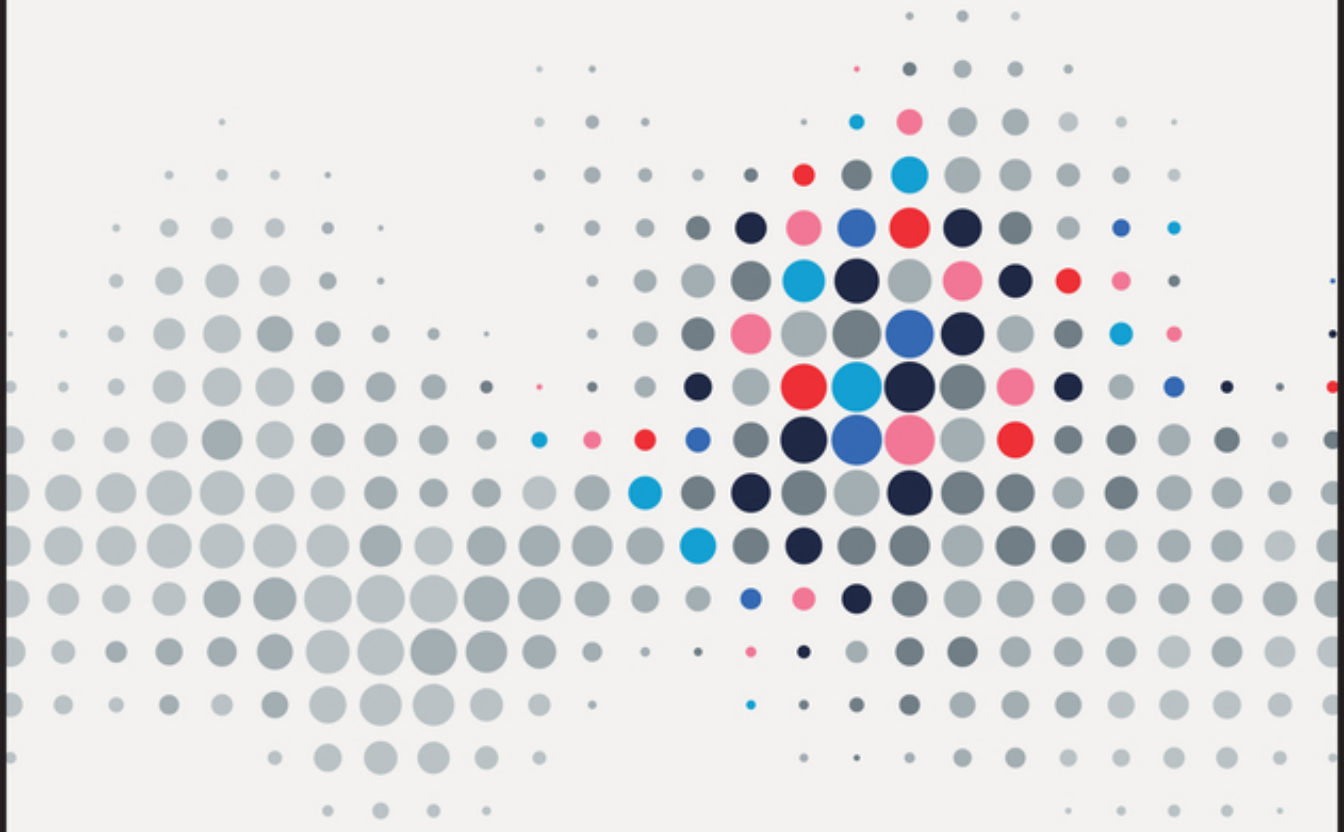
FILE SYSTEM FORENSICS



FERGUS TOOLAN

WILEY

File System Forensics

File System Forensics

Fergus Toolan
Norwegian Police University College
Ballina, Tipperary

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

The manufacturer's authorized representative according to the EU General Product Safety Regulation is Wiley-VCH GmbH, Boschstr. 12, 69469 Weinheim, Germany, e-mail: Product_Safety@wiley.com.

Library of Congress Cataloging-in-Publication Applied for:

Hardback ISBN: 9781394289790

Cover Design: Wiley
Cover Image: © aleksandarvelasevic/Getty Images

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

To Helen
Thanks for everything!

Contents

Preface

The prevalence of digital evidence in modern investigation has led to the need for more skilled analysts who can interpret digital evidence in a meaningful manner. Much digital evidence is stored in a file system and the correct recovery of this information is crucial to investigation. While there exist many tools that automate the recovery of data from file systems, there is a need for greater understanding of what these tools do. This allows the expert to verify the findings of file system forensic tools leading to greater trust in the recovered evidence.

The target audience of this title is anyone with an interest in digital investigation, who wishes to know how evidence is recovered from file systems. This includes University students taking modules or even entire programmes in the digital forensics and cybersecurity domains and law enforcement agents (or other investigators) who are tasked with recovering information from file systems and explaining its relevance. For both audiences the aim of this book is to provide an in-depth understanding of how information is recovered from common file systems.

Structure

This book is organised in five distinct parts. Part I provides the preliminaries that all digital forensic experts require. Parts II–IV provide the technical meat of the title. These parts focus on the common file systems for each of the most popular operating systems (Windows, Linux and macOS). Part V discusses the future of file system forensics and what new (and some old) challenges are ahead for the discipline.

Part I, Preliminaries, begins with an introduction to digital forensics in general and discusses some of the principles that govern the area. This chapter also introduces the reader to digital forensic methodologies and how they are used to streamline investigation. Chapter 2 describes the Linux operating system and how it can be used for file system forensics. Throughout the remainder of the text the examples will be given using the Linux command line, but there is no requirement for readers to follow this. Chapter 2 provides an introduction to Linux for those that wish to use it going forward. For those who do not intend to use (or already use) it, this chapter can be skipped. Chapter 3 discusses the topic of information representation. Computers are capable of processing and storing only binary data (ones and zeros). How these ones and zeros are interpreted as meaningful information is of vital importance. This chapter shows how numbers, text and time are represented in computing systems and how we interpret the raw hex data that we will encounter during file system forensics. The final chapter in this part introduces the reader to disk storage, partitions and file

systems. This chapter describes the common partitioning schemes in use today and shows how they can be located. It then introduces the file system and shows the various concepts that exist within this organisational structure. The chapter finishes with an introduction to disk acquisition – how we acquire the file systems that we will later process – and the analysis of these file systems.

Part II introduces the Windows family of file systems, although more accurately this might be called the Microsoft family. This includes FAT (Chapter 5), ExFAT (Chapter 6) and NTFS (Chapter 7). For many years FAT file system variants were the standard for removable media. While this is changing as ExFAT becomes dominant a large volume of devices with the FAT file system are still found during investigation. This, coupled with the relative simplicity of the FAT file system, means that it is an ideal choice for the first file system to be studied in this book. Subsequently the ExFAT file system is introduced. While some consider this to be another variant of the FAT file system, it is sufficiently more advanced to be deserving of its own chapter. This is the most common file system found on removable media today. Both FAT and ExFAT are supported by all major operating systems and, as such, can be found in many cases. The final Windows file system in this text is that of NTFS. The New Technology File System is the default on all Windows systems and as such is very commonly encountered in digital investigation.

Part III introduces the Linux file systems. This begins with the ext family of filesystems (Chapters 8 and 9). Many might wonder how important these chapters are as Linux is rarely encountered in digital investigation. However, these are some of the most important file systems in use today. The ext family are one of the default file systems found in Android phones, and as such, one of the most commonly encountered file systems in investigation. Knowledge of these file systems is of vital importance to all file system forensic investigators. File systems of the ext family are commonly encountered in many IoT devices. So while true to say that they are not commonly encountered in the home computer area, they are very important file systems for the analyst. This part of the book also introduces two less common Linux file systems: XFS (Chapter 10) and BtrFS (Chapter 11). These file systems are encountered in some Linux distributions, and are also found in large scale storage applications such as data centres. As device storage capacity increases these file systems will become more common in the home market.

Part IV examines the Apple file systems. For many years Apple devices utilised the HFS+ file system (Chapter 12). Since late 2017 Apple devices have begun to use the APFS file system (Chapter 13). This is a modern file system which can scale to modern device sizes in an efficient manner. Due to the popularlity of Apple devices (phones, tablets, computers etc.) this file system is encountered very frequently in modern digital investigation. The final part of this book looks to the future and in particular examines the challenges that will face the community over the next number of years.

Each file system chapter (Chapters 5–13) is organised in a similar manner. Initially a description of the general layout and the actual structures present in the file system is provided for each file system. This is followed by a manual analysis in which the reader can see all steps required to recover file content and metadata. Finally each chapter finishes with advanced topics for each file system. This always includes topics such as deleted and fragmented files along with topics specific to each file system.

Resources

Throughout this book example file systems are used to demonstrate the manual analysis and more advanced topics. Each file system generally has between three to five different image files that are

used. All of this data is available to the user through the book's supporting website. This data can be found at: https://www.fsforensics.com/book.

Please email any corrections to me at fergus@fsforensics.com.

Ireland, June 2024 FERGUS TOOLAN

Acknowledgements

When I began this project I never realised how difficult it would be to complete. It would never have been completed without the support of many people. To everyone who had any part in this thank you all very much! There are some who must be mentioned directly!

Firstly those that proof read the document itself. Thank you! To my technical proof readers Ray Genoe, Alan Browne, Ivar Friheim and Ulf Bergum who between them checked the code listings and exercises in the book. Many thanks for that! My father Dónal read every single word of the manuscript and offered countless corrections and suggestions for improvement.

Many thanks to the editorial team at John Wiley & Sons Ltd. Victoria Bradshaw, Vishal Paduchuru, and Aileen Storry for all of your help throughout the publication process.

Over the years I have been very fortunate to work with some amazing people in the area of digital forensics. In my current role I am part of an amazing team of academics and law enforcement officers. Thanks to Carlota, Georgina, Nina, Rune, Sara and Ulf. Also past colleagues including Ray, Cormac, Gerry, Kurt-Helge, Yves, Jørn Helge and Alexander. I thank the wonderful heads of the investigation section in the Norwegian Police University College: Ivar, John Ståle, Dag and Inger. All of you gave me the freedom to explore the topics that I was most interested in.

This book started as a resource for my students. I thank all of you that I taught over the years. Much of the information presented in this book is based on the challenging questions that you have asked me. Hopefully this book will answer some of them!

During my own formative years I had some wonderful teachers in University College Dublin. Special mention to Joe Carthy, John Dunnion, Nick Kushmerick and Henry McLoughlin. I would not have had the ability or the confidence to attempt this project without your inspirational teaching over the years.

This project would never have been completed without the constant belief and support of my family. Thanks to my parents, Dónal and Anna, who always encouraged me to pursue my dreams and supported me throughout.

Finally when I thought about giving up on this project, it was my partner Helen who encouraged me to keep going. She always believed that I could do this even on the days when I didn't believe that myself. Thank you!

Fergus Toolan

Part I

Preliminaries

1

Introduction

In recent years the volume of digital evidence in criminal investigations has increased dramatically. Consider the situation at the turn of the century when the standard computer was the desktop computer a device that resided on, as the name suggests, the desk. The majority of crime scenes did not involve a computer. When computers were involved they were generally relevant only in specific case types such as hacking/cybercrime; child abuse material; and fraud. Returning to the present, there is digital evidence in almost every case![1] The majority of people carry smartphones on their person at all times. Cars contain navigation, entertainment and camera systems. Homes and businesses have digital CCTV systems that run continuously. People communicate through social media. The end result for investigators[2] is that there has been a vast increase in the quantity of digital evidence encountered during investigation.

Almost all data in electronic storage media is held in files. A file is an object on a computing device that stores data, information, settings or applications. Every document, picture, spreadsheet, database, etc. on a computer system is composed of one or more files. Every computer system therefore needs a method of managing files. This is generally achieved through the use of a file system. File systems exist on every electronic storage device and provide a method of locating the actual file's content and also provide information about the file itself. An ability to access these files is of vital importance during investigation.

Investigators have many tools at their disposal which allow them to access this information. However, these tools suffer certain limitations including:

- **Unsupported File Systems:** There are many different file systems in existence. File system forensic tools generally support only the most common file systems, those that are found on the most common operating systems. However, there are many more that are sometimes encountered during an investigation. These might be impossible to process without knowledge of how file systems function.
- **Undisclosed Methods:** Most file system forensic tools are closed source[3] meaning users are unable to see exactly what actions are being performed. A knowledge of file system structures will allow the investigator to show how data is stored in a file system and therefore show possible means of recovering said data. It also supports verification of the results of closed-source tools.
- **Cost:** The majority of these tools are commercial tools with associated cost implications for users. Knowledge of how file systems function could ultimately allow an investigator to create their own tools.

1 The UK's National Police Chief's Council estimate that over 90% of crimes involved a digital element in 2020.
2 Investigator is used throughout this book to signify any party involved in criminal or corporate investigations.
3 One exception to this is the Sleuth Kit which will be used throughout this book to validate results when possible.

File System Forensics, First Edition. Fergus Toolan.

Hence it is necessary that investigators understand the structures that are utilised by file systems. This not only allows the investigator to analyse file systems which are not supported by the current tool but also allows them to explain possible means by which these tools work. Digital forensic analysts are often considered 'experts' in their field. Knowledge of file systems and their underlying structures will allow these analysts to more validly claim this title and stand over the evidence generated by file system forensic tools.

1.1 What is Digital Forensics?

In recent years the term digital forensics has become more familiar to all, especially anyone involved in investigation. Digital evidence is seen in TV shows on a regular basis. Numerous jobs are available in many areas which require skills in digital forensics, from e-discovery to incident response and cybersecurity. But what is digital forensics? That's a difficult question to answer!

There is no single accepted definition of digital forensics. In this section a number of definitions are presented and the common elements identified. This process will eventually culminate in the definition that is used throughout this book.

Interpol defines digital forensics as:

> … a branch of forensic science that focuses on identifying, acquiring, processing, analysing, and reporting on data stored electronically.
>
> ***Interpol (n.d.)***

Lang et al define digital forensics as:

> … the science of identifying, collecting, preserving, documenting, examining, analyzing and presenting evidence from computers, networks, and other electronic devices.
>
> ***Lang et al. (2014)***

Techopedia defines digital forensics as:

> … the process of uncovering and interpreting electronic data. The goal of the process is to preserve any evidence in its most original form while performing a structured investigation by collecting, identifying and validating the digital information for the purpose of reconstructing past events.
>
> ***Techopedia (n.d.)***

These, and the many other definitions that can be found, all share some common traits. For instance all of them mention electronic devices/data. All evidence in digital forensics is generated from electronic traces. These traces may be found on storage media, in network traffic, online, etc. Hence digital forensics is forensic analysis performed on electronically stored/transmitted information. Additionally both Lang et al. (2014) and Interpol mention science. Digital forensics is a branch of forensic science and as such should be based on scientific principles.

All of the above definitions attempt to define the process that is followed. Many definitions use similar wording to describe these processes. For instance words such as identifying, collecting (or acquiring), preserving, presenting (or reporting) or analysing (or interpreting) are used in the majority of definitions.

Hence, for the purposes of this book the following definition will be adopted.

> Digital forensics is the application of **scientific principles** to the **identification**, **preservation**, **collection**, **analysis** and **presentation** of evidence obtained from **electronic media**.

1.2 File System Forensics

File system forensics is a particular branch of digital forensics in which the electronic medium in question is the actual storage device (i.e. the disk). File systems are structures which organise information on disk. The file system is the structure that allows saved information to be retrieved at a later date. When a file is saved, not only is the content saved but information about the content (metadata) is also saved. This metadata provides much necessary information about the file content such as timestamps and file size, but also provides information on how to locate the content on disk.

Techopedia defines a file system as:

> … a process that manages how and where data on a storage disk, typically a hard disk drive (HDD), is stored, accessed and managed. It is a logical disk component that manages a disk's internal operations as it relates to a computer and is abstract to a human user.
>
> ***Techopedia (n.d.)***

File system forensics therefore involves the application of the scientific method to identify, preserve, collect, analyse and present evidence recovered from a file system. In order to be able to perform these tasks the analyst (whether human or software) must fully understand the structures on which the file system in question is based. Different file systems result in very different structures and hence different analysis methods.

For instance compare two commonly encountered file systems in digital forensics: The File Allocation Table (FAT) and the new Apple File System (APFS). FAT is an old file system and in comparison to modern file systems such as APFS it is very simple. Generally older file systems allow for the storage/retrieval of information from them and very little else. FAT contains three structures that are of interest to forensic examiners (the volume boot record, the file allocation table and directory entries) meaning that only a small amount of knowledge is required to analyse this file system effectively. Now compare this to APFS. APFS is a modern file system. It provides much more functionality than an older system such as FAT. This includes encryption, snapshots, compression, etc. The underlying structures are inherently more complex meaning that this file system is much more difficult to examine effectively.

1.3 Digital Forensic Principles

The United Kingdom's Association of Chief Police Officers[4] (ACPO) drafted the Good Practice Guide for Digital Evidence. Version 5 (released in 2012) is the latest version of this document. This document contains a set of four principles for the effective handling of digital evidence.

4 This organisation was dissolved in 2015 and replaced by the National Police Chief's Council (NPCC).

These ACPO principles, as they are often called, are based on the UK legal system and the rules of evidence in that system. However, the principles are almost universal and have been adopted in many countries over recent years. These principles are:

- **Principle 1:** No action taken by law enforcement agencies, persons employed within these agencies or their agents should change data which may subsequently be relied upon in court.
- **Principle 2:** In circumstances where a person finds it necessary to access original data, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions.
- **Principle 3:** An audit trail or other record of all processes applied to digital evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.
- **Principle 4:** The person in charge of the investigation has overall responsibility for ensuring that the law and these principles are adhered to.

These principles were originally drafted in the early days of digital evidence. At that stage most digital evidence resided on computer storage devices. The standard method was to power off the device, create an image and analyse that image. As the area of digital forensics has developed over the years this standard method of operation has also developed. Now it is no longer always advised to switch off the computer. Instead it is sometimes recommended to analyse running machines. Not all information is now stored locally, some will be stored remotely, even the crime scene itself may be remote. However, while some argue that the principles need to be updated, they are still fit for purpose.

The overall aim of these principles is to ensure that all digital evidence accessed during a criminal investigation is handled in such a manner that it can be used in court. Hence the first principle states that no changes should be made to the original data. The second principle ensures that only trained personnel will ever access original data and the third principle ensures that others will be able to recreate and validate the analysis of the evidential material. For traditional computer forensics these three principles are sufficient.

Now consider a modern scenario. First responders arrive at the home of a suspected cybercriminal. They use a warrant to enter the premises and seize all electronic evidence that is encountered. Upon entering the premises they discover a running computer. The traditional advice was to 'pull-the-plug'; however, with modern computers this might lose evidence. Most modern operating systems allow options for encrypted storage. If power is removed the data on the device will become inaccessible due to its encryption. Additionally many users use remote storage resources for files, emails, profiles, etc. Connections to these sites (and potentially access to the data they contain) will be lost if the power is removed. Hence live data forensics (LDF) is conducted, in which the running computer is analysed to determine if there is anything of evidential value present.

Returning to ACPO Principle 1 which states 'no action taken by law enforcement agencies … should change data which may subsequently be relied upon in court'. LDF immediately breaks this principle. Every action performed on a running computer system will leave traces in memory and on disk. Even the simple act of moving the mouse will have consequences, as it will change certain parts of the computer system. This has led to some researchers suggesting that the ACPO principles should be rewritten. However, they are still fit for purpose as while LDF breaks principle 1, combining this with principles 2 (competence) and 3 (audit trail), the *altered* data collected from the LDF process can still be used in a court setting.