Effective Digital Forensics Research is Investigator-Centric

Robert J. Walls    Brian Neil Levine    Marc Liberatore    Clay Shields†
Dept. of Computer Science, University of Amherst, MA
†Dept. of Computer Science, Georgetown University, Washington, D.C.
{rjwalls, brian, liberato}@cs.umass.edu    clay@cs.georgetown.edu

Abstract

Many technical mechanisms across computer security for attribution, identification, and classification are neither sufficient nor necessary for forensically valid digital investigations; yet they are often claimed as useful or necessary. Similarly, when forensic research is evaluated using the viewpoints held by computer security venues, the challenges, constraints, and usefulness of the work is often misjudged. In this paper, we point out many key aspects of digital forensics with the goal of ensuring that research seeking to advance the discipline will have the highest possible adoption rate by practitioners. We enumerate general legal and practical constraints placed on forensic investigators that set the field apart. We point out the assumptions, often limited or incorrect, made about forensics in past work, and discuss how these assumptions limit the impact of contributions.

1 Introduction

Digital forensics is the application of science to lawful investigation. It is a field strongly driven by practitioners who can readily adapt cutting-edge research. Consequently, researchers have an enormous opportunity for impact by transforming novel research results into techniques that can be used by investigators. The need is great: the National Academy of Sciences (NAS) recently published a scathing report calling for a scientific overhaul of digital forensics [33]. Further, prevalent forensic techniques do not scale and already the demand for forensic examination is much greater than current capacity [15]. Our group has taken up this call, and we work directly with many law enforcement organizations.

Unfortunately, experiences like ours in deploying well-used tools based on novel forensic research are rare. More typically, computer security research aimed towards forensic applications has little or no impact — often because the researchers are poorly acquainted with the real-world problems faced by forensic investigators and the constraints placed on solving them. Similarly, when forensic research is evaluated using the viewpoints held by computer security venues, the challenges, constraints, and usefulness of the work is often misjudged.

Adding to the confusion, the technical overlap between security and forensics can falsely color one’s view of the latter. For example, packet attribution techniques proposed by security researchers can be useful for determining the source IP address for network-level attacks, but as Clark and Landau [9] point out, such mechanisms are “neither as useful nor as necessary as it would appear” for investigations that require identification of a person rather than a machine. We generalize that statement further: many technical mechanisms across computer security for attribution, identification, and classification are neither sufficient nor necessary for forensically valid digital investigations. Developing a security mechanism for, say, remote identification of a device, and claiming it works for forensics is akin to developing a new cryptographic hash function and claiming it can be applied to many security problems: the claim is easy to make, but the impact is negligible.

In general, digital forensics is concerned with techniques (1) that support or refute a hypothesis that explains a person’s violation of law or organizational policy; (2) such that the investigator is limited by a defined set of procedural restrictions for gathering evidence; (3) where the value of evidence is defined by a qualitative context and not only quantitative measure; (4) and where the error rates and procedures of techniques are known and testable.

Contributions. We define many key aspects of digital forensics with the goal of ensuring that papers seeking to advance the discipline will have the highest possible impact on investigators. Our observations are based largely on our experience working directly with practitioners [26,42] and advancing work [28,44] that operates
within forensic constraints that we detail within this paper. While we focus primarily on investigations in the context of the U.S. legal system, our conclusions are applicable to most other forensic contexts. At a high level, these constraints are as follows.

- Digital forensics is investigator-centric, and unless developed with an understanding of the restrictions that investigators are under, most novel results cannot and will not be adopted. For example, prior to the issuance of a warrant, techniques for criminal investigators cannot violate plain view observation, and evidence gathered otherwise would be suppressed in court.
- The value of a new technique depends in part on its complexity and therefore it must be judged against simpler options available to investigators. Similarly, defenses against investigation should not be evaluated with an assumption that high precision is always needed. For example, civil investigations are often based on simple subpoena and mere demonstration of relevance; sophisticated investigative techniques may be needlessly restrictive or indirect compared to capabilities and information available after subpoena.
- Forensic techniques are most valuable when addressing the most common adversary, not the strongest; there is no correlation between technical savvy and dangerousness to society. It is not possible for one savvy criminal to destroy, hide, or obfuscate the evidence of everyone else. In contrast, security work must consider that one person can leverage a vulnerability to attack every computer that uses the flawed system.
- Finally, forensic investigations seek to find the person responsible rather than stopping at a machine or line of code. Consequently, the scope of forensics is often more broad than that of the traditional security domain. For example, most any policy or law requires consideration of a person’s intent, something that is often demonstrated indirectly through an amalgamation of facts and evidence.

Our position is that security venues should publish forensics research, but these works should be evaluated in the proper context. When selecting reviewers, ensure they can examine papers using the goals and principles of digital forensics and not just those of computer security. The reviewer himself should question if the authors are actually looking outside of the computer security problem when claiming an approach is applicable to forensics. Otherwise, the security community risks encouraging low-impact work while rejecting worthwhile solutions to forensic problems.

In Section 2 we begin by enumerating general legal and practical constraints placed on forensic investigators in the context of the U.S. legal system. We then move to briefly describing the wider scope of investigations forensics examiners face. In Section 3, we review a set of useful lessons for researchers regarding the applicability of techniques to digital forensics, and highlight recent work in this context.

2 Forensic Investigations

The highest impact work in forensics works within, and is evaluated under, the real constraints and goals of investigations. In this section, we detail general models for criminal and civil scenarios, and we describe how investigations are focused on people rather than systems.

In both civil and criminal contexts, digital forensics is concerned with techniques that address the four points stated in the introduction: (1) hypothesis testing, (2) procedural constraints, (3) evidentiary value, and (4) error rates. Practice provides another set of constraints. For example, all methods rely on acquisition of evidence, yet data is commonly destroyed, lost, stolen, or encrypted. The capacity of investigators to take on new cases is limited, and selection is based on many external factors. Even if a case is accepted, there is often too much data to image, store, and process, and an imperfect triage process is frequently necessary [15,31].

2.1 Criminal Investigations

Criminal investigations take place along two phases, a pre-warrant (plain view) phase, occurring prior to the issuance of a search warrant, and a post-warrant phase.

Pre-warrant phase. In this plain view phase, investigators are limited by U.S. law, stemming largely from the Fourth Amendment, that dictates what can acquired before a warrant is in place. The main goal of this phase is to meet the probable cause standard for obtaining a magistrate-issued warrant. This standard is is a qualitative measure often defined as meeting a “fair probability” that further evidence will be found in the location to be searched; see U.S. v. Sokolow, 490 U.S. 1 (1989). Probable cause does not require that the evidence is strong enough for conviction, merely that the evidence support a reasonable belief that the suspect committed a crime. The caveat is that the collected evidence must be in plain view and therefore not violate a person’s expectation of privacy. During this phase, general criminal activity is monitored, and then specific targets are selected among

A person’s expectation of privacy is established using the two-pronged Katz test. The first prong asks whether the person subjectively demonstrated an expectation of privacy, and the second prong asks if that expectation is objectively reasonable from the standpoint of society; see Katz v. U.S., 389 U.S. 347 (1967).
available suspects. The choice of target is a balance between dangerousness to society and efficiency of case execution.

Post-warrant phase. In this phase, investigators are limited only by a court-issued warrant. Warrants require particularity, which limits the place to be searched; for example, its unlikely to get a warrant for all apartments in a building, nor does having a warrant for one allow investigators to enter another. Warrants also require specificity which defines the type of item to be found. In digital contexts, its hard to violate specificity since any computing device is typically a viable target.

After obtaining a warrant, there are few legal restrictions on law enforcement’s technical approach. However, in order to obtain a conviction in court, the investigator must collect a higher standard of evidence. This evidence must help prove beyond a reasonable doubt that a suspect committed a crime.

Peer review of digital forensics research must include an analysis of the legal justification required to employ proposed techniques. Techniques that don’t require special privileges are applicable to the broadest number of settings and therefore are the most desirable. But such papers should provide the justification that a warrant or other restrictions, such as Kyllo v. U.S., do not apply; see our discussion of Admissibility below. If post-warrant capabilities are indeed required, the paper should detail why easier technical solutions aren’t available to obtain the same results. For example, it’s likely that watermarking the outgoing traffic of a user requires a wiretap; at that point, the criminal investigator will find it just as easy to get a warrant to install a covert key-logger or other device on the target’s computer. We return to this point in Section 3.

Exceptions to the Fourth Amendment. The law underlying realistic investigative models shifts frequently. For example, courts have long held that there are exceptions to the Fourth Amendment’s warrant requirement, but these exceptions are not always clear in the case of digital evidence. One exception to the warrant requirement is based on consent by a person to proceed without a warrant. Often the limit of that consent can be unclear, especially when the objects to be searched are digital devices. Recent events involving the ACLU and Michigan State Police [17] motivate the following example.

Imagine a person is stopped by police for a traffic violation and the person consents to an examination of their phone and browse through the contents of the address book or call log? Does it also extend as far as to allow the officer to use a special tool to extract and save all of the phone’s information, including deleted data? The answers to these questions depend highly on the specific circumstances and the courts are inconsistent on this issue [22].

Search at the U.S. customs border is another exception to the warrant requirement. This recently impacted the security community with several high profile border searches of Bradley Manning associates [16]; see also U.S. v. Howard Cotterman, 09-10139 (2011).

There are investigators and models that have implicit exceptions to the Fourth Amendment. For example, investigators working under the U.S. national security (FISA) rules may defer a request for a warrant until after a search, and there are analogous positions in other governments. Similarly, rogue investigators can elect to not follow any restrictions, risking that collected evidence is thrown out during trial. We assert that developing new techniques to function under such models is largely wasted effort due to their limited applicability in a civil society.

Admissibility and Validity. The exclusionary rule, set forth in Weeks v. U.S., 232 U.S. 383 (1914), in concert with the fruit of the poisonous tree doctrine, set forth in Silverthorne Lumber Co. v. U.S., 251 U.S. 385 (1920), dictates that illegally obtained evidence cannot be used in court, nor can any evidence further found from this illegal evidence be used. These rules ensure that techniques not valid within the plain view rules will not be commonly used by law enforcement in the U.S.

Further, we note that in the pre-warrant stage, law enforcement must take care in acquiring evidence from third parties. In addition to the problems of hearsay, even well-meaning third parties cannot repeatedly gather information for law enforcement. In doing so, the third party is acting under the color of law, and any evidence they collect is governed by the same rules as apply to law enforcement, including the need for warrants.

Another relevant ruling for computer scientists is Kyllo v. U.S. 533 U.S. 27 (2001), where the court ruled that using a technology that is not in “general public use” to gather evidence pre-warrant is a violation of a person’s expectation of privacy. This exact phrasing is important: source code available publicly on a researcher’s web site is not general public use. With regard to digital forensics, this has been interpreted by investigators to mean that tools can only use information provided by normal operation of the system being investigated. Recent cases have supported this view, including U.S. v. Borowy, 595 F.3d 1045 (9th Cir. 2010) and U.S. v. Gabel, 2010 WL 3927697, but the exact extent to which can investigators can exploit a network protocol to gather information remotely is unsettled law.

In order for a forensic investigator’s testimony to be
where ownership or contractual agreements (e.g., files on we elide much civil case law in this section. The general within the scientific community.

asked to gather evidence about how and why copied files can directly determine intent, but an investigator may be release it. Bradley Manning’s alleged theft of documents to copy a file for work use, but not to sell or otherwise determine directly. For example, a user may be allowed to copy a file for work use, but not to sell or otherwise release it. Bradley Manning’s alleged theft of documents falls in this category. No existing security mechanism can directly determine intent, but an investigator may be asked to gather evidence about how and why copied files were used. Similarly, possessing photos of children likely does not violate a security policy, but having pictures of children being sexually exploited is illegal. Systems that can generalizably differentiate between the two do not exist. Furthermore, possession requires a demonstration of knowing intent: unexamined images that are unknowingly and unintentional stored in a user’s spam folder are not illegal. Howard [20] provides a cogent discussion of indirect evidence of knowing possession. Researchers must be aware of the focus on people, not just computers. Systems that answer questions about user behavior can be beneficial to forensics investigators, even if no computer security problem is being addressed.

2.2 Civil Investigators

Our focus is on criminal forensics, and due to space limits we elide much civil case law in this section. The general rule is that there are somewhat fewer constraints and that the standard of evidence is lower, as the government is typically not using state force to prove guilt for a crime; rather, citizens seek the power of the court to discover evidence of a violation of civil law or a contract.

In civil cases, examiners have complete access to their own data, including network traffic and logs, and any files where ownership or contractual agreements (e.g., files on an employee’s work machine) permit access. However, examiners will need to subpoena information they are refused access to by other parties. Courts have a very low bar of relevance for subpoenaed access (see Federal Rule of Evidence 401), but this process can be adversarial. Courts may place limits on evidence acquisition if one party is concerned about legal liability from exposure of unrelated, proprietary information. Knowing this, many organizations choose not to keep data that can be subpoenaed, regardless of how useful it might be to their business.

2.3 Investigating People as a Goal

Computer security is centrally concerned with the enforcement or defeat of technically based and clearly delineated computer security policies [4]. Response to computer security failures is typically designed to identify the cause of the failure, which might provide clues to the identity of an attacker. However, intrusion response, which is familiar to security investigators, is but a small subset of digital forensics.

Forensic investigators instead often investigate suspected breaches of organizational policies or laws not reflected in any computer security policy. Indeed, such rules may be impossible to implement in a computer system. A user’s intent is often relevant but impossible to determine directly. For example, a user may be allowed to copy a file for work use, but not to sell or otherwise release it. Bradley Manning’s alleged theft of documents falls in this category. No existing security mechanism can directly determine intent, but an investigator may be asked to gather evidence about how and why copied files

2.4 Applicability and Impact

Many proposed forensic techniques are easily thwarted with only limited technical knowledge, but that doesn’t lessen their practical effectiveness. While security mechanisms have impact because they can address the worst case, forensic mechanisms have impact because they can address the common case. For example, the most realistic forensic model allows for any individual to erase information from storage; in this scenario, why should we expect new techniques to work at all? Surely criminals will seek to cover their tracks.

For example, investigators commonly identify images of child pornography on p2p networks by hash value. Criminals could easily change just one bit in shared images to escape detection; yet millions do not [26] for several reasons. First, they may not think that they need to make changes to evade investigators, or they may lack the skill to do so. Second, they may not think they’ll be caught, and percentage-wise that is largely true. Third, people are interested in sharing content, and in order to do so they give files very descriptive names including “illegal child pornography” ; at that point there is no legal reason to flip a bit.

Given that anyone could flip a bit but that millions do not, it is high impact to develop techniques that succeed in the common, rather than worst case. This fact is anathema to computer security researchers, even though, analogously, security systems with known flaws remain useful.

For example, despite power monitoring attacks [13, 23,34], most people do not use tamper-proof hardware. Various forms of the Sybil attack [12] succeed and are used against Google [1,3], EBay [8], and p2p file sharing networks [30,36] yet these systems enjoy great success. The Tor privacy network is architected to provide reasonable performance instead of perfect security against known attacks [11]. Similarly, the banking industry finds it more effective to allow “bad guys to take a cut” [5] than attempting to deploy a system where all attackers
are defeated. Further, the TSA admits it cannot defeat all terrorists, and instead simply mitigates risks [19].

Finally, we note that law enforcement are interested in catching the most dangerous people. Similarly, civil investigators are interested in catching the people that have caused the largest damages. In both cases, there is no evidence that such dangerousness is necessarily correlated with technical savvy. Furthermore, it is not possible for one savvy criminal to destroy, hide, or obfuscate the evidence of everyone else; in contrast, security work must consider that one person can leverage a vulnerability to attack every computer that uses the system.

3 Lessons to Learn

In this section, we review past security papers from a forensics viewpoint. Some explicitly invoke the concept of forensics; some do so implicitly. We point out the assumptions, often limited or incorrect, made about forensics in these papers, and discuss how these assumptions limit the impact of the contributions. Our goal is not to denigrate others’ contributions, but instead to show if and how these contributions fit within the forensic framework we’ve discussed in previous sections.

We separate our survey into several broad classes, corresponding to lessons learned about forensic practice in the previous sections: 1) Investigations are about people and their activities; 2) Forensic investigators are not all-powerful, and while the legal system can grant impressive powers, they are constrained in many ways; 3) A proposed system that depends on access to data across organizational boundaries may fail — as this access is not often permitted; 4) Proposed systems that expand the view of investigators can be useful, but simply expanding the amount of data collected by itself generally is not.

3.1 Problem Exists Between Keyboard and Chair

Forensics investigations of individual’s computers arise because of user’s actions. Investigators are therefore most interested in people and how they used the system. Security researchers, however, tend to focus on technical aspects of security system failures. Solutions that add additional information about system events [10,25,35] tend not to benefit investigators directly. Instead, researchers can have impact by focusing on mechanisms that support common investigations types, such as theft of intellectual property, violations of organizational misuse policies, and embezzlement [41]. These are common problems, and solutions will have high impact. Work in this area needs to consider the types of users who commit these acts. Most have weak computer skills, so approaches that might be trivially thwarted when used against security experts can be very effective, as we outline in Section 2.4. To paraphrase a police maxim, it is useful to catch the stupid ones.

3.2 Lines in the Sand

There are many clear lines lawful investigators cannot cross. Wiretapping and analogous inspection of traffic is not permitted without a warrant, nor, presumably are manipulations of Internet traffic beyond normal participation in protocols. Actively watermarking packets, by manipulating their contents or timing [14,29,45,46,48] to defeat anonymity systems crosses this line, and would poison the evidence acquired by a law enforcement. For civil cases, organizations might choose to use these techniques on their internal networks, but are unlikely to provide access, mark traffic, or help recover timing information for external parties. Further, many such techniques work best to confirm a suspicion, as they require both manipulation of traffic at a source, and the observation of traffic at a suspected endpoint. Evidence required for this level of suspicion may rise to the level that would lead to a warrant.

But if an investigator had sufficient evidence for a warrant, a more direct search would likely be preferable — if not, an electronic bug in the computer’s audio system almost certainly would. As evidence ultimately supports a court case, recordings of a suspect or copies of emails would be far more useful than a watermark. Systems that assume technically sophisticated attacks, such as inference based on clock skew, temperature or power consumption changes, and acoustic or electromagnetic measurements, or packet size [2,24,27,32,43] are similarly useless to a criminal investigator, due to Kyllo, and obsolete after court order or warrant.

On the flip side, many systems are built to withstand cursory investigation, such as disk encryption systems. It is an unsettled question of law as to whether cryptographic keys are a more analogous to a physical key, which an individual can be compelled to produce in both civil and criminal courts, or to testimony, which in the criminal context is protected by the Fifth Amendment. Failing that, systems focused on key recovery can be forensically valid [18].

Some systems attempt to sidestep this question by providing users with plausible deniability (for example, Rubehose and its successor TrueCrypt). These systems may be valid responses to overly intrusive governments. But they are inappropriate in more common use cases such as secure corporate record keeping, where allegations of wrongdoing will likely require court-ordered key revelation or else result in being found in contempt of court.
3.3 No Keys to the City

Investigators working in different organizations or with different goals may not be willing or able to collaborate with one another. Sharing of data across institutional boundaries is not always feasible or even legal. Wide-scale intrusion or anomaly detection, de-anonymization, or flow attribution systems [7,38–40,47] that require a network-wide view or deep packet inspection are akin to a massive surveillance campaign. Monitoring of this breadth could never be lawful for law enforcement without a court order — and even then, is likely too broad in scope. These systems are still useful, outside of the criminal investigation context: ISPs or private organizations could collect this data and use it internally to improve security and performance. They cannot collect it at police request, however, as that would turn these organizations into de facto agents of the law. In many cases, it would also be unlikely that the information would be willingly shared between organizations in civil cases. This is particularly true when it might expose the organization to legal jeopardy, such as showing that it acted as a gateway or stepping stone for attacks.

3.4 Don’t Grow the Haystack; But Do Find More Needles

A system that increases the amount of information available to investigators is often a double-edged sword. The investigator may benefit from additional information, but that benefit is directly proportional to the information’s quality. For example, Bratus et al. [6] and Piatek et al. [37] highlight how current practices in DMCA copyright enforcement focus on broad and highly automated techniques resulting in unacceptably high rates of false positives. This problem isn’t just limited to p2p investigations. Clark and Landau [9] also question the utility of packet attribution in forensics. We believe their criticisms extend to other areas such as de-anonymization. Poor information may lead to tangible costs such as wasted resources or intangible costs such as emotional distress for the falsely accused.

4 Conclusions

Computer security researchers have the potential to make significant contributions to digital forensics; however, they must first understand the forensics context and its differences with existing security models. Similarly, the onus is on computer security venues to support these efforts by recruiting knowledgeable reviewers who are familiar with the challenges and requirements of forensics; otherwise, the security community risks encouraging low-impact work while rejecting worthwhile solutions to forensic problems.

Acknowledgements. We are grateful for insightful comments from Emery Berger. This work was supported in part by NSF awards CNS-1018615, CNS-0905349, and DUE-0830876, and in part by NIJ award 2008-CE-K005.

References

As the primary aim of any digital forensics investigation, is to allow others to follow the same procedures and steps and still end with same result and conclusions, considerable effort must be spent on developing policies and standard operating procedures (SOP) in how to deal with each step and phase of the investigation. 2. Evidence Assessment. Principle. All sources of possible digital evidence should be thoroughly assessed with respect to the scope of the case. This will help establish the size of the investigation and determine the next steps. Special attention should be given to reviewin Digital Forensics investigation is one of the leading disciplines developing from the extensive field of forensic science. Digital forensics investigations may also be applied in the corporate sector, including during computer hacking investigations or internal corporate investigations. Here, the digital forensics analysts investigate the environment and degree of an unlawful network intrusion or system hack. The rapidly expanding field of digital forensics includes numerous branches related to databases, malware, firewalls, mobile devices, cloud, and network forensics. To find out more about digital forensics, read our in-depth article on what digital forensics is. What is the difference between Digital Forensics... constraints are as follows. â€œ Digital forensics is investigator-centric, and unless, developed with an understanding of the restrictions. that investigators are under, most novel results can-. not and will not be adopted. For example, prior to the.Â committed a crime. Peer review of digital forensics research must include. an analysis of the legal justï¬–cation required to employ. proposed techniques. Digital forensics or digital forensic science is a branch of forensic science focused on the recovery and investigation of material found in digital devices and cybercrimes. Digital forensics was originally used as a synonym for computer forensics but has expanded to cover the investigation of all devices that store digital data. As society increases reliance on computer systems and cloud computing, digital forensics becomes a crucial aspect of law enforcement agencies and businesses. Digital forensics is concerned with the identification, preservation, examination and analysis of digital evid Similarly, we can investigate the recent activity of any particular drive. We can also change the configurations or apply/remove any filters as per the requirement but these changes are to be done before starting the scan. To edit the configurations click on â€œConfigâ€ button located at the top right corner on recent activity window. Check/Uncheck the options as required or if required change the date/date range for a particular time-based activity and click OK.Â Deleted files recovery is one of the prime requirements for digital forensics. OSF offers a very simple and efficient deleted file recovery/search. To search the deleted files click on â€œDeleted files Searchâ€ and select the drive we want to search on from the drop-down.