SpringerBriefs on Cyber Security Systems and Networks

The series aims to develop and disseminate an understanding of innovations, paradigms, techniques, and technologies in the contexts of cyber security systems and networks related research and studies. It publishes thorough and cohesive overviews of state-of-the-art topics in cyber security, as well as sophisticated techniques, original research presentations and in-depth case studies in cyber systems and networks. The series also provides a single point of coverage of advanced and timely emerging topics as well as a forum for core concepts that may not have reached a level of maturity to warrant a comprehensive textbook. It addresses security, privacy, availability, and dependability issues for cyber systems and networks, and welcomes emerging technologies, such as artificial intelligence, cloud computing, cyber physical systems, and big data analytics related to cyber security research. The mainly focuses on the following research topics:

Fundamentals and Theories

- Cryptography for cyber security
- Theories of cyber security
- Provable security

Cyber Systems and Networks

- Cyber systems Security
- Network security
- Security services
- Social networks security and privacy
- Cyber attacks and defense
- Data-driven cyber security
- Trusted computing and systems

Applications and Others

- Hardware and device security
- Cyber application security
- Human and social aspects of cyber security

More information about this series at http://www.springer.com/series/15797

Darren Quick · Kim-Kwang Raymond Choo

Big Digital Forensic Data

Volume 1: Data Reduction Framework and Selective Imaging

Springer

Darren Quick
University of South Australia
Adelaide, SA
Australia

Kim-Kwang Raymond Choo
University of Texas at San Antonio
San Antonio, TX
USA

ISSN 2522-5561 ISSN 2522-557X (electronic)
SpringerBriefs on Cyber Security Systems and Networks
ISBN 978-981-10-7762-3 ISBN 978-981-10-7763-0 (eBook)
https://doi.org/10.1007/978-981-10-7763-0

Library of Congress Control Number: 2018934398

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
part of Springer Nature
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

"Work smarter not harder." So goes the age old adage. Unfortunately, we in the digital forensics community have continued to work harder and harder over the years—developing distributed computing models and tools, leveraging collaboration platforms, and mastering the task force approach to large scale investigation and analysis. This isn't to say the community hasn't benefited from increasingly "smart" tools and people. Our community is full of superb tool developers and a highly skilled, technically adept community of investigators. "Smart" surely isn't a quality lacking in our field, as I continue to be impressed with technically adept practitioners, researchers, and tool developers.

The future also looks bright as I peer years ahead, anticipating an increasingly smart suite of tools that more fully leverage the machine learning wave that's gaining momentum and finding its rightful place in our field. I look forward to the day, in the not too distant future, where the analyst's toolkit transcends the simple 'search, extract, and present' paradigm of old and begins to truly reduce the analytical burden and overhead that still plagues today's investigators. Yet today, we seem to be stuck at the dangerous intersection of "collect and search every literal bit of evidence" and "storage capacity and use eclipses modern, common place processing capabilities." And in my view, simply throwing additional compute at the problem, to index and present voluminous amounts of data to the investigator faster isn't the right solution. We are drowning in a deluge of data, more and more every day.

This book not only provides a great review and critical analysis of the current literature surrounding big data forensics, it provides useful and paradigm shifting frameworks for approaching the problem we face today—where the amount of data far eclipses the intelligence of our analytical platforms. Simply put, in this book, Drs. Quick and Choo provide transformative frameworks for selective imaging, quick analysis, and intelligence driven information fusion. This book provides mechanisms, backed up by empirical studies, to work smarter not harder in answering investigative questions today. Thankfully, they do so while remaining mindful of the need to preserve evidence for more in-depth analysis.

In this book, the authors provide useful frameworks that *augment* current approaches, not replace them. They provide frameworks that make investigations more effective and efficient. They provide a compelling argument for changing the way we currently do business. In short, anyone interested in advancing the field of big data forensics will find this book a great resource for surveying the field. I hope and expect this book will facilitate greater discussion of the big data challenges and solutions thereto in the very important field of digital investigations.

San Antonio, USA

Nicole Beebe, Ph.D., CISSP, CCFP, EnCE, ACE
Director, The Cyber Center for Security and
Analytics, Associate Professor of Cyber Security,
The University of Texas at San Antonio (Computer
Crime Investigator, U.S. Air Force Office of Special
Investigations 1998–2007)

Preface

Digital forensic analysis is the process of identification, preservation, analysis, and presentation of digital and electronic evidence in a manner that is legally acceptable. A major challenge to digital forensic analysis is the ongoing growth in the volume of data seized and presented for analysis. This is a result of the continuing development of storage technology, consumer devices, and cloud storage, which has led to increasing backlogs of evidence awaiting analysis, often many months to years, affecting even the largest digital forensic labs.

There have been many calls for research to address the volume challenge. While more people are needed to undertake analysis, there is also a potential to develop innovative methods to collect relevant data to conduct analysis, reducing the time a practitioner spends sorting the wheat from the chaff, or looking for needles in ever-growing haystacks. Data mining is a process of knowledge discovery which may offer a faster way to understand the ever-increasing volume of data. Applying the process of data mining to digital forensic data may lead to a methodology to assist practitioners in analyzing the vast volumes of data.

The research outlined herein involved collecting and assembling a corpus of test data from a range of devices: mobile phones, portable storage, and computers, as well as other sources of digital forensic data. Research was then undertaken using the collected data in relation to applying data reduction and intelligence analysis methodologies to determine which, if any, are applicable to digital forensic analysis.

In the following book, a framework for data mining and data reduction is outlined, including a methodology for data reduction, which paves the way for Volume 2, in which a process of quick analysis, in-depth analysis, semi-automated information and entity extraction, and link charting in conjunction with link analysis is outlined. In Volume 1, the framework is explained, and then applied to a test data corpus and real-world data to ensure the process is valid and applicable to real-world data and investigations.

The data experiments observed a reduction to 0.206% of the original source data volume, as an example, from 8.57TB of data to 12.3GB in a digital forensic data subset. The data subsets were then used in Volume 2 to explore processes of quick analysis, and semi-automated information and entity extraction, including a process of value adding to the data subset with open-source information, with positive results.

The proposed digital forensic data reduction and data mining framework provides forensic practitioners with a methodology to guide through the process of digital forensic analysis, allowing for instances where the practitioner can decide whether to undertake full forensic imaging and analysis, or rapidly collect data which is then processed and reviewed in a timely manner. Should the reduction and review process not discover evidence or intelligence of value, the framework provides for this by including an ability to traverse between rapid collection and timely review, and full forensic imaging and analysis. This is not seen in other digital forensic frameworks for analysis.

Adelaide, Australia Darren Quick
San Antonio, USA Kim-Kwang Raymond Choo

Acknowledgements

The permission of Elsevier (Chap. 2) and Springer (Chap. 4) for permission to reprint the respective copyrighted material is acknowledged. The views and opinions expressed in this book are those of the authors alone and not the organizations with whom the authors have been associated or supported.

Contents

Abbreviations

$MFT	Windows Master File Table
ACPO	Association of Chief of Police Officers
AD1	AccessData Logical Evidence File
CSV	Comma Separated Value
CTR	X-Ways Logical Image Container
DOCX	Windows Document Format
E01	Encase Physical Evidence Format
EXIF	Exchangeable image file format
EXT3/4	Linux Extended File System
FAT	File Allocation Table
FTK	Forensic Tool Kit
HD	Hard Drive
HFS/+	Apple Hierarchal File System
HTML	Hypertext Markup Language
ICT	Information and Communication Technology
IEF	Internet Evidence Finder
iOS	Apple iPhone Operating System
IP	Internet Protocol
ISO	International Organization for Standardization
IT	Information Technology
JPG	Joint Picture Group
L01	Encase Logical Evidence Format
LT	Laptop
MD5	Message Digest
NIJ	National Institute of Justice
NIST	National Institute of Standards and Technology
NTFS	New Technology File System
OS	Operating System
OSX	Apple Operating System
PC	Personal Computer

PDF	Portable Document Format
PLIST	Property List
PPTX	Microsoft PowerPoint format
RAM	Random Access Memory
RTF	Rich Text Format
SHA	Secure Hash Algorithms
UFED	Forensic Software from Cellebrite for mobile device analysis
URL	Uniform Resource Locator
USB	Universal Serial Bus
VM	Virtual Machine
VMDK	Virtual Machine Disk
XLSX	Microsoft Spreadsheet format
XRY	Forensic software from MSAB for mobile device analysis

Keywords

Big Data · Big Forensic Data · Computer Forensics · Criminal Intelligence
Data Mining · Data Reduction · Data Volume · Digital Evidence
Digital Forensic Analysis · Digital Forensics · Evidence Discovery
Forensic Challenges · Forensic Computer Analysis · Forensic Computing
Forensic Intelligence · Intelligence Analysis · Knowledge Management
Mobile Device Forensic Extracts · Open Source Intelligence · Quick Analysis
Selective Imaging