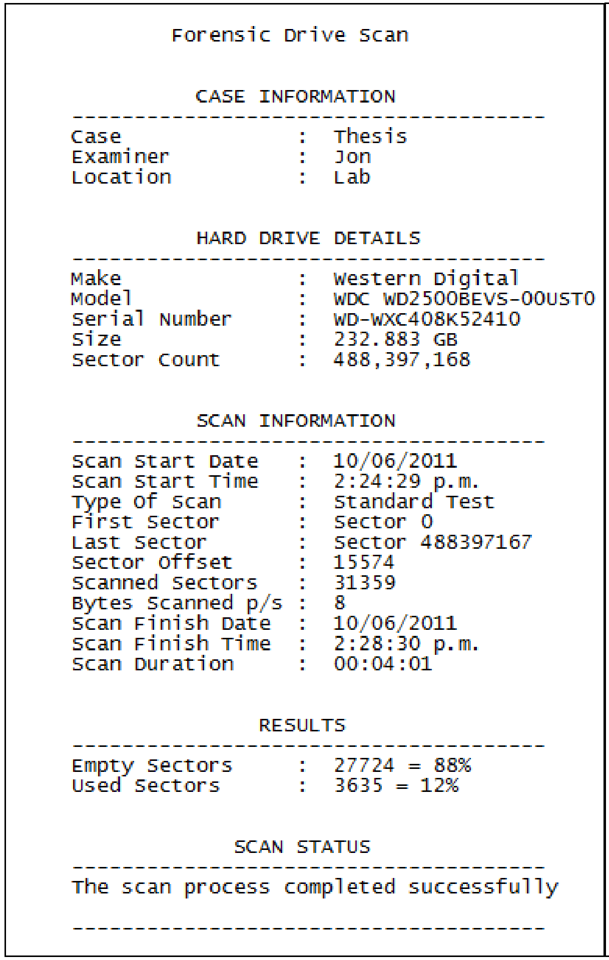


```
              Forensic Drive Scan

               CASE INFORMATION
      --------------------------------------
      Case                 :  Thesis
      Examiner             :  Jon
      Location             :  Lab

              HARD DRIVE DETAILS
      --------------------------------------
      Make                 :  Seagate
      Model                :  ST3120022A
      Serial Number        :  4JT07EAV
      Size                 :  111.788 GB
      Sector Count         :  234,441,648

               SCAN INFORMATION
      --------------------------------------
      Scan Start Date      :  10/06/2011
      Scan Start Time      :  9:30:12 a.m.
      Type Of Scan         :  Standard Test
      First Sector         :  Sector 0
      Last Sector          :  Sector 234441647
      Sector Offset        :  11612
      Scanned Sectors      :  20189
      Bytes Scanned p/s    :  512
      Scan Finish Date     :  10/06/2011
      Scan Finish Time     :  9:32:13 a.m.
      Scan Duration        :  00:02:01

                    RESULTS
      --------------------------------------
      Empty Sectors        :  2136 = 11%
      Used Sectors         :  18053 = 89%

                  SCAN STATUS
      --------------------------------------
      The scan process completed successfully

      --------------------------------------
```

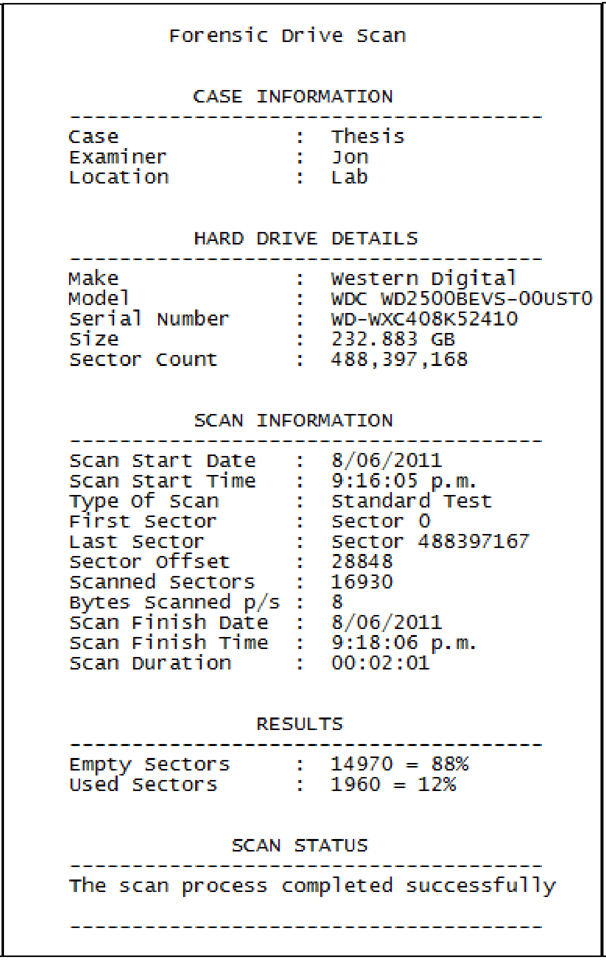


```
              Forensic Drive Scan

               CASE INFORMATION
      --------------------------------------
      Case                 :  Thesis
      Examiner             :  Jon
      Location             :  Lab

              HARD DRIVE DETAILS
      --------------------------------------
      Make                 :  Seagate
      Model                :  ST3120022A
      Serial Number        :  4JT07EAV
      Size                 :  111.788 GB
      Sector Count         :  234,441,648

               SCAN INFORMATION
      --------------------------------------
      Scan Start Date      :  6/06/2011
      Scan Start Time      :  5:46:42 p.m.
      Type Of Scan         :  Full Sector Scan
      First Sector         :  Sector 0
      Last Sector          :  Sector 234441647
      Sector Offset        :  1
      Scanned Sectors      :  234441648
      Bytes Scanned p/s    :  512
      Scan Finish Date     :  7/06/2011
      Scan Finish Time     :  6:46:22 a.m.
      Scan Duration        :  12:59:40

                    RESULTS
      --------------------------------------
      Empty Sectors        :  24109201 = 10%
      Used Sectors         :  210332447 = 90%

                  SCAN STATUS
      --------------------------------------
      The scan process completed successfully

      --------------------------------------
```

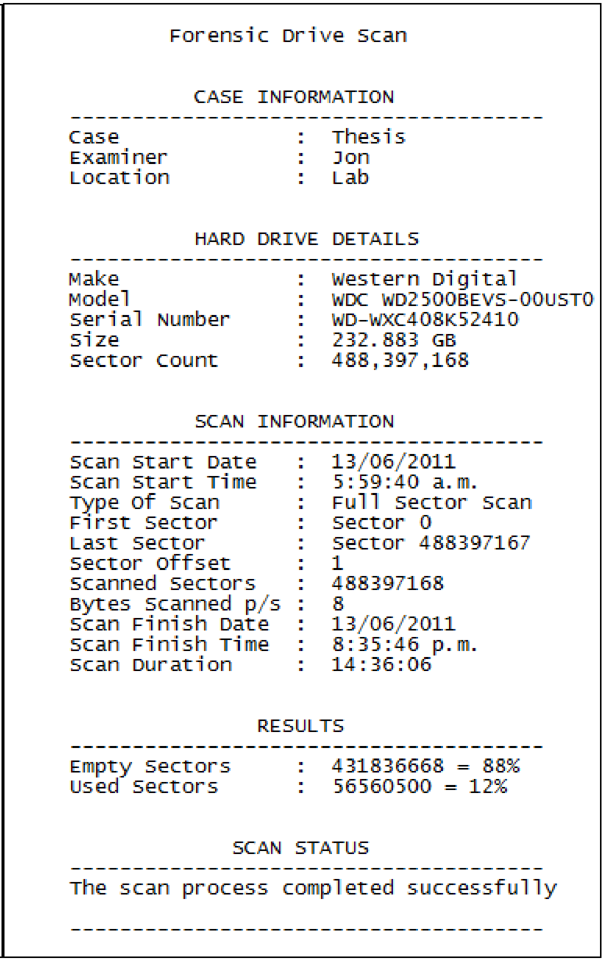
Eight Byte Test Reports for Test Subject 6



```
            Forensic Drive Scan

             CASE INFORMATION
--------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

            HARD DRIVE DETAILS
--------------------------------------
Make                 :  Western Digital
Model                :  WDC WD2500BEVS-00UST0
Serial Number        :  WD-WXC408K52410
Size                 :  232.883 GB
Sector Count         :  488,397,168

             SCAN INFORMATION
--------------------------------------
Scan Start Date      :  10/06/2011
Scan Start Time      :  2:24:29 p.m.
Type Of Scan         :  Standard Test
First Sector         :  Sector 0
Last Sector          :  Sector 488397167
Sector Offset        :  15574
Scanned Sectors      :  31359
Bytes Scanned p/s    :  8
Scan Finish Date     :  10/06/2011
Scan Finish Time     :  2:28:30 p.m.
Scan Duration        :  00:04:01

                  RESULTS
--------------------------------------
Empty Sectors        :  27724 = 88%
Used Sectors         :  3635 = 12%

                SCAN STATUS
--------------------------------------
The scan process completed successfully

--------------------------------------
```

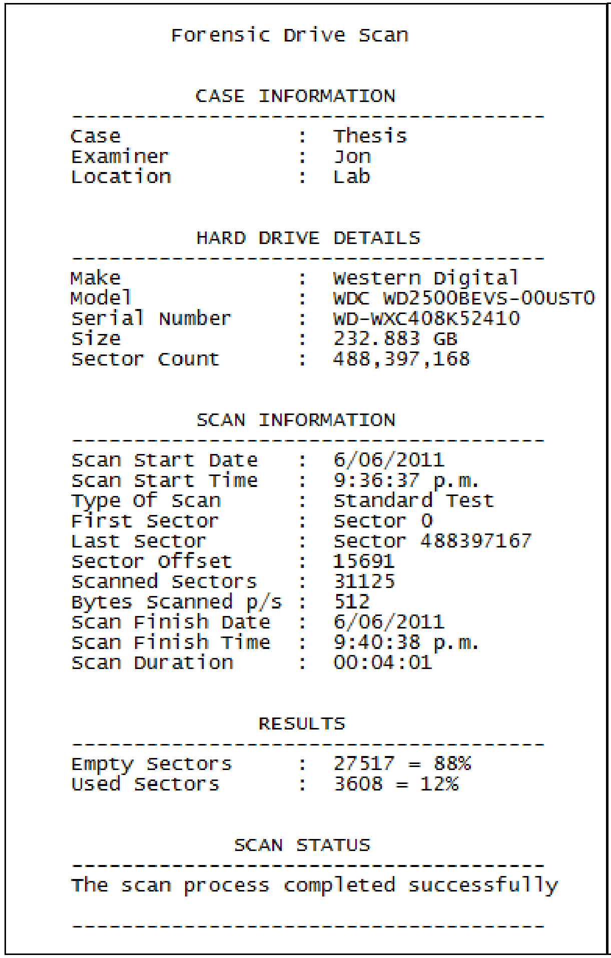


```
            Forensic Drive Scan

             CASE INFORMATION
--------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

            HARD DRIVE DETAILS
--------------------------------------
Make                 :  Western Digital
Model                :  WDC WD2500BEVS-00UST0
Serial Number        :  WD-WXC408K52410
Size                 :  232.883 GB
Sector Count         :  488,397,168

             SCAN INFORMATION
--------------------------------------
Scan Start Date      :  8/06/2011
Scan Start Time      :  9:16:05 p.m.
Type Of Scan         :  Standard Test
First Sector         :  Sector 0
Last Sector          :  Sector 488397167
Sector Offset        :  28848
Scanned Sectors      :  16930
Bytes Scanned p/s    :  8
Scan Finish Date     :  8/06/2011
Scan Finish Time     :  9:18:06 p.m.
Scan Duration        :  00:02:01

                  RESULTS
--------------------------------------
Empty Sectors        :  14970 = 88%
Used Sectors         :  1960 = 12%

                SCAN STATUS
--------------------------------------
The scan process completed successfully

--------------------------------------
```

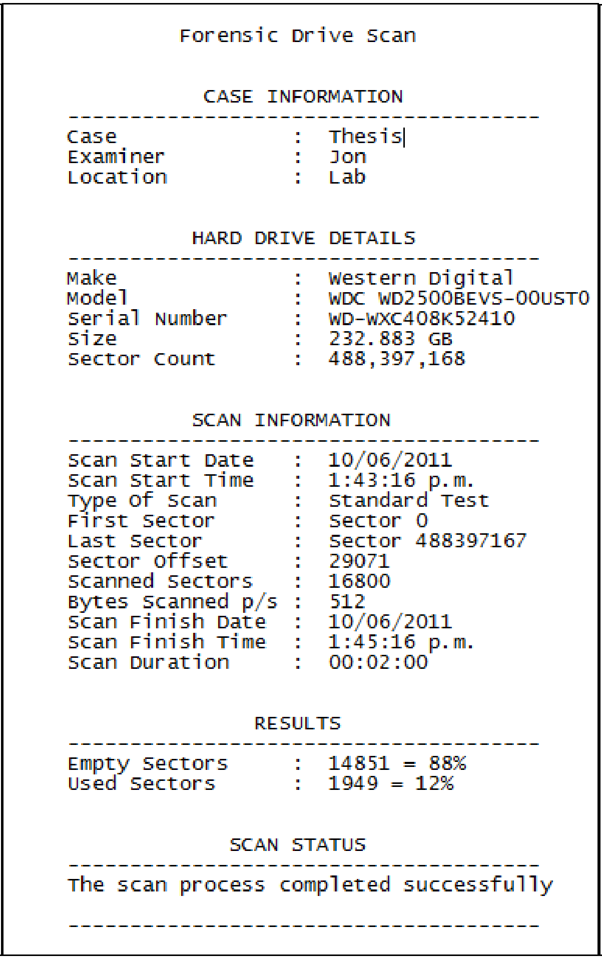


```
            Forensic Drive Scan

             CASE INFORMATION
--------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

            HARD DRIVE DETAILS
--------------------------------------
Make                 :  Western Digital
Model                :  WDC WD2500BEVS-00UST0
Serial Number        :  WD-WXC408K52410
Size                 :  232.883 GB
Sector Count         :  488,397,168

             SCAN INFORMATION
--------------------------------------
Scan Start Date      :  13/06/2011
Scan Start Time      :  5:59:40 a.m.
Type Of Scan         :  Full Sector Scan
First Sector         :  Sector 0
Last Sector          :  Sector 488397167
Sector Offset        :  1
Scanned Sectors      :  488397168
Bytes Scanned p/s    :  8
Scan Finish Date     :  13/06/2011
Scan Finish Time     :  8:35:46 p.m.
Scan Duration        :  14:36:06

                  RESULTS
--------------------------------------
Empty Sectors        :  431836668 = 88%
Used Sectors         :  56560500 = 12%

                SCAN STATUS
--------------------------------------
The scan process completed successfully

--------------------------------------
```

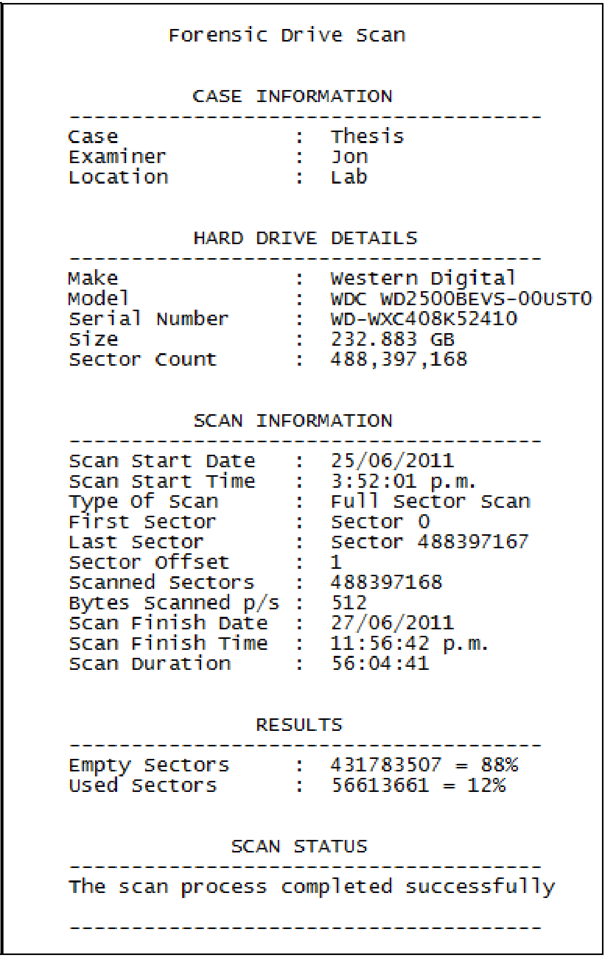
512 Byte Test Reports for Test Subject 6



```
            Forensic Drive Scan

             CASE INFORMATION
     -------------------------------------
     Case                :  Thesis
     Examiner            :  Jon
     Location            :  Lab

             HARD DRIVE DETAILS
     -------------------------------------
     Make                :  Western Digital
     Model               :  WDC WD2500BEVS-00UST0
     Serial Number       :  WD-WXC408K52410
     Size                :  232.883 GB
     Sector Count        :  488,397,168

             SCAN INFORMATION
     -------------------------------------
     Scan Start Date     :  6/06/2011
     Scan Start Time     :  9:36:37 p.m.
     Type Of Scan        :  Standard Test
     First Sector        :  Sector 0
     Last Sector         :  Sector 488397167
     Sector Offset       :  15691
     Scanned Sectors     :  31125
     Bytes Scanned p/s   :  512
     Scan Finish Date    :  6/06/2011
     Scan Finish Time    :  9:40:38 p.m.
     Scan Duration       :  00:04:01

                  RESULTS
     -------------------------------------
     Empty Sectors       :  27517 = 88%
     Used Sectors        :  3608 = 12%

                SCAN STATUS
     -------------------------------------
     The scan process completed successfully

     -------------------------------------
```

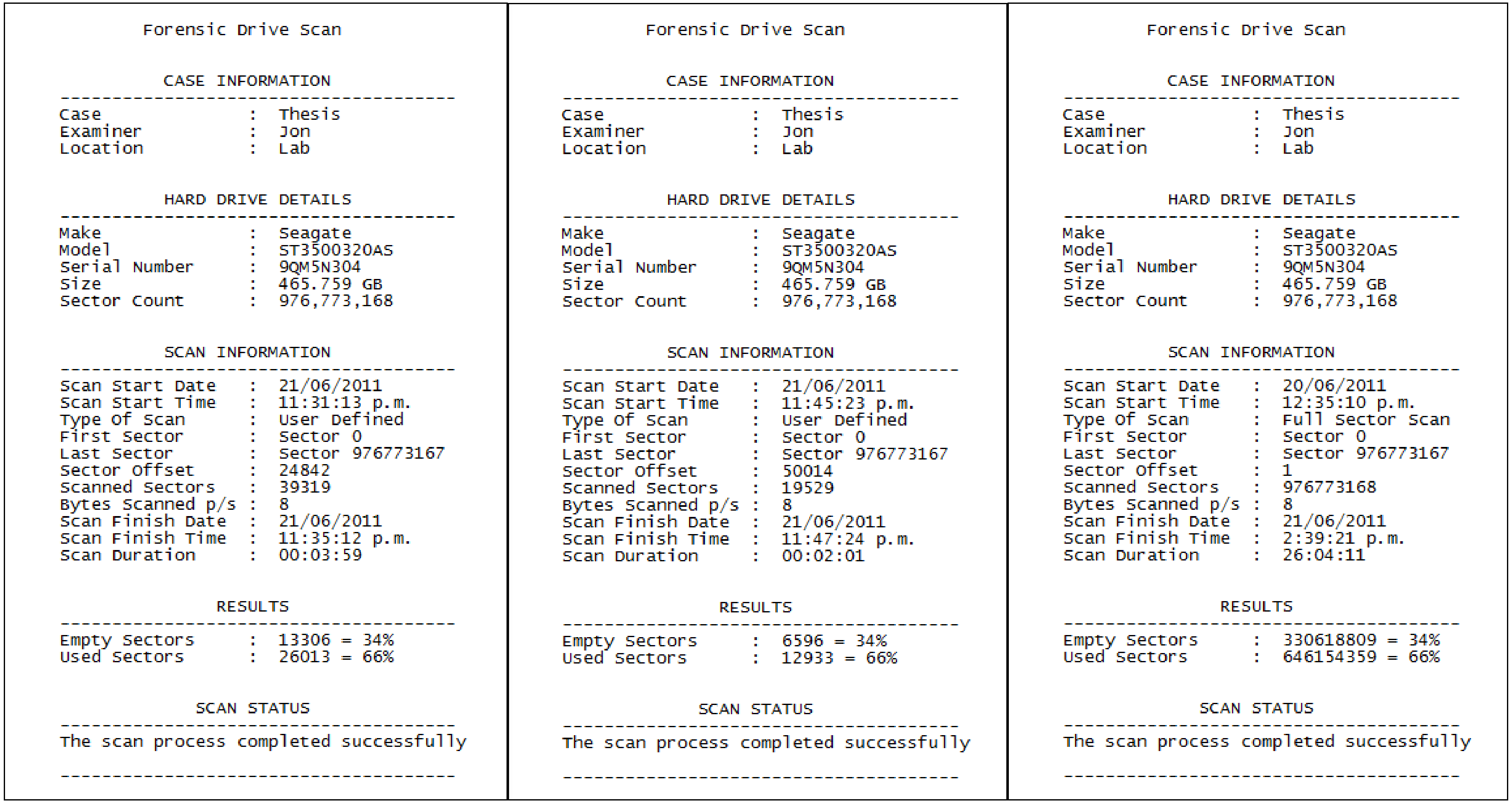


```
            Forensic Drive Scan

             CASE INFORMATION
     -------------------------------------
     Case                :  Thesis
     Examiner            :  Jon
     Location            :  Lab

             HARD DRIVE DETAILS
     -------------------------------------
     Make                :  Western Digital
     Model               :  WDC WD2500BEVS-00UST0
     Serial Number       :  WD-WXC408K52410
     Size                :  232.883 GB
     Sector Count        :  488,397,168

             SCAN INFORMATION
     -------------------------------------
     Scan Start Date     :  10/06/2011
     Scan Start Time     :  1:43:16 p.m.
     Type Of Scan        :  Standard Test
     First Sector        :  Sector 0
     Last Sector         :  Sector 488397167
     Sector Offset       :  29071
     Scanned Sectors     :  16800
     Bytes Scanned p/s   :  512
     Scan Finish Date    :  10/06/2011
     Scan Finish Time    :  1:45:16 p.m.
     Scan Duration       :  00:02:00

                  RESULTS
     -------------------------------------
     Empty Sectors       :  14851 = 88%
     Used Sectors        :  1949 = 12%

                SCAN STATUS
     -------------------------------------
     The scan process completed successfully

     -------------------------------------
```



```
            Forensic Drive Scan

             CASE INFORMATION
     -------------------------------------
     Case                :  Thesis
     Examiner            :  Jon
     Location            :  Lab

             HARD DRIVE DETAILS
     -------------------------------------
     Make                :  Western Digital
     Model               :  WDC WD2500BEVS-00UST0
     Serial Number       :  WD-WXC408K52410
     Size                :  232.883 GB
     Sector Count        :  488,397,168

             SCAN INFORMATION
     -------------------------------------
     Scan Start Date     :  25/06/2011
     Scan Start Time     :  3:52:01 p.m.
     Type Of Scan        :  Full Sector Scan
     First Sector        :  Sector 0
     Last Sector         :  Sector 488397167
     Sector Offset       :  1
     Scanned Sectors     :  488397168
     Bytes Scanned p/s   :  512
     Scan Finish Date    :  27/06/2011
     Scan Finish Time    :  11:56:42 p.m.
     Scan Duration       :  56:04:41

                  RESULTS
     -------------------------------------
     Empty Sectors       :  431783507 = 88%
     Used Sectors        :  56613661 = 12%

                SCAN STATUS
     -------------------------------------
     The scan process completed successfully

     -------------------------------------
```

Eight Byte Test Reports for Test Subject 7

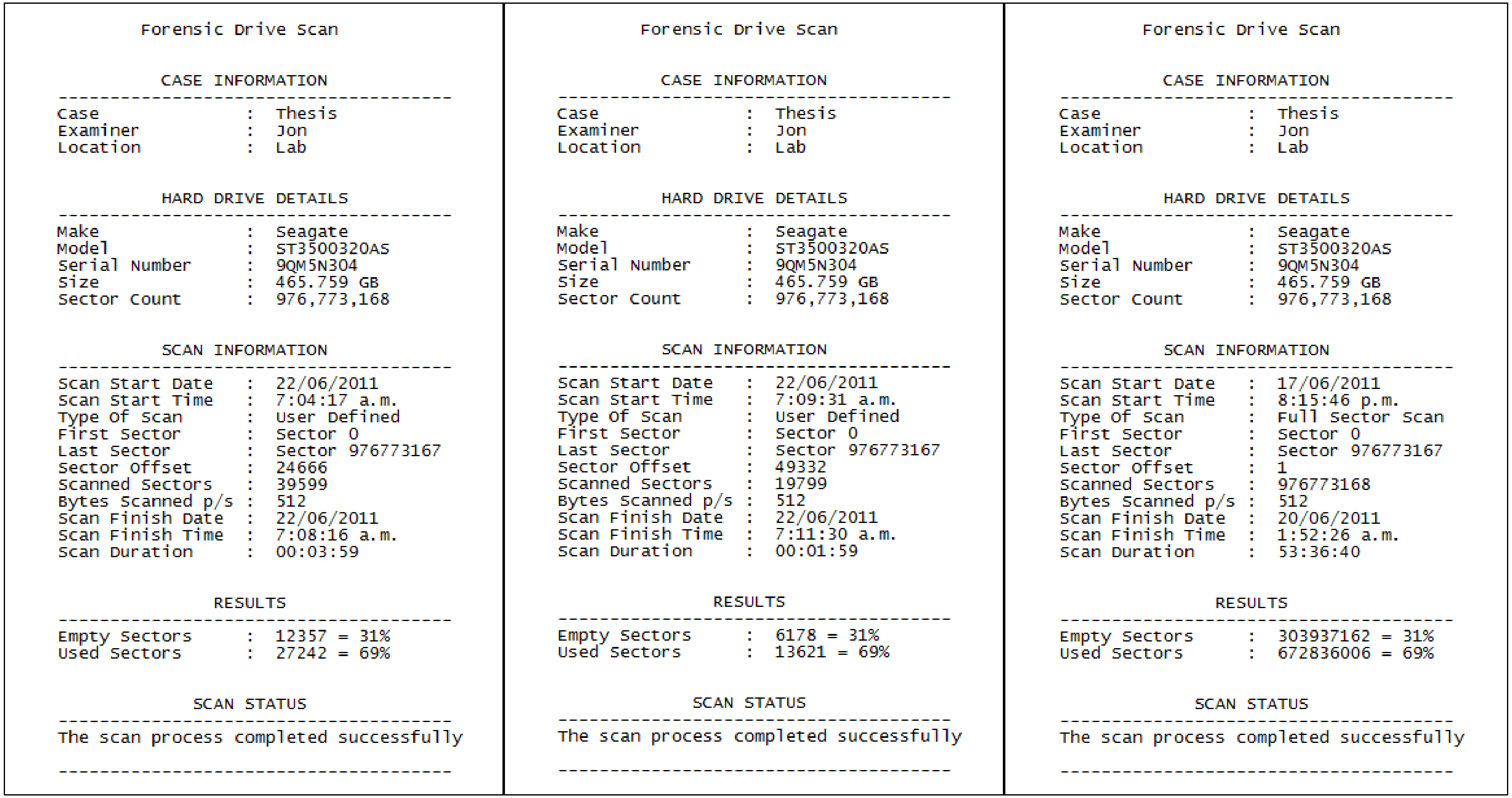


```
            Forensic Drive Scan

             CASE INFORMATION
 ----------------------------------------
 Case                 :  Thesis
 Examiner             :  Jon
 Location             :  Lab

            HARD DRIVE DETAILS
 ----------------------------------------
 Make                 :  Seagate
 Model                :  ST3500320AS
 Serial Number        :  9QM5N304
 Size                 :  465.759 GB
 Sector Count         :  976,773,168

             SCAN INFORMATION
 ----------------------------------------
 Scan Start Date      :  21/06/2011
 Scan Start Time      :  11:31:13 p.m.
 Type Of Scan         :  User Defined
 First Sector         :  Sector 0
 Last Sector          :  Sector 976773167
 Sector Offset        :  24842
 Scanned Sectors      :  39319
 Bytes Scanned p/s    :  8
 Scan Finish Date     :  21/06/2011
 Scan Finish Time     :  11:35:12 p.m.
 Scan Duration        :  00:03:59

                  RESULTS
 ----------------------------------------
 Empty Sectors        :  13306 = 34%
 Used Sectors         :  26013 = 66%

                SCAN STATUS
 ----------------------------------------
 The scan process completed successfully

 ----------------------------------------
```

```
            Forensic Drive Scan

             CASE INFORMATION
 ----------------------------------------
 Case                 :  Thesis
 Examiner             :  Jon
 Location             :  Lab

            HARD DRIVE DETAILS
 ----------------------------------------
 Make                 :  Seagate
 Model                :  ST3500320AS
 Serial Number        :  9QM5N304
 Size                 :  465.759 GB
 Sector Count         :  976,773,168

             SCAN INFORMATION
 ----------------------------------------
 Scan Start Date      :  21/06/2011
 Scan Start Time      :  11:45:23 p.m.
 Type Of Scan         :  User Defined
 First Sector         :  Sector 0
 Last Sector          :  Sector 976773167
 Sector Offset        :  50014
 Scanned Sectors      :  19529
 Bytes Scanned p/s    :  8
 Scan Finish Date     :  21/06/2011
 Scan Finish Time     :  11:47:24 p.m.
 Scan Duration        :  00:02:01

                  RESULTS
 ----------------------------------------
 Empty Sectors        :  6596 = 34%
 Used Sectors         :  12933 = 66%

                SCAN STATUS
 ----------------------------------------
 The scan process completed successfully

 ----------------------------------------
```

```
            Forensic Drive Scan

             CASE INFORMATION
 ----------------------------------------
 Case                 :  Thesis
 Examiner             :  Jon
 Location             :  Lab

            HARD DRIVE DETAILS
 ----------------------------------------
 Make                 :  Seagate
 Model                :  ST3500320AS
 Serial Number        :  9QM5N304
 Size                 :  465.759 GB
 Sector Count         :  976,773,168

             SCAN INFORMATION
 ----------------------------------------
 Scan Start Date      :  20/06/2011
 Scan Start Time      :  12:35:10 p.m.
 Type Of Scan         :  Full Sector Scan
 First Sector         :  Sector 0
 Last Sector          :  Sector 976773167
 Sector Offset        :  1
 Scanned Sectors      :  976773168
 Bytes Scanned p/s    :  8
 Scan Finish Date     :  21/06/2011
 Scan Finish Time     :  2:39:21 p.m.
 Scan Duration        :  26:04:11

                  RESULTS
 ----------------------------------------
 Empty Sectors        :  330618809 = 34%
 Used Sectors         :  646154359 = 66%

                SCAN STATUS
 ----------------------------------------
 The scan process completed successfully

 ----------------------------------------
```

512 Byte Test Reports for Test Subject 7

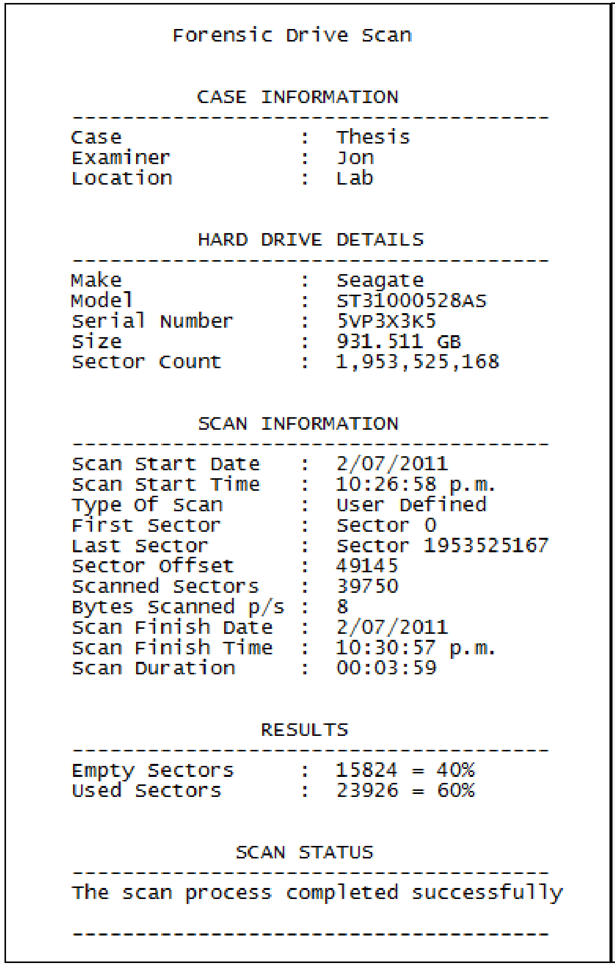


```
           Forensic Drive Scan

            CASE INFORMATION
---------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

           HARD DRIVE DETAILS
---------------------------------------
Make                 :  Seagate
Model                :  ST3500320AS
Serial Number        :  9QM5N304
Size                 :  465.759 GB
Sector Count         :  976,773,168

            SCAN INFORMATION
---------------------------------------
Scan Start Date      :  22/06/2011
Scan Start Time      :  7:04:17 a.m.
Type of Scan         :  User Defined
First Sector         :  Sector 0
Last Sector          :  Sector 976773167
Sector Offset        :  24666
Scanned Sectors      :  39599
Bytes Scanned p/s    :  512
Scan Finish Date     :  22/06/2011
Scan Finish Time     :  7:08:16 a.m.
Scan Duration        :  00:03:59

                 RESULTS
---------------------------------------
Empty Sectors        :  12357 = 31%
Used Sectors         :  27242 = 69%

               SCAN STATUS
---------------------------------------
The scan process completed successfully

---------------------------------------
```

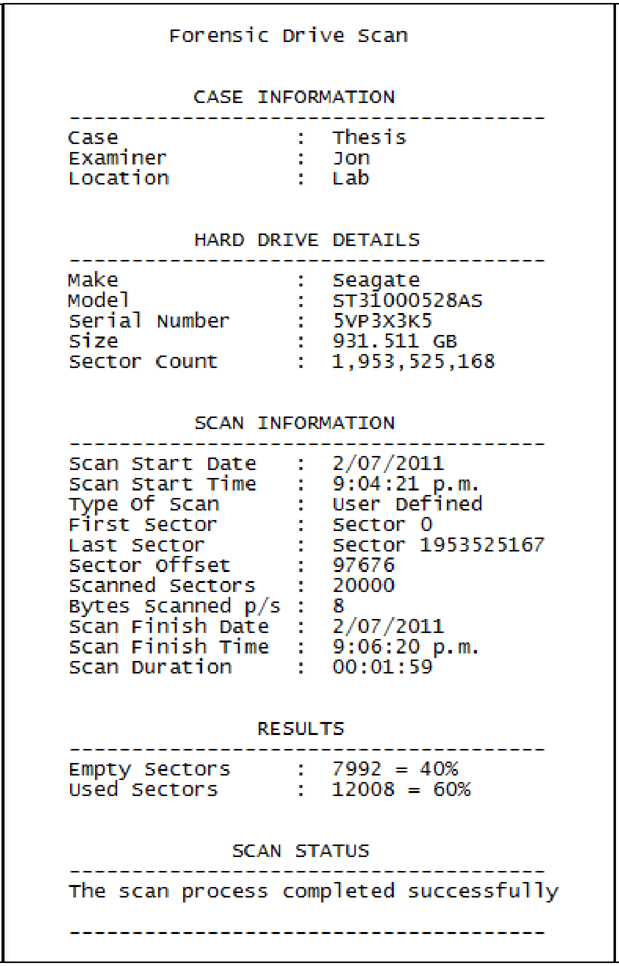
```
           Forensic Drive Scan

            CASE INFORMATION
---------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

           HARD DRIVE DETAILS
---------------------------------------
Make                 :  Seagate
Model                :  ST3500320AS
Serial Number        :  9QM5N304
Size                 :  465.759 GB
Sector Count         :  976,773,168

            SCAN INFORMATION
---------------------------------------
Scan Start Date      :  22/06/2011
Scan Start Time      :  7:09:31 a.m.
Type of Scan         :  User Defined
First Sector         :  Sector 0
Last Sector          :  Sector 976773167
Sector Offset        :  49332
Scanned Sectors      :  19799
Bytes Scanned p/s    :  512
Scan Finish Date     :  22/06/2011
Scan Finish Time     :  7:11:30 a.m.
Scan Duration        :  00:01:59

                 RESULTS
---------------------------------------
Empty Sectors        :  6178 = 31%
Used Sectors         :  13621 = 69%

               SCAN STATUS
---------------------------------------
The scan process completed successfully

---------------------------------------
```

```
           Forensic Drive Scan

            CASE INFORMATION
---------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

           HARD DRIVE DETAILS
---------------------------------------
Make                 :  Seagate
Model                :  ST3500320AS
Serial Number        :  9QM5N304
Size                 :  465.759 GB
Sector Count         :  976,773,168

            SCAN INFORMATION
---------------------------------------
Scan Start Date      :  17/06/2011
Scan Start Time      :  8:15:46 p.m.
Type of Scan         :  Full Sector Scan
First Sector         :  Sector 0
Last Sector          :  Sector 976773167
Sector Offset        :  1
Scanned Sectors      :  976773168
Bytes Scanned p/s    :  512
Scan Finish Date     :  20/06/2011
Scan Finish Time     :  1:52:26 a.m.
Scan Duration        :  53:36:40

                 RESULTS
---------------------------------------
Empty Sectors        :  303937162 = 31%
Used Sectors         :  672836006 = 69%

               SCAN STATUS
---------------------------------------
The scan process completed successfully

---------------------------------------
```

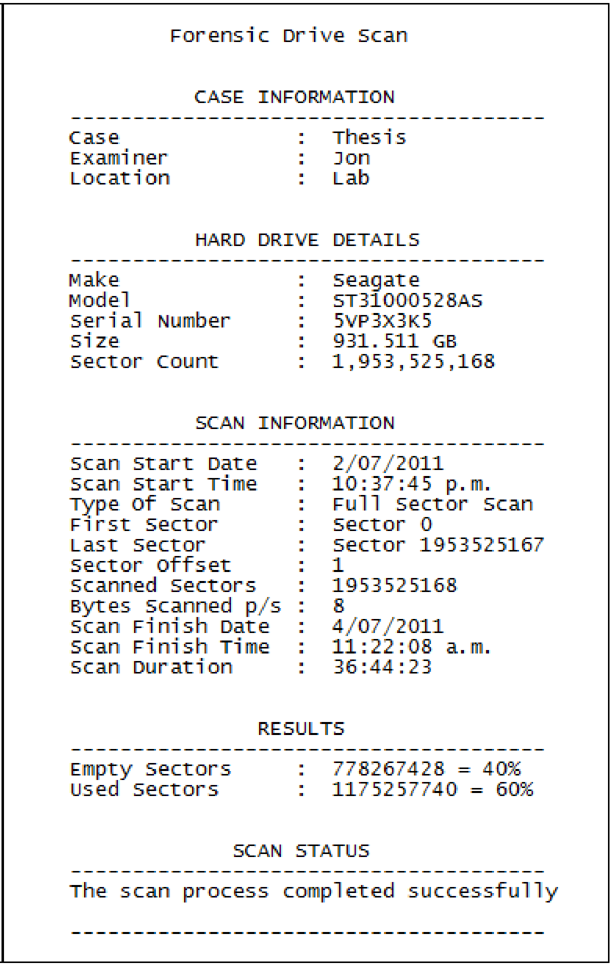
Eight Byte Test Reports for Test Subject 8



```
          Forensic Drive Scan

            CASE INFORMATION
----------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

           HARD DRIVE DETAILS
----------------------------------------
Make                 :  Seagate
Model                :  ST31000528AS
Serial Number        :  5VP3X3K5
Size                 :  931.511 GB
Sector Count         :  1,953,525,168

            SCAN INFORMATION
----------------------------------------
Scan Start Date      :  2/07/2011
Scan Start Time      :  10:26:58 p.m.
Type Of Scan         :  User Defined
First Sector         :  Sector 0
Last Sector          :  Sector 1953525167
Sector Offset        :  49145
Scanned Sectors      :  39750
Bytes Scanned p/s    :  8
Scan Finish Date     :  2/07/2011
Scan Finish Time     :  10:30:57 p.m.
Scan Duration        :  00:03:59

                 RESULTS
----------------------------------------
Empty Sectors        :  15824 = 40%
Used Sectors         :  23926 = 60%

               SCAN STATUS
----------------------------------------
The scan process completed successfully

----------------------------------------
```

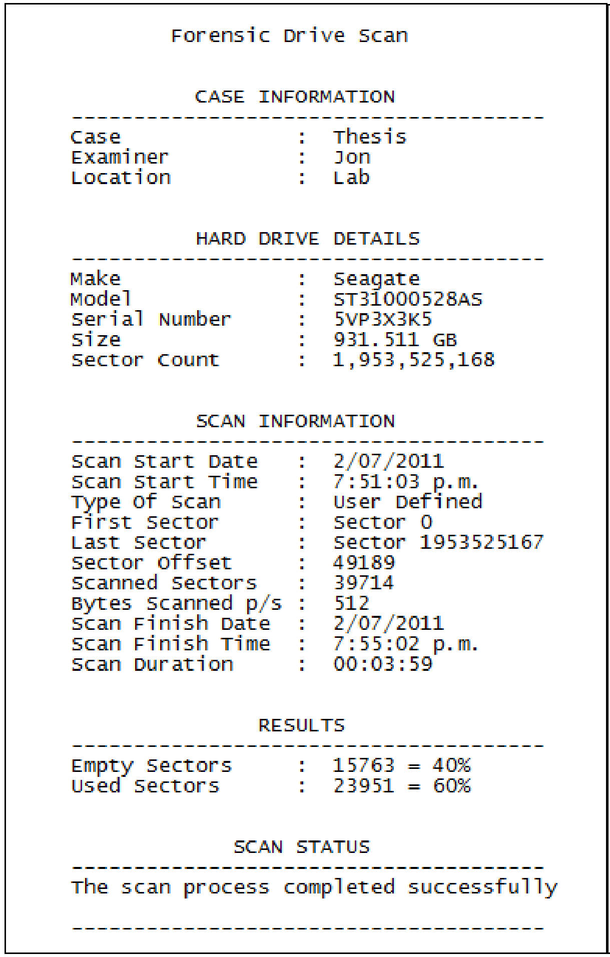


```
          Forensic Drive Scan

            CASE INFORMATION
----------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

           HARD DRIVE DETAILS
----------------------------------------
Make                 :  Seagate
Model                :  ST31000528AS
Serial Number        :  5VP3X3K5
Size                 :  931.511 GB
Sector Count         :  1,953,525,168

            SCAN INFORMATION
----------------------------------------
Scan Start Date      :  2/07/2011
Scan Start Time      :  9:04:21 p.m.
Type Of Scan         :  User Defined
First Sector         :  Sector 0
Last Sector          :  Sector 1953525167
Sector Offset        :  97676
Scanned Sectors      :  20000
Bytes Scanned p/s    :  8
Scan Finish Date     :  2/07/2011
Scan Finish Time     :  9:06:20 p.m.
Scan Duration        :  00:01:59

                 RESULTS
----------------------------------------
Empty Sectors        :  7992 = 40%
Used Sectors         :  12008 = 60%

               SCAN STATUS
----------------------------------------
The scan process completed successfully

----------------------------------------
```

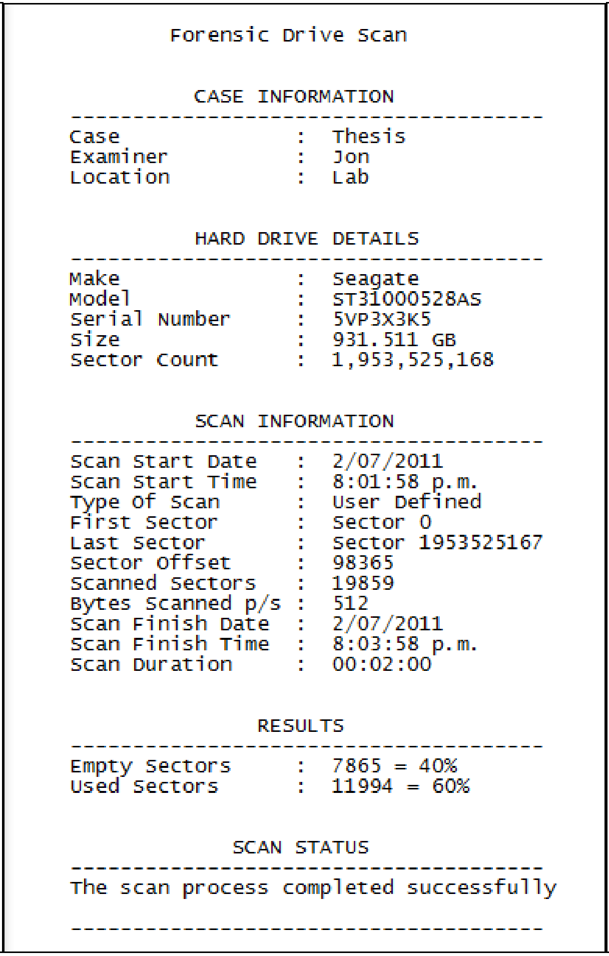


```
          Forensic Drive Scan

            CASE INFORMATION
----------------------------------------
Case                 :  Thesis
Examiner             :  Jon
Location             :  Lab

           HARD DRIVE DETAILS
----------------------------------------
Make                 :  Seagate
Model                :  ST31000528AS
Serial Number        :  5VP3X3K5
Size                 :  931.511 GB
Sector Count         :  1,953,525,168

            SCAN INFORMATION
----------------------------------------
Scan Start Date      :  2/07/2011
Scan Start Time      :  10:37:45 p.m.
Type Of Scan         :  Full Sector Scan
First Sector         :  Sector 0
Last Sector          :  Sector 1953525167
Sector Offset        :  1
Scanned Sectors      :  1953525168
Bytes Scanned p/s    :  8
Scan Finish Date     :  4/07/2011
Scan Finish Time     :  11:22:08 a.m.
Scan Duration        :  36:44:23

                 RESULTS
----------------------------------------
Empty Sectors        :  778267428 = 40%
Used Sectors         :  1175257740 = 60%

               SCAN STATUS
----------------------------------------
The scan process completed successfully

----------------------------------------
```

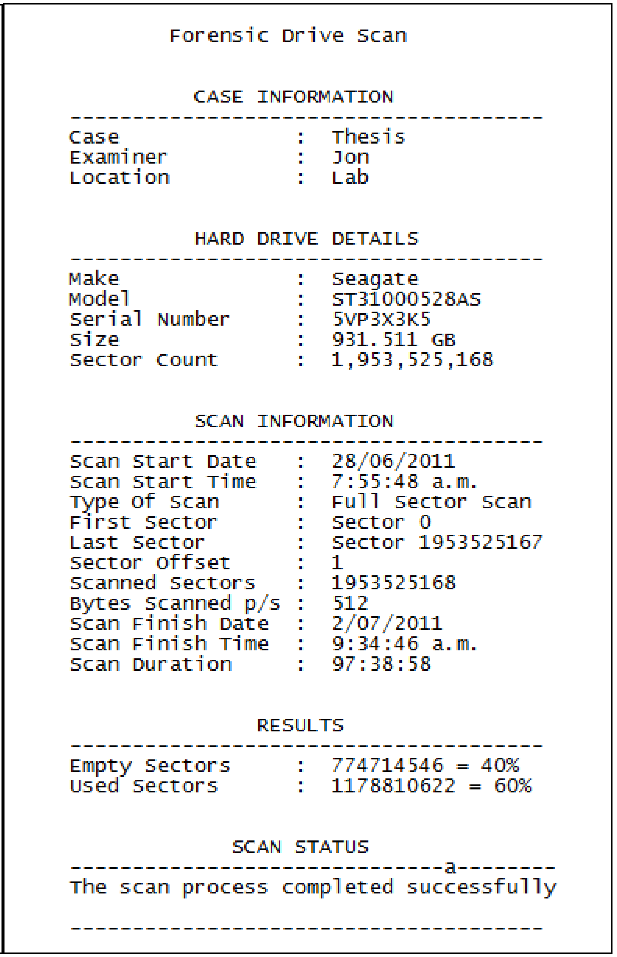
512 Byte Test Reports for Test Subject 8



```
           Forensic Drive Scan

             CASE INFORMATION
 ---------------------------------------
 Case                 :  Thesis
 Examiner             :  Jon
 Location             :  Lab

            HARD DRIVE DETAILS
 ---------------------------------------
 Make                 :  Seagate
 Model                :  ST31000528AS
 Serial Number        :  5VP3X3K5
 Size                 :  931.511 GB
 Sector Count         :  1,953,525,168

             SCAN INFORMATION
 ---------------------------------------
 Scan Start Date   :  2/07/2011
 Scan Start Time   :  7:51:03 p.m.
 Type Of Scan      :  User Defined
 First Sector      :  Sector 0
 Last Sector       :  Sector 1953525167
 Sector Offset     :  49189
 Scanned Sectors   :  39714
 Bytes Scanned p/s :  512
 Scan Finish Date  :  2/07/2011
 Scan Finish Time  :  7:55:02 p.m.
 Scan Duration     :  00:03:59


                  RESULTS
 ---------------------------------------
 Empty Sectors        :  15763 = 40%
 Used Sectors         :  23951 = 60%


                SCAN STATUS
 ---------------------------------------
 The scan process completed successfully

 ---------------------------------------
```



```
           Forensic Drive Scan

             CASE INFORMATION
 ---------------------------------------
 Case                 :  Thesis
 Examiner             :  Jon
 Location             :  Lab

            HARD DRIVE DETAILS
 ---------------------------------------
 Make                 :  Seagate
 Model                :  ST31000528AS
 Serial Number        :  5VP3X3K5
 Size                 :  931.511 GB
 Sector Count         :  1,953,525,168

             SCAN INFORMATION
 ---------------------------------------
 Scan Start Date   :  2/07/2011
 Scan Start Time   :  8:01:58 p.m.
 Type Of Scan      :  User Defined
 First Sector      :  Sector 0
 Last Sector       :  Sector 1953525167
 Sector Offset     :  98365
 Scanned Sectors   :  19859
 Bytes Scanned p/s :  512
 Scan Finish Date  :  2/07/2011
 Scan Finish Time  :  8:03:58 p.m.
 Scan Duration     :  00:02:00


                  RESULTS
 ---------------------------------------
 Empty Sectors        :  7865 = 40%
 Used Sectors         :  11994 = 60%


                SCAN STATUS
 ---------------------------------------
 The scan process completed successfully

 ---------------------------------------
```



```
           Forensic Drive Scan

             CASE INFORMATION
 ---------------------------------------
 Case                 :  Thesis
 Examiner             :  Jon
 Location             :  Lab

            HARD DRIVE DETAILS
 ---------------------------------------
 Make                 :  Seagate
 Model                :  ST31000528AS
 Serial Number        :  5VP3X3K5
 Size                 :  931.511 GB
 Sector Count         :  1,953,525,168

             SCAN INFORMATION
 ---------------------------------------
 Scan Start Date   :  28/06/2011
 Scan Start Time   :  7:55:48 a.m.
 Type Of Scan      :  Full Sector Scan
 First Sector      :  Sector 0
 Last Sector       :  Sector 1953525167
 Sector Offset     :  1
 Scanned Sectors   :  1953525168
 Bytes Scanned p/s :  512
 Scan Finish Date  :  2/07/2011
 Scan Finish Time  :  9:34:46 a.m.
 Scan Duration     :  97:38:58


                  RESULTS
 ---------------------------------------
 Empty Sectors        :  774714546 = 40%
 Used Sectors         :  1178810622 = 60%


                SCAN STATUS
 ------------------------------a---------
 The scan process completed successfully

 ---------------------------------------
```