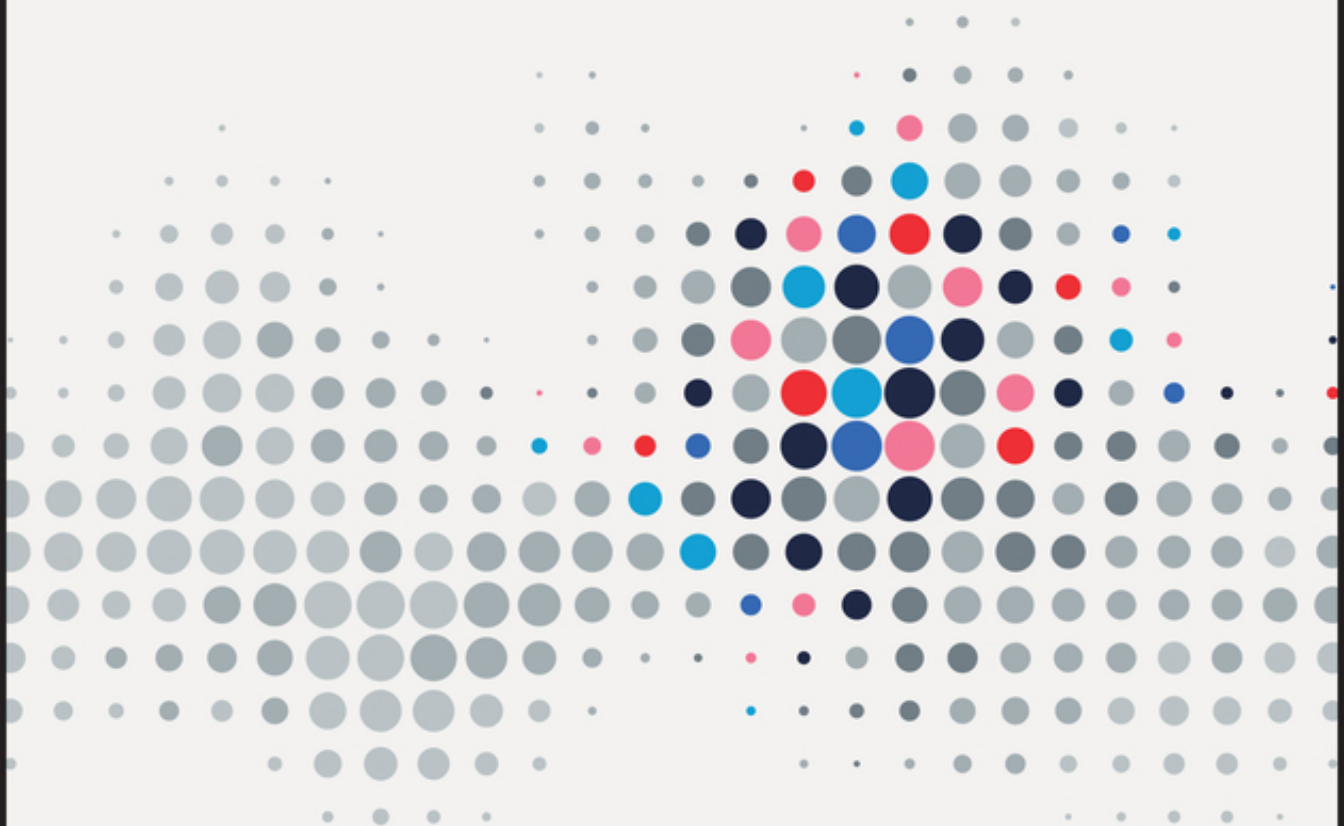


FILE SYSTEM FORENSICS



FERGUS TOOLAN

WILEY

File System Forensics

File System Forensics

Fergus Toolan

Norwegian Police University College
Ballina, Tipperary

WILEY

Copyright © 2025 by John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial technologies or similar technologies.

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

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

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.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.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 *xvii*

Acknowledgements *xxi*

Part I Preliminaries *1*

- 1 Introduction** *3*
 - 1.1 What is Digital Forensics? *4*
 - 1.2 File System Forensics *5*
 - 1.3 Digital Forensic Principles *5*
 - 1.4 Digital Forensic Methodology *7*
 - 1.4.1 Preparation *8*
 - 1.4.2 Localisation/Preservation *8*
 - 1.4.3 Acquisition *8*
 - 1.4.4 Processing *9*
 - 1.4.5 Analysis *9*
 - 1.4.6 Reporting *9*
 - 1.4.7 Quality Assurance *10*
 - 1.4.8 Evidence Return *10*
 - 1.5 About This Book *10*
 - 1.5.1 Who Should Read This Book? *11*
 - 1.6 Book Structure *12*
 - 1.7 Summary *13*
 - Exercises *13*
 - Bibliography *14*
- 2 Linux as a Forensic Platform** *17*
 - 2.1 Open-Source Software *17*
 - 2.1.1 Advantages of Open-Source Software *19*
 - 2.1.2 Open Source $\neq$ Free *20*
 - 2.2 Open-Source Software in Digital Forensics *20*
 - 2.3 What is Linux? *21*
 - 2.3.1 The Anatomy of the Linux OS *22*
 - 2.3.2 Linux Distributions *27*

2.3.3	A (very) Brief History of Linux	28
2.4	Using Linux	29
2.4.1	User Accounts	30
2.4.2	Basic Linux Commands	32
2.4.2.1	Navigating the File System	32
2.4.2.2	Getting Help	34
2.4.2.3	Viewing/Editing Text Files	34
2.4.2.4	Managing Directories	35
2.4.2.5	Redirection and Pipes	35
2.5	Linux as a Forensic Platform	36
2.5.1	Commands for Digital Forensics	36
2.5.1.1	Hashing	36
2.5.1.2	Hex Viewers	38
2.5.1.3	Archiving/Compression	39
2.5.1.4	The file Command	40
2.5.1.5	The strings Command	40
2.5.1.6	Text Searching with (e)grep	41
2.6	Summary	42
	Exercises/Discussion Topics	42
	Bibliography	43
3	Mathematical Preliminaries	45
3.1	Bits and Bytes	45
3.2	Number Systems	48
3.2.1	Notational Conventions	48
3.2.2	Decimal	48
3.2.3	Binary	49
3.2.4	Hexadecimal	50
3.2.5	Number Conversions	51
3.2.6	Number Conversion with Bash	51
3.2.7	Negative Numbers	53
3.2.8	Floating-Point Numbers	53
3.3	Representing Text	56
3.3.1	ASCII	56
3.3.2	ISO-8859	57
3.3.3	Unicode	59
3.3.4	UTF-8	60
3.3.5	UTF-16	61
3.4	Representing Time	62
3.4.1	Unix Time	63
3.4.2	The Linux date Command	64
3.5	Endianness and Raw Data	64
3.6	Summary	66
	Exercises	67
	Bibliography	68

4	Disks, Partitions and File Systems	69
4.1	Disk Storage	70
4.1.1	Traditional Rotational Hard Drives	71
4.1.1.1	Optical Media	72
4.1.2	Flash Drives	73
4.1.3	Solid-State Drives	73
4.2	Partitions	74
4.2.1	Creating Partitions/File Systems on Linux	74
4.2.1.1	Mounting File Systems on Linux	77
4.2.2	Master Boot Record	78
4.2.3	GUID Partition Table	80
4.3	File Systems	83
4.3.1	File System Concepts	83
4.3.2	Comparison of File Systems	86
4.4	Acquisition of File System Data	88
4.4.1	Logical vs Physical Acquisition	88
4.4.2	Acquisition Under Linux	88
4.4.2.1	The dd Family	89
4.4.2.2	Expert Witness Format (EWF)	90
4.4.2.3	guymager	91
4.5	Analysis of File Systems	92
4.5.1	The Sleuth Kit	92
4.5.1.1	Determine the Partition Layout	93
4.5.1.2	Determine the File System Type	93
4.5.1.3	List the Files	94
4.5.1.4	Recover File Metadata	95
4.5.1.5	Recover File Content	95
4.5.1.6	Other TSK Commands	95
4.5.2	Data Carving	96
4.6	Summary	97
	Exercises	97
	Bibliography	98

Part II Windows File Systems 99

5	The FAT File System	101
5.1	On-Disk Structures	101
5.1.1	Layout	102
5.1.2	Volume Boot Record	102
5.1.3	File System Information (FSINFO)	102
5.1.4	File Allocation Table	104
5.1.5	Directory Entries	105
5.1.6	FAT Date and Time	108
5.1.7	Mapping Clusters to Sectors	109
5.2	Analysis of FAT32	109

5.2.1	Creating FAT32 File Systems	109
5.2.2	Supplied FAT32 Image Files	110
5.2.3	FAT32 Manual Analysis	110
5.2.3.1	Process the VBR	111
5.2.3.2	Process the Root Directory	112
5.2.3.3	Process Sub-directories	113
5.2.3.4	Recover Metadata/Content	113
5.3	FAT32 Advanced Analysis	115
5.3.1	Deleted Files	116
5.3.2	The Volume Label	117
5.4	Summary	117
	Exercises	118
	Bibliography	118
6	The ExFAT File System	121
6.1	On-Disk Structures	121
6.1.1	Volume Boot Record	122
6.1.2	File Allocation Table	123
6.1.3	Directory Entries	125
6.1.3.1	Allocation Bitmap (Type: 0x81)	127
6.1.3.2	Up-Case Table (Type: 0x82)	128
6.1.3.3	Volume Label (Type: 0x83)	128
6.1.3.4	File (Type: 0x85)	129
6.1.3.5	Volume GUID (Type: 0xA0)	130
6.1.3.6	Stream Extension (Type: 0xC0)	130
6.1.3.7	Filename Extension	131
6.1.3.8	Other Directory Entries	132
6.2	Analysis of ExFAT	132
6.2.1	Creating ExFAT File Systems	132
6.2.2	Supplied ExFAT Image Files	132
6.2.3	ExFAT Manual Analysis	132
6.2.3.1	Step 1: Process the VBR	133
6.2.3.2	Step 2: Process the Root Directory	133
6.2.3.3	Step 3: Process Subdirectories	136
6.2.3.4	Step 4: Recover Metadata	137
6.2.3.5	Step 5: Recover Content	137
6.3	ExFAT Advanced Analysis	139
6.3.1	Long File Names	139
6.3.2	Deleted Files	140
6.3.3	Fragmented Files and Large Directories	141
6.4	Summary	142
	Exercises	143
	Bibliography	143
7	The NTFS File System	145
7.1	On-Disk Structures	146
7.1.1	\$Boot	146

7.1.2	Indexes	147
7.1.3	Fixup Arrays	149
7.1.4	Time in NTFS	150
7.1.5	Master File Table	151
7.1.6	MFT Record Structure	152
7.1.6.1	MFT Record Header	152
7.1.6.2	Browsing Attributes	155
7.1.6.3	\$STANDARD_INFORMATION (0x10)	155
7.1.6.4	\$ATTRIBUTE_LIST (0x20)	156
7.1.6.5	\$FILENAME (0x30)	156
7.1.6.6	\$OBJECT_ID (0x40)	157
7.1.6.7	\$SECURITY_DESCRIPTOR (0x50)	159
7.1.6.8	\$VOLUME_NAME (0x60)	162
7.1.6.9	\$VOLUME_INFORMATION (0x70)	162
7.1.6.10	\$DATA (0x80)	163
7.1.6.11	\$INDEX_ROOT (0x90)	163
7.1.6.12	\$INDEX_ALLOCATION (0xA0)	165
7.1.6.13	\$BITMAP (0xB0)	165
7.1.6.14	\$REPARSE_POINT (0xC0)	166
7.1.6.15	\$EA_INFORMATION (0xD0) and \$EA (0xE0)	167
7.2	Analysis of NTFS	167
7.2.1	Creating NTFS File Systems	168
7.2.2	Supplied NTFS Image Files	168
7.2.3	NTFS Manual Analysis	168
7.2.3.1	Process \$Boot	169
7.2.3.2	Recover \$MFT	171
7.2.3.3	Process Directories	173
7.2.3.4	Recover File Metadata	177
7.2.3.5	Recover File Content	182
7.3	NTFS Advanced Analysis	185
7.3.1	Further File System Information	185
7.3.2	Deleted Files	186
7.3.3	Fragmented Files	187
7.3.4	Alternate Data Streams	190
7.3.5	Large MFT Records	191
7.4	Summary	194
	Exercises	194
	Bibliography	195

Part III Linux File Systems 197

8	The EXT2 File System	199
8.1	On-Disk Structures	200
8.1.1	The Superblock	201
8.1.2	The Block Group Descriptor Table	204
8.1.3	The Inode Table	205

8.1.3.1	Mode/Permissions	207
8.1.3.2	Inode Flags	208
8.1.3.3	Block Pointers	208
8.1.4	The Data and Inode Bitmaps	209
8.1.5	Locating an Inode	209
8.2	Analysis of ext2	210
8.2.1	Creating ext2 File Systems	210
8.2.2	Supplied ext2 Image Files	210
8.2.3	Ext2 Manual Analysis	211
8.2.3.1	Process the Superblock	211
8.2.3.2	Map the Block Groups	213
8.2.3.3	Process Root Directory Inode	216
8.2.3.4	Process the Root Directory	217
8.2.3.5	Process Directories	219
8.2.3.6	Process Files	219
8.3	Ext2 Advanced Analysis	222
8.3.1	Fragmented Files	222
8.3.2	Links	223
8.3.3	Deleted Files	225
8.4	Summary	226
	Exercises	226
	Bibliography	227
9	The EXT3/EXT4 File Systems	229
9.1	Supplied Image Files	229
9.2	The ext3 File System	229
9.2.1	The Ext Journal	230
9.2.2	HTree Directory Indexing	237
9.3	The Ext4 File System	241
9.3.1	Large Inodes	241
9.3.1.1	Timestamps	241
9.3.2	Ext4 Data Storage	244
9.3.2.1	Extent-Based Storage	244
9.3.2.2	Inline Storage	248
9.3.2.3	Symbolic Links	248
9.3.3	File Deletion in Ext4	249
9.3.4	Extended Attributes	252
9.3.5	Ext4 Block Group Descriptors	255
9.3.6	Flexible Block Groups	255
9.4	Summary	258
	Exercises	259
	Bibliography	260
10	The XFS File System	263
10.1	On-Disk Structures	264
10.1.1	Allocation Groups	264
10.1.2	Addressing	266

10.1.2.1	Inode Addressing	266
10.1.3	XFS B+ Trees	267
10.1.4	The Superblock	268
10.1.4.1	Locating Superblocks	268
10.1.5	XFS Signatures	271
10.1.6	XFS Inodes	271
10.1.7	Directories	273
10.1.8	Extents	274
10.1.9	Time in XFS	275
10.2	Analysis of XFS	275
10.2.1	Creating XFS File Systems	275
10.2.2	Supplied XFS Image Files	275
10.2.3	XFS Manual Analysis	276
10.2.3.1	Process the Superblock	276
10.2.3.2	Locate the Root Directory	277
10.2.3.3	Process the Root Directory	279
10.2.3.4	Process the Subdirectories	281
10.2.3.5	Recover File Content/Metadata	281
10.3	XFS Advanced Analysis	282
10.3.1	AG Free Space Management	283
10.3.1.1	AG Free List	285
10.3.2	AG Inode Management	286
10.3.3	Deleted Files	289
10.3.4	Extended Attributes	290
10.3.5	Links	291
10.3.6	The XFS Journal	292
10.4	Summary	300
	Exercises	301
	Bibliography	301
11	The Btrfs File System	303
11.1	On-Disk Structures	304
11.1.1	The Superblock	305
11.1.2	Btrfs Trees	305
11.1.3	Btrfs Tree Structure	307
11.1.3.1	Node Header Structure	307
11.1.3.2	Internal Node Structure	309
11.1.4	Btrfs Keys	309
11.1.5	Btrfs Items	310
11.1.6	Time in Btrfs	315
11.1.7	Logical and Physical Addressing	315
11.2	Analysis of Btrfs	317
11.2.1	Creating Btrfs File Systems	317
11.2.2	Supplied Btrfs Image Files	318
11.2.3	Btrfs Analysis Methodology	318
11.2.4	Manual Analysis of a Single Device File System	320
11.2.4.1	Process the Superblock	320

11.2.4.2	Process the CHUNK_ARRAY	321
11.2.4.3	Locate the CHUNK_TREE	322
11.2.4.4	Process the CHUNK_TREE	323
11.2.4.5	Locate the Root Tree	326
11.2.4.6	Locate the FS_TREE	327
11.2.4.7	Processing the FS_TREE	328
11.2.4.8	Process Directories	329
11.2.4.9	Recovering Metadata	335
11.2.4.10	Recovering File Contents	336
11.3	Btrfs Advanced Analysis	338
11.3.1	File Deletion	338
11.3.2	Analysis of Internal Nodes	342
11.3.3	Multiple Device Configuration	343
11.3.4	Subvolumes and Snapshots	346
11.4	Summary	350
	Exercises	350
	Bibliography	351

Part IV Apple File Systems 353

12	The HFS+ File System	355
12.1	On-Disk Structures	355
12.1.1	Forks	357
12.1.2	Time in HFS+	357
12.1.3	Volume Header	358
12.1.4	B-Trees	358
12.1.5	Catalog File	362
12.1.6	HFS+ Permissions	363
12.1.7	Text Encoding	365
12.1.8	Extents Overflow File	365
12.1.9	Allocation File	366
12.1.10	HFS+ Journal	367
12.2	Analysis of HFS+	369
12.2.1	Creating HFS+ File Systems	369
12.2.2	Supplied HFS+ Image Files	370
12.2.3	HFS+ Manual Analysis	370
12.2.3.1	Process the Volume Header	370
12.2.3.2	Locate the Catalog File	371
12.2.3.3	Process the Catalog B-Tree	373
12.2.3.4	Gather Metadata	377
12.2.3.5	Recover File Content	377
12.3	HFS+ Advanced Analysis	380
12.3.1	Deleted Files	380
12.3.2	Index Nodes	381
12.3.3	Fragmented Files	383
12.3.4	Links	387

12.4	Summary	390
	Exercises	391
	Bibliography	391
13	The APFS File System	393
13.1	On-Disk Structures	394
13.1.1	Time in APFS	394
13.1.2	Objects	394
13.1.3	B-Trees	396
13.1.4	Containers and Volumes	399
13.1.5	Container Superblock	400
13.1.6	Volume Superblock	402
13.1.7	Object Maps	404
13.1.8	File-Related Structures	405
13.1.8.1	File System Keys	406
13.1.8.2	Inode	407
13.1.8.3	Directory Record	408
13.1.8.4	Extent	410
13.1.9	Checkpoints	410
13.1.10	Other APFS Structures	412
13.2	Analysis of APFS	412
13.2.1	Creating APFS File Systems	412
13.2.2	Supplied APFS Image Files	413
13.2.3	APFS Manual Analysis	413
13.2.3.1	Process the Container Superblock	414
13.2.3.2	Process the Container Object Map	415
13.2.3.3	Process the Volume Superblock	418
13.2.3.4	Process the Volume Object Map	418
13.2.3.5	Process the File System Tree	419
13.3	APFS Advanced Analysis	425
13.3.1	Deleted Files	425
13.3.2	Checkpoint Recovery	426
13.3.3	Multi-Level B-Trees	427
13.3.4	Multiple Volumes	429
13.3.5	Extended Attributes	430
13.3.6	Links	431
13.4	Summary	433
	Exercises	433
	Bibliography	434
Part V	The Future	435
14	Future Challenges in Digital Forensics	437
14.1	Challenges in Digital Forensics	437
14.1.1	Data Volume	438
14.1.2	Multi-Source Correlation	439

14.1.3	New File Systems	440
14.1.4	Encryption	440
14.1.5	Cloud Storage	441
14.1.6	Lack of Resources	441
14.1.6.1	Human Resources	441
14.1.6.2	Software Resources	442
14.1.6.3	Hardware Resources	442
14.1.7	Tool Validation/Datasets	443
14.1.8	Lack of Standardisation	444
14.1.9	Legal/Scientific Challenges	444
14.1.10	Presentation of Evidence	445
14.1.11	Human Error/Bias	446
14.2	Where Do We Go from Here?	447
14.2.1	Training/Education	448
14.2.2	Free Open-Source Software (FOSS)	448
14.2.3	Triage	449
14.2.4	Artificial Intelligence (AI)	449
14.2.5	Live Data Forensics	450
14.2.6	Legal Solutions	451
14.2.7	Data Set Development/Tool Testing	452
14.2.8	Standardisation	452
14.2.9	Information Sharing	453
14.2.10	Virtualisation	453
14.3	Summary	454
	Bibliography	454

Index	457
--------------	-----

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 popularity 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 Story 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 investigators² 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 source³ 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.

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 Officers⁴ (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.