CSIT (March 2013) 1(1):91–114
DOI 10.1007/s40012-012-0008-7

ORIGINAL RESEARCH

Digital forensic research: current state of the art

Sriram Raghavan

Received: 9 October 2012 / Accepted: 30 October 2012 / Published online: 13 November 2012

Abstract Digital forensics is the process of employing scientific principles and processes to analyze electronically stored information and determine the sequence of events which led to a particular incident. In this digital age, it is important for researchers to become aware of the recent developments in this dynamic field and understand scope for the future. The past decade has witnessed significant technological advancements to aid during a digital investigation. Many methodologies, tools and techniques have found their way into the field designed on forensic principles. Digital forensics has also witnessed many innovative approaches that have been explored to acquire and analyze digital evidence from diverse sources. In this paper, we review the research literature since 2000 and categorize developments in the field into four major categories. In recent years the exponential growth of technological has also brought with it some serious challenges for digital forensic research which is elucidated. Within each category, research is subclassified into conceptual and practical advancements. We highlight the observations made by previous researchers and summarize the research directions for the future.

Keywords Digital forensics · Taxonomy · Digital forensic acquisition · Digital forensic examination · Digital forensic analysis · Digital forensic process modeling

S. Raghavan (✉)
Information Security Institute, Queensland University of Technology, 126 Margaret Street, Brisbane, QLD 4000, Australia
e-mail: s.raghavan@qut.edu.au

Present Address:
S. Raghavan
No. 24, New Moti Bagh, Delhi 110023, India

1 Introduction

Digital forensics is a branch of science that involves the application of scientific principles to the investigation of artifacts present in one or more digital devices in order to understand and reconstruct the sequence of events that must have transpired in generating the said artifacts. Digital forensics pertains to acquiring, examining, analyzing, and possibly documenting and presenting these artifacts and the reconstructed sequence of events as evidence in a court of law. Digital forensics developed as an independent field in the late 1990s and early 2000s when computer based crime started growing with the increasing usage of computers and more so, the Internet. In early days, it was called computer forensics since the evidence collected was restricted to computers. However, in recent years, with several technological advances, this restriction is no longer true. Consequently, the process of conducting forensic investigations involving contemporary digital evidence has become more challenging.

Computer forensics developed as an independent field in late 1990s and early 2000 when computer based crime started growing with the increasing popularity of computers and especially the Internet. Of the approximately half of respondents who experienced at least one security incident last year, fully 45.6 percent of them reported they'd been the subject of at least one targeted attack. According to the 2010/11 CSI Computer Crime Survey [60], almost 46 % of the respondents were affected by at least one form of computer crime. According to 2010 Gallup Computer Crime survey [73], 11 % of American adults report that they were a victim of a computer or Internet crime on their home computer in the past year, up from the 6 to 8 % levels found in the previous 7 years. The 2012 Indian Risk survey [71] indicates that Computer and Internet crime

remains the single largest source of national threat at 10.81 % closely followed by terrorism at 10.43 %. The 2006 Australian Computer Crime Survey [12] has estimated computer facilitated financial fraud and proprietary information breaches at over A$ 2,000,000 in lost revenue. With the recent proliferation of newer digital devices in the markets and the increasing frequency of discovering such devices in investigations, a new term called *digital forensics* was coined. This new term now refers to investigating any type of media capable of storing digital information as part of a forensic investigation. The Digital Forensic Research Workshop (DFRWS) Technical committee [63] has defined *digital forensic science* as below:

> *The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations.*

1.1 Digital forensics: the process

Digital forensics is multi-staged process starting with the identification of digital media from a scene (possible criminal) as potential evidence to the stage where it is presented as evidence by an expert witness in a court of law. The sequence of activities is illustrated at a high level in Fig. 1.

The very first stage of the digital forensic process is the *identification* of relevant digital evidence. This involves the identification of one or more sources of digital storage capable of storing digital information associated with the investigation at hand. Some examples of hardware that can provide digital evidence include hard disks on computer systems, random access memory cards, USB and other external sources of secondary storage, mobile phones, PDAs and so on. Once identified, evidence is acquired from the devices and forensically preserved.

By *acquisition*, we refer to the process of obtaining a binary bitwise copy of the entire contents of all digital media that are identified. The evidence thus acquired is *preserved* and standard hash signatures like MD5 or SHA1 is used to verify integrity of the digital evidence.

In a digital forensics investigation, investigators deal with acquiring digital records for examination. Digital records can vary in form and type. Documents on a computer, telephone contact list, lists of all phone calls made, trace of signal strength from the base station of a mobile phone, recorded voice and video files, email conversations, network traffic patterns and virus intrusions and detections are all examples of different types of digital records. In short, digital evidence encompasses:

a. User data
b. Metadata associated with user data
c. Activity logs; and possibly
d. System logs

User data pertains to data directly created or modified or accessed by one or more users involved in an investigation. Metadata pertains to data providing context of how, when, who and in what form the user data was created or modified or accessed. Activity logs are records of user activity by a system or application or both detailing specific actions conducted by one or more users and system logs pertain to variations in system behavior from the normal based on one or more actions conducted by the users.

Once the digital evidence is acquired, it is always necessary to make copies and conduct all forensic tests on such read-only copies, lest any activity tamper the data stored within the original sources [58, 59]. The digital evidence is then examined using one or more forensic tools. These forensic tools generally provide some form of file system abstraction to the digital evidence, such that their contents may be examined for trace of evidence. This stage is called *evidence examination* where the digital evidence sources are examined for their contents and possibly indexed for conducting searches. This definition is in accordance with

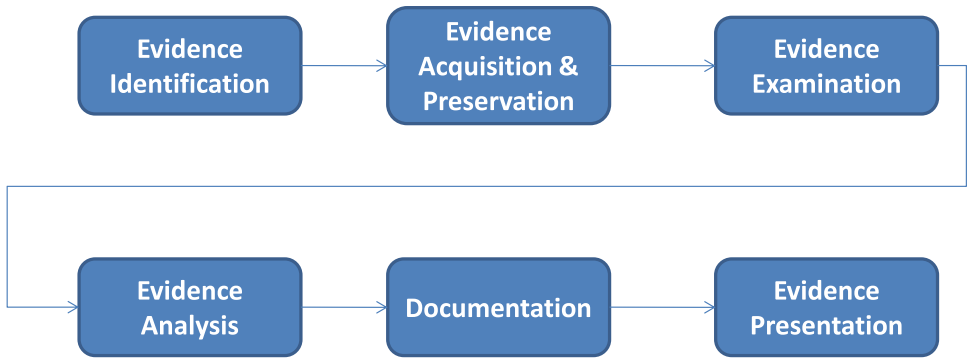

Fig. 1 Illustrating the digital forensic multi-staged process

Casey's view of the digital forensic examination process. Casey [44] defines forensic examination as the process of extracting information from digital evidence and making it available for analysis. In some cases, the examination of digital evidence may reveal some hidden or otherwise not explicit information which has to be extracted and subsequently analyzed. The act of identifying such information is termed evidence discovery.

After evidence examination and discovery, *forensic analysis* begins where the evidence sources and the discovered data are analyzed to determine the sequence of events leading to the reported crime under investigation. Casey [44] defines forensic analysis as the application of scientific methods and critical thinking to address the fundamental questions in an investigation: what, who, why, how, when and where. The individual stages are thoroughly documented and this *documentation* is presented in a court of law. Oftentimes, the presentation of digital evidence in court may be accompanied by an expert witness for testifying.

1.2 Research challenges

In a digital investigation, investigators deal with acquiring digital records for examination. Digital records can vary in forms and types. Documents on a computer, telephone contact list, list of all phone calls made, trace of signal strengths from base station of a mobile phone, recorded voice and video files, email conversations, network traffic patterns and virus intrusions and detections are all examples of different types of digital records. In the last decade, a large number of new digital devices have been introduced with advancements in digital technology. Lalis et al.'s [115] article on wearable computing provides a flavor for this changing digital scenario. These advances in digital technology and the relatively gradual progress in digital forensics have resulted in five major challenges [34, 74, 169]. They are:

1. Complexity problem
2. Diversity problem
3. Consistency and correlation;
4. Quantity or volume problem; and
5. Unified time-lining problem

Digital forensics has developed primarily a reactive field [74] which is a prime cause for these challenges, viz., advancements in digital forensics were triggered by crime first being committed on a computer or any digital device. Consequently, the field apparently seems to follows the trend rather than leading it.

Primarily, digital evidence is acquired in raw binary form which is too difficult for humans to understand and this leads to the complexity problem [34]. Forensic tools are hence used to interpret the raw digital evidence to address this problem. But currently, there is abundance in the number of forensic tools to interpret binary data in digital evidence and consequently, complexity has taken a backseat.

Of late, the amount of data collected during investigations has been steadily growing and it is becoming ineffective to analyze every single byte. The volumes and the heterogeneity of digital evidence have called for the application of data reduction techniques by grouping data into larger chunks or by removing known and irrelevant data prior to analysis. Garfinkel [74] also acknowledges the growing volumes of storage devices and makes an additional observation that in the presence of the multiple operating systems, file formats and devices, there is no standard way to examine and analyze all types of digital evidence—this has led to the diversity problem. Besides, with digital investigations often having to deal with multiple sources, investigators are required to examine consistency and correlate the evidence discovered across these sources leading to the consistency and correlation challenge. Garfinkel [74] observes that as there are no standards in data representation across these devices, many of which are proprietary, forensic examination and analysis become a significant challenge. Besides, the forensic tools currently in existence are designed to find pieces of digital evidence but not assist in investigations [78]; hence, majority of the analysis is conducted manually. Since different sources require different forensic tools, this has resulted in the *diversity problem*.

Despite this seemingly common structure of many file systems, these file systems are customized in the manner in which they store and process files. As a result, a file system partition which is defined as NTFS cannot process an EXT or a HFS partition. Another example of such a seemingly common structure for potential evidence sources is among logs; all log files have a set of fields and corresponding set of values, and they are used to record activities to tracking system behavior or users' activities. Nevertheless, not all logs can be interpreted the same way. Each log is customized to track specific activities and hence the events of a system log and a network can never be merged together. In other words, the semantics of the log is embedded in the log type which is lost when they are merged. Moreover, when multiple sources of digital evidence are identified for investigation, not only is it essential to analyze them, it is also essential to corroborate and correlate the data between these sources for consistency. For instance, if a user has visits a webpage, the visit creates a record in the user's browser history as well as the cookies. If the user accessed the webpage via a proxy, the proxy will also contain an entry corresponding to the visit. Hence, multiple logs may require to be corroborated during forensic analysis. This is the consistency and correlation problem.

With the rapid increase in the sizes of storage media, the volumes of digital evidence collected these days are

tremendously large [36]. Investigators are required to analyze enormous volumes of data in any given investigation and in the absence of sufficient automation, it is tedious work. Richard and Roussev [169] and Garfinkel [74] have also separately acknowledged the growing volume of digital evidence as a major concern. This growing volume of digital evidence is known simply as the *volume problem*. Marziale et al. [125] recognize the need to have efficient computing systems to run forensic tools, especially in distributed computing scenarios and propose a multi-threaded solution to carve[1] data from digital evidence. Liebrock et al. [122] propose a preliminary design for a terabyte dataset framework to accommodate the growing volumes of digital evidence by storing them in RAID arrays. The XIRAF architecture [6] automatically indexes content in digital evidence allowing investigators to query evidence.

Generating such a unified timeline across multiple sources of digital evidence presents many challenges [25, 30, 116, 182, 198]. Broadly speaking, some of these challenges are:

1. Time zone reference and timestamp interpretation
2. Clock skew, clock drift and synchronization; and
3. Syntax aspects

We refer to this as the *unified time-lining problem*. Coutaz et al. [61] argue that capturing contextual information retains the key to integrating different technology services. Context would allow the system to decide the most relevant evidence to be retained. The aim is to limit the investigation space by drawing boundaries on the evidence categories to restrict the tests conducted on these classes of evidence.

An increasing number of digital systems are getting integrated and there is a need to access and interpret the data from these systems in a uniform and consistent manner. Fundamentally, owing to the variety that the current digital systems exhibit, it is integral to identify or establish a common platform for digital data integration and analysis. Given the ever growing volumes of digital investigation cases, the success of this approach hinges on the ability to automate the process. The paper is organized as follows. In Sect. 2, we classify digital forensic research literature into 4 main categories and each subsequent section explores the details of the published works. We conclude in Sect. 7 with a brief summary and in Sect. 8, take a look at some areas which hold much promise.

[1] Carving is the process of identifying the file types using a string of bytes, called *magic numbers*, from an image and matching with a database of known magic numbers to recover deleted or partially deleted files [71].

2 Classification of research literature

In this section, we review the broad area of digital forensics to inform us of the state of the art developments and best practices in the field and with an aim of identifying unresolved research challenges in the field. Hosmer [99] calls for the need to standardize the concept of digital evidence to provide a common platform for investigators to perform forensic analysis. Drawing parallel from physical evidence acquisition process, Hosmer suggests adopting a methodology that is similar to how physical evidence are stored and organized. However, since digital evidences can be altered, copied or erased, he proposes the 4-point principles of authentication, integrity, access control and non-repudiation while handing digital evidence. Mercuri [131] outlines some of the major challenges facing the field of digital forensics:

i. scaling technology and the need to adapt scalable architectures
ii. need to adopt uniform certification programs and courses in digital forensics
iii. need for changes in the digital evidence permissibility laws in courts

Casey [45] discusses recent challenges set by network intrusions and suggests steps to manage security breaches. He calls for sophisticated digital evidence acquisition techniques, efficient methods to preserve evidence over long periods of time, effective tools for analysis and development of forensic theories to lay a stronger foundation for future analysis. Adelstein [1] presents an argument for the need to adopt new acquisition and analysis techniques for the growing number of live memory forensic analysis. Trends indicate that it is infeasible to always bring down a system to image the system and often investigators must rely on their ability to reliably image the memory and available storage drives for examination during an investigation. Increasingly, it appears that forensics must quickly learn to bridge the gap between what is necessary and what is available. However, in order to tackle such dynamic variety in digital data, there is need to abstract the evidence model and analyze its characteristics before further challenges can be identified.

Turner [202] states that when devices become more specialized, forensic examiners will require acquaintance with as many different processing tools to interpret the data they contain. This is owing to the fact that forensics is limited today as it can process captured information only as a single entity. Existing digital forensic tools are typically fine-tuned to capture and extract data from specific storage media. Some tools like EnCase and the Forensic Toolkit have sufficient intelligence built-into understand and

interpret a few different types but there is no tool in existence to date that can interpret all types of data. The common digital evidence storage format working group [58] has re-iterated the drawbacks with current forensic analysis tools in terms of not being able to cope with multiple proprietary image formats. The group emphasizes the need for introducing a common digital evidence storage format that is common to variety of evidence sources including hard disk images, network logs, proxy cache data, memory dumps, etc.

Current research in digital forensics can be classified into 4 major categories, viz. evidence acquisition and representation, evidence discovery and examination, digital forensic analysis and digital forensic process modeling. *Evidence acquisition* is concerned with identifying and acquiring digital data in a forensically secure manner from a variety of digital devices. This branch examines the forensic scope of data from different devices and presents new techniques and tools (both hardware and software) to acquire data from the field. The data so acquired is then carefully imaged into secure drives for data discovery and examination. *Evidence examination and discovery* deals with techniques to discover relevant data within the acquired sources and the software support needed to examine the contents using one or more forensic tools. Evidence examination deals with the extraction of information from digital evidence and makes it available of analysis [44]. The different forensic tools used generally provide some form of file system or schema support to the digital evidence sources enabling investigators to navigate through the sources examining their contents. *Digital forensic analysis* is the application of the scientific method and critical thinking to address the fundamental questions in an investigation: who, what, where, when, how and why [44]. The process involves the analysis of artifacts from one or more sources of digital evidence to determine the sequence of events and answer these fundamental questions in order to solve the crime that is being investigated.

Forensic analysis also involves using the fundamental principles underpinning the creation, modification, tamper and deletion of digital data on storage media and coming up with a logical sequence of events to explain the state of data in acquired evidence. *Digital forensic process modeling* deals with establishing theoretical backgrounds on the forensic process and defining procedures and processes that must be in place while guaranteeing integrity of evidence throughout an investigation. The modeling process also defines fundamental forensic principles for the development of new tools in forensics examination and analysis. In the following sections, we will deal with each category separately identifying the different published research in them.

3 Evidence acquisition and representation

Evidence acquisition, being the first step in a digital investigation has been thoroughly studied to understand where there is scope for data (potential digital evidence) and how it can be extracted. Several national governmental agencies have recognized the need to deal with increasing use of digital data and participated in efforts to define guidelines for their use and handling.

3.1 Standards and guidelines

The *National Institute of Justice* (NIJ) and the *Department of Justice* (DoJ) in the *United States of America* have laid down principles for first responders, where to search for evidence in a crime scene and how to go about acquiring data. The *National Institute of Standards and Technology* (NIST) has supported many such initiatives and has provided both tools and tool testing capability [147, 150–154] for evidence acquisition. The Association of Chief Police Officers (ACPO) [11] has published the *Good Practice Guide for Computer based Electronic Evidence* in the United Kingdom and Standards Australia [196] has laid down guidelines for the management of IT evidence in Australia. While there has been a general growth in awareness for acquiring digital evidence and different national standards have been published, the underlying principle in evidence acquisition remains the same. Typically, when a hard disk must be acquired, it is connected to a forensic system via a *write-blocker* and a binary image of the entire disk is taken. A write blocker is a hardware device or software tool that allows read-only access to the suspect disk to avoid tampering evidence and maintains data integrity. While it is a safe and secure method for hard disk acquisition and is applicable to all disk formats, the sheer volumes of hard disks today render the process tedious. Further, if a disk was purchased in a secondary market, as in many cases, often investigators acquire and analyze far too much data than necessary which amounts to precious lost time in an investigation. This can be attributed to the fact that such disks could contain irrelevant data, deleted, lost or otherwise, which would be captured by the acquisition tool. In such cases, improper formatting of secondary disks and possibly improper magnetization in the disks could result because of aging. Since in most cases the data are acquired in raw binary format, there are no reliable means to compress the size of the acquired data which renders the process cumbersome. Since then, however, several proprietary formats have been engineered to compress these images and manage size of data [59].

Since initially recognizing the need to acquire digital data and use it in digital investigations, research has paved the way for several new acquisition techniques and tools in

the public domain for evidence in different types of devices. Lyle [123] describes the functions of a hardware write blocker and describes how the NIST had come up with testing tools to validate their functionality. Garfinkel [79, 80] notes in many cases often investigators acquire and analyze far too much data than necessary which amounts to precious lost time in an investigation. This can be attributed to the fact that certain sources of digital evidence could contain irrelevant or deleted data which would be captured by the acquisition tool.

Since initially recognizing the need to acquire digital data and use it in digital investigations, research has paved the way for several new acquisition techniques and tools in the public domain for evidence in different types of devices. While acquisition was recognized as a straightforward process, it involved gathering a variety of different devices and data in several different formats, viz., raw binary format, expert witness format (EWF), advanced forensic format (AFF), Encase image file format and so on. The raw binary format is a purely binary image of the source. The EWF is the basis of the image file format created by EnCase. The Encase image file format is relatively compressed but proprietary image format used by Encase forensic tools.

3.2 AFF

Garfinkel [79] developed the AFF which is an opensource format exclusively for hard disk images. The AFF is partitioned into two-layers providing both abstraction and extended functionality. AFF's lower *data storage layer* describes how a series of name/value pairs are stored in one or more disk files in a manner that is both operating system and byte-order independent. AFF's upper *disk presentation layer* defines a series of name/value pairs used for storing images and associated metadata. It presents an opportunity for an investigator to capture all the information related to a disk and also allows recording of case related metadata. Garfinkel has developed the *afflib*[2] open source library to support AFF format, integrated in many open source forensic tools. However, the AFF is primarily designed for forensic images of hard disks and does not account for raw sources of digital evidence, such as files and logs and regular sources such as memory dumps and network packet captures. Cohen et al. [56] proposed the AFF4 by redesigning the AFF model to accommodate out-of-band information. AFF4 supports multiple secondary storage devices, new data types (including network packets and memory images), extracted logical evidence, and forensic workflow. The investigator may choose to extract relevant evidence from the encapsulation and conduct evidence examination and analysis. The raw binary format or EWF are the most popular imaging formats but they do not provide effective means to compress the acquired data which renders handling digital evidence in the later stages rather unwieldy.

3.3 Digital evidence bags (DEB)

Turner [202] proposes the DEB, an abstraction model for evidence when multiple source types are involved. This model accounts for including volatile memory and other forms of data that were being acquired in some investigations. DEB is a hierarchical evidence model, illustrated in Fig. 2. It consists of an open-ended TAG file containing information about:

1. *The collector of evidence,*
2. *Meta information about the evidence capture process,*
3. *List of evidence units (EUs) contained,*
4. *An expandable tag continuity block (TCB) providing history of evidence transaction record; and*
5. *A hash signature of the evidence bag.*

The index extension files list the files and folders contained in evidence and record the metadata information like file creation times, access times, modification dates and folder paths. Alternatively, it could contain the make, model, serial number and the manufacturer details as metadata for storage disks. The bag extension files contain the actual files obtained from the evidence site. These include the contents of the files in raw format. Turner demonstrates the use of the model in a network investigation and system administration task [205] and uses it for selective intelligent acquisition [204] from static storage devices. Masters and Turner [126] describe a method to represent data from magnetic swipe card readers using the DEB model.

Trends indicate that it is infeasible to always bring down a system to image it and often investigators must rely on their ability to reliably image the memory and available

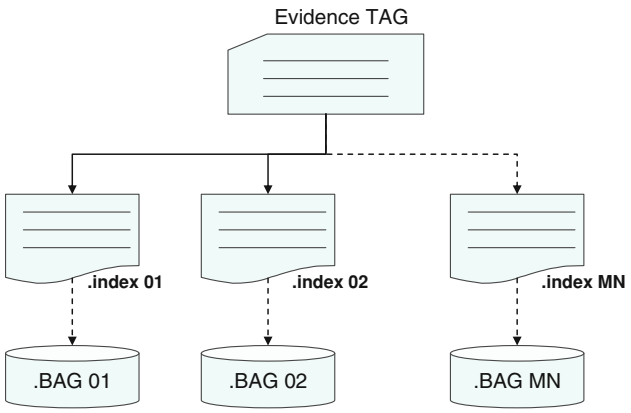

Fig. 2 Digital evidence bags

[2] http://www.afflib.org/.

storage drives for examination during an investigation using *online imaging*. This recent development is a significant detour from McKemmish's 4-point model [128] which assumes that digital evidence is always acquired after the system is turned off.

While DEB is an integral concept to maintain provenance information regarding digital evidence, it omits any reference to time zone information and timestamp representation, for instance when dealing with multiple digital evidence sources, from several time zones. Being able to abstract the representation of digital evidence and the time-correlation of multiple sources is one of the key research focal points for the future. While the TAG file definition is novel and helps verify evidence integrity, it is mainly intended to help human interpretation.

3.4 Secure digital evidence bags

Schatz and Clark [181] observe that the DEB model is monolithic in nature and proposed a representation approach to integrate metadata with evidence information and introduced an open DEB architecture called *sealed digital evidence bags* (SDEB), illustrated in Fig. 3. The SDEB, however, assumes the pre-existence of forensic domain ontology in the context of the case. The model is defined using resource description framework using Universal Resource Indicator (URI) [17] to tag the evidence bags. Each tag is uniquely identified with an identifier and they are immutable. When the analysis of primary evidence results in secondary evidence, a new evidence bag is created into which the details are stored. Hence, the existing evidence bags are untouched and unlikely to undergo modifications. In addition, each tag is also associated with a tag signature which records and stores the hash signature like SHA1 or MD5 to verify SDEB integrity before using it.

Each of the digital evidence models discussed above have provided significant advance over earlier research with regard to representing digital evidence and addressing the diversity problem to varying degrees. The DEB and the SDEB models emphasize data integrity, not forensic examination and analysis. Consequently, these models fail to address the volume, consistency and correlation and the unified time-lining problems. While the AFF4 addresses the diversity problem and provides certain level of consistency among digital evidence, it is designed using the ZIP specification as the container format does not record accurate time zone information.

Analysis requires a framework that superstructures the forensics investigation process and enables the inclusion of new tools to understand and interpret the data in a holistic manner. Such a framework would naturally support an abstraction model which retains data integrity while also allowing enhanced analysis capabilities by providing easier access to evidence encapsulated in these models.

3.5 Forensic acquisition tools

There have been several other efforts in advancing the state of the art in techniques for data acquisition from electronic devices. Gillam and Rogers [87] present the *FileHound* "field analysis" software for first responders. Adelstein and Joyce [2] propose *File Marshal* for automatic extraction of P2P data over a network. Kornblum [114] presents a methodology for forensic acquisition of Linux disk with odd number of sectors. LaVelle and Konrad [118] propose the *FriendlyRoboCopy* as a method for forensic preservation while acquiring data from a network. Carrier and Grand [38] describe a hardware based memory acquisition procedure while Schatz [180] presents a software based volatile memory capture using *BodySnatcher*. Schuster [183, 185]

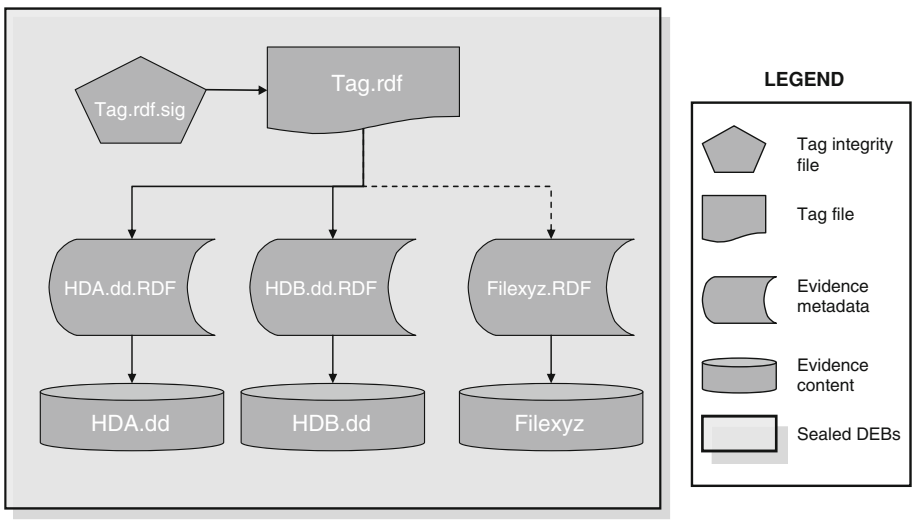

Fig. 3 Sealed digital evidence bags

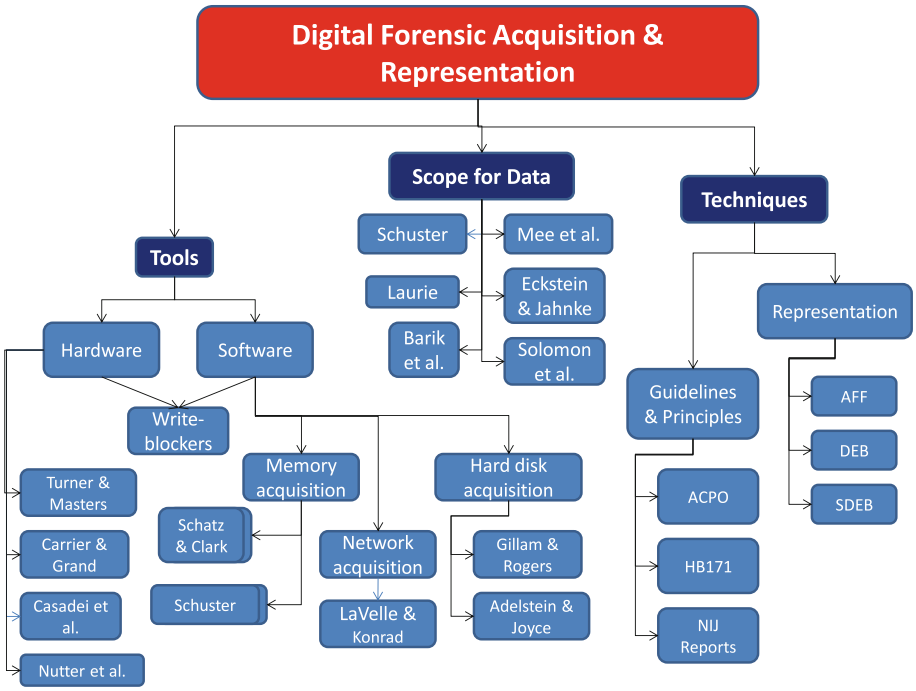

Fig. 4 Taxonomy of digital forensic acquisition and representation

examines the presence of processes and threads in Windows memory dumps and examines memory allocation strategies in Windows Operating systems, Solomon et al. [195] analyze user data persistence in physical memory and Mee et al. [130] have examined the Windows registry as a forensic artifact. Schuster [184] describes the Microsoft Vista event log format and studies its forensic capabilities and Murphey [145] presents a methodology for automated event log forensics combining multiple logs. Hargreaves et al. [93] describe the Windows Vista format and examine the challenges it poses to forensics while Park et al. [159] study data concealment and detection in Microsoft Office 2007 files. Eckstein and Jahnke [68] present a study on data hiding in journaling file systems, Gupta et al. [92] study hidden disk areas in a hard disk, Barik et al. [13] propose a methodology to preserve authentic date and timestamps in EXT2 file system for forensic purposes and Schatz et al. [182] propose a method for establishing timestamp provenance in digital evidence by corroborating system timestamps with a universal source such as NTP timestamps. Kenneally and Brown [107] present a risk sensitive approach to evidence collection while adhering to a legal framework and Johnston and Reust [106] highlight the importance of evaluating evidence in a network intrusion case study. Casadei et al. [42] present an overview of the SIM card forensics, Laurie [117] analyzes the forensic scope for Bluetooth technology and Nutter [155] examines TomTom records for identifying locations. Figure 4 illustrates the taxonomy of digital forensic acquisition and representation. The author recognizes that the figure is not exhaustive for space constraints but the author has tried to fit in as many literature works as possible.

4 Evidence discovery and examination

During Evidence examination, digital evidence sources are interpreted using one or more forensic tools. These forensic tools essentially provide a file system abstraction to the digital evidence source as defined by Carrier's forensic tool abstraction layers [34] which bridges the gap between the definition of a forensic process model and the development of associated forensic tools in aiding an investigation. Evidence discovery involves the process of *reliably*[3] recovering encrypted, hidden, lost or deleted data from the acquired evidence for further examination. Since raw data from digital evidence is often very difficult to understand, the data are translated through one or more layers of abstraction using forensic tools until they can be understood. The directory is an example of a file system abstraction while ASCII is a non-file system binary abstraction. The abstraction layer concept has been instrumental in the development of many forensic tools. The tool abstraction model proposed by Carrier is illustrated in Fig. 5.

[3] This involves the process of obtaining data as it is represented in a digital evidence source, without having to manipulate or modify any information contained on that evidence source.

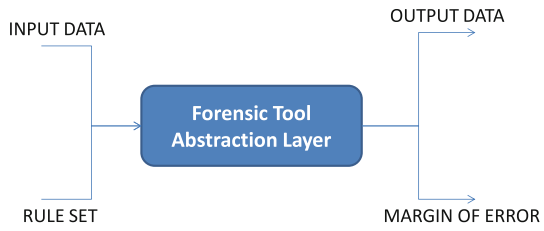

Fig. 5 Carrier's forensic tool abstraction concept

According to Carrier, the abstraction layers used in forensic tools can introduce two types of errors, namely, *tool implementation error* introduced by tool design errors and *abstraction error* which is introduced by the simplifications used to generate the tool. Pan and Batten [158] study the reproducibility of digital evidence that builds on this abstraction layer concept.

4.1 Forensic tools and frameworks

The forensic community has also witnessed the advent of many other tools for examining digital evidence from hard disk images, logs, network captures, memory dumps, mobile phones and so on. Corporates like *AccessData* and *Guidance software* came up with their own proprietary evidence storage formats and introduced the Guidance EnCase and AccessData FTK forensic tool suites for examining digital evidence. Sleuthkit [35], Pyflag [55], Wireshark [57], log2timeline,[4] tcpdump[5] and volatility[6] are a few equivalents from the opensource community.

Typically the evidence is represented as a binary large object with associated metadata and the tools provide read-only interfaces to the data at both text and hexadecimal abstraction level. With the development of research in this category, such tools are since supporting a variety of evidence formats. The common digital evidence storage format working group presents a comparative study [59] on the different evidence storage formats and the software support that they have on a suite of forensic toolkits.

The challenge we face today is that the forensic tools currently in existence are highly specialized for certain types of digital evidence. Sleuthkit for instance, only examines forensic images of hard disks and Wireshark can only examine network captures. There is an increasing level of awareness of this issue and we have some exceptions in our midst. Pyflag, for example, can examine forensic images, memory dumps, logs and network captures; and log2timeline can examine and timeline different types of logs. Nonetheless, even such tools haven't been able to achieve complete integration of multiple sources. Pyflag, for instance, can examine forensic hard disk images, memory dumps, network captures and logs. However, while Pyflag can support the examination of multiple sources of digital evidence, each source must be examined separately and the analysis must to be conducted by manually collating the different reports that Pyflag generates. The wide diversity amongst the various forensic tools for conducting forensic examination has also made it very cumbersome to integrate information and perform analysis of multiple sources of digital evidence in a unified manner.

Existing digital forensic tools are typically fine-tuned to capture and extract data from specific storage media. Some tools like EnCase and the Forensic Toolkit (FTK) support multiple file system types and a few digital evidence formats. Microsoft released a forensic toolkit called COFEE[7] to extract evidence from Windows computers but exclusively for law enforcement agencies. X-Ways Forensics[8] is a computer forensics software which integrates file system examination for multiple file systems, but it is proprietary and based on licensing. Nevertheless, current challenges in forensic analysis scale far beyond the realms of hard disk analysis and must also account for evidence provided from system and network logs, network captures, memory dumps and a large number of other digital devices.

Table 1 lists some of the most popular forensic toolkits and compares their capabilities. Of these, the Encase and FTK are commercial varieties while the remaining three are open source. While there are several other specialized tools to access and examine the contents of network packets, volatile memory, deleted files, and logs, they are not as comprehensive as the five listed in the table to qualify for a toolkit. Some of these specialized tools are tcpdump, Wireshark, volatility, rifiuti, Vinetto, and log2timeline.

The table is classified into 4 major categories based on interpretation and analysis of digital evidence, viz., binary access, representation and interpretation, metadata and evidence composition and within each category, the specific capabilities are listed. Binary access corresponds to the abstraction of digital evidence data at its lowest level and provides binary level access. Moreover, this access is read-only to ensure the integrity of the source of digital evidence during all stages of an investigation. Representation and interpretation corresponds to providing file level abstractions in file systems, log record abstractions in logs, process level abstractions in memory dumps and network packet abstractions in network packet captures. Many forensic tools also index the text at this abstraction level for subsequent querying and searching. Metadata corresponds

[4] http://log2timeline.net/.

[5] http://www.tcpdump.org/.

[6] https://www.volatilesystems.com/default/volatility.

[7] https://cofee.nw3c.org/.

[8] http://www.x-ways.net/forensics/index-m.html.

Table 1 Comparison of contemporary forensic toolkits and frameworks

	Binary access		Representation and interpretation			Metadata	Evidence composition	
	Binary abstraction	File system interpretation	Log analysis	Network analysis	Text indexing and Search	Metadata extraction	Multiple sources of DE (examination and analysis)	Identify correlations
Encase	√	√	×	×	√	Only FS metadata	Only FS images (exami-nation)	×
FTK	√	√	×	×	√	Only FS metadata	Only FS images (exami-nation)	×
Sleuthkit	√	√	×	×	√	Only FS metadata	Only FS images (exami-nation)	×
PyFlag	√	√	√	√	√	Only FS metadata	Only examination	×
OCFA	√	√	×	×	√	Only FS metadata	Only FS images (exami-nation)	×

to extracting the metadata from the relevant file and log abstractions for subsequent analysis and evidence composition corresponds to handling multiple and heterogeneous sources of digital evidence in order to facilitate the conduction of a holistic investigation.

Binary abstraction of digital evidence was established by Carrier [34] to overcome the complexity problem and all these forensic toolkits support it. In fact, all forensic tools must provide this basic support. File system based forensics is fairly established and so is the act of text querying and searching; all these forensic toolkits support these two functionalities.

In the metadata category, the identification of file system metadata, especially MAC timestamps has been deep-seated and hence the ability to extract file system metadata is common to these toolkits; however other types of metadata have been sparingly accessed or used, even on other forensic tools. Over the last decade, the design of forensic toolkits has principally been from the point of view of extracting all types of digital evidence that can be identified on a source [74, 169], consequently much of the task of putting the information discovered from evidence together and interpreting the semantics has been left to an investigator.

Recent advent of such tools, especially in the open-source community, is an acknowledgement of the importance associated with developing solutions that can integrate increasingly more number of digital evidence sources to tackle technological diversity. However, the ability to analyze and cross correlate information derived from one source across other sources is not supported in the architectures they build on. The examination and forensic analysis of digital evidence hence remain disconnected and analysis continues to be performed manually.

Carrier developed the Sleuthkit [35] that exports results to a browser interface (Autopsy) as HTML output. Cohen [55] extended the functionality of Sleuthkit and developed the Pyflag framework that can operate on forensic images, memory dumps, logs and network captures. Sleuthkit addresses the integration of file system analysis across multiple file systems and Pyflag integrates the examination of file systems, memory dumps, network packet captures and logs into a single framework. The *open computer forensic architecture*[9] (OCFA) developed by the Dutch National Police Agency[10] is another example on an integrated forensic architecture. However, OCFA only integrates the forensic image formats such as RAW, EnCase and EWF for file system examination. All these tools, i.e., Sleuthkit, Pyflag and OCFA, allow multiple sources of digital evidence to be examined simultaneously. However, the analysis needs to be conducted manually by an investigator using a search and browse interface.

4.2 Data carving

In several cases, it was found that deleted data or partial file data could help an investigation which gave rise to the new field of data carving. Carving is the process of identifying the file types using a string of bytes, called *magic numbers*, from an image and matching with a database of known magic numbers to recover deleted or partially deleted files [63]. The magic number is a constant used to identify a file format and is hence unique to each format. The DFRWS report of 2001 [63] defines,

> *Data carving is the process of extracting a collection of data from a larger data set. Data carving techniques frequently occur during a digital investigation when the unallocated file system space is analyzed to extract files. The files are "carved" from the unallocated space using file type-specific header and footer values. File system structures are not used during the process.*

[9] http://ocfa.sourceforge.net/.

[10] http://www.politie.nl/KLPD/.

Carving is done on a disk when the unallocated file system space is analysed to extract files because data cannot be identified due to missing of allocation info, or on network captures where files are "carved" from the dumped traffic using the same techniques [78, 79]. One drawback of this process on disks or images is that file-carving tools typically contain many false positives, hence tests must be done on each of the extracted files in order to check its consistency. A huge repository of such file types and headers are then incorporated into each forensic tool which then examines the section of data that need to be carved with the reference file signatures. Garfinkel has proposed a technique by controlling state space explosion to carve from AFF images [81]. Richard and Roussev [168] describe a high performance file carver called *Scalpel* carving files from hard disk images. The paper compares its performance in terms of speed and memory requirements with Foremost, a popular Linux file carver. Marziale et al. [125] propose a hyper threading scheme to improve digital forensic tool performance. The hyper threading architecture performance is analyzed in terms of time taken to carve a set of large volume hard disk images. Garfinkel [83] studies forensic feature extraction using file carving across 750 hard disk images and attempts to determine cross drive correlation. In [82], Garfinkel proposes a method for continuous fragmented file carving using fast object validation. Alvarez [7] proposes a method for using EXIF file headers for file carving in images. Since 2004, the opensource community[11] has been actively promoting the use of several forensic tools which perform specific tasks and can be operated in conjunction with one another. However, the analysis of digital evidence, especially in an automated manner, has continued to evade the forensic community.

4.3 Data hiding and steganography

We mentioned earlier that evidence examination is often accompanied by discovery of new information from within digital evidence and this is called evidence discovery. One such evidence discovery technique is the discovery of steganographic information. Steganography is the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message. Digital steganography may include hiding information inside document files, image files, programs or protocols. Media files are ideal for steganographic transmission because of their large size. Hosmer and Hyde [98] introduce the challenges posed by steganography and propose the saturation view technique to detect steganographic information from digital images. Lee et al. [119] present an approach for detecting image anomalies by combining computer graphics principles and AI reasoning. Image forgery has been classified into four categories, viz. deletion, insertion, photomontage and false captioning. The approach operates on the premise that if key objects (*known* a priori) can be identified in an image then reasoning can be employed to determine whether or not it has been tampered with. The approach segments a given image, computes the *importance map* on *regions of importance* and employs a rule based reasoning component to determine forgery status. While this work presents a novel combination of graphics and AI, such techniques are also equally important in detecting anomalies in other types of evidence data. Mead [129] from NIST examines the techniques used at the national software reference library for building a corpus of known software, file profiles and file signatures used by law enforcement. The Scientific Working Group on Digital Evidence has explored scope for digital evidence in Windows operating systems [178, 179].

4.4 Metadata in forensics

Metadata refers to data about the data that is stored within a source of digital evidence. Metadata can be defined at many levels, such as system metadata, file system metadata, application metadata, document metadata, email metadata, business metadata, geographical metadata and many more. Each type of metadata contains information describing aspects pertaining to the type they are attributed to. Metadata of a particular type provides certain context information that enables easy handling and management of the data contained and is hence very informative. For instance, file system metadata describes certain attributes as recorded by a file system regarding a particular file, such as its location, MAC timestamps, file size, owner and permissions. Similarly, application metadata records context as recorded by the application handling that file or artifact such as author, application version, format, and encoding. Thus, the term *metadata* is an umbrella definition to encompass all such different types of metadata. According to the *Sedona Principles for Addressing Electronic Document Production* [187],

> *metadata includes information about the document or file that is recorded by the computer (or digital device) to assist in storing and retrieving the document or file. The information may also be useful for system administration as it reflects data regarding the generation, handling, transfer and storage of the document or file within the computer (or digital device). Much of the metadata is neither created by nor normally accessible to a computer user.*

Broadly, file system metadata and application metadata are also often referred to as *external* and *embedded*

[11] http://www.opensourceforensics.org/tools.

metadata [188] since file system metadata is stored external to the document or file it describes and application metadata is embedded into it. In the traditional sense, metadata are only native to files and documents that reside on file systems. Nevertheless, log records and network packets also have some associated information that can be attributed the term metadata. Although logs and network packet captures themselves reside as files in a file system, the entries they contain are independent units that correspond to specific events. For instance, an entry in the IE history log, *index.dat*, would correspond to visiting a web page characterized by a URI. The attributes corresponding to this entry contain the timestamp of web page visit, the domain, the host server IP address, and so on. Similarly, an entry in a network packet capture corresponds to a network packet that was observed by the network capture sensor on a particular network belonging to a specific protocol containing a source and destination address. A network packet can be associated with a timestamp, source and destination IP addresses, the protocol for transfer and payload size. Such information may be treated as metadata for a log record or a network packet, as the case may be. Shankaranarayanan and Even [192] have discussed the different semantics metadata can possess under various contexts and how valuable it is to researchers despite the enigma it poses. Carrier and Spafford [39] have noted that metadata can be treated as the characteristics of a digital object. According to Carrier and Spafford, every digital object which is a discrete collection of digital data, is an evidence of at least one event and the metadata is a partial representation of the state of a digital object. Buchholz and Spafford [28] have examined the role of file system metadata in digital investigations and they note that despite the lack of quality and quantity of information stored in file system metadata, it plays a crucial role in reconstructing events.

Boutell and Luo have used EXIF metadata in digital photographs for classifying based on camera specifications [21] and to perform scene classification [23] and Alvarez [7] uses EXIF metadata in digital photographs to verify authenticity of a picture and determine whether it has been altered. Bohm and Rakow [20] discuss the different aspects of classifying multimedia documents based on document metadata. Multimedia documents can be classified into six orthogonal categories, viz., representation of media type, content description, content classification, document composition, document history and document location.

Castiglione et al. [49] highlight the type of information that can be obtained from document metadata on Microsoft Compound Document File Format (MCDFF) which may be of relevance in digital investigations. Garfinkel and Migletz [76] develop a tool for automatic metadata extraction from digital evidence. Rowe and Garfinkel [174] develop a tool that uses directory and file metadata to determine anomalous files on a large corpus. The tool uses fiwalk [78] to traverse the corpus and compute statistical characteristics on the numerical metadata and generate 204 output files based on which anomalous files such as misnamed files and duplicate copies of files were identified. Garfinkel et al. [75] propose an automated solution for the multi-user carved data ascription problem using file location information from the OS and training a classifier.

4.5 Digital timestamps and time-lining

A timestamp has a physical realization and a temporal interpretation [67]. The physical realization is an encoding as a pattern of bits while the temporal interpretation stipulates the meaning of the bit pattern, the calendar date and time to which the pattern corresponds. A timestamp is the record of the time, according to some reference clock, associated an event. Allen [4, 5] discusses the different representations of timestamps adopted in literature, including one where timestamps are logical timestamps only, merely a sequential numbering of events on a system.

With regard to metadata in logs and network packet captures, timestamps are the most popular type of metadata used in generating timelines [25, 57]. Often in network packet captures, the packets are organized according to the IP addresses and protocol in investigations involving network intrusion detection. Zander et al. [217] classify IP traffic based on statistical flow characteristics by filtering based on destination address and port. Snort[12] intrusion detection tool allows IP packets to be monitored and sequenced according to IP addresses. Jiang et al. [105] have proposed a coloring scheme to identify a remotely accessible server or process to detect provenance aware self-propagating worm contaminations. This scheme associates a unique color as a system-wide identifier to each remote server or process and that is inherited by all spawned child processes. The color also diffuses to other processes that interact with a colored process through read/write operations.

Weil [214] presents a method for correlating times and dates contained within a file to the modified, accessed, and created/change of status (MAC) times of the file. The method attempts to standardize the apparent file MAC times to the actual time. According to Weil, dynamic date and time stamp analysis relies on external independent sources of time within a file and the MAC times at a singular point in time. In the case study presented in this work, Weil correlates the MAC timestamps with the timestamps within the body of HTML pages. Increasing the number of independent sources enhances the reliability of the data and minimizes

[12] http://www.snort.org/.

CMOS limitations. While the method proposed is feasible for small sets of timestamps during analysis, a more comprehensive method is needed to address this challenge across multiple heterogeneous sources of digital evidence.

Boyd and Forster [25] describe the timestamp interpretation challenges associated with the Internet Explorer and time zone translations between UTC and local time. In their paper, Boyd and Forster describe a case study where investigators were wrongly accused of tampering with computer evidence based on misinterpreted timestamps. They discuss the Microsoft Internet Explorer time structures together with local and UTC time translation issues and suggest a checklist for examiners while interpreting timestamps. This work reinforces our expectations with regard to the challenges in timestamp interpretation and behavior across time zones.

Lamport [116] provides a precise characterization of causality in distributed systems (called the clock consistency condition) and a framework for explaining and reasoning about partial event ordering in distributed systems. The simplest way to implement the clock consistency condition is with "logical clocks" that Lamport introduced.

Gladyshev and Patel [90] formulate the event time-bounding problem and propose an algorithm for solving it when the causal order is known. They propose a sandwich algorithm to time bound an event when its causal relationship is known with respect to other events whose timestamps are available. Further, they attempt to shorten the time bound $[T^{B}_{max}, T^{B}_{min}]$ to the smallest value within which the event would have occurred. Willassen [215] proposes a similar formal approach using hypothesis based testing on timestamps to detect antedating.

Stevens [198] proposes the unification of timestamps from different sources by accounting for factors affecting the behavior of system clocks with respect to a global clock. Stevens proposes a global clock model that can account for these factors is used to simulate the behaviour of each independent clock. The clock models are used to remove the predicted clock errors from the time stamps to get a more realistic indication of the actual time at which the events occurred. All the time stamps from different sources are then unified using this global clock model onto a single time-line. In order to be able to unify all the digital events, two sets of information are required. Firstly, one needs to identify all the different clocks that were used and which time stamps were produced by each clock. Secondly, one needs to know the complete behaviour of each clock over the relevant time period. It is also necessary to have a full understanding of how time stamps are generated and their semantics.

Not all system clocks are always accurate. Since system clocks are based on a low frequency CMOS transistor, the clock drifts over several charging and discharging cycles and 1 s count no longer remains exactly 1 s. Schatz et al. [182] and Buchholz and Tjaden [31] have independently analyzed clock skew and clock drift across a system of clocks and their impact in determining exact time when recording system events. Koen and Olivier [111] discuss the information deficiency problem and the use of file timestamps from a UNIX file system in digital forensics. Chow et al. [53] propose a method for systematic evaluation of timestamp behavior on the NTFS file system.

Sarmoria and Chapin [177] present an approach for monitoring access to shared memory mapped files and Schuster [185] examines the impact of Windows memory allocation strategies on process and context persistence in memory. van Baar et al. [206] describe a method for recovering files mapped in memory and to link mapped file information process data. The paper presents a case for extracting such data which reduces the amount of unidentified data in memory dumps. The paper claims that 25 % of pages in memory dumps could be identified as part of mapped file. Morgan [144] examines the cause for deleted registry data and proposes a technique for recovering deleted data from Windows registry. Dolan-Gavitt [66] examines the structure of Windows registry and explores the use of tools to extract this data from memory dumps. The paper also describes a compelling attack that modifies cached registry and proposes a method to detect such attacks by examining memory. Petroni et al. [161] propose the *FATKit*, an extendable framework for extraction and analysis of volatile system memory. Harms [94] investigates system restore points in Windows XP and Arasteh and Debbabi [8] use the process logic to model extracted properties of memory stack and verify against model generated from program assembly code. Arasteh et al. [9] propose a model checking approach to the formalization of forensic analysis of logs. Properties of the model, attack scenarios and event sequences are expressed as formulae of a logic having dynamic, linear, temporal and modal characteristics. The model was then applied to a variety of system, user and network logs to detect an intrusion on a corporate network. Jansen and Ayers [103] provide an overview of PDA forensics and compare different present day tools in their capabilities and limitations.

4.6 Indexing and querying digital evidence

Alink et al. [7] propose XIRAF, a new XML based indexing and retrieval of stored digital evidence. The XIRAF architecture indexes into raw disk images storing them in annotated XML format. A query engine called XQuery is used to query into the XML database for evidence related information. However, this architecture is designed only to index and retrieve digital evidence and does not support any means for combining information from multiple types. Further, the architecture lacks

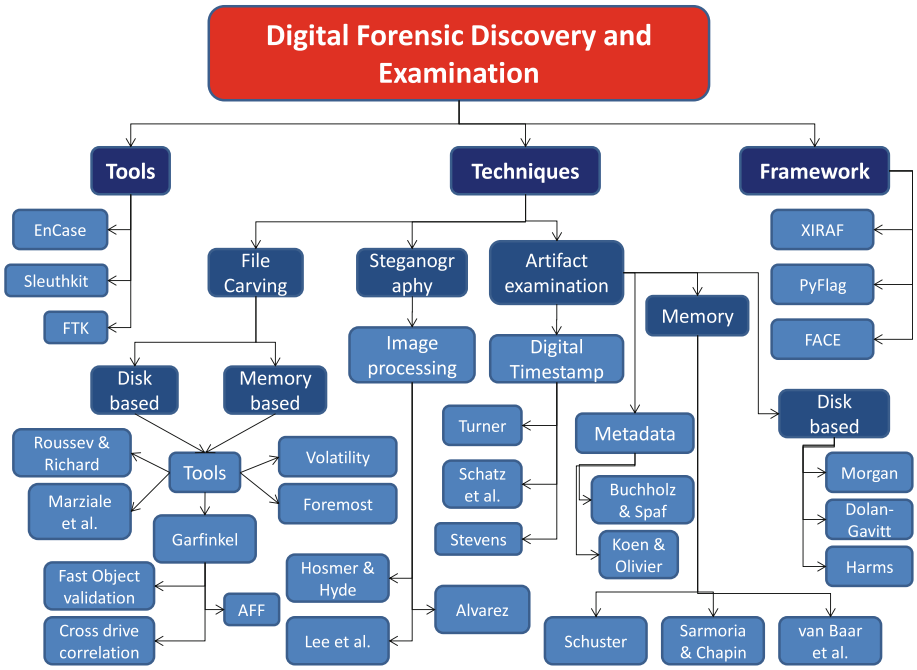

Fig. 6 Taxonomy of evidence discovery and examination

flexibility to extract content which may provide key correlations in data elsewhere. Richard et al. [171] propose the forensic discovery auditing module for storing and manipulating digital evidence using digital evidence containers. Beebe and Clarke [15] propose a new methodology to categorize forensic string search results by thematically clustering them to improve information retrieval effectiveness. The approach uses Kohonen self-organizing maps (SOM), an unsupervised neural network which learns the case themes and associated strings from an expert volunteer. The results of the search output are then evaluated based on query precision and recall ratios. While this approach provides significant benefits with respect to string searching, it is infeasible to have an expert classify each case prior to performing analysis. Besides, such an unsupervised model could take a long time learn the themes which is again not within practical boundaries.

Lee et al. [120] present a hardware base approach for improving performance of digital forensic tools. They propose the use of the Tarari content processor to implement a high speed search engine. They also describe the characteristics of the processor and how it can be exploited in the context of digital forensic analysis. Carrier [37] proposes a new methodology for volume analysis using multiple disk volumes. The paper addresses the concerns in accessing large data sets stored across several disks typically in RAID format.

Research in this area has independently addressed some of the challenges in evidence examination and discovery but continues to remain widely scattered. Besides, many of the implementations are customized to a particular operating platform. As a result, the extrapolation of an approach to make it generic is a rather long leap. There is a need to build upon forensic data discovery and examination techniques to develop new mechanisms for the integrated analysis of evidence and determine the sequence of events which would explain the state of data so acquired. Another aspect of research in the area is the lack of availability of forensic datasets for evaluation and validation. Often researchers have to rely on their ability to develop hypothetical case studies or synthetic datasets to validate research contributions. Since much of the research is developed on customized operating platforms, using a particular case study developed by one group to cross validate has not been very successful. Garfinkel [77] acknowledges this absence and has called for conscious efforts on the part of researchers to develop extensive datasets and contribute to the community through various channels. Garfinkel and his group at the Naval Postgraduate School have since developed the Digital Corpora[13] which is available to academics on request. Figure 6 illustrates the taxonomy of digital forensic discovery and examination.

5 Digital forensic analysis

In recent years there is widespread acknowledgement to focus research efforts in this area [74, 169]. While

[13] www.digitalcorpora.org/.

published research remains sparse, it hold much promise and most likely to witness developments in the years to come. One of the main aims of forensic analysis of digital evidence is the determination of possible reconstruction scenarios. In event reconstruction, the contents of digital evidence sources are analyzed to use the timestamps to set the time windows within which certain activities might have occurred. Additionally, the metadata or data about the contents are used in determining who created or accessed the contents and how they may have been created or accessed. Taking into consideration any pre-conditions that are essential for the existence of the contents would also contribute towards determining how and who could have created or accessed them. Such an exhaustive analysis eventually leads to mapping out a set of possible scenarios from which the investigator identify the most appropriate scenario based on other leads they may have.

5.1 Finite state approach and parametric reconstruction

Gladyshev and Patel [90] propose a finite state model approach for event reconstruction. They demonstrate that even a simple printer investigation problem can have exponential state space for analysis. In the context of current cases, clearly such a system is impractical and newer methods are needed to simplify the state space analysis. Carrier and Spafford [41] propose a method for analysis using the computer history model. However, like in the finite state model case, the application is not practical to current case complexities. Jeyaraman and Atallah [104] present an empirical study of automatic reconstruction systems and Khan et al. [110] propose a framework for post event timeline reconstruction using neural networks. Both research works use a set of network logs and train a neural network to learn the attributes of the logs in order to integrate the events into a single timeline.

5.2 Correlation and corroboration

Kornblum [113] presents a novel approach to identifying almost identical files using context triggered piecewise hashing. The aim of this paper is to automate detection of visual similarity between two files. The approach combines the rolling hash with the spamsum algorithm to compare the resultant signature and determine if similarity exists. Pal et al. [157] present an approach to detect file fragmentation and use sequential hypothesis testing to identify the fragmentation point. By using serial analysis the approach aims to minimize errors in detection. The approach maintains a base fragment as reference and determines whether a data block is joined or separated. Calhoun and Coles [33] examine the performance of Fisher's linear discriminant and longest common subsequence methods for predicting the type of file fragment. The work is aimed at improving file carvers by being able to reconstruct files when directory information is lost or deleted. The algorithms were compared across a set of 100 files whose header bytes were partially deleted or lost and the results are reported. Bogen and Dampier [19] propose a case domain modeling approach for large scale investigations and define case specific ontology using UML. Wang and Daniels [213] propose an evidence graph approach to network forensic analysis and build a correlation graph using network captures. Brinson et al. [27] a cyber forensics ontology and focuses on identifying the correct layers for specialization, certification and education within the domain. While the paper discusses the ontology to up to 5 levels in hierarchy, it is determined that this structuring is insufficient for forensic analysis which is far more diverse. Fei et al. [69] introduce SOM, which is an unsupervised neural network, for detecting anomalous human behavior in controlled networks. However, its immediate applicability to integrated forensic analysis is unclear.

Case et al. [43] propose the *FACE* framework for performing automatic correlations in forensic investigation. However, the framework is structured to only consider static and known relations in data (for example, linking network socket in memory to TCP requests in packet capture) especially when signification case detail is available a priori. Cohen [56] describes the PyFlag network forensic architecture, which is an open-source effort in providing a common framework for integrating forensic analysis from diverse digital sources. PyFlag, however sorely needs an analysis architecture to make the analysis more cohesive. Raghavan et al. [165] propose the forensic integration architecture and describe how to integrate evidence from different sources irrespective of the logical type of its contents. The integration architecture exploits existing technologies and brings varied evidence sources together under one roof to perform unified forensic analysis. The architecture attempts to provide multifarious interpretation and analysis of varied evidence types in a uniform manner independent of origination source and storage formats. It conceptualizes how to integrate evidence using content information from diverse evidence sources which is demonstrated through a case study.

In brief, the analysis category appears to be the most promising among the different categories. However, literature has only witnessed widely scattered efforts in different aspects of forensic analysis. We believe that a consistent and concerted effort by integrating different aspects of digital forensic analysis will perhaps be the future course of research. Figure 7 illustrates the taxonomy of digital forensic analysis.

6 Digital forensic process modeling

A digital forensic investigator typically has to contend with several digital image formats during an investigation. There can be general lack of cohesiveness in the manner in which the evidence acquisition, examination and analysis are handled. DFRWS 2001 [63] presents a consolidated report calling out the challenges facing the field in the coming years and demanding specific actions to advance the field and develop a better understanding of the digital forensic process.

Many digital forensic process models have been proposed in the literature. Primarily, these models deal with the definition of the general stages in a digital forensic investigation. According to McKemmish [128], the four broad stages involved in a digital forensic investigation are:

1. Identification of digital evidence;
2. Preservation of digital evidence;
3. Analysis of digital evidence; and
4. Presentation of digital evidence.

Carrier [34] introduces the forensic tool abstraction layer which classifies abstraction layers as lossy or lossless. Carrier and Spafford [39] study the digital investigation process and compare its functioning with a physical investigation and highlight the similarities. Carrier and Spafford [40] set up an event based investigation framework where the abstraction layer concept is extended to other sections of a digital investigation. Pan and Batten [158] study the reproducibility of digital evidence that builds on Carrier's abstraction layer concept. Gerber and Leeson [86] observe that computer-based input–output processes have not been thoroughly understood. They define computer-based IO simply as a sequence of translations followed by transport of data. They propose the layered Hadley model for IO which follows a layered abstracted approach. While the Hadley model accurately models the IO on a single computer, current trends in digital forensics have called for extension of this concept to multiple computers simultaneously and over distributed networks. Besides, digital evidence includes data collected from network logs, proxy caches, memory dumps and therefore, a more comprehensive framework is essential to provide holistic understanding.

As early as 2003, Mocas [142] has identified three main challenges that researchers need to overcome to advance the field of digital forensics from a theoretical standpoint. These challenges are:

1. Scaling technology and the need to adapt scalable architectures;
2. Need to adopt uniform certification programs and courses in digital forensics;
3. Need for changes in the digital evidence permissibility laws in courts.

Leighland and Krings [121] present formalization to digital forensics using a hierarchical elements model. Beebe and Clarke [14] argue the need for an objective based framework for digital forensic process and divide the process into six stages and propose a 2-tier hierarchical objectives framework. The six stages defined by this work are

1. *preparation,*
2. *incident response,*
3. *data collection,*
4. *data analysis,*
5. *presentation of findings; and*
6. *incident closure*

The framework further breaks down these stages into sub-stages (called sub-phases) and list out objectives at each stage for typical investigations.

Carrier and Spafford's computer history model [41] was one of the first works that attempted to formalize digital forensics using a finite state automaton. However, they concluded that the model is computationally infeasible

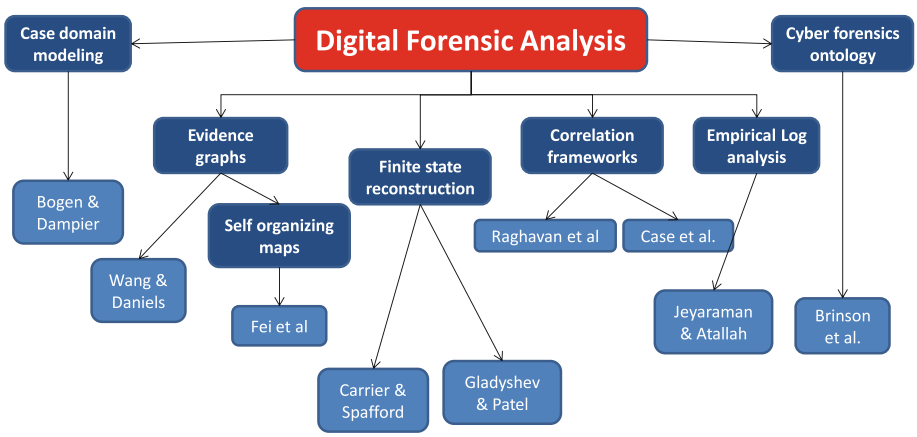

Fig. 7 Taxonomy of digital forensic analysis

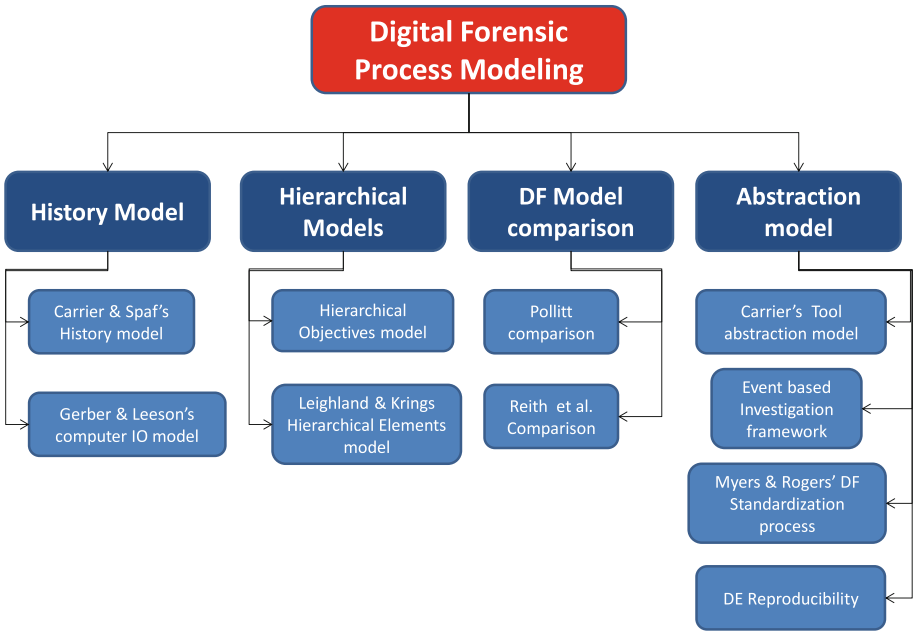

Fig. 8 Taxonomy of digital forensic process modeling

owing to state space explosion. Hosmer [98, 99] emphasizes the importance of chain-of-custody equivalents in the digital world and calls for auditing every operation conducted on digital evidence from digital devices. Since data on digital devices can be altered, copied or erased, Hosmer proposes the following 4-point principles:

- authentication,
- integrity,
- access control; and
- non-repudiation

…while handing digital evidence. The significance of this concept is reinforced by Turner's DEB [202]. Turner focuses on these 4 aspects from the standpoint of forensic acquisition and draws parallel from physical investigations to define DEBto record provenance information.

Myers and Rogers [146] can for the need to standardize the forensic investigation process and present an argument for achieving this through education and certification. Pollitt [162] presents an annotated bibliography of the different digital forensic models and examines their legal constraints while Reith et al. [166] present an independent examination of the digital forensic models and analyze its implications in the context of the report from DRFWS 2001. Figure 8 illustrates the taxonomy of digital forensic process modeling.

7 Summary

In summary, the digital forensic literature has diversified significantly since its importance was first recognized in early 2000 as a distinct field of study. The digital forensic process is multi-staged which involves the collection of digital evidence from one of multiple crime scenes, called as evidence acquisition. This is followed up by digital forensic examination of the contents of the evidence using forensic toolkits which provide various levels of abstractions to data. This process also serves to discover hidden, deleted or lost data within the contents and detect and decrypt encrypted data. Each file has associated metadata that can be extracted using software support for analysis [44]. The digital forensic analysis covers the realm of analyzing data to understand the set of possible explanations and associated logical sequences of events which explain the state of data in digital evidence. Digital forensic process modeling has attempted to provide overall growth to the area by proposing new theories and principles for the developments of methodologies and forensic tools in the digital investigation process. The overall taxonomy is illustrated in Fig. 9. The author acknowledges that it is not exhaustive for space reasons but the figure attempts to provide a generic categorization of the different areas of research and identify significant classes of published research.

The recent technological advancements have resulted in significant increase in the volumes of digital evidence being acquired and analyzed in a digital investigation. As a result, the process is not only getting tiresome but humanly impossible. There is urgent need for methodologies and approaches that automate much of the preliminary stages. Both the opensource community and proprietary vendors have recognized this growing need and developed a suite of

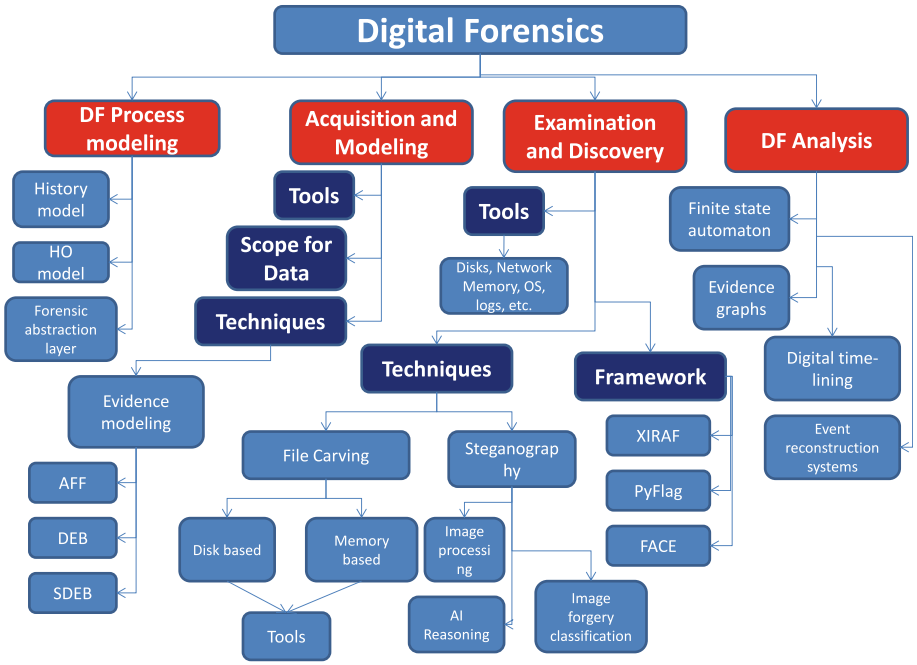

Fig. 9 Summarizing the taxonomy of digital forensic research

tools that can work with each other, however, the extent of cooperation needs to be expanded further. It is evident that research in this space cannot remain isolated, viz., already researcher have started developing tools and methods that combine acquisition with examination or some examination with preliminary analysis of digital evidence. Researchers are presently attempting to integrate multiple forensic and other analysis tools, primarily within a single framework to achieve this task and it has been a major victory. We are sure that in future, more such methodologies and tool development are there to be witnessed.

8 The road ahead…

Going forward, the volume challenge [34, 74, 169] remains the single largest challenge to conquer in the near future as we get accustomed to deal with terabytes of digital evidence on a daily basis. While computation power and hardware performance continues to grow, it is unlikely to challenge the growing volumes of digital evidence. Researchers have acknowledged the need to move from the space of developing tools to extract all data, to the space where the evidence is correlated towards solving an investigation [74, 169]. While the types of digital devices have grown exponentially, methinks there is significant potential to identify correlations in such information. One of the ways to manage the volume challenge would be to recognize this correlation across multiple digital evidence sources and automatically associate such evidence items which could lead to a reduction in the number of items for individual analysis. Furthermore, this approach could also facilitate automatic corroboration which is an integral component of forensic analysis. Notwithstanding the targeted and customized applications in the past, this approach has shown promising results [37, 80].

The literature recognizes the need for a comprehensive analysis framework which can adopt and support interpretation of a variety of digital evidence sources. There is an abundance of metadata in today's digital systems and literature recognizes its value to digital forensics, particularly with regard to event reconstruction. Metadata provides what can be considered to be situational information to determine under what context events transpired. Besides, metadata transcends data and formats and hence can bridge the diversity challenge naturally. Sequencing these events across multiple diverse digital evidence sources can also provide a solution to the unified time-lining challenge and provide an investigator a holistic view of all the events across all digital evidence sources which can be very valuable during an investigation.

References

1. Adelstein F (2006) Live forensics: diagnosing your system without killing it first. Commun ACM 49(2):63–66
2. Adelstein F, Joyce RA (2007) FileMarshal: an automatic extraction of peer-to-peer data, digital investigation. In: Proceedings of the 7th annual digital forensic research workshop (DFRWS'07). Digit Investig 4(Supplement 1):S43–S48

3. Agrawal N, Bolosky WJ, Douceur JR, Lorsch JR (2007) A five-year study of file system metadata. ACM Trans Storage 3(3): 9:1–9:32
4. Allen J (1983) Maintaining knowledge about temporal intervals. Commun ACM 26(11):832–843
5. Allen J (1991) Time and time again: the many ways to represent time. Int J Intell Syst 6(4):1–14
6. Alink W, Bhoedjang RAF, Boncz PA, de Vries AP (2006) XIRAF—XML-based indexing and querying for digital forensics. In: The proceedings of the 6th annual digital forensic research workshop (DFRWS'06). Digit Investig 3(Supplement 1):50–58
7. Alvarez P (2004) Using extended file information (EXIF) file headers in digital evidence analysis. Int J Alvarezal Evidence 2(3):1–5
8. Arasteh A R, Debbabi M (2007) Forensic memory analysis: from stack and code to execution history, digital investigations. In: Proceedings of the 7th annual digital forensic research workshop (DFRWS'07). Digit Investig 4(Supplement 1):S114–S125
9. Arasteh AR, Debbabi M, Sakha A, Saleh M (2007) Analyzing multiple logs for forensic evidence, digital investigations. In: Proceedings of the 7th annual digital forensic research workshop (DFRWS'07). Digit Investig 4(Supplement 1):S82–S91
10. Arthur K, Olivier M, Venter H (2007) Applying the biba integrity model to evidence management. paper presented at the digital forensics; advances in digital forensics III. In: IFIP international conference on digital forensics, Orlando
11. Association of Chief Police Officers (ACPO) (2003) Good practice guide for computer based electronic evidence. NHTCU Publications, London, pp 1–51
12. Australian Computer Emergency Response Team (AusCERT) (2006) 2006 Australian Computer Crime and Security Survey. AusCERT & Australian High Tech Crime Center (AHTCC). ISBN 1-86499-849-0
13. Barik MS, Gupta G, Sinha S, Mishra A, Mazumdar C (2007) Efficient techniques for enhancing forensic capabilities of Ext2 file system. Digit Investig 4(Supplement 1):55–61
14. Beebe NL, Clark JG (2005) A hierarchical, objectives-based framework for the digital investigations process. Digit Investig 2(2):147–167
15. Beebe NL, Clark JG (2007) Digital forensic text string searching: improving information retrieval effectiveness by thematically clustering search results. Digit Investig 4(Supplement 1):49–54
16. Berghel H (2007) Hiding data, forensics and anti-forensics. Commun ACM 50(4):15–20
17. Berners Lee T, Fielding R, Masinter L (1998) Uniform resource identifiers (URI), general syntax http://www.ietf.org/rfc/rfc2396.txt. Accessed 20 Mar 2008
18. Bogen AC, Dampier DA (2005) Unifying computer forensics modeling approaches: engineering perspective. In: Proceedings of the first international workshop on systematic approaches to digital forensic engineering (SADFE'05). IEEE Publication, Taipei
19. Bogen AC, Dampier DA (2005) Preparing for large scale investigations with case domain modeling. Paper presented at the 5th annual digital forensic research workshop (DFRWS'05), New Orleans
20. Bohm K, Rakow TC (1994) Metadata for multimedia documents. SIGMOD Rec. 23(4):21–26
21. Boutell M, Luo J (2004) Photo classification by integrating image content and camera metadata. In: 17th international conference on pattern recognition (ICPR'04), vol 4, Cambridge, pp 901–904
22. Boutell M, Luo J (2004) Incorporating temporal context with content for classifying image collections. In: 17th international conference on pattern recognition (ICPR'04) vol 2, Cambridge, pp 947–950
23. Boutell M, Luo J (2004) Bayesian fusion of camera metadata cues in semantic scene classification. In: IEEE computer society conference on computer vision and pattern recognition (CVPR'04), vol 2, Washington, pp 623–630
24. Boutell M, Luo J (2005) Beyond pixels: exploiting camera metadata for photo classification. Pattern Recognit Image Underst Photogr 38(6): 935–946. doi:10.1016/j.patcog.2004.11.013
25. Boyd C, Forster P (2004) Time and date issues in forensic computing—a case study. Digit Investig 1(1):18–23
26. Brand A, Daly F, Meyers B (2003) Metadata demystified. The Sheridian and NISO Press, http://www.niso.org/standards/resources/Metadata_Demystified.pdf, pp 1–19. ISBN: 1-880124-59-9
27. Brinson A, Robinson A, Rogers M (2006) A cyber-forensics ontology: creating a new approach to studying cyber forensics. Digit Investig 3(Supplement 1): S37–S43
28. Buchholz F, Spafford EH (2004) On the role of system metadata in digital forensics. Digit Investig 1(1):298–309
29. Buchholz F, Spafford EH (2007) Run-time label propagation for forensic audit data. Comput Secur 26(2007):496–513
30. Buchholz F (2007) An improved clock model for translating timestamps, JMU-INFOSEC-TR-2007-001. James Madison University, Madison
31. Buchholz F, Tjaden B (2007) A brief history of time. In: Proceedings of the 7th annual digital forensic research workshop (DFRWS'07). Digit Investig 4S:S31–S42
32. Burke P, Craiger P (2007) Forensic analysis of Xbox consoles. Paper presented at the digital forensics. Advances in digital forensics III. In: IFIP international conference on digital forensics, Orlando
33. Calhoun WC, Coles D (2008) Predicting the types of file fragments. In: Proceedings of the 8th annual digital forensic research workshop (DFRWS'08). Digit Investig 5(1):S14–S20
34. Carrier BD (2003) Defining digital forensic examination and analysis tools using abstraction layers. Int J Digit Evidence (IJDE) 1(4):1–12
35. Carrier BD (2003) Sleuthkit. http://www.sleuthkit.org/sleuthkit/. Accessed 12 July 2011
36. Carrier BD (2005) File system forensic analysis. Addison Wesley, Upper Saddle River. ISBN 0-32-126817-2
37. Carrier BD (2005) Volume analysis of disk spanning multiple volumes. Digit Investig 2(1):78–88
38. Carrier BD, Grand J (2004) A hardware-based memory acquisition procedure for digital investigations. Digit Investig 1(1):50–60
39. Carrier BD, Spafford EH (2003) Getting physical with the digital investigation process. Int J Digit Evidence 2(2):1–20
40. Carrier BD, Spafford EH (2004) An event-based digital forensic investigation framework. Paper presented at the 4th annual digital forensic research workshop (DFRWS'04), Lafayette
41. Carrier BD, Spafford EH (2006) Categories of digital investigation analysis techniques based on the computer history model. In: The proceedings of the 6th annual digital forensic research workshop (DFRWS'06). Digit Investig 3(Supplement 1):121–130
42. Casadei F, Savoldi A, Gubian P (2006) Forensics and SIM cards: an overview. Int J Digit Evidence 5(1):1–21
43. Case A, Cristina A, Marziale L, Richard GG, Roussev V (2008) FACE: automated digital evidence discovery and correlation. In: Proceedings of the 8th annual digital forensic research workshop (DFRWS'08). Digit Investig 5(Supplement 1):S65–S75

44. Casey E (2011) Digital evidence and computer crime: forensic science, computers and the internet. Academy Press Publications, London. ISBN 978-0-12-374268
45. Casey E (2006) Investigating sophisticated security breaches. Commun ACM 49(2):48–54
46. Casey E (2007) Digital evidence maps—a sign of times. Digit Investig (Editorial) 4(1):1–2
47. Casey E (2007) What does "forensically sound" mean? Digit Investig (Editorial) 4(1):49–50
48. Casey E (2009) Timestamp misinterpretations in file systems. http://blog.cmdlabs.com/tag/timestamps/. Accessed 12 July 2011
49. Castiglione A, De Santis A, Soriente C (2007) Taking advantages of a disadvantage: digital forensics and steganography using document metadata. J Syst Softw 80(5):750–764
50. Choo Kim-Kwang R (2010) Cloud computing: challenges and future directions. Trends and issues in crime and criminal justice No. 400. Australian Institute of Criminology, Canberra. ISSN 1836-2206
51. Choo Kim-Kwang R (2011) Cyber threat landscape faced by financial and insurance industry. Trends and issues in crime and criminal justice No. 408. Australian Institute of Criminology, Canberra. ISSN 1836-2206
52. Choi Kan-San, Lam EY, Wong KKY (2006) Source camera identification using footprints from len aberration. Proceedings of the SPIE-IS&T Electronic Imaging SPIE 6069:60690J-1–60690J-8
53. Chow K, Law F, Kwan M, Lai P (2007) The rules of time on NTFS file system. In: Proceedings of the 2nd international workshop on systematic approaches to digital forensic engineering, Seattle
54. Ciardhuain SO (2004) An extended model for cybercrime investigations. Int J Digit Evidence 3(1):1–22
55. Cohen MI (2008) PyFlag—an advanced network forensic framework. In: Proceedings of the 8th annual digital forensic research workshop (DFRWS'08). Digit Investig 5(Supplement 1):S112–S120
56. Cohen MI, Garfinkel S, Schatz B (2009) Extending the advanced forensic format to accommodate multiple data sources, logical evidence, arbitrary information and forensic workflow. In: Proceedings of the 9th annual digital forensic research workshop (DFRWS'09). Digit Investig 6:S57–S68
57. Combs G (1998) Wireshark—network protocol analyzer. http://www.wireshark.org/about.html. Accessed 12 July 2011
58. Common Digital Evidence Storage Format Working Group (CDESF-WG) (2006) Standardizing digital evidence storage. Commun ACM 49(2):67–68
59. Common Digital Evidence Storage Format Working Group (CDESF-WG) (2006) Survey of disk image storage formats. Paper presented at the 6th annual digital forensic research workshop (DFRWS'05), New Orleans, pp 1–18
60. Computer Security Institute (2010/11) Computer crime and security survey. In: 15th Annual Computer Crime survey (2010, GoCSI). https://cours.etsmtl.ca/log619/documents/divers/CSIsurvey2010.pdf. Accessed 8 Oct 2012
61. Coutaz J, Crowley JL, Dobson S, Garlan D (2005) Context is key. Commun ACM 48(3):49–53
62. Dennen VP (2005) Looking for evidence of learning: assessment and analysis methods for online discourse. Paper presented at the cognition and exploratory learning in digital age: CELDA, Lisbon
63. DFRWS Technical Committee (DFRWS) (2001) A road map for digital forensic research: DFRWS Technical Report. DTR-T001-01 FINAL
64. Denecke K, Risse T, Baehr T (2009) Text classification based on limited bibliographic metadata. In: Proceedings of the fourth IEEE international conference on digital information management, ICDIM 2009, Ann Arbor, pp 27–32. ISBN 978-1-4244-4253-9
65. Ding X, Zou H (2011) Time based data forensic and cross reference analysis. In: Proceedings of the ACM symposium on applied computing 2011, TaiChung, Taiwan, pp 185–190. ISBN: 978-14503-0113-8
66. Dolan-Gavitt B (2008) Forensic analysis of windows registry in memory. In: Proceedings of the 8th annual digital forensic research workshop (DFRWS'08). Digit Investig 5(Supplement 1):S26–S32
67. Dyreson CE, Snodgrass RT (1993) Timestamps semantics and representation. J Inf Syst 18(3):143–166
68. Eckstein K, Jahnke M (2005) Data hiding in journaling file systems. Paper presented at the 5th annual digital forensic research workshop (DFRWS'05), New Orleans
69. Fei BKL, Eloff JHP, Olivier MS, Venter HS (2006) The use of self-organising maps for anomalous behaviour detection in a digital investigation. In: Forensic science international 17th triennial meeting of the international association of forensic sciences 2005, Hong Kong. Forensic Sci Int 162(1–3), 33–37
70. Fernandez E, Pelaez J, Larrondo-Petrie M (2007) Attack patterns: a new forensic and design tool. Paper presented at the digital forensics: advances in digital forensics III: IFIP international conference on digital forensics, Orlando
71. FICCI Indian Risk Survey (2012) FICCI & Pinkerton C&I India Ltd. 2012 Risk Survey. www.ficci.com/SEDocument/20186/IndiaRiskSurvey2012.pdf. Accessed 8 Oct 2012
72. Fu Z, Sun X, Liu Y, Li Bo (2011) Forensic investigation of OOXML format documents. Digit Investig 8(1):48–55
73. Gallup Politics (2010) 2010 Gallup computer crime survey. http://www.gallup.com/poll/145205/new-high-households-report-computer-crimes.aspx. Accessed 8 Oct 2012
74. Garfinkel SL (2010) Digital forensic research: the next 10 years. In: Proceedings of the 10th annual conference on digital forensic research workshop (DFRWS'10). Digit Investig 7:S64–S73
75. Garfinkel SL, Parker-Wood A, Huynh D, Migletz J (2010) An automated solution to the multiuser carved data ascription problem. IEEE Trans Inf Forensics Secur 5(4):868–882
76. Garfinkel SL, Migletz J (2009) New XML-based files: implications for forensics. IEEE Secur Privacy Mag 7(2):38–44
77. Garfinkel SL, Farrell P, Roussev V, Dinolt G (2009) Bringing science to digital forensics with standardized forensic corpora. In: Proceedings of the 9th annual conference on digital forensic research workshop (DFRWS'09). Digit Investig 6:S2–S11
78. Garfinkel SL (2009) Automating disk forensic processing with Sleuthkit, XML and Python. In: Proceedings of the 2009 fourth international IEEE workshop on systemmatic approaches to digital forensic engineering (SADFE 2009), Berkeley, pp 73–84. ISBN: 978-0-7695-3792-4
79. Garfinkel SL (2006) AFF: a new format for storing hard drive images. Commun ACM 49(2):85–87
80. Garfinkel SL (2006) Forensic feature extraction and cross drive analysis. Digit Investig 3(Supplement 1):S71–S81
81. Garfinkel SL, Malan D, Dubec K, Stevens C, Pham C (2006) Advanced forensic format: an open extensible format for disk imaging. In: Olivier M, Shenoi S (eds) Proceedings of the second annual IFIP WG 11.9 international conference on digital forensics, advances in digital forensics II. Springer, Boston, pp 17–31. ISBN: 0-387-36890-6
82. Garfinkel SL (2007) Carving contiguous and fragmented files with fast object validation. Digit Investig 4(Supplement 1):S2–S12
83. Garfinkel SL (2009) Providing cryptographic security and evidentiary chain-of-custody with the advanced forensic format library and tools. Int J Digit Crime Forensics 1(1):1–28

84. Gehani A, Reif J (2007) Super-resolution video analysis for forensic investigations. Paper presented at the digital forensics: advances in digital forensics III: IFIP international conference on digital forensics, Orlando
85. Geiger M (2005) evaluating commercial counter forensic tools. Paper presented at the 5th annual digital forensic research workshop (DFRWS'05), New Orleans
86. Gerber M, Leeson J (2004) Formalization of computer input and output: the Hadley model. Digit Investig 1(3):214–224
87. Gillam WB, Rogers M (2005) FileHound: a forensics tool for first responders. In: Proceedings of the 5th annual digital forensic research workshop (DFRWS'05), New Orleans
88. Gilligan J (2001) Beating the daylight savings Time bug and getting the correct file modification times. Code project—date and time. http://www.codeproject.com/KB/datetime/dstbugs.aspx. Accessed 12 July 2011
89. Gladney HM (2006) Principles for digital preservation. Commun ACM 49(2):111–116
90. Gladyshev P, Patel A (2004) Finite state machine approach to digital event reconstruction. Digit Investig 1(2):130–149
91. Gloe T, Bohme R (2010) The Dresden Image database for benchmarking digital image forensics. In: Proceedings of the ACM symposium on applied computing 2010 (SAC 2010), Sierre. ISBN 978-1-60558-639-7
92. Gupta MR, Hoeschele MD, Rogers MK (2006) Hidden disk areas: hPA and DCO. Int J Digit Evidence 5(1):1–8
93. Hargreaves C, Chivers H, Titheridge D (2008) Windows vista and digital investigations. Digit Investig 5(1):34–48
94. Harms K (2006) Forensic analysis of system restore points in microsoft windows XP. In: Proceedings of the 6th annual digital forensic research workshop (DFRWS'06). Digit Investig 3(1):151–158
95. Hartong M, Goel R, Wijeskera D (2007) A framework for investigating railroad accidents. Paper presented at the digital forensics; advances in digital forensics III: IFIP international conference on digital forensics, Orlando
96. Hearst MA (2006) Clustering versus faceted categories for information exploration. Commun ACM 49(4):59–61
97. Hoepmann J-H, Jacobs B (2007) Increased security through open source. Commun ACM 50(1):79–83
98. Hosmer C, Hyde C (2003) Discovering covert digital evidence. Paper presented at the 3rd annual digital forensic research workshop (DFRWS'03), Cleveland
99. Hosmer C (2006) Digital evidence bag. Commun ACM 49(2):69–70
100. Huang H-C, Fang W-C, Chen S-C (2008) Copyright protection with EXIF metadata and error control codes, security technology. In: International conference on security technology 2008, Sanya, pp 133–136
101. Ieong RSC (2006) FORZA—digital forensics investigation framework that incorporate legal issues. In: The proceedings of the 6th annual digital forensic research workshop (DFRWS'06). Digit Investig 3(Supplement 1):29–36
102. Jain AK, Ross A (2004) Multibiometric systems. Commun ACM 47(1):34–40
103. Jansen W, Ayers R (2005) An overview and analysis of PDA forensic tools. Digit Investig 2(2):120–132
104. Jeyaraman S, Atallah MJ (2006) An empirical study of automatic event reconstruction systems. In: Proceedings of the 6th annual digital forensic research workshop (DRFWS'06). Digit Investig 3(Supplement 1):S108–S115
105. Jian X, Walters A, Xu D, Spafford E, Buchholz F, Wang Y (2007) Provenance-aware tracing of worm break-in and contaminations: a process coloring approach. In: Proceedings of the 24th IEEE international conference on distributed computing systems, (ICDCS 2006), Lisbon. ISBN: 0-7695-2540-7
106. Johnston A, Reust J (2006) Network intrusion investigation—preparation and challenges. Digit Investig 3(1):118–126
107. Kenneally EE, Brown CLT (2005) Risk sensitive digital evidence collection. Digit Investig 2(2):101–119
108. Kee E, Farid H (2010) Digital image authentication from thumbnails. In: Proceedings of the SPIE symposium on electronic imaging, San Jose
109. Kee E, Johnson MK, Farid H (2011) Digital image authentication from JPEG headers. IEEE Trans Inf Forensic Secur 6(3):1066–1075
110. Khan MNA, Chatwin CR, Young RCD (2007) A framework for post-event timeline reconstruction using neural networks. Digit Investig 4(3–4):146–157
111. Koen R, Olivier M (2008) The use of file timestamps in digital forensics. In: Proceeding of the information security of South Africa (ISSA 2008), Pretoria, pp 1–16
112. Kornblum JD (2008) Using JPEG quantization tables to identify imagery processed by software. In: Proceedings of the 8th annual digital forensic research workshop (DFRWS'08). Digit Investig 5:S21–S25
113. Kornblum JD (2006) Identifying almost identical files using context triggered piecewise hashing. In: Proceedings of the 6th annual digital forensic research workshop (DFRWS'06). Digit Investig 3(Supplement 1):S91–S97
114. Kornblum JD (2004) The linux and the forensic acquisition of hard disks with odd number of sectors. Int J Digit Evidence 3(2):1–5
115. Lalis S, Karypidis A, Savidis A (2005) Ad-hoc composition in wearable and mobile computing. Commun ACM 48(3):67–68
116. Lamport L (1978) Time, clocks, and the ordering of events in a distributed system. Commun ACM 21(7):558–565
117. Laurie A (2006) Digital detective. Digit Investig 3(1):17–19
118. Lavelle C, Konrad A (2007) FriendlyRoboCopy: a GUI to robocopy for computer forensic investigators. Digit Investig 4(1):16–23
119. Lee S, Shamma DA, Gooch B (2006) detecting false captioning using common sense reasoning. In: Proceedings of the 6th annual digital forensic research workshop (DFRWS'06). Digit Investig 3(Supplement 1):S65–S70
120. Lee J, Un S, Hong D (2008) High-speed search using tarari content processor in digital forensics. In: Proceedings of the 8th annual digital forensic research workshop (DFRWS'08). Digit Investig 5(Supplement 1):S91–95
121. Leighland R, Krings AW (2004) A formalization of digital forensics. Int J Digit Evidence 3(2):1–32
122. Liebrock LM, Marrero N, Burton DP, Prine R, Cornelius E, Shakamuri M et al. (2007) A preliminary design for digital forensics analysis of terabyte size data sets. Paper presented at the symposium on applied computing (SAC'2007), Seoul
123. Lyle JR (2006) A strategy for testing hardware write block devices. Paper presented at the 6th annual digital forensic research workshop (DFRWS'06). Digit Investig 3(Supplement 1):S3–S9
124. Marchionini G (2006) Exploratory search: from finding to understanding. Commun ACM 49(4):41–46
125. Marziale L, Richard III GG, Roussev V (2006) Massive threading: using GPUs to increase performance of digital forensic tools. Paper presented at the 6th annual digital forensics research workshop (DFRWS'06). Digit Investig 4:73–81
126. Masters G, Turner P (2007) Forensic data discovery and examination of magnetic swipe card cloning devices. In: The proceedings of the 7th annual digital forensic research workshop (DFRWS'07). Digit Investig 4(Supplement 1):S16–S22
127. McGrew R, Vaughn R (2007) Using search engines to acquire network forensic evidence. Paper presented at the digital

forensics; advances in digital forensics III: IFIP international conference on digital forensics, Orlando
128. McKemmish R (1999) What is forensic computing? Trends and issues in crime and justice, vol 188. Australian Institute of Criminology, Canberra, pp 1–6. ISBN 0-642-24102-3
129. Mead S (2006) Unique file identification in the national software reference library. Digit Investig 3(1):138–150
130. Mee V, Tryfonas T, Sutherland I (2006) The windows registry as a forensic artefact: illustrating evidence collection for Internet usage. Digit Investig 3(3):166–173
131. Mercuri RT (2005) Challenges in forensic computing. Commun ACM 48(12):17–21
132. Metadata Working Group (2010) Guidelines for handling metadata, Ver 2.0. http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf. Accessed 12 July 2011
133. Microsoft Developer Network Library (2011) SYSTEMTIME Structure, MSDN Microsoft Corporation. http://msdn.microsoft.com/en-us/library/ms724950(v=VS.85).aspx. Accessed 12 July 2011. Microsoft Developer Network Library, TIME_ZONE_INFORMATION Structure, MSDN Microsoft Corporation. http://msdn.microsoft.com/en-us/library/ms725481(v=VS.85).aspx. Accessed 12 July 2011
134. Microsoft Developer Network Library (2011) DYNAMIC_TIME_ZONE_INFORMATION structure, MSDN Microsoft Corporation. http://msdn.microsoft.com/en-us/library/ms724253(v=VS.85).aspx. Accessed 12 July 2011
135. Microsoft Developer Network Library (2011) File times, MSDN Microsoft Corporation. http://msdn.microsoft.com/en-us/library/ms724290(v=VS.85).aspx. Accessed 12 July 2011
136. Microsoft Developer Network Library (2011) Local time, MSDN Microsoft Corporation. http://msdn.microsoft.com/en-us/library/ms724493(v=VS.85).aspx. Accessed 12 July 2011
137. Microsoft Developer Network Library (2011) DateTime. ToUniversalTime Method, MSDN Microsoft Corporation. http://msdn.microsoft.com/en-us/library/system.datetime.touniversaltime.aspx. Accessed 12 July 2011
138. Microsoft Support (2011) Time stamps change when copying from NTFS to FAT, Article ID 127830, Microsoft Corporation. http://support.microsoft.com/kb/127830. Accessed 12 July 2011
139. Microsoft Support (2011) Description of NTFS date and Time stamps for file and folders. Article ID 299648, Microsoft Corporation. http://support.microsoft.com/kb/299648. Accessed 12 July 2011
140. Microsoft Support (2011) Interpreting timestamps on NTFS file systems. Article ID 158558, Microsoft Corporation. http://support.microsoft.com/kb/158558. Accessed 12 July 2011
141. Miskelly GM, Wagner JH (2005) Using spectral information in forensic imaging. Forensic Sci Int 155(2–3):112–118
142. Mocas S (2004) Building theoretical underpinnings for digital forensics research. Digit Investig 1(1):61–68
143. Mohay GM, Anderson A, Collie B, de Vel O, McKemmish R (2003) Computer and intrusion forensics. Artech House Publications, London. ISBN 1580533698, 9781580533690
144. Morgan TD (2008) Recovering data from the windows registry. In: Proceedings of the 8th annual digital forensic research workshop (DFRWS'08). Digit Investig 5(Supplement 1):S33–S41
145. Murphey R (2007) Automated Windows Event Log Forensics. Paper presented at the 7th annual digital forensic research workshop (DFRWS'07). Digit Investig 4(Supplement 1):S92–S100
146. Myers M, Rogers M (2004) Computer forensics: a need for standardization and certification. Int J Digit Evidence 3(2):1–11
147. National Institute of Justice (NIJ) (2001) Electronic crime scene investigation guide: a guide for first responders. National Institute of Justice, Department of Justice (DoJ) 2001. http://www.ncjrs.gov/pdffiles1/nij/187736.pdf
148. Nikkel BJ (2006) Improving evidence acquisition from live network sources. Digit Investig 3(2):89–96
149. NISO (2004) Understanding metadata. NISO Press, pp 1–20. ISBN: 1-880124-62-9, http://www.niso.org/publications/press/UnderstandingMetadata.pdf
150. NIST (2007) Test results for hardware write block device: Tableau Forensic SATA Bridge T3u. NIST, Gaithersburg (Unpublished manuscript)
151. NIST (2002) Hard disk hardware write block tool specification. NIST, Gaithersburg (Unpublished manuscript)
152. NIST (2003) Hard disk software write block tool specification. NIST, Gaithersburg (Unpublished manuscript)
153. NIST (2001) General test methodology for computer forensic tools. NIST, Gaithersburg (Unpublished manuscript)
154. NIST (2001) Disk imaging tool specification. NIST, Gaithersburg (Unpublished manuscript)
155. Nutter B (2008) Pinpointing TomTom location records: a forensic analysis. Digit Investig 5(1):10–18
156. Olievier MS (2008) On metadata context in database forensics. Digit Investig 5(1):1–8
157. Pal A, Sencar HT, Memon N (2008) Detecting file fragmentation point using sequential hypothesis testing. In: Proceedings of the 8th annual digital forensic research workshop (DFRWS'08). Digit Investig 5(Supplement 1):S2–S13
158. Pan L, Batten LM (2005) Reproducibility of digital evidence in forensic investigations. Paper presented at the 5th annual digital forensic research workshop (DFRWS'05), New Orleans
159. Park B, Park J, Lee S (2009) Data concealment and detection in microsoft office 2007 files. Digit Investig 5(3–4):104–114
160. Pering T, Ballagas R, Want R (2005) Spontaneous marriages of mobile devices and interactive spaces. Commun ACM 48(9):53–59
161. Petroni J, Nick L, Walters A, Fraser T, Arbaugh WA (2006) FATKit: a framework for the extraction and analysis of digital forensic data from volatile system memory. Digit Investig 3(4):197–210
162. Pollitt MM (2007) An Ad-hoc review of digital forensic models. In: Proceedings of the second international workshop on systematic approaches to digital forensic engineering (SADFE'07). IEEE Publication, Washington
163. Poolsapassit N, Ray I (2007) Investigating computer attacks using attack trees. In: Pollitt M, Shenoi S (eds) Proceedings of the third annual IFIP WG 11.9 international conference on digital forensics; advances in digital forensics III: IFIP international conference on digital forensics. Springer, Orlando. ISBN: 978-0-387-73741-6
164. Popescu AC, Farid H (2004) Statistical tools for digital forensics. In: Proceedings of sixth international workshop on information hiding, Toronto
165. Raghavan S, Clark AJ, Mohay G (2009) FIA: an open forensic integration architecture for composing digital evidence. In: Proceedings of the ICST second annual international conference on forensic applications and techniques in telecommunications, information and multimedia (e-Forensics 2009), Adelaide
166. Reith M, Carr C, Gunsch G (2002) An examination of digital forensic models. Int J Digit Evidence 1(3):1–12
167. Reyes A, O'Shea K, Steele J, Hansen JR, Jean BR, Ralph T (2007) Digital forensics and analyzing data, cyber crime investigations. Syngress, Burlington, pp 219–259
168. Richard III GG, Roussev V (2005) Scalpel: a frugal high performance file carver. Paper presented at the 5th annual digital forensics research workshop (DFRWS'05), New Orleans
169. Richard GG III, Roussev V (2006) Next-generation digital forensics. Commun ACM 49(2):76–80
170. Richard III GG, Roussev V, Marziale L (2006) In-place file carving. In: Proceedings of the second annual IFIP WG 11.9

international conference on digital forensics, advances in digital forensics II. Springer, Boston, pp 1–12. ISBN: 0-387-36890-6
171. Richard GG III, Roussev V, Marziale L (2007) Forensic discovery auditing of digital evidence containers. Digit Investig 4(2):88–97
172. Rossev V, Chen Y, Bourg T, Richard III GG (2005) md5Bloom: forensic filesystem hashing revisited. Paper presented at the 5th annual digital forensics research workshop (DFRWS'05), New Orleans
173. Roussev V, Richard GG III, Marziale L (2007) Multi-resolution similarity hashing. Digit Investig 4(Supplement 1):105–113
174. Rowe NC, Garfinkel S (2011) Finding anomalous and suspicious files from directory metadata on a large corpus, to appear. In: Proceedings of the third international conference on digital forensics and cyber crime, ICDF2C 2011, Dublin
175. Rui Y, Huang TS, Shih-Fu Chang (1998) Image retrieval: current technologies, promising directions and open issues. J Vis Commun Image Represent (IJVCIR) 10:39–62
176. Sanderson P (2006) Identifying an existing file via KaZaA artefacts. Digit Investig 3(3):174–180
177. Sarmoria CG, Chapin SJ (2005) Monitoring Access to shared memory mapped files. Paper presented at the 5th annual digital forensic research workshop (DFRWS'05), New Orleans
178. Scientific Working Group on Digital Evidence (2009) technical notes on microsoft windows vista. SWGDE Technical Notes, pp 1–25
179. Scientific Working Group on Digital Evidence (2010) Technical notes on microsoft windows 7. SWGDE Technical Notes, pp 1–20
180. Schatz B (2007) BodySnatcher: towards reliable volatile memory acquisition by software. Digit Investig 4(Supplement 1):126–134
181. Schatz BL, Clark AJ (2006) An open architecture for digital evidence integration. In: Proceedings of the AusCERT R&D Stream, AusCERT 2006, Gold Coast, pp 15–29
182. Schatz B, Mohay G, Clark A (2006) A correlation method for establishing provenance of timestamps in digital evidence. In: The proceedings of the 6th annual digital forensic research workshop (DFRWS'06). Digit Investig 3(Supplement 1):98–107
183. Schuster A (2006) Searching for processes and threads in microsoft windows memory dumps. In: The proceedings of the 6th annual digital forensic research workshop (DFRWS'06). Digit Investig 3(Supplement 1):10–16
184. Schuster A (2007) Introducing the microsoft vista event log file format. Digit Investig 4(Supplement 1):65–72
185. Schuster A (2008) The impact of microsoft windows pool allocation strategies on memory forensics. In: Proceedings of the 8th annual digital forensic research workshop (DFRWS'08). Digit Investig 5(Supplement 1):S58–S64
186. Schraffel MC, Wilson M, Russel M, Smith DA (2006) MSpace: improving information access to multimedia domains with multimodal exploratory search. Commun ACM 49(4):47–49
187. The Sedona Conference Working Group (2007) The Sedona principles: best practices recommendations & principles for addressing electronic document production (2nd edn.) http://www.thesedonaconference.org/content/miscFiles/TSC_PRINCP_2nd_ed_607.pdf. Accessed 12 July 2011
188. The Sedona Conference Working Group (2010) The Sedona conference glossary: e-discovery & digital information management (3rd edn.) www.thesedonaconference.org/dltForm?did=glossary2010.pdf. Accessed 12 July 2011
189. The Sedona Conference Working Group (2011) The Sedona Conference: Commentary on ESI Evidence & Admissibility (2008). http://www.thesedonaconference.org/dltForm?did=ESI_Commentary_0308.pdf. Accessed 12 July 2011
190. Sencar HT, Memon N (2009) Identification and recovery of JPEG files with missing fragments. Digit Investig 6(4):S88–S98
191. Sencar HT, Memon N (2008) Overview of state-of-the-art in digital image forensics, part of Indian statistical institute platinum jubilee monograph series titled statistical science and interdisciplinary research. World Scientific Press, Singapore
192. Shankaranarayanan G, Even A (2006) The metadata enigma. Commun ACM 49(2):88–94
193. Shannon MM (2004) Forensic relative strength scoring: aSCII and entropy scoring. Int J Digit Evidence 2(4):1–19
194. Slewe T, Hooenboom M (2004) Who will rob you on the digital highway? Commun ACM 47(5):56–60
195. Solomon J, Huebner E, Bem D, Szezynska (2007) User data persistence in physical memory. Digit Investig 4(1):68–72
196. Standards Australia (2003) HB171-guidelines for the management of IT evidence
197. Steele J (2007) Digital forensics and analyzing data: alternate data storage forensics. Syngress, Burlington, pp 1–38
198. Stevens MW (2004) Unification of relative time frames for digital forensics. Digit Investig 1(1):225–239
199. Teerlink S, Erbacher R (2006) Improving the computer forensic process through visualization. Commun ACM 49(2):71–75
200. Toyama K, Logan R, Roseway A, Anadan P (2003) Geographic location tags on digital images. In: Proceedings of ACM multimedia 2003, Berkeley, pp 156–166. ISBN: 1-58113-722-2
201. Turnbull B, Blundell G, Slay G (2006) Google desktop as a source of digital evidence. Int J Digit Evidence (IJDE) 5(1):1–12
202. Turner P (2005) Unification of digital evidence from disparate sources (digital evidence bags). Digit Investig 2(3):223–228
203. Turner P (2005) Digital provenance—interpretation, verification and corroboration. Digit Investig 2(1):45–49
204. Turner P (2006) Selective and intelligent imaging using digital evidence bags. In: The proceedings of the 6th annual digital forensic research workshop (DFRWS'06), Digit Investig 3(Supplement 1):59–64
205. Turner P (2007) Applying a forensic approach to incident response, network investigation and system administration using digital evidence bags. Digit Investig 4(1):30–35
206. van Baar RB, Alink W, Van Ballegooji AR (2008) Forensic memory analysis: files mapped in memory. In: Proceedings of the 8th annual digital forensic research workshop (DFRWS'08). Digit Investig 5(Supplement 1):S52–S57
207. Venter J, de Waal A, Willers C (2007) Specializing CRISP-DM for evidence mining. Paper presented at the digital forensics; advances in digital forensics III: IFIP international conference on digital forensics, Orlando
208. Vlastos E, Patel A (2007) An open source forensic tool to visualize digital evidence. Comput Stand Interfaces 29(6):614–625
209. Vlastos E, Patel A (2008) An open source forensic tool to visualize digital evidence. Comput Stand Interfaces 30(1–2):8–19
210. Wang G, Chen H, Atabakhsh H (2004) Automatically detecting deceptive criminal identities. Commun ACM 47(3):71–76
211. Wang S-J, Kao D-Y (2007) Internet forensics on the basis of evidence gathering with peep attacks. Comput Stand Interfaces 29(4):423–429
212. Wang S-J (2007) Measures of retaining digital evidence to prosecute computer-based cyber-crimes. Comput Stand Interfaces 29(2):216–223
213. Wang W, Daniels TE (2005) Network forensic analysis with evidence graphs. Paper presented at the 5th annual digital forensic research workshop (DFRWS'05), New Orleans
214. Weil MC (2002) Dynamic time and date stamp analysis. Int J Digit Evidence 1(2):1–6
215. Willassen S (2008) Finding evidence of antedating in digital investigations. In: Proceedings of the third international conference on availability, reliability and security, ARES, Barcelona, pp 26–32

216. Xu J, Chen H (2005) Criminal network analysis and visualization. Commun ACM 48(6):100–107
217. Zander S, Nguyen T, Armitage G (2005) Automated traffic classification and application identification using machine learning. In: Proceedings of the IEEE conference on local computer networks, IEEE LCN 2005, Sydney, pp 250–257. ISBN: 0-7695-2421-4