Conversational Digital Forensics Analysis

A ConversationalGeek® Book

Sponsored by ACCESSDATA

MINI Edition

Learn about:

- The complexities of digital intelligence that can threaten a case
- New forensics technologies that are evolving investigative workflows

By Derek Smith
(Certified Information Systems Security Professional)

Sponsored by AccessData

Conversational Digital Forensics Analysis

by Derek A. Smith

ConversationalGeek®

Conversational Digital Forensics Analysis

Published by Conversational Geek Inc.
www.conversationalgeek.com

Trademarks

Warning and Disclaimer

Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied. The information provided is on an “as is” basis. The author and the publisher shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or programs accompanying it.

Additional Information

For general information on our other products and services, or how to create a custom Conversational Geek book for your business or organization, please visit our website at ConversationalGeek.com

Publisher Acknowledgments

All of the folks responsible for the creation of this book:

Author:	Derek Smith
Project/Copy Editor:	Steven Zimmerman
Content Reviewer(s):	J. Peter Bruzzese

The “Conversational” Method

We have two objectives when we create a “Conversational” book: First, to make sure it’s written in a conversational tone so that it’s fun and easy to read. Second, to make sure you, the reader, can immediately take what you read and include it into your own conversations (personal or business-focused) with confidence.

“Geek in the Mirror” Boxes

We infuse humor and insight into our books through both cartoons and light banter from the author. When you see one of these boxes it’s the author stepping outside the dialog to speak directly to you. It might be an anecdote; it might be a personal experience.

Within these boxes I can share just about anything on the subject at hand. Read ’em!

Conversational Digital Forensics Analysis

HOW TO CATCH THE BAD GUY:
1. Collect tons of data
2. Sort all the data
3. Analyze every last word
4. Find time to catch the bad guy

Crime and misconduct are almost as old as time itself. Remember Cain and Abel? With the invention of misdeeds came the equally brilliant notion of the cover-up – after all, who wants to get caught? Since then, there have been those tasked with discovering and sorting through the evidence to determine “who done it?”

During the 1970s the work of forensic investigation went digital. Since then, the demand for digital forensics has grown exponentially. In our highly interconnected world, it is next to impossible to avoid leaving a digital trail of a misdeed; unless, of course, the suspected miscreant has lived and worked exclusively in the backwaters of the Amazon basin for the last 30 years.

I’ll bet there’s even a digital trail to find my car keys!

We create a digital footprint every time we use a computer for online or offline activities. Moreover, we increase our footprint with every phone call, every email, every text message, and with each use of our credit or debit card. Even walking by an ATM camera increases our digital footprint.

Just how big is the average digital footprint?

Consider that, in 1860, a pony express rider carried up to 1,280 messages weighing no more than ¼ ounce from New York to San Francisco. The journey took 10 days and carried the outrageous expense of $50/month for each rider. That was 1860, and those 1,280 letters were roughly the equivalent of 640k of data. Today we can send terabytes of data – roughly a billion times more than the rider carried – to multiple simultaneous destinations around the world in a matter of minutes, if not seconds.

Consider also that our digital technology is capable of producing and collecting more than a terabyte of data each week for every individual on the planet – at least one-third of which is aggregated, analyzed, processed, and stored in the cloud. With this level of data production, it is virtually impossible to hide a misdeed. Somewhere in that vast digital ocean is evidence with the suspect’s name on it, proving “who done it.”

The challenge with that much data, of course, isn’t whether or not the digital evidence exists – *it does*. The challenge is isolating the necessary evidence from the rest of the information. Today’s digital investigations bring a whole new meaning to the old cliché of “a needle in a haystack.”

The amount of data generated by each person is still increasing with no inflection point in sight. And we have yet to find an acceptable cure for crime and misdeeds. What is left, then, is the need to find greater efficiencies for digital investigations.

Digital Investigators are Overwhelmed

Broadly speaking, a digital forensic investigation is the application of forensic sciences to digital information. The digital forensic investigation is concerned with identifying, collecting, examining, and analyzing digital evidence in a way that maintains a strict chain of custody while preserving data integrity so that the evidence is admissible in legal courts and other settings. The work of the investigator is to find the information that others either didn’t know was there or that someone tried to destroy.

Back in the 70s, when digital forensic investigations were getting started, computers booted from one 5 ¼” floppy and stored their programs and data on another. Each 5 ¼” floppy could hold up to 110 KB of data. It’s incredible that we were able to digitally function back then – today you can’t get by with less than a gigabyte of memory just to turn on a Windows 10 machine. The relatively small amount of data and virtually non-existent connectivity (as measured by today’s standards) meant that only two or three digital investigators could well-enough handle the entire workload of a large law firm or regional office of a federal agency.

Each year since then, the workload for digital investigators has rapidly increased. Just from 2007 to 2008, for example, FBI digital investigators increased the amount of data processed by 27%. By 2010, the average FBI digital investigation caseload was 0.4 terabytes. And by 2013, the average case size increased to 1.0 terabytes with 20% of cases exceeding five terabytes.

Five terabytes of data per case – it could take months to go through all that!

Today, of course, with the amount of digital information generated by each individual, caseloads exceeding five terabytes is common. According to Kathryn Seigfried-Spellar, Purdue University computer science professor, “Almost every type of crime – whether it’s homicide, arson, or a computer crime – is going to have some sort of digital evidence associated with it.” Recently, for example, digital forensic investigators working for Target were asked to look into the chain’s shrinkage of Blu-Ray inventory. The investigators went to work pouring over mountains of digital information.

Before long, the investigation centered on a specific suspect. Through analysis of the suspect’s digital footprint, investigators learned where the suspect lived, identified the car he drove, and located where he resold the stolen Blu-Ray inventory. $15,000 worth of missing Blu-Ray inventory was tied to the suspect, and the case was turned over to local law enforcement who charged the suspect with grand larceny.

While it is technically possible for a well-trained digital investigator to pare down the mountain of data collected in a given investigation to just the information which is most likely to prove relevant, that doesn’t help the overworked investigator get through the proverbial haystack any faster.

Many agencies require the digital forensic investigators to analyze all the collected data, as this could lead to more or expanded charges being filed against the suspect. The result, says Seigfried-Spellar, is a significant backlog – which can be as long as years in some cases – of data and devices waiting for digital forensic analysis at police labs.

The amount of data collected and required to be forensically analyzed creates several pain points for public sector, as well as private, organizations. Common challenges include:

- The legal requirement to gather, analyze, and keep terabytes of forensic information produces monstrous data sets that must be stored;
- Colossal storage requirements, on top of burgeoning caseloads for analysis, strain agency and organizational budgets as they try to

maintain hardware, infrastructures, and equipment;

- Growing workloads, with HR budgets that cannot keep pace, are producing significant backlogs of forensic data to investigate;
- As datasets grow more cumbersome, finding the proverbial needle in the haystack – and making critical connections between disparate sets of data, particularly when data is gathered from different sources – becomes increasingly difficult;
- Larger datasets from diverse sources means greater difficulty to collaborate reviews; and
- The inability of forensic hardware to keep up with, and search, new or uncommon data types.

Digital investigators, whether working for federal agencies, local law enforcement, or for-profit organizations like Walmart and American Express, know the drill. They may be called upon to investigate corporate theft, insider cyber threats, or online fraud or terrorism and track the flow of information back to its source. The investigation might require defeating encryption, overcoming passwords or physically damaged equipment, or detailed anlysis of exceptionally large graphics.

The investigator may need to analyze activity across an extensive network and through the cloud to piece together, from disparate sources, which specific employee, for example, has taken sensitive corporate information to sell to the organization’s competitors.

Moreover, equipping and running a digital forensics lab can be expensive, requiring constant capital investment, as hardware-based solutions are antiquated continuously by new technology. For example, extraction of data from mobile devices such as smartphones needs to be re-engineered each time the latest version of the device hits the market with stronger security protocols.

An average of nine new smartphones come out each quarter – that's a lot of technology to keep up with!

Cloud Computing Adds Additional Investigative Challenges

Cloud computing – and the proliferation of IoT devices which send an endless stream of data to the cloud, especially survillence devices – add to the overwhelming workload of digital forensic investigators. Here's how:

- Data that lives in the cloud is no longer isolated to a specific drive or server. It can physically live anywhere in the world, which could raise the issue of privacy laws if the data lives in certain countries;

- Establishing the chain of custody, authenticity, and integrity of information may be difficult or impossible, depending on the cloud provider's available forensic services;

- The cloud platform or the service provider may make it difficult to identify and preserve relevant data subject to a legal hold, to conduct fast and accurate searches of data, or even to ensure a proper approach to data retention has been taken.

It may be noted that maintaining a proper chain of evidence and obtaining forensically sound logs and snapshots are possible when following cloud (AWS / Azure) best practices. However, according to renowned cloud security and cyber risk analyst Dr. Keyun Ruan – who first coined the term “cloud forensics” while working on her Ph.D. –digital forensics in the cloud are far more challenging than just finding where the data lives or working with cloud providers. “There are [unique] challenges with multi-tenant hosting, synchronization problems, and techniques for segregating the data in the logs.”

Moreover, says Martin Novak, a physical scientist at the National Institute of Justice, “Many users will have access to a particular cloud. How can law enforcement seize only that portion of the media where the evidence may exist? How will they know if they have gotten everything that they will need during the analysis, interpretation, documentation and presentation phases?”

A Better Way: Creating Efficiencies in Investigative Workflows

What would happen if your agency or organization were made responsible for an investigation with five or more terabytes of data to analyze and hundreds of gigabytes of digital survillence to process for crucial evidence? That would correlate to about 5.2 million documents to scan, correlate, analyze, and process in a forensically defensible manner.

You could hire an army of analysts – but this may not be a feasible solution for agencies or organizations without unlimited financial resources.

You could take years to sort through the data – but this may not be a realistic solution as most investigations are required to be handled in a “reasonable” time frame.

You could file for early retirement – but that doesn’t solve the problem; it only shifts the responsibility to someone else.

Efficient workflows are the key to success when dealing with terabytes of data per case!

Faced with finite resources, time constraints, and a sense of responsibility, the only way to handle an investigative load of this magnitude is to develop scalable and cost-effective workflow efficiencies.

When dealing with an average case size of 5 terabytes, forensic workflow efficiency is a top priority. However, the additional challenges and constraints within the public and government sectors conspire to make timely, efficient processing of cases a truly Sisyphean task.

Let’s examine the numbers which characterized a recent case handled by a federal criminal investigation agency, for which the explosive growth of digital devices over the past five years resulted in the agency collecting 700% more data from 300% more devices.

Requirements to collaborate with other public sector organizations, and to train their personnel, added to the mounting workload.

By the time the federal agency turned to outside solutions for resolution, their case backlog had grown to nine months and was still expanding at an alarming rate. Under mandate to accelerate the processing of this workload without adding new staff, the agency's burden only grew greater.

Every device deemed "of interest" in an investigation has to be processed, and all the data stored on the device – even data which users attempted to erase – needs to be captured and fixed. Yet digital data is, by nature, volatile, and capturing the data on a device of interest to a case requires creating a forensic image – a bit-for-bit duplication of the data on the device or network.

Traditionally, a remote digital image would be captured with a USB 2.0 write blocker. The average sustained transfer rate of a USB 2.0 device is 32.8 Mbps. At that rate, it would take about 13 ½ hours process a 160 GB image. Nonstop work on creating digital images for the average caseload of 5 TB would consume 17 ½ days – during which time the rest of the investigative process may be at a standstill. Even if the hardware were upgraded to USB 3.0 – which can sustain speeds roughly twice as fast as FireWire 800 – processing the digital images for a caseload would still consume 5 ½ days.

And the challenges don't stop there. Once the data of interest is captured as forensic images, it has to be processed to isolate the evidenciary needles in the proverbial haystack. For an average 5 TB caseload, the required data processing would be equivalent to carefully examining 3.7 million documents, determining which information is relevant to the investigation, and then collating and indexing the relevant data for reference.

Let us, once more, put this in perspective.

If a team of ten investigators manually worked nonstop 24 hours a day to fully process 2 documents an hour (reading and extracting the relevant information, then collating and correlating it), getting through 3.7 million documents would consume *21 years*!

Multiply the challenges of investigating just one case, from imaging through to final processing, by the number of cases for a federal criminal investigation agency is responsible and it is easy to understand why organizations need specialized resources in their digital forensics toolkit.

In this instance, the aforementioned federal criminal investigation agency implemented an enterprise solution to improve their workflow efficiency. The sreamlined workflow allowed them to:

- Reduce their case backlog from nine months to just two weeks;
- Reduce the number of staff dedicated to processing and reviewing evidence by 25%; and
- Reduce the cost of their digital forensic investigations by $500,000 over five years.

Making Hay from Digital Trails

Since investigations of one sort or another will be with us until the end of time, the role of digital forensic investigators will continue to be an increasingly important part of the process.

The incredible amount of data produced, transmitted, processed, and stored in our digitally-driven world isn’t about to decrease. If anything, the amount of information generated will continue to proliferate at an accelerated rate.

This data is generated in our homes as we browse the Internet, place calls, and text friends. Data is generated as we work, whether from home, while traveling, or in office. Even if our job is ditch-digging, we leave a digital footprint every time we clock in or out from the job, get paid, or deposit our paycheck in the bank.

Our digital footprint is increased every time we interact with an IoT-enabled sensor or camera, whether we know it or not. This doesn’t even include the normal data generated by businesses and agencies conducting operations every day.

That’s why specialized third-party vendors have engineered solutions which deliver a scalable, resource-effective way to create significant efficiencies in investigative workflows; because sizing up, breaking down, diving into, and overcoming mountains of data is a task no one should do unassisted.

This is excellent news for those persons responsible for handling the cases under investigation; cases can be investigated in a timely, forensically defensible way without costing an arm and a leg. It’s also great news for those doing the leg work on the investigation, now able to focus more on finding their suspect and less on finding a reasonable way to digest piles of data.

By leveraging specialized collection, preservation, processing, and data assessment solutions, investigators’ days can be much more fruitful and efficient, so they can go home happier at night.

NOTES

NOTES

The data available to modern investigators may be a threat to criminals; communications, surveillance, transactions, and more comprise a backlog that can take years to go through. This book explores the processes, standards, and tools necessary for today's investigators to accelerate their search and discover more-impactful intel.

About Derek Smith

With over 30 years in the security industry, Derek A. Smith is a former government agent, cybersecurity SME, holds a variety of certifications (CISSP, CEH, C/CISO, Security+, etc.), eight college degrees, is a published author, conference speaker, cybersecurity analyst for several international and local television news stations, government program manager, and more. Follow him on Twitter @DerekASmith1