The Future of Digital Forensics:

Challenges and the Road Ahead

Luca Caviglione | National Research Council of Italy
Steffen Wendzel | Worms University of Applied Sciences
Wojciech Mazurczyk | Warsaw University of Technology

Today's huge volumes of data, heterogeneous information and communication technologies, and borderless cyberinfrastructures create new challenges for security experts and law enforcement agencies investigating cybercrimes. The future of digital forensics is explored, with an emphasis on these challenges and the advancements needed to effectively protect modern societies and pursue cybercriminals.

Modern society increasingly depends on communication networks, mobile appliances, Internet of Things (IoT) solutions, cyber-physical system (CPS) technologies, and cloud-based services. By taking advantage of the most cutting-edge information and communication technology (ICT) features, commercial activities, business transactions, and government services have grown, transforming the lifestyle of nearly all individuals. The tight coupling of the physical world with ICT technologies has led to indubitable benefits. At the same time, it accounts for the proliferation of new threats and cybersecurity issues such as identity theft, cyberbullying, data leakage exploiting social engineering or information hiding, malicious software turning IoT nodes into "zombies," distributed denial of service (DDoS) orchestrated through botnets, and malware targeting specific appliances such as those affecting VoIP devices and smart vehicles.[1] As a consequence, cyberattacks can have a significant socioeconomic impact on both global enterprises and individuals. Therefore, cybercriminals should be promptly identified, and high-quality evidences of the attacks should be made available in the courtroom.

Unfortunately, pursuing cybercrimes often requires complex investigations that can span international borders and be subjected to different jurisdictions and legal systems. This issue, jointly with the huge volume and richness of information, highly heterogeneous ICT technologies, and complex modern hardware/software frameworks, raises new challenges, especially in the field of digital forensics. In fact, it's common to have to trace an attacker through the Internet, or to collect clues to illegal digital activities buried in large volumes of data. Therefore, law enforcement agencies (LEAs) and security professionals must update and often rethink their approach to digital forensics. For instance, forensic investigators must be fully prepared to gather effective and unambiguous digital evidences, while LEAs should be ready to "handle" a multidisciplinary process that spans several fields, including law, computer science, finance, networking, data mining, and criminal justice.

We provide a compact survey of the most important challenges to be considered when designing or developing modern tools for digital forensics. This article's main contribution is the identification of the most important

 Copublished by the IEEE Computer and Reliability Societies 1540-7993/17/$33.00 © 2017 IEEE

mid- and long-term opportunities and issues to be considered both by security experts and LEAs working in the field of digital investigations.

The State of Digital Forensics

Although digital forensics might seem to be a novel discipline, its roots date back to 1970, when engineers recovered the only copy of a database that had been inadvertently deleted.[2] From such a starting point, digital forensics rapidly evolved. Nowadays, it's possible to address a variety of aspects concerning digital investigations, for instance, undeleting data or dumping network traffic to reconstruct attacks offline. The standard toolbox used for digital forensics covers all the different aspects of the cyber investigation procedure. However, as hinted at earlier, evolutions in the ICT domain reduce its efficiency and pose hazards. More specifically, modern forensic procedures and their main limitations can be summarized as follows.

Stored Data and Filesystem Analysis

Digital forensics experts are expected to maintain the integrity of seized evidence, such as hard drives. Thus, safe practice consists of accessing the device in read-only mode and using forensic tools to create working copies, or *forensic images*. For instance, forensic images can be used to analyze files and installed applications, or to inspect unused space to spot remnants of deleted contents. Unfortunately, the increasing number of devices and the volume of available data make creating and analyzing such working copies highly time consuming. A recent approach triages a live system, which allows forensics examiners to acquire evidences potentially hidden in volatile digital artifacts (for example, the contents of RAM or the clipboard, the list of open files, running processes, active network connections, or mapped drives). This aspect is also vital to avoid losing evidence as a consequence of a reboot, which could cause encryption of the filesystem or deletion of temporary data. In fact, an increasing number of OSs make it easy to encrypt the filesystem including the swap area (see, for example, popular Unix-based OSs such as Linux and macOS). Although this contributes to the security and privacy of end users and enterprises, it also represents a hurdle for forensic investigations. Due to the integration and almost per-click setup in some environments, the percentage of encrypted filesystems can be expected to increase, eventually becoming the default scenario in the long-term future.

Network Forensics

Typically, network forensics requires collecting and analyzing the traffic produced by a host, an intermediate node, or an entire portion of a network. Forensic analysts rely on traffic captures as well as on the available logs of network services and security appliances such as intrusion detection systems. To this aim, different granularities are possible (for example, see Suleman Khan and his colleagues' recent survey on the topic[3]). The most fine-grained techniques perform per-packet inspections, where the threat/attack is traced back by analyzing header fields or IPv4/IPv6 addresses. Unfortunately, collecting and storing each packet of large network infrastructures is economically infeasible and could saturate resources of routers or dedicated security tools such as firewalls. Moreover, the amount of data could be impossible to evaluate using commodity hardware, making the forensics analysis too expensive, especially if real-time requirements must be satisfied. A more coarse-grained approach exploits information collected by ad hoc tools such as firewalls, intrusion detection systems, or honeypots. Also in this case, the huge traffic volume, the heterogeneity of the used technologies (for example, proprietary communication protocols deployed in automation environments mixed with the TCP/IP suite if data has to be transmitted through the Internet), the availability of information-hiding-capable cyberthreats, traffic encryption, or obfuscating mechanisms such as onion routing can render network forensics efforts highly challenging.

Reverse Engineering

Reverse engineering is performed by inspecting the binary of a malware sample, the recordings of network traffic, or other execution traces, such as log files generated by the guest OS. However, the effectiveness of such approaches is limited for a subset of modern threats deploying anti-forensics countermeasures such as code obfuscation, or multistage loading architectures, in which each stage of the threat is hidden and encrypted. The same holds for information-hiding-capable malware that covertly communicates with a remote command and control facility.[4]

Key Challenges

Despite the field's quick evolution, advancements in digital forensics are now more difficult to achieve. Its continuous evolution is heavily challenged by the increasing popularity of digital devices and the heterogeneity of the hardware and software platforms being used. For example, the plethora of file formats and OSs impedes the creation of unified or standard tools, and the advent of smartphones that extensively use cryptography or embed digital rights management/trusted computing frameworks makes collecting evidence a complex task (see, for example, "Forensic Analysis Techniques for Fragmented Flash Memory Pages in Smartphones"[5]).

The presence of a mixed set of technologies is not the only factor increasing the complexity of the problem space to be faced when performing digital investigations.

Specifically, cybercrime's evolution has led to the significant challenge of novel "business models," including crime as a service (CaaS), which guarantees to potential adversaries easy access to the tools, programming frameworks, and services required to carry out cyberattacks. A notable example is Tox, a ransomware construction kit discovered McAfee Labs on the dark web in May 2015. Briefly, the Tox framework can be customized and used to spread and coordinate infections in return for 20 percent of every ransom paid.[6]

Cloud computing is one of the most controversial emerging areas of digital forensics. According to Keyun Ruan and her colleagues' survey,[7] professionals can't agree on whether cloud computing makes forensics harder or easier. As regards its cons, cloud platforms reduce physical control over data and its location. Moreover, they require handling the lack of standard interfaces when developing tools; present issues due to multiple ownerships, tenancies, and jurisdictions; and make the investigation of complex attacks more difficult, mainly because of the lack of collaboration among providers.

Concerning cloud computing's pros, the way data is managed by the cloud infrastructure offers many advantages. Briefly, information is distributed across different datacenters to enhance performance and allow load-balancing, scalability, and redundancy features. To this aim, information has to be properly indexed to, for example, allow retrieval and perform optimization to avoid duplication, which could inflate storage requirements. Therefore, evidence left by criminals is harder to destroy because it could be mirrored in multiple places, or already be hashed and indexed, making the collection of artifacts and their analysis simpler.

Another important aspect challenging digital forensics is the availability of different paradigms for delivering cloud services. Specifically, investigating the data of an infrastructure-as-a-service (IaaS) user can be done without too many restrictions, but in the case of customers using software-as-a-service (SaaS) resources, access to information might be minimal or entirely absent.

Thus, modern digital forensics faces many challenges, from both ethical and technological viewpoints:

- *High speed and volumes.* The availability of gigabit class links and multimedia-rich contents accounts for an explosion in the volume of data to be stored and processed for collecting clues or detecting incidents. This is of particular relevance in the case of live network analysis, as the investigator might not be able to capture and store all the necessary traffic. Nevertheless, issues related to acquiring, storing, and processing large amounts of data for forensic purposes have been causing problems for at least a decade,[8] and are now exacerbated by the ubiquitous availability and massification of digital information.
- *Explosion of complexity.* The technological advances in and proliferation of novel services account for a dramatic increase in the complexity that forensics professionals must manage. Specifically, evidence is no longer confined within a single host but, rather, is scattered among different physical or virtual locations, such as online social networks, cryptocurrency wallets, CaaS machinery, cloud resources, and personal network–attached storage units. For this reason, more expertise, tools, and time are needed to completely and correctly reconstruct evidence. This explosion in complexity also impacts the length of digital investigations, including the degree of occupancy of resources involved (such as the manpower of forensics experts or third parties, including specialized professionals hired by LEAs). Such issues could be partially solved by automating some tasks. However, this has been highly criticized by the digital investigation community,[9] because it could quickly deteriorate both the quality of the investigation and the knowledge of forensics experts.
- *Development of standards.* As mentioned earlier, digital forensics requires the ability to handle several hardware and software entities, ranging from RAM to USB solid-state mass storage. Despite technological advances, files are still the most popular digital artifacts to be collected, categorized, and analyzed. Thus, the research community has tried to agree on standard formats, schema, and ontologies—but without much success.[2] Investigations of cutting-edge cybercrimes might require processing information in a collaborative manner or using outsourced storage and computation. Therefore, a core step for the digital forensics community will be the development of proper standard formats and abstractions (see "Automating Disk Forensic Processing with SleuthKit, XML and Python"[10] for an example using XML).
- *Privacy-preserving investigations.* Nowadays, people bring into cyberspace many aspects of their lives, primarily through online social networks or social media sites. This has dramatically boosted social engineering–based attacks, which should be promptly reported and investigated. Unfortunately, collecting information to reconstruct and locate an attack can severely violate users' privacy and is linked to other hurdles when cloud computing is involved.
- *Legitimacy.* Modern infrastructures are becoming complex and virtualized, often shifting their complexity at the border (such as in fog computing) or delegating some duties to third parties (such as in platform-as-a-service frameworks). Thus, an important challenge for modern digital forensics will be

executing investigations legally, for instance, without violating laws in borderless scenarios.

- *Rise of antiforensics techniques.* As defensive measures become increasingly efficient, more aggressive deployment of antiforensics methods can be envisioned. These encompass encryption, obfuscation, and cloaking techniques, including information hiding. For example, a challenge for filesystem analysis is steganographic configurations, which allow hiding information in unused areas of the hard disk or in metadata, such as timestamps.[11] With the exception of binary obfuscation, such mechanisms aren't yet widespread, but they could become relevant for digital forensics investigation in the mid-term future.

Future Paradigms

Many of the digital forensics tools tailored to discovering evidence are expected to reside on the suspect's device, but offer limited features for investigating unknown and complex environments, including big data–like sources.[2,3,8] Consequently, the majority of forensic software is unsuitable for identifying anomalies in an automatic or unattended way. One of the major challenges to be addressed in the near future, therefore, is the creation of tools and techniques to analyze the bulk of data and report possible digital clues to the examiner for further investigation. Alas, such tools and techniques' engineering, including proper visualization features to help the forensic examiner, is a complex task, particularly because of the lack of unified standards and the nontrivial computational requirements.

Fortunately, digital investigation can leverage the features of cloud computing, for instance, to offload the most demanding operations of digital forensics procedures, such as log analysis, data indexing, and multimedia processing. From this perspective, one of the most interesting aspects of the cloud is the opportunity to exploit a new paradigm in which forensics is provided as a utility, à la forensics as a service (FaaS). For example, Guiseppe Totara and his colleagues developed a tool for indexing forensic disk images that can be easily used by investigators through a web interface.[12] An additional benefit of pursuing an FaaS paradigm is the possibility of concentrating the software in a single point, which makes updates and improvements easier. This can also hide complexity from end users, allowing professionals to concentrate on the investigation. Similarly, digital investigations can leverage the proliferation of software-defined networking techniques, which offer additional layers of abstraction useful for analyzing attacks or infections without the need for resource-consuming traffic analysis campaigns.

Lastly, digital forensics could quickly become essential even in new and unforeseen scenarios. IoT usage creates a point of interaction between the cyber and physical worlds, making digital IoT forensics an effective way to collect information about the nondigital environment as well. For example, IoT nodes can provide evidence of when a person was present in a room by investigating in-door presence sensor values.[13,14] Obviously, such investigations are linked to further privacy issues,[14,15] especially as sensors might be influenced not only by a single user but by an undefined set of influencers: several individuals could trigger a presence sensor in a room each day, not just the potential criminal.

Because personal devices, appliances, and IoT nodes are starting to "reverse the fate" of several court trials, in the following section, we focus on how IoT and CPS can impact digital forensics, in terms of both challenges and opportunities.

The Case of IoT and CPS

Due to their dramatic diffusion, IoT and CPS deployments are gaining particular attention from cybercriminals, security experts, LEAs, and other professionals dealing with digital forensics. Besides its increasing usage for leisure and home automation (domotics), the IoT is a core building block for the envisioned Industry 4.0 revolution. IoT and CPS frameworks will account for the optimization and revenue maximization of many manufacturing processes, but at the same time, they create new opportunities for ad hoc attacks, for example, industrial espionage or cybersabotage. Digital forensics will have to be able to investigate IoT and CPS components, which can range from nodes deployed in small scenarios (such as smart objects and smart watches) to those deployed in large ones (such as smart cities and smart factories). This includes the analysis of attacks on IoT devices as well as the digital forensics–aided investigations of physical-world crimes. In fact, IoT devices contain several controllers and pieces of monitoring equipment, as well as a large number of sensors and actuators.[14] Whereas sensors measure the environment (for example, the temperature in a room), actuators influence the physical environment (for example, by opening or closing a window, or adjusting the level of a heater). Unfortunately, historic sensor data and actuator states aren't necessarily persistent or always accessible to an investigator. Consequently, IoT environments highly differ from classical network and computing deployments, leading to new challenges, including the following:

- In IoT scenarios, persistent recording isn't easily achieved due to resource constraints in embedded systems, or smart objects with limited memory and computing performance.
- The diffusion of many IoT devices with proprietary interfaces might lead to difficulty accessing stored

Table 1. Properties of Internet of Things and cyber-physical systems devices and their effect on modern digital forensics.

Property	Relation to (digital) forensic investigation	Exemplary challenges
Density of device deployment	Influences the resolution of events that took place in a physical environment.	Reconstruct physical events on the basis of unevenly deployed devices (for example, not all the space has the same density of Internet of Things [IoT] nodes and information). This can also vary according to the considered environment.
Device type	Influences the type of information (sensor data such as temperature, humidity, pressure, or motion detection as well as the history of actuator states such as door status, pump pressure, or heating level).	Provision of a computer-aided, evidence-driven, and court-proof frameworks for the reconstruction of events. Such software should be able to take into account a mixed/increasing set of devices, for instance by means of a plug-in architecture.
Device location	Influences physical accessibility of the device for a digital forensic investigation (the device might be placed behind country borders) and influences which area of a physical environment was covered by the device (the part of a forensic site that has been influenced).	Develop a cost–benefit analysis to determine whether IoT devices located in hard-to-reach areas are worth accessing. One idea is to use databases that point out device properties useful for forensics investigations and additional details such as the accuracy of onboard sensors.
Recording history	All available information on an IoT device can be recorded locally or in the cloud. Local storage is usually limited; thus, the number of recorded sensor values/actuator states is kept under a certain threshold. Older data might not be accessible.	Automatic integration of IoT devices into the reconstruction process of physical events. This requires fetching the recording history of sensors and correctly placing it within the time frame of the event to be reconstructed. It could entail the support of visual analytics tools capable of handling devices that provide data with inconsistent or inaccurate timing and spatial positions.
Device interfaces	The interfaces used to access evidence highly influence the amount of information that can be retrieved. Some types of information might not be provided by certain interfaces while others are. In several cases, interfaces might be undocumented by the vendor.	Provision of a unified metainterface for IoT forensics covering a large spectrum of different devices and low-level interfaces of several vendors. This can likely be adequately addressed by larger community projects.

values, possibly requiring the investigator to perform a non-negligible reverse-engineering effort.
- Reduced energy (for instance, solar-powered nodes) often leads to intermittent and partially incomplete information.

It's unclear how well various IoT devices can be investigated in terms of cost–benefit analyses. In this sense, "cost" can be of a juridical, technical, or economical nature. Additionally, IoT devices differ not only in their type (for example, sensors in a plant or actuators in a smart home) but also in accessibility and interfaces, vendor-specific features, and data storage strategy (local versus cloud-based and persistent versus volatile). We point out that the number of IoT devices (the density of deployed devices per square or cubic meter), their location, their type (for example, smart watch, pressure sensor, or pump controller), and their accessibility attributes must be understood and considered individually for forensic investigations. Moreover, many IoT devices frequently change their location because they're mobile (such as fitness trackers, smart clothing, or smart watches) or because they're part of larger movable objects (such as drones, bikes, or cars). Gaining access to such nodes can be especially challenging. For example, if the laws in one state grant LEAs access to smart devices, criminals might simply park their smart cars behind the closest border, or mail their smart watches another country, making them inaccessible to LEAs—or at least delaying access. Table 1 summarizes IoT and CPS device properties, how they influence investigations, and potential challenges to be faced.

As discussed, digital forensics is fundamental to investigations performed in a reality that's often tightly coupled with its cyberextension. Modern digital societies are subject to cybercriminal activities and fraud leading to economic losses or hazards for individuals. Therefore, the new wave of forensics tools should be engineered to support heterogeneous investigations, preserve privacy, and offer scalability, just to mention the most important desiderata. Use of the IoT and CPS poses several new problems, which are linked to the mixed flavor of nodes and their differences in data storage, accessibility, and investigative tradeoffs. New tools are necessary to improve IoT forensics, especially because antiforensic techniques will continue to become increasingly sophisticated. This becomes more significant in light of emerging trends such as CaaS, which allows (almost) anyone to execute serious cyberattacks. CaaS requires collecting evidence form different parts of the Internet, potentially leading to huge amounts of incompatible information being stored and processed. Luckily, cloud computing can be combined with intelligent automation and visualization tools to aid forensic analysts in highly complex environments. ■

Acknowledgments

The authors would like to thank Simson L. Garfinkel for his valuable ideas and fruitful discussions.

References

1. J. Jang-Jaccard and S. Nepal, "A Survey of Emerging Threats in Cybersecurity," *J. Computer and System Sciences*, vol. 80, no. 5, 2014, pp. 973–993.
2. S.L. Garfinkel, "Digital Forensics Research: The Next 10 Years," *Digital Investigation*, vol. 7 supplement, 2010, pp. S64–S73.
3. S. Khan et al., "Network Forensics: Review, Taxonomy, and Open Challenges," *J. Network and Computer Applications*, vol. 66, May 2016, pp. 214–235.
4. W. Mazurczyk and L. Caviglione, "Information Hiding as a Challenge for Malware Detection," *IEEE Security & Privacy*, vol. 13, no. 2, 2015, pp. 89–93.
5. J. Park, H. Chung, and S. Lee, "Forensic Analysis Techniques for Fragmented Flash Memory Pages in Smartphones," *Digital Investigation*, vol. 9, no. 2, 2002, pp. 109–118.
6. "Meet 'Tox': Ransomware for the Rest of Us," McAfee Labs, 23 May 2015; blogs.mcafee.com/mcafee-labs/meet-tox-ransomware-for-the-rest-of-us.
7. K. Ruan et al., "Cloud Forensics Definitions and Critical Criteria for Cloud Forensic Capability: An Overview of Survey Results," *Digital Investigation*, vol. 10, no. 1, 2013, pp. 34–43.
8. G.G. Richard III and V. Roussev, "Next-Generation Digital Forensics," *Comm. ACM*, vol. 49, no. 2, 2006, pp. 76–80.
9. J.I. James and P. Gladyshev, "Challenges with Automation in Digital Forensic Investigations," Computing Research Repository (CoRR), arXiv.org, 2013; abs/1303.4498.
10. S.L. Garfinkel, "Automating Disk Forensic Processing with SleuthKit, XML and Python," *Proc. 4th Int'l IEEE Workshop Systematic Approaches to Digital Forensic Engineering*, 2009, pp. 73–84.
11. S. Neuner et al., "Time Is on My Side: Steganography in Filesystem Metadata," *Digital Investigation*, vol. 18 supplement, 2016, pp. S76–S86.
12. G. Totaro et al., "ISODAC: A High Performance Solution for Indexing and Searching Heterogeneous Data," *J. of Systems and Software*, vol. 118, Aug. 2016, pp. 115–133.
13. S. Buhr, "An Amazon Echo May Be the Key to Solving a Murder Case," TechCrunch, Dec. 2016; techcrunch.com/2016/12/27/an-amazon-echo-may-be-the-key-to-solving-a-murder-case.
14. S. Wendzel et al., "Cyber Security in Smart Buildings," *Security and Privacy in Cyber-Physical Systems: Foundations and Applications*, H. Song et al., eds., Wiley, 2017.
15. T. Mudth, F. Krüger, and T. Wolleberg, "Who Refuses to Wash Hands? Privacy Issues in Modern House Installation Networks," *Proc. 7th Int'l Conf. Broadband, Wireless Computing, Communication and Applications*, 2012, pp. 271–277.

Luca Caviglione is a researcher with the Institute of Intelligent Systems for Automation, National Research Council of Italy. His research interests include P2P systems, wireless communications, cloud architectures, and network security. Caviglione received a PhD in electronics and computer engineering from the University of Genoa. He is an associate editor of the *Transactions on Emerging Telecommunications Technologies*. Contact him at luca.caviglione@ge.issia.cnr.it.

Steffen Wendzel is a professor for information security and computer networks at Worms University of Applied Sciences. He wrote five books and his research focuses on information hiding and security in the Internet of Things. Wendzel received a PhD in computer science from the University of Hagen. Contact him at wendzel@hs-worms.de.

Wojciech Mazurczyk is an associate professor at the Institute of Telecommunications in the Faculty of Electronics and Information Technology at the Warsaw University of Technology. His research interests include network security, information hiding, and network forensics. Mazurczyk received a PhD and a DSc in telecommunications from the Warsaw University of Technology. He is also an associate technical editor for *IEEE Communications Magazine*. Contact him at wmazurczyk@tele.pw.edu.pl.