Current Privacy Concerns with Digital Forensics

Neil C. Rowe
US Naval Postgraduate School, United States

ABSTRACT

Digital forensics is a rapidly advancing technology for examining the contents of computers and digital devices. It raises many challenges to conventional notions of privacy because it involves more detailed search of digital data than is possible with other techniques, and it can be done surreptitiously. However, there are analogies to homes and the rights of individuals to be free from unwarranted searches and seizures in their private spaces. Even though commercial software and data comprises most of digital space, there are clearly enclaves of data that should be kept private. We discuss the techniques of digital forensics and investigative targets. We identify key challenges to privacy, and outline both the legal protections and the technical protections available. Unfortunately, privacy laws are ineffective in most countries, and users need to supplement them with their own measures to protect themselves.

Keywords: Investigation, Cybercrime, Searching, Media Acquisition, Metadata, Encryption, Antiforensics, Secondary Storage, General Data Protection Regulation, Steganography, Wiping, Metadata, Network, Cloud

INTRODUCTION

Digital forensics analyzes the files and memory of computers and digital devices (Jones, Bejtlich, & Rose, 2006). It is primarily focused on the secondary storage such as magnetic disks and flash drives (containing information that remains after power is turned off), but can also examine more volatile information such as main memory and network data. Its main uses are to obtain evidence in legal proceedings both criminal (Greengard, 2012) and civil (Matthews, 2010), intelligence gathering about businesses or governments, routine monitoring of systems, and “malware” analysis when malicious software has infected a machine or device. Important legal uses establish whereabouts of suspects, prove or disprove alibis, document criminal conspiracies, find child pornography, prove financial fraud, and prove piracy of copyrighted material. Digital forensics can be done under court or administrative order, when computers and devices are brought in for repair, maintenance, or discard, or even by malicious software that gains a foothold on a computer.

This chapter will focus on privacy of personal information from government, business, or other organizations. Digital forensics raises privacy concerns because it permits very detailed and thorough examinations of digital data. It can retrieve any digital artifacts left behind on a computer or device including deleted data and fragments of data, not just those retrievable by the operating system or software, since it bypasses the usual access controls on computers. This means that data that users thought was hidden or long gone could be found in a forensic investigation, more so than with Web postings and electronic mail. Unscrupulous forensic investigators could sell discovered personal data to businesses for personal gain; they could get information needed for bullying or stalking; or they could seek embarrassing information for harassment of blackmail. Furthermore, digital forensics can find private information more quickly than navigating the Internet could because it investigates more concentrated sources of data.

Even when obtained responsibly, digital-forensic data can easily be misjudged. That is because there is so much of it, interpreting it is highly technical, and context is often missing unlike with, say, police searches of houses. This means that evidence is more likely to be wrongly judged suspicious by digital-forensic than by conventional investigation, leading to unfair legal consequences. Furthermore, the increasing use of server machines holding data of many people makes it difficult to confine a forensic investigation to one person when a search warrant is issued. In addition, digital-forensics searches are invisible unlike a physical search of a house, so privacy may be violated without the victim being aware of it. Cybercriminals and spies can secretly break into large organizations to efficiently steal private information of many individuals simultaneously using digital-forensic techniques, as in recent "data breaches" at Yahoo, Equifax, eBay, Heartland Payment Systems, and Target affecting hundreds of millions of people each (Trend Micro, 2018).

Clearly we need more legal protections, since current laws are not keeping up with the technology (National Research Council, 2007). The European Union implemented in 2018 a broad approach in the General Data Protection Regulation that is starting to have a significant effect on practices involving private data (Tiku, 2018). The U.S., on the other hand, has a competing set of privacy laws whose applicability is often unclear (Rodriguez, 2014). It will help to enumerate some privacy principles for cyberspace to which most people can agree and which can serve as a basis for laws (Bernal, 2014). In the meantime, however, there are technical measures that individuals can take to effectively protect their privacy such as encryption, obfuscation, and systematic erasure.

In this chapter, we will first explain the techniques of digital forensics and its targets. Then we will enumerate the key privacy challenges, the legal solutions to these, and the technical solutions. We conclude with a discussion of future directions.

THE TECHNIQUES OF DIGITAL FORENSICS

Digital forensics is a technical discipline whose main stages are media acquisition, media searching, and aggregation of results. Additional methods are necessary for forensics on main memories, servers, and networks.

Media Acquisition

Raw material for digital forensics is called "media" in the specialized sense of storage media. It may be acquired by inspection of hardware, retrieving from Internet sites, or observing network traffic. An increasing portion of the world's activity takes place in cyberspace, so digital forensics has much potential raw material. Usually acquisition is focused on the secondary storage ("drives") of a computer or device. For computers this is usually a magnetic disk drive, and for devices it is usually a large flash memory. Drives can be physically removed from their computer or device, and then usual access controls enforced by the operating systems can be bypassed. Data can also be retrieved by special network protocols or by logging into computers or devices as an administrator. Acquisition is more difficult for smaller devices such as sensors or those of the "Internet of Things" (Plachkinova, Vo, & Alluhaidan, 2016).

It is important for criminal investigations to prove that the evidence has not been tampered with. So most media acquisition copies data so the copy can be inspected repeatedly without risking modifications to the original. Copying the entire drive is important since it is often difficult to predict where useful data will be found, and it may be in fragments that must be pieced together. Tampering possibilities are further reduced by using media acquisition hardware and software with "write blockers" that prevent accidental modification of most (but not all) data on a drive during access. Beyond this, additional rules about conduct of forensic investigations can be enforced in court cases (Manes and Downing, 2010).

Media acquisition can also be done remotely across a network. It can be done with “remote desktop” software that permits logging in to a remote computer or device, mainly for use by system administrators of networks. It can also be done with more restricted programs that retrieve only files, or more powerful programs much like malware (Abel, 2009; Elisan, 2012). Remote acquisition is common when police departments extract data from cell phones, for instance. However, the operating system of the target computer or device controls remote access, and ease of access can vary considerably depending on its configuration. Typically, remote acquisition provides a “logical image” of what the operating system thinks is there and will not extract deleted files.

Media Searching

The next phase of digital forensics involves search within both data and metadata for useful information. Metadata is data describing other data, and includes names of files, creation and modification times, and sizes. Metadata is often considerably smaller than data, so examining it first often saves time in finding useful data. Usually known operating-system and application files can be excluded. A benchmark reference is the National Software Reference Library from the U.S. National Institute of Standards and Technology, which identifies the files of most standard software. Then the remaining files are a richer source of data for most forensic investigations.

Then an investigator can use programs to search the contents of the media for keywords and structured strings that are relevant in an investigation. Personal names are common keywords, and structured strings such as email addresses and phone numbers are also often sought, as many investigations focus on connections between people. Databases can be particularly rich sources of personal information (Stahlberg, Miklau, & Levine, 2007).

Deleted files and file fragments may also be retrievable from drives by forensic techniques. Upon deleting a file, most operating systems merely mark it as such in the file directory and do not erase its bits. Similarly, deleted material in some documents such as those of Rich Text Format (RTF) and Adobe Portable Document Format (PDF) can remain in the document (Catiglione, De Santis, & Soriente, 2010). Parts of deleted files may be reused for other files, but fragments that remain may contain useful data. File fragments may also occur outside the file system in parts of storage media that have been marked as faulty, either because of hardware failures or deliberately by users wishing to conceal data. File fragments can be reassembled by techniques known as “file carving” if there are enough of them, and this is aided if there were multiple copies of a file. Deliberate deletion is a statement by a user that they want to conceal data, and concealed data can be important in investigations.

Results Aggregation

Once forensic data has been extracted, subsequent analysis includes comparing it and aggregating it into summaries. For instance, a criminal investigation could involve collecting electronic mail that shows a pattern of malfeasance. Timeline analysis to suggest causes and effects is essential part of this aggregation. Cross-drive forensics compares data between different drives to see patterns, and it is increasingly important with the multitude of devices that people use. Visualization techniques in the form of charts and tables are important in data aggregation.

Main-Memory Forensics

If a computer or device is accessed from a keyboard or a network while it is running, its main (volatile) memory can also be investigated. A number of tools can convert main-memory contents into a file that then can be searched. This may catch data that has been concealed in secondary storage by encryption since it must be decrypted in main memory to be used. Main-memory forensics is essential in analyzing today’s sophisticated malware by running it to see what it does.

Server Forensics

Computers called servers provide centralized network resources such as Web pages, email, and backups. When implemented as distributed systems, they are often called “cloud architectures”. Servers can be investigated with many of the same techniques as for traditional computers. However, they usually include data from multiple users. Limiting searches to a single user at a time is usually important for compliance with privacy laws and search warrants, and specialized tools for server forensics can enforce such constraints. Forensics on server usually is done remotely, so it produces a logical image of the server and not a bit-for-bit copy (physical image).

Forensics on cloud data has additional problems (Bagby, 2013). Many cloud servers move data between machines without the user’s knowledge to optimize access, so law enforcement may be forced to make a long list of machines to subpoena. Cloud servers may contain the data of many users, and a careless investigation may violate the privacy of individuals not its targets (Jahankhani & Hosseinian-Far, 2017). However, copies of cloud data may often be found in single-user machines and devices too (Koppen et al, 2012) and that may be a better place to start a search.

Network Forensics

Network packets of data can also be subjected to forensic analysis, but there are many challenges. Usually this includes a considerable amount of uninteresting communications overhead and packets do not have helpful names, so considerable searching is required. Usually there is insufficient room to store more than a few hours’ worth of data on a busy organization’s servers, so analysis needs to be done quickly after the data is generated if it is to be done at all. Packet data is usually sent in fragments of a limited size that an investigator would need to reassemble, and fragments are often intermixed between users. Packet data may also need decoding or decryption to be read because of data-security measures. Finally, addresses in packet data may not be reliable because it is quite easy to fake (“spoof”) them or use uninteresting “front” sites for them to conceal one’s identity. For instance, the evidence presented to the public so far that links North Korea to cyberattacks on Sony Corporation in 2014 is weak because it could have been spoofed (Fox-Brewster, 2015).

Peer-to-peer file-sharing networks are important targets of digital forensics since they are convenient for illegally obtaining copyrighted material such as music and movies. It may be difficult to find who is originally responsible for posting unauthorized content on such a network, but software for the networks can modified to make this easier (Myneedu and Guan, 2012).

FORENSIC TARGETS

Whatever its methods, digital forensics generally focuses on a small set of files (Akhgar, Staniforth, & Bosco, 2014). Files created by humans are important including documents, mail messages, photographs, video, Web downloads, Web-page caches, and spreadsheets (Cohen, 2007). This is often not many files; the average Windows-system drive in our test corpus, the Real Drive Corpus, had 141,000 files, of which 6.3% were text documents and 4.5% were audio and video, and many of the text, audio, and video were uninteresting software-support files. Files of interest can often be defined by file extensions (the file-type identifiers) and directories (the containing folders). Nonetheless, digital-forensic investigations can still be time-consuming (Pearson, 2010).

A consequence is that law enforcement prefers to collect easier types of data, data in obvious places that is easy to find and classify. For instance, pornography tends to be sought more often than messages of criminal conspiracies that require reading files carefully. Unfortunately, ease of access is rarely consistent with mandated priorities of police work. Between 1994 and 2006, the number of U.S. Federal cases involving sexual abuse increased 1% while the number of cases involving child pornography increased 82% (Motivans and Kyckelhahn, 2007). Most sexual-abuse cases involve children. So it does not appear that significantly increased investigation of child pornography uncovers an increasing amount of sexual abuse.

Broadly defined investigations such as anti-terrorism (Shipler, 2011) create further challenges for forensics. When terrorist suspects are identified, everything they do and everyone they contact could merit some degree of scrutiny. But many of these contacts are likely to be innocent people, as terrorists need to buy groceries too. Although there have been no terrorist attacks in the United States since 2001, terrorism investigations consume a large amount of U.S. resources in law enforcement and intelligence gathering and the likelihood of concomitant abuses is high. Some U.S. mechanisms for protecting privacy in terrorism monitoring are in place (National Academy of Sciences, 2008), but they are not very effective as discussed elsewhere in this book.

Forensics in intelligence gathering differs in important ways from forensics in criminal investigations (Burd, Jones, & Seazzu, 2011). Finding actionable intelligence is the key, warrants are rarely required, and going to court later is rarely considered. Investigators are usually well-trained professionals who are looking for narrowly focused information that is often not about individuals. Nonetheless, government and corporate intelligence gathering has frequently been criticized on general grounds (Toomey, 2018) and can easily be abused by governments in totalitarian states.

Malware investigations usually focus on key executables of a system and detailed examination of their contents. However, malware can lie concealed anywhere, so a thorough investigation needs to at least check the rest of a computer or device including user documents and downloads. Therefore, malware investigations can risk privacy abuses too.

PRIVACY CONCERNS IN DIGITAL FORENSICS

From the previous discussion, it is clear that digital forensics raises serious issues in regard to privacy:

- **Centralization of Data**: Forensic methods usually see all the digital data on a computer or device. This tempts investigators to violate user privacy since they may find other interesting things not originally sought or authorized during their searches (Hong et al, 2013). There are analogous limits in the United States about what police can search for in situations like traffic stops (Shipler, 2011).
- **Misjudgment of Data**: Forensic investigators often have limited understanding of all the variety of data they encounter and may be overly suspicious of innocent data. For instance, it may be difficult to tell if people in a photograph are at least 18 years of age; it may be difficult to tell if an email-documented meeting between suspects is part of a conspiracy or a social event; or it may be difficult to tell just which financial-transaction documents indicate an alleged Ponzi scheme. Investigators may just open a large number of files because they do not understand what they are looking at, and may see things outside the bounds of their authorization.
- **Unwarranted Reporting of Forensic Findings**: Because of the difficulty of judging data, there is a serious risk that investigators may cause harm by reporting incorrect results. For instance, child abuse is often difficult to ascertain in photographs. However, if a forensic investigator reports suspected child abuse in the U.S., there can be serious consequences such as legal proceedings and loss of access by parents to children, whether or not any child abuse is later proved.

- **Violating Privacy of Third Parties**: Investigation of a shared resource such as a server computer, cloud site, or even a family computer may see data owned by different people. If only one person is subject of the investigation, that could violate the privacy of the others. For instance, it is unjustified for governments to search all the data on cloud servers for terrorism clues, since terrorists are very rare and the benefits of searching are likely to be small compared to the privacy risks.
- **Surreptitious Searches**: A key issue with digital forensics is that the owner of the data may not be aware what is being searched, unlike with the search of a physical object, for which in the U.S. the owner must be present and be served a warrant. A drive can be seized and carried off for investigation; it can be investigated remotely with the right protocols; and even if the owner of the drive is watching an investigation, they may not understand what is going on and cannot tell if their privacy is being violated.
- **Selling of Private Forensic Data**: Since private user data has monetary value, unscrupulous investigators could sell it to the many Internet brokers of user information like the customers of Google. This could considerably broaden the damage of a privacy violation. Governments are unlikely to do this, but businesses and individuals like computer-repair personnel could.
- **Criminal Use of Digital Forensics**: Internet-of-Things systems like home-monitoring systems can be exploited with a little digital forensics to tell thieves when people are present or absent at a location (Plachkinova, Vo, & Alluhaidan, 2016). Unscrupulous investigators could use private data they find for personal gain, such as using bank-card numbers they find to steal from bank accounts, using passwords found to break into systems, or using embarrassing private information they find for blackmail. The Chinese government appears to be doing digital forensics on a large scale to steal technology secrets from U.S. corporate computer systems (Surowiecki, 2014), and these same techniques can be used against individuals.
- **Difficulty of Assessing Damage to Privacy**: It is difficult to assign monetary damages to privacy violations in general, as often the damage is not financial and is difficult to measure. However, class-action lawsuits in the case of large breaches are possible as a deterrent against abuses.
- **Lack of Support for Privacy Management by Forensic Tool Vendors**: The major forensic software tools of SleuthKit, FTK, and EnCase currently provide little support for tracking privacy issues during a forensic investigation.

LEGAL PROTECTIONS AGAINST DIGITAL FORENSICS

Several kinds of laws apply to privacy and digital forensics (Ryan and Shpantzer, 2010), though many are limited in that they only apply after the damage has been done or have conditions that are hard to prove. Formulating laws by citizen consensus can be difficult because citizens do not always accurately assess the dangers of sharing their private information even for important data like medical records (Damschroder et al., 2007), or individuals may differ considerably in when they consider data to be private as with information about their location (Cottrill & Thakuriah, 2015).

Certification of Forensic Professionals

Control over who can do forensic investigation reduces the chances of the worst forensic abuses. In the United States, most states have jurisdiction and many require a private investigator license for digital forensic investigation in criminal cases, something that can be enforced when investigators testify in court. Countries of Europe are similar (Manes and Downing, 2010). Certification methods have been inconsistent, but we are seeing increasing use of standardized tests such as those of the American College of Forensics Examiners Institute. Tests cover legal issues as well as technical issues. On the other hand, there are no official requirements for investigators conducting forensics not related to legal cases.

Commercial Companies in the United States

Laws in the United States are not especially favorable for protecting user privacy from digital forensics by businesses and nongovernmental organizations. A key problem is that privacy of data is not well defined in U.S. law, although respect for physical property is well defined and would seem to provide a precedent (Hoebich, 2008). U.S. privacy law has been termed a "patchwork" of partial solutions (National Research Council, 2007).

This means that businesses and other nongovernment organizations in the United States can freely view, exploit, or sell data generated by citizens whether or not it is related to their business (Greengard, 2010). They can also sell their data to governments, which can be good for governments because they may not have the resources to collect that data themselves. They can also extensively monitor the digital activities of their own personnel for purposes such as quality assurance. However, they do need to inform customers and employees in general terms about how they share information with other businesses and organizations, and they can still be sued in civil court for damages if their use harms someone. The U.S. Government announced in 2008 that it would provide a mechanism for consumers to limit the data collected on them, but this has been blocked from implementation by business interests (Campbell, 2014).

Cookies are a special case of forensic data that is often exploited by businesses, and represent typically data from Web activity. For instance, they can implement "shopping carts" of items selected for purchase along with customer-identification data (Pinsent Masons LLP, 2015). Cookies can hold any private data they choose without telling a user what they are doing, and often the data is encoded or encrypted for security purposes so the user cannot read it. If the cookies are not encrypted, someone other than the creator may be able to read them ("third-party cookies") though better Web browsers prohibit this.

Giving a computer or device to a third party for repair is currently interpreted under U.S. law as giving consent to perform digital forensics on it (Jarrett et al, 2014). This means that if a computer is virus-infested and needs to be cleaned out, service personnel can report any child pornography or terrorist documents they find to law enforcement. They can also report anything that they don't like. However, case law is unclear as to the precise terms of this authority to search (Lonardo et al, 2011), so there will undoubtedly be legal challenges to it in the future.

Businesses and organizations can be subject to data breaches in which criminals break into them using malware. So even though a business has a privacy-protection policy, cybercriminals may break in, steal the information, and make their policy moot. Current protection against malware is not good as is witnessed by the breaches frequently reported by the news media. Most breaches are due to flaws in commercial software that victims are slow to fix, so targets of breaches must share some of the blame for any privacy violations.

Governments in the United States

Citizens worry about abuse of their privacy by governments, but in Western countries there are usually more controls on governments than on businesses and non-governmental organizations. (Jarrett et al, 2014) summarizes the legal issues with digital artifacts in criminal prosecution in the United States, with much attention to forensic artifacts. For law-enforcement operations, there is an important distinction between investigation without and with a search warrant. Investigation without a warrant requires "probable cause" of criminal activity to a standard similar to that for entry into a house for searching it. This means that most computers and devices cannot be searched by the U.S. government without a warrant. With a warrant, there are still limitations since the targets of the search must be indicated in the warrant. For instance, a warrant to search for evidence of drug laundering in financial records does not also permit a search for child pornography. Inconsistent standards for warrants have been apparent between different jurisdictions (Losavio and Keeling, 2014).

There are a number of exceptions to the general protections on privacy of computer and digital devices, however. If the computer or device is not analogous to a "closed container" in a person's home, it can be searched freely. This applies to public terminals, government-owned devices, information given to a third party such as a business, information obtained during a lawful arrest, information in "plain view" or easily visible during a search, and voluntarily revealed information. It is generally interpreted as not including remote shared storage services such as servers, since these are often intended as a backup for secondary storage of privately owned computers and devices which are protected from random searches.

The U.S. government also recently claimed an exception to general privacy protection in regard to metadata, data describing other data, as for example phone numbers called by someone. However, metadata can be sensitive information just like data, and it is unclear if this claimed exception will stand (Schneier, 2015). There are also exceptions in the other direction, to the ability of law enforcement to execute warrants when freedom of expression is involved, in regard to journalism, medical, and legal records. Additional approvals are necessary prior to such searches.

Electronic communications in the form of signals are protected by another set of laws relating to wiretapping. Generally signal metadata is not protected against broad court orders, but access to the contents of the signals is more tightly restricted.

The main problem with digital forensics by governments is that they often have an unbalanced view of what they should be investigating. We mentioned earlier the overemphasis on anti-terrorism by U.S. law enforcement. Another crime extensively investigated by law enforcement with digital forensics in the U.S. is possession of child pornography, with violent crime and financial fraud being less investigated. The evidence is poor that the possession of child pornography is harmful, as opposed to its creation, much as the ubiquity of violent crime on television and in movies does not appear to increase those crimes. So governments may chase headline-grabbing issues and risk violating privacy.

European Privacy Protection

Privacy laws in Europe and other countries have important differences from those of the U.S. (Fink, 2014). The General Data Protection Regulation (GDPR) of the 28 countries of the European Union (European Union, 2016; Eugdpr.org, 2019) which became legal on May 25, 2018 goes well beyond current U.S. law in applying to all organizations having data on EU citizens. The GDPR focuses on the rights of individuals to control their data and how it is collected (Tiku, 2018). Precedents were the European Convention on Human Rights and the Data Protection Directive.

The GDPR addresses several kinds of issues:

- It focuses on "personal data", defined as "information relating to an individual, whether it relates to his or her private professional, or public life. It can be anything from a name, a home address, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer's IP address."
- There must be a lawful basis for processing private data. This can be (1) consent of a subject, (2) to fulfill contractual obligations of a subject, (3) to fulfill legal obligations of a data controller, (4) to protect vital interests of a subject, (5) to perform a task in the public interest, or (6) for the legitimate interests of a data controller or third party providing they are not overridden by interests of the subject. Informed consent must be a "freely-given, plainly-worded, and unambiguous affirmation", something impossible with the traditional long terms-of-service consent forms required by many software vendors before their software is enabled. Digital forensics for law-enforcement and national-security purposes could be covered under items (3) and (5), but many other current uses of digital forensics are not.

- Organizations handling private data must grant users certain rights. This includes informing subjects of the extent of data collection, enabling subjects to see data that has been collected about them, informing subjects how long data will be retained, informing subjects of third parties who will share the data, enabling subjects to revoke consent to process their data, enabling subjects to delete their personal data, and enabling subjects to file complaints with a "Data Protection Authority". Additional restrictions apply to data about race, religion, political affiliation, and sexual orientation. These rights are difficult to enable for digital forensics for several reasons. One is that the people analyzing the data are often unconnected to the people acquiring the data and the subjects of the data. Allowing users to revoke consent to process data or completely delete it is especially difficult to implement given the previously discussed methods of extraction of deleted data using digital forensics. Informing subjects in advance of third parties who will see the data is also very difficult since personal data is a marketable product and it is difficult to anticipate who will want to buy it.
- Data must be protected "by design and by default". That means use of standard practices such as encryption, anonymization, and access controls with defaults being to protect rather than to reveal. Data protection is done in some digital-forensic investigations but not uniformly. Thorough anonymization can be difficult because it can be hard to distinguish personal names in documents from regular words of languages.
- Organizations processing data must keep records of their activities to prove they are obeying the law. They must appoint "Data Protection Officers" and "EU Representatives" to monitor compliance. Data breaches involving private data must be reported to supervisory authorities quickly.

Sanctions for noncompliance are an important part of the GDPR. Article 83 says they can be (1) warnings in cases of first and unintentional non-compliance, (2) periodic audits of data protection mechanisms, and (3) a fine up to 20 million euros or up to 4% of the annual world "turnover" in the case of a business, whichever is greater. The company Google has already been fined 50 million euros under the GDPR (Satariano, 2019) so these regulations may make important changes in the way data-sharing companies do business in Europe. A large conflict may be developing.

Other Governments

Privacy outside of the Western world is treated inconsistently. Governments of totalitarian states like Russia and China think they have the right to investigate many details of personal lives. However, some of the things they do, and use digital forensics to accomplish, violate the U.N. Universal Declaration of Humans Rights (United Nations, 1948):

- "Article 12: No one shall be subjected to arbitrary interference with his privacy, family, home, or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks." Russia clearly violates this in attempts to manipulate politics around the world.
- "Article 18: Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief, and freedom, either alone or in community with others and in public or private, to manifest this religion or belief in teaching, practice, worship and observance." Russia and China clearly violate this in the endless harassment of their own citizens.
- "Article 27: Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author." China clearly violates this in their continuing campaign to steal U.S. technology.

Individuals versus Other Individuals

Individuals can also violate the privacy of other individuals in cyberspace using digital forensics. This could be done for purposes of extortion, blackmail, stalking, harassment, or some other motivation. Laws regulating trespass also apply in cyberspace, and individuals can be sued in civil court for harms, including mental, that they cause because of privacy violations.

Proving Harm to an Individual

An important question is to what extent these laws are enforceable against digital-forensic investigation. Since forensics is generally done remotely on copies of data, a user may not have any clue their forensic information is being inspected. Sometimes it will be apparent if actions show knowledge of data that could only have been obtained forensically, as when a police department uses forensics to track someone. However, a victim of identity theft will have difficulty proving who revealed their data. Businesses do not store most data for long since most is unnecessary for the conduct of business; so by the time that a subpoena is issued, the critical data may well be deleted. For this reason, most users do not need to worry about long-term storage of their locations. It is true that many data breaches are never discovered, and many are never reported for fear of harming the reputation of an organization. However, violations of privacy law could be revealed when organizations are forced to acknowledge their data-handling practices by court orders. Class lawsuits may be necessary; a group of customers harmed by an announced data breach at a business or government can sue collectively for damages.

TECHNICAL PROTECTION METHODS AGAINST DIGITAL FORENSICS

Although laws may be inadequate to protect user privacy from digital forensics, technical means can provide some protection. Some simple steps that users can take to increase their privacy are to delete cookies, delete browser and file caches, stop scripting for Web pages, and run user rather than administrator mode (Bailey, 2015). Software vendors can also facilitate user privacy actions by enabling users to encrypt data and set access rights easily. There are also more elaborate countermeasures to digital forensics called "anti-forensics" (Shanmugam, Powell, & Owens, 2011). Businesses and governments can try to restrict use of these techniques. However, "if policy choices require that individuals shoulder the burden of protecting their own privacy, law and regulation should support the individual in doing so" (National Research Council, 2007, p. 14).

Forensic Transparency

Users cannot adequately protect their data unless they know what is being stored, which is why this concept of "data transparency" is mandated by the European GDPR. Many software tools such as database systems currently only show users some of their data (Stahlberg, Miklau, & Levine, 2007), the "official" data. This may not include header information and previous versions which may also contain private data. For instance, the author of a document is stored in the header of a Microsoft Word document, but users rarely see it. Part of good software design should be to make it clear to users what private data of theirs is hidden or stored externally.

A related issue is that users should know why particular data is being collected and how it will be combined with other data (Dehghantanha & Franke, 2014). Many online services in the U.S. ask for private information like social-security numbers without justification when service-specific numbers would do as well. Other services fail to acknowledge when data will be used for marketing.

Forensic Privacy Policies

Users can often be better judges of whether information should be private than organizations, and helpful organizations can make it easier for them to so designate in an individualized "privacy policy". Also, non-binary policies on privacy can provide finer gradations of control. For instance, a multilevel model of privacy like that of (Halboob et al, 2015) can distinguish data that all investigators can access from data that only authorized investigators can access and from data for which access is controlled by a filtering mechanism. This can be enforced in design of forensic tools.

In some cases, software could automatically identify possibly private information and block its release by policy. This is easiest with formatted data that can be recognized such as government-identification numbers, bank-card numbers, and telephone numbers, which could be automatically blocked from outgoing messages. Data from people irrelevant to an investigation ("third parties") could be blocked automatically when too different from that of known investigation targets (Van Staden, 2013).

Removing Private Data

When a user suspects that a drive will be subjected to digital forensics, they can remove private data from it in advance. This is useful if a computer or device is being sold, transferred, or discarded, but it can be done any time for obsolete data or the entire drive to be sure no copies persist. Software "wiping" tools can do this by writing over data with zeros. Note that many "factory resets" on mobile devices do an incomplete job of removing data, most notably on Android devices (Schwamm and Rowe, 2014). Small amounts of magnetism previously could persist on a magnetic disk after being overwritten, but this not a serious problem with modern magnetic disks.

Data need not persist forever by default. Policies can be set that all copies of private information will be automatically deleted after a particular time period (Yu & Tun, 2017). This makes sense, for instance, for online purchases, accounting records, and localization records of people for which data value decreases quickly over time anyway.

To reduce the chance of surreptitious access to data on a computer or device, users can put the private data in more protectable places like portable storage. For instance, a full Web browser can be kept on a thumb drive and plugged into computers when needed, to reduce cookies and accessibility of its browsing history (Marrington et al, 2013). Portable storage could still be seized or stolen, but it could be a better way to keep data private than putting it on a network that is vulnerable to data breaches.

Encrypting or Obfuscating Private Data

The traditional and strong way to keep data private is by encrypting it so that only the owner can read it. Encryption uses substitution and rearrangement methods to convert data into something unreadable, then permits a user possessing a data "key" to reverse the process and recover the original data. Software to encrypt files and drives is readily available and has many legitimate purposes. It does however put an extra burden on the user to remember and enter a key every time they want to access some data.

Correctly implemented modern cryptographic techniques provide a negligible chance that encrypted data can be deciphered when stolen without the key. However, correct implementation is not easy. Temporary files containing the unencrypted data before they were encrypted may be found using digital forensics if the implementation was poor. The encryption keys may be found in main memory if they were not deleted quickly enough. Or the encryption implementation itself may be designed to deliberately leak data (a charge frequently leveled against encryption methods created by the U.S. National Security Agency with, however, no evidence to support it). Another disadvantage of encryption is that it its use is often detectable since encryption software and encrypted files are easy to spot with digital forensics. Law enforcement can then demand that users give them the decryption keys as part of serving a warrant. Still, encryption creates major obstacles for most people trying to steal data, and is a good defense method for anyone who values their privacy.

Obfuscation of data rather than encryption can be done, meaning encoding the data in some other difficult-to-decipher way. Obfuscation can be accomplished by mapping data to a special set of encodings, hiding data in obscure places (as with steganography discussed below), or mixing it with other data. However, it tends not to be as effective as encryption at hiding data with today's improved tools for analyzing patterns in data.

Steganography and Watermarks for Private Data

An alternative to encryption is to use steganography to conceal data within other data so it does not even appear to exist and law enforcement cannot ask for it. Text examples would be storing data in the numbers representing the lengths of lines in a file, in every 137th letter of a file, or in the fragmentation pattern of a disk (Khan et al, 2011). Image examples would be storing a message in the least-significant bits of a picture, or in the lower right corner of every 137th frame in a video. Using steganography requires special programs to store and retrieve the hidden data.

Steganography can also involve spreading the data in small pieces over many files so it is hard for forensic investigators to assemble all the pieces (Khan et al, 2011). For instance, 64-bit sequences of a file could be distributed over 64 machines so that each machine gets one bit of each sequence. This requires effort to set up, and can be deciphered by a persistent investigator, but it will deter most investigators. Peer-to-peer file sharing is another method makes investigation more difficult but not impossible, and BitTorrent Sync or Resilio (Farina, Scanlon, & Kechadi, 2014) makes it quite difficult. Peer-to-peer file sharing does not protect privacy directly, but makes it more difficult to find data origins and thus attribute it.

Steganography can also be used to prove theft of private data by attaching digital "watermarks" to a file. Watermarks are concealed and encoded messages attached to data which specify its origin, often in the form of a cryptographic "signature". For instance, this could prove theft of private photos of celebrities.

Steganography can be detected by noting unusual statistics in a file and checking for a series of steganographic techniques. However, there are so many steganographic techniques that an investigator trying to find it would have innumerable options to check, and the effort to do such investigations is not often worth it.

Modifying Metadata and Data

Since metadata is essential to forensic investigations, investigation can be impeded by obfuscating it in some way. For instance the Metasploit Framework, open-source tools for malware (Anonymous, 2014), provides ways to modify file and directory names and times on files to confuse investigators. Obfuscated file names prevent easy classification of files, and modified times prevent cause-effect analysis, so they are quite effective at discouraging forensics. Modified file names, however, can cause software to malfunction, and modified times may confuse backup and updating software.

Booby-Trapping

Fancier methods of anti-forensics involve modifying a computer or device so that attempts to copy its contents will result in false data. This can be done, for instance, by recognizing requests for copying large blocks of consecutive storage addresses, something that occurs with systematic forensics and rarely with any other software. The disk controller can be modified to return all zeros instead. This is particularly useful in defending against sophisticated espionage.

Internet Proxy Servers

To prevent forensics on network data, individuals can use proxy services like Tor (Anonymous, 2015) to conceal their Internet and email addresses. Such services replace user addresses by randomly generated ones drawn from a set, and then connect to destinations through those addresses. This makes activities by a user much more difficult to trace back. Tests have shown some metadata about packet flows can be attributed, but not the contents of the packets, which means that Tor is quite good at protecting privacy. However, it cannot protect data of online transactions such as purchases where the user must explicitly identify themselves to accomplish personal goals.

Reverse Forensics

Forensics can also be done on violators of privacy, the businesses and governments that steal personal information. An excellent way to show that privacy violations have occurred is to subpoena computers collecting data, and show by watermarks that data that could not obtained by other than illegal means. This works well against governments since they tend to be reluctant to destroy data, and may even be required to store data for a minimum period of time as with government research in the U.S. (Executive Office of the President of the United States, 2013).

FUTURE DIRECTIONS

It is unclear what will happen with privacy issues in the future. Much of the impetus for protection is driven by news reports of specific abuses, and reports occur irregularly and with varying subjects. It is likely that some cases involving privacy violation will reach the courts and will help establish precedents. However, much of the privacy violation will remain invisible and not subject to public debate. With increased use of countermeasures by organized crime and malware authors, the value of digital forensics in fighting major crime will decrease. This means that most of what law enforcement will find with digital forensics will be various kinds of petty crime. For citizens, technical methods for protecting their privacy will be necessary in the absence of effective laws, similar to the insufficiently regulated environments in the Western part of the U.S. in the 19th century (the “Wild West”). However, Europe may be different.

We can, however recommend some principles relating to digital forensics that citizens need to defend, extending (Bernal, 2014):

- The right to keep private data on computers and devices, data that cannot be viewed by anyone except whom the creator chooses.
- The right to delete their data so that it cannot be viewed by anyone ever again.
- The right to view any Internet site and engage in any Internet commerce.
- The right to be protected from search for their data on a computer or device beyond the kind of data specified in a legal authorization.
- The right to demand and correct personal data acquired by any entity, and confirm that it was authorized for release.
- The right to prevent their data being disseminated to third parties without their knowledge.

CONCLUSION

Digital forensics provides a powerful set of tools for investigating details of people’s private lives, more so than inspecting their public documents and postings. Digital forensics is difficult to do routinely due to the difficulty of obtaining access and the large volume of the data that must be inspected. However, it can provide useful information whenever an investigator has a narrowly defined target. Explicit legal limits need to be set on digital forensics just as with other violations of privacy. The Europeans appear to be ahead of the U.S. in this regard. But progress on this will be slow, and in the meantime, individuals must take proactive technical steps to prevent violations of their privacy.

ACKNOWLEDGEMENT

This chapter expresses the views of the author and does not represent the views of the U.S. Government. This work was supported in part by the U.S. National Science Foundation under the Secure and Trustworthy Cyberspace program.

REFERENCES

Abel, W. (2009, March-July). Agents, Trojans, and tags: the next generation of investigators. *Intl. Review of Law, Computers, & Technology*, 23 (1-2), 99-108.

Anonymous (2014). *Metasploit*. Retrieved December 31, 2014 from www.metasploit.com.

Anonymous (2015). *Tor*. Retrieved January 4, 2015 from www.torproject.org.

Akhgar, B., Staniforth, A., & Bosco, F. (eds.) (2014). *Cyber crime and cyber terrorism investigator's handbook.* Waltham, Massachusetts, US: Syngress.

Bagby, J. (2013). On resolving the cloud forensics conundrum. *Journal of Digital Forensics, Security, & Law*, Conference Supplement, p. 21.

Bailey, M. (2015). *Complete guide to Internet privacy, anonymity, & security, second edition*. New York: Nerel Online.

Bernal, P. (2014). *Internet privacy rights: rights to protect autonomy*. Cambridge, UK: Cambridge University Press.

Burd, S., Jones, D., & Seazzu, A. (2011). Bridging differences in digital forensics for law enforcement and national security. In *Proc. 44th Hawaii Intl. Conf. on System Sciences* (pp. 1530-1605/11). New York: IEEE Press.

Catiglione, A., De Santis, A., & Soriente, C. (2010). Security and privacy issues in the Portable Document Format. *Journal of Systems and Software*, 83 (10), 1813-1822.

Campbell, F. (2014). The slow death of "do not track". *The New York Times*, December 26. Retrieved January 5, 2015 from www.nytimes.com/2014/12/27/opinion/the-slow-death-of-do-not-track.html.

Cohen, C. (2007). The growing challenge of computer forensics. *The Police Chief*, 74 (3), March.

Cottrill, C., & Thakuriah, P. (2015). Location privacy preferences: A survey-based analysis of consumer awareness, trade-off, and decision-making. *Transportation Research Part C*, 56, 132-148.

Damschroder, L, Pritts, J., Neblo, M., Kalarickal, R., Crewsell, J., & Hayward, R. (2007). Patients, privacy, and trust: Patient's willingness to allow researchers to access their medical records. *Social Science and Medicine,* 64, 223-235.

Dehghantanha, A., & Franke, K. (2014). Privacy-respecting digital investigation. In Proc. 12th Annual Conf. on Privacy, Security, and Trust, Toronto, CA, July 2014, pp. 129-138.

Elisan, C. (2012). *Malware, rootkits, and botnets: a beginner's guide*. New York: McGraw-Hill Osborne Media.

EU GDPR.org (2019). *The EU General Data Protection Regulation*. Retrieved April 16, 2019 from https://eugdpr.org.

European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Retrieved April 22, 2019 from eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN.

Executive Office of the President of the United States (2013, February 22). Increasing access to the results of federally funded scientific research. Retrieved January 4, 2015 from www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Farina, J., Scanlon, M., & Kechadi, M. (2014, May). BitTorrent Sync: first impressions and digital forensic implications. *Digital Investigation*, 11 Supplement, S77-S86.

Fink, U. (2014). Protection of privacy in the EU, individual rights and legal instruments. In Witoleb, N., Lindsay, D., Paterson, M. & Rodrick, S. (eds.), *Emerging challenges in privacy law: comparative perspectives* (pp. 75-91). Cambridge, UK: Cambridge University Press.

Fox-Brewster, T. (2015, January 7). Why you still shouldn't totally trust FBI claims on North Korean hacking of Sony. Retrieved January 13, 2015 from www.forbes.com/sites/thomasbrewster/2015/01/07/fbi-claims-on-north-korea-sony-hack-still-questionable.

Greengard, S. (2012, September). Advertising gets personal. *Communications of the ACM*, 55 (8), 18-20.

Greengard, S. (2012, November). On the digital trail. *Communications of the ACM*, 55 (11), 19-21.

Halboob, W., Mahmod, R., Udzir, N., & Abdullah, M. (2015). Privacy levels for computer forensics: Toward a more efficient privacy-preserving investigation. Proc. Intl. Workshop on Cyber Security and Digital Investigation, *Procedia Computer Science* 56, 2015, 370-375.

Hoebich, M. (2008). Are your computer files protected under the Fourth Amendment? *Information Security Journal: A Global Perspective*, 77, 143-150.

Hong, I., Yu, H., Lee, S., & Lee, K. (2013, September). A new triage model conforming to the needs of selective search and seizure of electronic evidence. *Digital Investigation*, 10 (2), 175-192.

Jahankhani, H., & Hosseinian-Far, A. (2017). Challenges of cloud forensics. In Chang, V. et al. (Eds.), *Enterprise Security LNCS 10131*, Springer, 2017, 1-18.

Jarrett, H., Bailie, M., Hagen, E., & Judish, N. (2014). *Searching and seizing computers and obtaining electronic evidence in criminal investigations*. Washington, DC, USA: Office of Legal Education Executive Office for United States Attorneys.

Jones, K., Bejtlich, R., & Rose, C. (2006). *Real digital forensics: computer security and incident response*. Upper Saddle River, NJ, USA: Addison-Wesley.

Khan, H., Javed, M., Khayam, S., & Mirza, F. (2011). Designing a cluster-based covert channel to evade disk investigation and forensics. *Computers & Security*, 30 (1), 35-49.

Koppen, J., Gent, G., Bryan, K., DiPippo, L., Dramer, J., Moreland, M., & Fay-Wolfe, V. (2012, October). Identifying remnants of evidence in the cloud. *Proc. 4th Intl. Conf. on Digital Forensics & Cyber Crime*, Springer Lecture Notes in Computer Science (vol. 42013, pp. 42-57). New York: Springer.

Losavio, M., & Keeling, D. (2014). Evidentiary power and the propriety of digital identifiers and the impact on privacy rights in the United States. *Journal of Digital Forensics, Security, & Law,* 9 (2), 197-203.

Lonardo, T., Martland, T., White, D., & Rea, A. (2011). Legal issues regarding digital forensic examiners third party consent to search*. Journal of Digital Forensics, Security, & Law*, 6 (4), 19-34.

Manes, G., & Downing, E. (2010). What security professionals need to know about digital evidence. *Information Security Journal: A Global Perspective*, 19, 124-131.

Marrington, A., Baggili, I., Al Ismail, T., & Al Kaf, A. (2012). Portable Web browser forensics. Proc. Intl. Conf. on Computer Systems and Industrial Informatics, Sharja, UAE, December 2012.

Matthews, D. (2010). eDiscover versus computer forensics. *Information Security Journal: A Global Perspective*, 19, 118-123.

Motivans, M., & Kyckelhahn, T. (2007, December). Federal prosecution of child sex exploitation offenders, 2006. Retrieved January 29, 2015 from www.bjs.gov/content/pub/ pdf/fpcseo06.pdf.

Myneedu, T., & Guan, Y. (2012). Evidence collection in peer-to-peer network investigations. In *Proc. IFIP Conf. Advances in Digital Forensics VIII* (pp. 215-230). New York: Springer.

National Research Council of the National Academies [U.S.] (2007). *Engaging privacy and information technology in a digital age*. Washington, DC, US: The National Academies Press.

National Research Council of the National Academies [U.S.] (2008). *Protecting individual privacy in the struggle against terrorists*. Washington, DC, US: The National Academies Press.

Pearson, S. (2010). *Digital triage forensics: processing the digital crime scene*. New York: Syngress.

Pinsent Masons LLP (2015). AboutCookies.org. Retrieved January 5, 2015 from www.aboutcookiest.org.

Plachkinova, M., Vo, A., & Alluhaidan, A. (2016). Emerging trends in smart home security, privacy, and digital forensics. Proc. 22nd Americas Conf. on Information Systems, San Diego.

Rodriguez, K. (2014, March 10). EFF to the United Nations: Protect individuals right to privacy in the digital age. Retrieved January 2, 2015 from www.eff.org/deeplinks/2014/02/eff-un.

Ryan, D., & Shpantzer, G. (2010). Legal aspects of digital forensics. Retrieved December 31, 2014 from euro.ecom.cmu.edu/program/law/08-732/Evidence/RyanShpantzer.pdf.

Satariano, A. (2019, January 21). Google is fined $57 million under Europe’s data privacy law. *The New York Times*, p. B1.

Schneier, B., (2015, March 25). NSA doesn’t need to spy on your calls to learn your secrets. Retrieved May 18, 2015 from www.wirec.com/2015/03/data-and-goliath-nsa_metadata-spying-your-secrets.

Schwamm, R., & Rowe, N. (2014). Effects of the factory reset on mobile devices. *Journal of Digital Forensics, Security, & Law*, 9 (2), 205-220.

Shanmugam, K., Powell, R., & Owens, T. (2011). An approach for validation of digital anti-forensics evidence. *Information Security Journal: A Global Perspective*, 20, 219-230.

Shipler, D. (2011). *The rights of people: how our search for safety invades our liberties*. New York: Knopf.

Stahlberg, P., Miklau, & Levine, B. (2007). Threats to privacy in the forensic analysis of database systems. In Proc. ACM SIGMOD International Conference on Management of Data, Beijing, CN, June 2007, pp. 91-102.

Surowiecki, J. (2014, June 9). Spy vs. spy. *The New Yorker*. Retrieved May 29, 2015 from www.newyorker.com/magazine/20154/06/09/spy-vs-spy-3.

Tiku, N. (2018, March 19). Europe’s new privacy law will change the Web, and more. *Wired*. Retrieved April 16, 2018 from www.wired.com/story/europes-new-privacy-law-will-change-the-web-and-more.

Toomey, P. (2018, August 22). The NSA continues to violate Americans' Internet privacy rights. Retrieved April 21, 2019 from www.aclu.org/blog/national-security/privacy-and-surveillance/nsa-continues-violate-americans-internet-privacy.

Trend Micro (2018). Data breaches 101: How they happen, what gets stolen, and where it all goes. Retrieve April 21, 2019 from www.trendmicro.com/vinfo/us/security/news/cyber-attacks/data-breach-101.

United Nations (1948). Universal declaration of human rights. Retrieved April 29, 2019 from www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf.

Van Staden, W. (2013). Protecting third party privacy in digital forensic investigations. In *Proc. IFIP Conf. Advances in Digital Forensics IX* (pp. 19-31). New York: Springer.

Yu, Y., & Tun, T. (2017). Snap forensics: A tradeoff between ephemeral intelligence and persistent evidence collection. In Proc. 1st ACM SIGSOFT Intern. Workshop on Software Engineering & Digital Forensics, Paderborn, DG, September 2017, pp. 10-11.

KEY TERMS AND DEFINITIONS

Cloud Computing: Distributed storage of user data and programs in large servers accessed by the Internet.

Cookies: Personal data stored by a Web page on a computer in a small package.

Data Breach: Revelation of personal data of many people simultaneously, usually due to digital exploits.

Digital Forensics: Methods for extracting data from digital media such as computers, mobile devices, storage devices, and networking devices.

Encryption: Encoding data so that it can only be decoded by encoder by means of a data string called a key (a "decryption" process). Software tools can accomplish it.

File Carving: Analysis of pieces of files left in storage or memory and reconstruction of how those pieces go together.

General Data Protection Regulation (GDPR): The European's Union's comprehensive privacy laws.

Media or Drive: A computer magnetic disk, flash memory, or optical disk (CD) when used for storage of data.

Metadata: Data such as file names and creation times describing other data.

Peer-to-Peer: Methods of automatic file sharing between computers or devices without going through a central site.

Steganography: Methods of concealing information inside other information.