

Advances in Digital Forensics Frameworks and Tools: A Comparative Insight and Ranking

Muhammad Abulaish¹, Nur Al Hasan Haldar²

¹Department of Computer Science, South Asian University, New Delhi, India

²Center of Excellence in Information Assurance, King Saud University, Riyadh, KSA

Abstract: Digital forensics science is a well-known initiative to unearth computer-assisted crimes. The thriving criminal activities using digital media have changed the typical definition of a traditional crime. Meanwhile, the means and targets of criminal activities have been transformed in a broader context due to the diverse nature of offenses associated with the multiple crime categories, affecting the way of investigations as well. In order to withstand the difficulties caused due to the crime complexity, forensics investigation frameworks are being tuned to adjust with the nature and earnestness of the felonies being committed. This paper presents an in-depth comparative survey of fourteen popular and most cited digital forensics process models and various forensics tools associated with different phases of these models. The relationships among these forensics process models and their evolutions are analyzed and a graph-theoretic approach is presented to rank the existing process models to facilitate investigators in selecting an appropriate model for their investigation tasks.

Keywords: Digital Forensics, Digital Investigation, Digital Forensics Framework, Digital Forensics Tools, Process Models

1. Introduction

Digital forensics (also known as computer forensics) is a systematic process of uncovering a crime through investigating the media components found in associated digital devices. The investigation practice follows a list of scientifically derived and justified mechanism towards gathering and illustrating the evidences of a crime scene. A forensic science integrates the scientific knowledge and methodology to a legal problem and criminal investigation. Over the last few years, digital forensics has been given much importance where electronic devices are used for executing an offense. Though the initial focus of digital forensics investigations was based on the crimes perpetrated using computers only, the field nowadays has been extended to incorporate different other digital devices like camera, smart phones, etc. Any digital information stored in such devices can be inspected and identified for various types of criminal activities [1].

Forensics is a very different business when it comes to technology. Compared with traditional forensic science, digital forensics differs significantly and also poses some substantial challenges. The traditional forensics analysis involves the investigation using tangible, physical items found around the crime scene, whereas the digital forensics encompasses with various operations like extraction, storage and analysis of digital data using scientifically derived and proven methods. A traditional forensic analysis can logically progress step-by-step, with a common intention with widely accepted forensic practices. It is generally dependent upon the laboratory setting and on-field activities. However, in general, it comes with the widely accepted physical forensics practices. In comparison, a computer forensic science is almost technology and market driven, independent of laboratory

environment and settings [10]. The digital examinations and analysis present a unique variations in different investigations. In case of sample accumulation for investigation, traditional forensics attempts to gather as much information as possible from an evidence sample, whereas digital forensics attempts to discover only the relevant information from a large volume of heterogeneous digital data.

In digital forensic research workshop, Palmer [17] defined digital forensics as “*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*”. This definition is frequently cited and also accepted to be an all-inclusive definition [1]. Willassen et al. [18] defined digital forensics in a broader way as “*the practice of scientifically derived and proven technical methods and tools towards the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of after-the-fact digital information derived from digital sources for the purpose of facilitating or furthering the reconstruction of events as forensic evidence*”. The main change in this definition in comparison to the Palmer’s definition is that Willassen et al. have removed the criminal events and unauthorized actions. As a result, this definition extends the scope of application to include digital forensics in various types of investigation, such as commercial investigation [1].

However, with the arrival of new technologies, some notable changes along with challenges have been observed in the digital investigation processes. Since a contemporary crime may be introduced due to the current age digital technologies, an investigation process model should be particularly flexible and intelligent enough to deal with such unfamiliar incidents. Technology has impacted the way evidences are gathered, analyzed, and presented in courts. A large number of digital forensics investigation process models, tools and equipments have been developed for facing challenges that are raised due to technology advancements. It is important for every country or organization to develop its own digital forensics investigation mechanism based on its specified laws, rules, and policies. In general, digital forensics investigation follows a number of processes like *identification, preparation, preservation, analysis, and presentation* to exhaust a proper investigation. Depending on the type and intensity of a crime such processes can be divided into various phases. The *identification* phase recognizes the incident type and tasks to be accomplished throughout the investigation. The *preparation* phase is involved in analyzing the organizational infrastructure, requirements and tools to be used to carry out the investigation. The *preservation* phase is followed by *preparation* phase, where digital data are extracted from device and preserved for future analysis. One of the usages of data preservation is to cross check and validate the identified evidences. In *analysis* phase, data are examined to identify evidence patterns and findings from crime scene. Finally, the *presentation* phase generates the reports and findings and produces them in front of jury or respective management units.

The main focus of digital forensics is to exhibit digital evidences and relate them to the crime scene. The criminal evidences may be hidden within massive amount of digital data stored in various digital storage devices. The digital investigation process follows some systematic steps to extract evidences from such vast amount of digital data in such a way that they should be admissible, as well as authentic by the court of law [2]. One of the main objectives of digital forensics is to preserve any evidence in its most original form. In order to reconstruct past events, digital forensics targets to perform an analytical investigation by identifying, collecting and justifying the digital information. In a broader scenario, digital forensics incorporates the field of computer science with legal practices while investigating a crime. After recovering and analyzing data from digital storage devices, the investigators must follow a legal procedure from beginning to the end of an investigation process so that the digital evidences produced by the investigators could be legally admissible in the court of

law. But, unfortunately, laws cannot adequately access the techniques used in computer search as it is written before the computer forensics era and mostly they are outdated [3]. The inability of law to keep consistency with technological improvement may eventually lower the forensics evidence outcome in court [4]. The current judiciary systems have already begun to question the “scientific” validity of many of the specific procedures and methodologies and they are also demanding proof of some sort of the theoretical foundation and scientific difficulties [5]. In this context, there are significant issues and challenges that have been identified in digital forensics investigation process. These challenges can be broadly mentioned as legal challenges, technical challenges, resource challenges, and data related challenges [6].

While investigating, an investigator must follow proper guidelines to guarantee that any evidence should satisfy its legal requirements, i.e., it should be authentic, reliable, complete, believable, and admissible [7]. Legal challenges can also be claimed from laws and legal tools needed to inspect crimes, however these are lagging behind the technological and structural advances. Technical complicatedness of digital evidence often makes it more challenging, as it is quite difficult for the court of law to understand the importance of the evidence. Resources are also an important part of investigation. Under-resourced, overburdened complex cases are tedious to execute an investigation properly within estimated time with perfection. A resource may be in terms of personnel support, economic aid, or even reliable tools. On the other hand, forensics analyses and evidence presentation are sometimes perplexed by unskilled and immature individuals, which is further aggravated by faulty case management. In addition, digital investigation may face numerous challenges due to insufficient financial supports to both investigation team and suspects. Due to the limited financial resources, it is hard for a suspect to rebut the evidences. However, the high cost of escalating a defense using forensic specialist is often beyond the financial reach of many defendants. Forensic tools are another important resource and most of them are often commercial products. Due to profit-driven mindset of corporates, the available tools may not fulfil real scientific forensic needs.

Another crucial challenge in forensics investigation comes from the data side [12]. In recent years, there has been immense growth in the volume of data on digital forensics which leads to big data issues [15]. This has a significant impact on not only for the data acquisition and imaging techniques used, but also more importantly on the way data is analyzed [8]. The increasing volume of digital evidences requires an intelligent analysis and storage methodology to support the various phases of digital forensic Investigations [13, 14]. The data privacy is another important issue in investigation process and due to its ignorance the human rights of a suspect may be ruptured. The possible privacy violation that could happen in this context is the access to all irrelevant information, which is private to the suspect. Hence, it is a challenging task to develop a most appropriate policy which can restrict access to irrelevant personal data to the investigator.

The rest of the paper is organized as follows. Section 2 presents a review of the state-of-the-art digital forensics frameworks. It also presents a summarized view of the digital forensics process models along with the number of proposed phases and sub-phases. A graph-theoretic approach is also presented in this section to establish the similarity and ranking of the existing forensics models. Section 3 presents a review of the privacy-preserving digital forensics models. Section 4 presents a review of the popular digital forensics tools and their mapping to various phases and sub-phases of the forensics models. Finally, section 5 concludes the paper with future directions of work.

2. Digital Forensics Frameworks

A Digital Forensics Framework (DFF), also termed as Digital Forensics Process Model (DFPM), is a sequence of defined steps, along with their sub-steps, inputs, outputs, requirements, order, and standards [31]. Over the past few years, digital forensics has reached

the top as an increasingly important method for identifying and prosecuting criminals [11]. Digital forensics investigation follows a sequence of scientifically proven methods to collect, preserve, search, and analyze evidences in order to determine a crime, whereas a digital forensics framework helps the forensics professionals to investigate an offence without compromising the systems, data, and other circumstances. Such framework can be defined as an architecture to support a successful forensics investigation [11]. A well-structured framework for digital investigation process is required so that any investigation can be conducted in an integrated and timely manner. A framework dedicated to digital forensics can easily be reconstructed while simulating a root cause of an investigation. Such framework can facilitate a common starting place from where computer science theory can be significantly applied to digital forensics science field. In order to collect, preserve and expose digital evidence in a systematic way, a well-defined DFF is needed.

Over the years, a number of digital forensics frameworks exist in literature. However, the process and terminologies associated with the contemporary forensics framework have not been accordingly standardized [1]. Some of the investigation models bears a very specific scenario whereas others can be applied to a vast scope with high-level phase-wise descriptions. Moreover, some of the models tend to be in detail and others may be too general. Such scenarios may be a bit difficult or even confusing for a forensic investigator to choose the most suitable and appropriate model for investigation purpose. Therefore, it is required to analyze various well-known forensics frameworks and compare their advances properly. Some detailed phase descriptions can be seen in [21], [24], and [25] where authors have described their phase-wise process models. In order to discuss various process models in a comparative manner, the original terminology should be kept same as it was termed by the respective authors. For example, some authors have used “analysis” as one of the processes, whereas others termed it as “examination”. In our discussions, the original terms associated with the respective process model are kept intact. However, when comparing and identifying common characteristics of the process models, we standardize them with the conventional terms. A brief survey of the popular and most cited digital forensics process models is presented in the following sub-sections.

2.1. Computer Forensic Process [16]

At the early age in 1995, Pollitt [16] proposed a four-step process (*acquisition, identification, evaluation, and admission*) to relate how digital media can describe the legal requirements for acceptability of paper-based evidence. In *acquisition* phase, evidences are obtained in adequate manner with proper technical and legal liabilities. The task of *identification* phase is to identify the digital components from acquired evidences. *Evaluation* phase evaluates the relevancy of an evidence. It involves both technical and legal judgments. The last phase, *admission*, presents evidence in the court of law. The importance of digital information storage for future is identified in this process model with a base for dealing with potential digital evidence. However there is no such concrete statement about how to acquire a document as both legal and technical evidences. On the other hand, a gap in communication between legal expert and forensics specialist exists in the model.

2.2. Investigative Process for Digital Forensic Science [17]

In 2001, first Digital Forensic Research Workshop (DFRWS) was held in Utica, New York, which was intended to summarize discussions among academics, investigators and experienced-practitioners to establish a research community dedicated to digital forensics [17]. The investigative road map of the workshop was mainly focused on framework development for forensics science and it was emphasized that digital forensics frameworks need to be more specific and flexible enough to support future technologies. In that workshop, a general-purpose digital forensics process model was concluded which comprises six phases – *identification, preservation, collection, examination, analysis, and presentation*. The

proposed DFRWS framework is a common starting place for forensics theory, however it is non-specific and generic in nature.

2.3. Scientific Crime Scene Investigation Model [19]

Henry Lee et al. [19] described the fundamental elements for a successful crime scene investigation. Though Lee's model focused on physical evidence, it can be fitted to include evidence found in a digital crime scene investigation [1]. Such model represents four phases – *recognize*, *identify*, *individualize*, and *reconstruct*. In *recognize* phase, suspicious items or patterns are identified to be potential evidences. This phase also has sub-phases like *collection* and *preservation*. The next phase, *identification*, is for labelling various types of evidences that may be classified as physical, chemical, biological, and so on. In *individualization* phase, the evidences are associated with a particular individual or events, whereas *reconstruction* phase constructs possible event sequences to be reported at end. Though the steps in [19] refer to a part of the forensic investigation process, they basically fall within the 'investigation' stage of a general process and there is no 'preparation' or 'presentation' stage included in such model [11]. This model focuses on an organized and methodical way of investigating any digital crime cases. But it bears some limitations in the digital forensic investigation, as it does not focus much on data acquisition, preparation, and presentation [30].

2.4. Abstract Digital Forensics Model [20]

Inspired by the DFRWS model [17] of forensics investigation, Reith et al. [20] proposed an enhanced model named Abstract Digital Forensics Model (ADFM) which consists of nine process components. In this model three significant phases, i.e., *preparation*, *approach strategy*, and *returning evidence* were newly introduced. Activities like, preparing tools, search warrant, identifying techniques, etc. are considered to be complete in *preparation* phase. The *approach strategy* phase is introduced with the objective to maximize the collection of innocent evidence and at the same time to minimize negative impact to the victim. The *returning evidence* phase ensures about the evidence to the legitimate person. The other processes strategies are almost same as the DFRWS model. This model is highly accepted by modern investigators, but it is open to one criticism – the *approach strategy* phase is somehow an extension of the *preparation* phase, because at the time of preparation of crime response, the *approach strategy* should be planned [23].

2.5. An Integrated Digital Investigation Process [21]

In 2003, Carrier and Spafford [21] proposed Integrated Digital Investigation Process (IDIP) model based on the theories and techniques from physical forensics investigation procedures. This model combines both the physical and digital forensics investigation processes. This process model has a total 17 sub-phases generalized into five main phases. The main phases are *readiness*, *deployment*, *physical crime scene investigation*, *digital crime scene investigation*, and *review*. The *readiness* phase checks whether the operations and infrastructures are able to fully support an investigation or not. The *deployment* phase provides a mechanism for the incidents to be detected and confirmed. The *digital* and *physical investigation* phases occur simultaneously in parallel. These two phases consist of six sub-phases in such a way that the digital crime scene focuses on the digital evidence in digital environment, whereas the physical crime scene is associated with the physical environment. Finally, in *review* phase, the investigation process is rechecked to identify the possible areas of improvement; hence ultimately building a mechanism for efficient forensic examinations.

2.6. Event-Based Digital Forensics Investigation Framework [22]

In 2004, Carrier and Spafford [22] simplified their previously proposed IDIP framework [21] as an Event-Based Digital Forensic Investigation Framework (EDFIF), which is based on the causes and effects of events. The model gives much more impression on digital forensics

investigation rather than physical investigation. This model is also more spontaneous and flexible for developing requirements for each phase. There are some important differences between earlier proposed IDIP model and event-based EDFIF model. In IDIP model, the presentations of physical and digital investigations are proposed in parallel in two distinct ways, whereas in case of EDFIF, the evidences are collected from both physical and digital investigation and thereafter the presentation of findings are produced together in a *complete* phase. The main phases of this process model are *readiness*, *deployment*, *physical crime scene investigation*, *digital crime scene investigation*, and *representation*. The target of *readiness* is same as proposed in IDIP model. It trains appropriate people, tests tools, and configures the infrastructure to ensure investigation capabilities. The *readiness* phase includes two sub-phases – *operations readiness* phase and *infrastructure readiness* phase. The *deployment* phase includes two sub-phases – *detection and notification* phase, and *confirmation and authorization* phase. The *detection and notification* phase detects incidents and investigators are notified, whereas *confirmation and authorization* phase confirms the permission to investigating team. The *physical crime scene phase* deals with the physical investigation, wherein the sub-phases are *physical evidence collection*, *search of physical evidence*, *reconstruction of physical events*, and so on. The *digital crime scene investigation* phase investigates the digital data for evidences. It has three sub-phases – *digital crime scene preservation and documentation*, *digital evidence searching and documentation*, and *digital event reconstruction and documentation*. In EDFIF framework, all digital crime scene investigation sub-phases are associated with proper documentation. The evidence searching and evidence reconstruction sub-phases are iterative process. Such activity replaces the review phase as one of the earlier model. The final phase of EDFIF is *presentation*, which presents the findings to either corporate audience or court of law [22]. In terms of flexibility, EDFIF model is more adaptable than IDIP, but it does not have any concept of physical evidence or physical object preservation.

2.7. Enhanced Integrated Digital Investigation Process [23]

The IDIP process model [21] also has some criticism about its practicality. It does not suggest adequate specificity to differentiate primary (suspect) and secondary (victim) crime scene investigation. Such shortcoming in IDIP model is figured out in [23] by including an investigation of primary and secondary crime scene [1]. The Enhanced Integrated Digital Investigation Process (EIDIP) model, proposed by Baryamureeba and Tushabe [23], makes clear difference between physical and digital crime scene processes, with additional two supplementary phases – *traceback* and *dynamite*. In *traceback* phase, the physical crime scene is tracked to determine the location of physical devices used in the crime, whereas *dynamite* phase investigates the primary crime scene by collecting and analysing the evidence items. This phase has four sub-phases – *physical crime investigation*, *digital crime scene investigation*, *reconstruction*, and *communication*. The objective of the first two phases *readiness* and *deployment* of Enhanced IDIP (EIDIP) model is same as the IDIP model. The *readiness* phase includes the plan of sufficient resource and infrastructure management to conduct the investigation, whereas the *deployment* phase provides a mechanism to detect and confirm an incident. The most significant addition in the EIDIP is that the phases are iterative rather than linear which allows to backtrack the previous phases. Such practice enables investigator to improve the outcome of each phase. However, Perumal [30] criticised the EIDIP model for missing of essential elements such as the “chain of custody”.

2.8. Extended Model of Cyber Crime Investigation [24]

The process model defined by Ciardhuáin [24] is one of the most comprehensive models for cybercrime investigations, which attempts to capture the investigative process including the digital evidence activities as much as possible. This model captures the full scope of an investigation, instead of only processing the evidences. In this process model, phases are

termed as activities where a total number of 13 activities are mentioned as *awareness*, *authorization*, *planning*, *notification*, *search and identification*, *collection*, *transportation*, *storage*, *examination*, *hypothesis*, *presentation*, *proof/defense*, and *dissemination*. It follows a linear waterfall model and executes the activities in a sequence once after the completion of the previous activities. However, certain sequence of examination-hypothesis-presentation-proof/defense sequence of activities usually can be repeated when evidence set grows [24].

The first activity of this investigation model is *awareness* where investigators are made aware about a crime reported by an authority. After analyzing the necessity of investigation, the next activity, *authorization*, requests for consent from concerned authorities to carry out the investigation. *Planning* is influenced by internal policies, strategies, and external organizational rules and regulations. The planning activity may need to backtrack for further authorization if full requirement of the investigation is not included in planned scope [1]. In *notification* activity, the stakeholder or concerned parties are informed that an investigation is going to take place. The *search and identification* activity deals with the location of evidences. *Collection* activity occurs when investigator takes possession of evidence in a form that can be preserved and analyzed. In this task, Ciardhuáin suggested hard disk imaging and seizing of computer. After collection, the evidences should be *transported* to a suitable and safe location in such a way that the integrity of evidence could not be affected. The *storage* of evidence is necessary because examination doesn't take place immediately just after evidence collection. Storage activity should also maintain the evidence integrity. *Examination* is the task to analyze evidences. In order to interpret voluminous evidences, a large number of automated techniques are required to support investigators. In *hypothesis* task, an investigator must conduct a theory, based on the examination of the evidences. Backtracking from this activity to previous activity (i.e., *Examination*) is expected as investigators try to develop better observation of an event. *Presentation* is the task where hypothesis is presented to people other than investigators. A decision is taken based on the findings of presentation. The *proof/defense* is a phase where investigators have to prove the effectiveness of the investigation hypothesis and to defend criticisms, if any. This stage can also backtrack the presentation stage to construct a better hypothesis. *Dissemination* is the final activity, which takes place with the publication of the descriptions of the investigation and its outcome [24].

Though the model proposed by Ciardhuáin incorporates whole investigation rather than only processing of digital evidences, it needs some standardized terminologies [1]. For example, the last activity, *dissemination*, is the same process as *presentation*, used in other models [27]. The other three activities (i.e., awareness, transport, and storage) along with dissemination are considered as irrelevant according to a survey conducted by Ciardhuáin itself [1].

2.9. Hierarchical, Objective-Based Framework [25]

Beebe et al. [25] proposed a hierarchical objective-based framework (HOF) dedicated to digital forensics investigation. To achieve usability and acceptability, HOF is incorporated with phases, sub-phases, principles, and objectives. The HOF framework encapsulates various phases and activities of well-known investigation process models, and it is a multi-tier model where the common phases are kept in first-tier. The phases of this model are *preparation*, *incident response*, *data collection*, *data analysis*, *presentation of findings*, and *incident closure*. The *preparation* phase ensures the availability of digital evidences, whereas the *incident response* phase detects, validates, and determines a response strategy. In *data collection* phase, information is collected from digital devices in a forensically sound manner. The purpose of *data analysis* phase is the confirmatory analysis and/or event reconstruction [25]. The next phase is *presentation of findings*, which communicates relevant technical and non-technical findings to appropriate persons like legal personnel, technical personnel, management teams, and so on. The last phase, *incident closure*, aims to conclude and preserve all information related to the incident.

The *data analysis* phase consists of objective-based sub-phases, which are placed in the second-tier to provide specificity and granularity, guided by principles and objectives [26]. The sub-phases including *survey*, *extract*, and *examine* are associated with each phase of the first-tier. The primary purpose of *survey* sub-phase is data mapping and data extraction from digital objects. The important activities in *extract* sub-phase are keyword searching, hidden data mining, filtering, pattern matching, and file signature analysis. Various analytical techniques are applied in *examine* sub-phase to examine the extracted data to accomplish a positive goals. However, this model focuses only on investigation and does not consider the legal requirements and significance of digital forensics readiness in an organization. According to the authors, although all phases should have sub-phases, they only focused on the *data analysis* sub-phase.

2.10. Investigative Framework [11]

Köhn et al. [11] proposed a reasonable complete digital forensics framework which merges the already existing frameworks mentioned previously. It has three broad stages – *preparation*, *investigation*, and *presentation*. The *preparation* stage includes various activities like training, legal advice, notification, documentation, planning/approach strategies, etc. The *investigation* stage is the core of the proposed framework. This stage incorporates evidence search, collection, transportation, storage, examination (with proper tools), and analysis. The final stage presents and concludes the analysis and findings of investigation. The legal requirement of specific system and documentation of all steps are associated as default with this investigation model [11]. According to the authors, the framework can be expanded to include any number of additional phases required in the future. However, the processes of this model are abstract in nature and do not provide proper descriptions. There is also no mention of forensics tools to be used for examination and analysis purposes.

2.11. Forensics Zachman Framework [28]

In order to break the technical barrier between information technologists, legal practitioners, and investigators a technical-independent framework is required [28]. Jeong proposed a framework, FORensics ZACHman (FORZA), which incorporates legal issues into a bigger picture of digital forensics investigation process. The process flow of the framework is connected into eight different layers. The *contextual investigation layer* is the first layer, which defines investigation objectives. The next layer is *contextual layer*, which is optional and associated with the business objectives. The *legal advisory layer* comes after *contextual layer*, which defines the legal procedures and identify preliminary issues. The *conceptual security layer* is responsible for designing the information system and relevant security controls. After legal requirement and investigation objectives are confirmed, the technical strategies should be confirmed. The *technical presentation layer* designs the forensics strategy model with the help of digital forensics specialists. In *data acquisition layer*, the investigator procures log files and disk image from the compromised machine and the victim machine. The relevant information are extracted and reviewed by group of digital forensics analysts in *data analysis layer*, which maintains some hypothetical procedures. Based on the analysis report, the final layer, *legal presentation layer*, can determine whether the incident can be taken to litigation process or to be closed when sufficient evidence has been collected [28].

2.12. Common Process Model for Incident Response and Computer Forensics [29]

A common process model for incident response and computer forensics was proposed by Freiling et al. in [29]. This investigation model combines the conceptions of incident response and computer forensics and it consists of three major phases – *pre-analysis*, *analysis*, and *post-analysis*. However, all these major phases consist of various sub-phases. The *pre-analysis* phase deals with the activities that are essential to be performed before the actual

analysis starts. The detection of incident and initial response strategy formulation are the most important tasks of this phase. The actual computer forensics is held in *analysis* phase, which includes some important tasks including data duplication, data recovery, and analysis. The *post-analysis* phase deals with writing a precise report that describes the incident. Such report should be understandable to non-technical readers or executives and meet the legal standard for admissibility in court [29].

2.13. Digital Forensic Model Based on Malaysian Investigation Process [30]

While conducting an investigation for potential criminal violations of the law, the local cyber law should be included in investigation process. A country-specific law-based investigation process model is proposed by Perumal [30], which is based on Malaysian cyber laws. The major stages of this process model are *planning, identification, reconnaissance, transport and storage, analysis, proof and defense, and archive storage*. The *planning* stage has two sub-processes – *authorization* and *search warrant collection*. The *identification* phase identifies the electronic equipment used by the suspect and also finds fragile evidences. The *reconnaissance* stage is a new process in this model. In this stage, gathering necessary evidences is very important. In *transport and storage* stage, all collected evidences are located in a safe place without their tempering. The *analysis* stage is the most complicated process where data are analyzed to discover crime. In *proof and defense* phase, an adverse hypothesis and supporting evidences are produced in front of jury. Any invalid findings have to rollback the process and to construct a new report. The *archive storage* stage stores all evidences that may need to be used in future or for training purpose. This model emphasizes on acquisition of both live and static digital data.

2.14. Integrated Digital Forensics Process Model [1]

One of the latest process models proposed by Kohn et al. [1] is termed as Integrated Digital Forensics Process Model (IDFPM). It incorporates processes and sub-processes of six investigation models out of which some have been discussed previously. This model consists of six processes – *preparation, incident, incident response, physical investigation, digital investigation, and presentation*. Another associated process, *documentation*, is continuously linked with all of the six processes. The primary goal of the *documentation* process is to document all requirements and outcomes of the investigation process. In *preparation* phase, the organization prepares itself to deal effectively with various types of incidents [1]. In *incident* phase, an incident is identified and authorized by various sub-processes. After successful completion of this phase, *incident response* is initiated. In first sub-phase of the *incident response*, approach strategy is planned on the basis on the types of incident to identify promising and potential digital evidences from incident scene. The other sub-tasks under this process include seizure of digital evidence, preservation of physical and digital evidence, transportation to secure location for storage, and so on. The next main process of this process model is *digital forensics investigation*. This process is divided into a number of sub-processes like, *collect, authenticate, examination, harvest, identify, hypothesis, analysis, reconstruct, communicate, review*, and so on. Most of the sub-phases are similar to previously discussed model. The *harvest* sub-phase is introduced in this model to produce logical structure of partially deleted files and folders. The *digital investigation process* occurs in parallel with *physical investigation process*. The target of physical investigation is to analyze DNA, fingerprints, and other physical items found in incident location. In *presentation* phase, the investigators present the hypothesis to jury or management team. A decision is made based on the presentation report. The final activity of the IDFPM is *dissemination*, which was first introduced by Ciardhuáin [24]. The activity of this sub-process is to review the existing policies and procedure of the organization.

Table 1 summarizes the year-wise development of digital forensics process models along with the number of proposed phases and sub-phases.

Table 1. Digital forensics process models along with the number of proposed phases and sub-phases

Process model #	Authors, Year [Reference]	Process model name	# Phases/ Sub-phases
PM1	Pollit., 1995 [16]	Computer Forensic Investigative Process	4 / NA
PM2	Palmer, 2001 [17]	Investigative Process for Digital Forensic Science	6 / NA
PM3	Lee, 2001 [19]	Scientific Crime Scene Investigation Model	4 / 8
PM4	Reith, 2002 [20]	Abstract Digital Forensics Model	9 / NA
PM5	Carrier, 2003 [21]	Integrated Digital Investigation Process Model	5 / 17
PM6	Carrier, 2004 [22]	Event-based Digital Forensics Investigation Framework	5 / 10
PM7	Baryamureeba, 2004 [23]	Enhanced Integrated Digital Investigation Process	5 / 13
PM8	Ciardhuáin, 2004 [24]	Extended Model of Cyber Crime Investigation	13 / NA
PM9	Beebe, 2005 [25]	Hierarchical, Objective-based Framework	6 / 3
PM10	Köhn, 2006 [11]	Investigative Framework	3 / NA
PM11	Ieong, 2006 [28]	FORensics ZAchman framework (FORZA)	8 / NA
PM12	Freiling, 2007 [29]	Common Process Model for Incident Response and Computer Forensics	3 / 12
PM13	Perumal, 2009 [30]	Digital Forensic Model Based on Malaysian Investigation Process	7 / 6
PM14	Kohn, 2013 [1]	Integrated Digital Forensics Process Model	6 / 36

Based on the discussions of the selected digital forensics process models, it is quite obvious that each model has different strategies and aims. A comparative summary of the process models discussed above is presented in Table 2. In this table, rows present different phases and sub-phases, which are identified as necessary for each process model. The first column represents the phases and sub-phases that are available in process models in general. The onward columns are process model index. The sub-phases available in a particular process model is given corresponding entries in the table.

Table 2. A comparative summary of the phases of digital forensics process models

Phases, sub-phases and activities	Process model # and their phases/sub-phases/activities													
	PM1	PM2	PM3	PM4	PM5	PM6	PM7	PM8	PM9	PM10	PM11	PM12	PM13	PM14
Identification (1.1), Detection (1.2)	1.1	1.1	1.1	1.1	1.2	1.2	1.2	1.1	-	-	1.1	1.2	1.1	1.2
Preparation (2.1), Readiness (2.2), Planning (2.3), Approach Strategy (2.4), Incident response (2.5), Formulation of response strategy (2.6), Notification (2.7), Documentation (2.8)	-	-	-	2.1 2.4	2.2	2.2	2.2	2.3	2.1 2.5	2.1 2.3 2.4 2.7	2.3	2.1 2.6	2.3	2.1 2.2 2.4 2.5
Preservation (3.1), Data duplication (3.2), Storage (3.3)	-	3.1	3.1	3.1	3.1	3.1	3.1 3.2	3.3	-	3.3	-	3.2	3.3	3.1 3.3
Acquisition (4.1), Collection (4.2), Gathering (4.3), Harvesting (4.4)	4.1	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.1	4.4	4.3	4.2

Analysis (5.1), Evaluation (5.4), Examination (5.3), Interpret (5.4), Investigation (5.5), Reconstruction (5.6), Hypothesis (5.7)	5.2	5.1 5.3	5.2 5.4 5.6	5.1 5.3	5.5 5.6	5.5 5.6	5.5 5.6	5.3 5.7	5.1 5.3 5.6	5.1 5.3 5.5	5.1 5.6	5.1	5.1	5.1	5.2 5.3 5.4 5.5 5.6 5.7
Presentation (6.1), Report (6.2), Admission (6.3), Proof & Defense (6.4)	6.3	6.1	6.1 6.2	6.1	6.1	6.1	6.1	6.1 6.4	6.1	6.1	6.1	6.2	6.4	6.1	6.2
Review (7.1)	-	-	-	-	7.1	-	7.1	-	7.1	-	-	-	-	-	7.1

3. Process Models Evolution and Ranking

In this section, the evolution of well-known digital forensics process models is discussed. In the early age, the nature and intensity of digital crimes were mostly elementary and straightforward. But, with the advancement of technology and tools, the magnitude of crime has gained a new extent. Day by day, a crime scene is approached to a more versatile and complex position. Hence, it is hard to handle a proper investigation based on fixed superannuated, stagnant digital forensics frameworks. We need a flexible and improved model such that it can conform to the current age technological advancements. Keeping this in mind, crime investigators and agencies tried to enhance old antecedent process models by adding or removing some activities to improve and tune their forensics experiments. Evolution is a process in which a subject matter passes by a level to a more advanced and mature stages than the previous. The evolution study of forensics process models is necessary for the better understanding of a particular model and to identify the changes in one frame. Such study can also compare the already available forensics frameworks and inspect their effectiveness with the current age requirements. A process model evolution can track the inheritance of phases and activities from old process models into the new one.

As discussed earlier, the phases and sub-phases of various process models follow a particular structure to carry out a proper digital forensics investigation. The major phases of renowned process models do not contradict much with the other models, and the differences are basically due to sub-phase activities and implementation strategies. In some process models, phases are well defined, whereas other models describe those in a general and abstract manner. Among various phases, the *data collection* phase is responsible for acquiring data from concerned digital devices. The acquired data are forensically analysed using business strategy, laws, and policies. As a major, the data collection and analysis phases are common to all digital forensics models discussed earlier. Other major phases, like identification and preparation are available in most of the process models. However the activities in major phases are quite distinct in most of the selected process models.

On analysis, it is found that there exists an intimate relationship between the forensics process models available in literatures. The relation can be defined as a weighted association between the models performing the same activities. The approach strategy of model execution may be different but we ignore the approach sequences. For example, in some forensics process model, review is associated with internal phases, whereas some model executes review as a standalone phase at the end of an investigation. The closeness of two different process models is calculated using *Cosine* similarity. As shown in Table 2, the first column represents the activities available in all process models. In other words, the first column can be mentioned as the union of all possible activities (including phases and sub-phases) available in selected 14 process models. In order to study the associativity of different process models, a contingency table, shown in Table 3, is constructed for each pair of models, wherein U is set of all activities, and u_i is the set of all activities available in process model i (i.e. PM_i). For example, the set of activities for the process model PM_1 is $u_1 = \{\text{identification,}$

evaluation, acquisition, admission}. Obviously, $= \bigcup_{i=1}^n u_i$, where n is 14 (total number of process models).

In the contingency table, a represents the number of activities that are common to both PM_i and PM_j , b represents the number of activities available in PM_i , c represents the number of available activities in process model PM_j , and d represents the number of unique activities available in either PM_i or PM_j .

Table 3. Contingency table for a pair of forensics process models

	PM_j	U
PM_i	a	b
U	c	d

A cosine similarity based association measure is defined which decide the associativity between a pair of process models using the contingency table. The Cosine similarity is used to measure the strength of closeness between a pair of objects having same dimension of feature vectors, and defined using equation 1, where X and Y is the set of feature values representing two different objects. Assuming various activities as features, the process models can be transformed into feature vectors. Using contingency table, the association between a pair of process models PM_i and PM_j , $\mu(PM_i, PM_j)$, can be determined using equation 2, where a , b , and c have same interpretation as discussed earlier in this section. Table 4, presents the values of association measures determined between all possible pairs of process models. The association measure lies in the interval $[0, 1]$, wherein a score approaching to 1 confirms the intimacy and resemblance of activities of the respective process models, and a score near 0 represents the fact that models' operational strategies and activities are quite dissimilar.

$$Cosine(X, Y) = \frac{|X \cap Y|}{\sqrt{|X|} * \sqrt{|Y|}} \quad (1)$$

$$\mu(PM_i, PM_j) = \frac{a}{\sqrt{b} * \sqrt{c}} \quad (2)$$

Table 4. Association measure values between the pair of process models

	PM1	PM2	PM3	PM4	PM5	PM6	PM7	PM8	PM9	PM10	PM11	PM12	PM13	PM14
PM1	1	0.20	0.38	0.18	0	0	0	0.18	0	0	0.41	0	0.20	0.12
PM2	-	1	0.46	0.87	0.43	0.46	0.41	0.58	0.58	0.41	0.50	0.15	0.33	0.48
PM3	-	-	1	0.40	0.40	0.43	0.38	0.40	0.40	0.25	0.46	0.14	0.15	0.53
PM4	-	-	-	1	0.38	0.40	0.35	0.50	0.63	0.59	0.43	0.27	0.29	0.58
PM5	-	-	-	-	1	0.94	0.94	0.38	0.50	0.24	0.29	0.13	0	0.67
PM6	-	-	-	-	-	1	0.88	0.40	0.40	0.25	0.31	0.14	0	0.62
PM7	-	-	-	-	-	-	1	0.35	0.47	0.22	0.27	0.25	0	0.63
PM8	-	-	-	-	-	-	-	1	0.38	0.59	0.43	0	0.58	0.42
PM9	-	-	-	-	-	-	-	-	1	0.47	0.43	0.27	0.14	0.67
PM10	-	-	-	-	-	-	-	-	-	1	0.27	0.13	0.27	0.55
PM11	-	-	-	-	-	-	-	-	-	-	1	0.15	0.50	0.29
PM12	-	-	-	-	-	-	-	-	-	-	-	1	0.15	0.36
PM13	-	-	-	-	-	-	-	-	-	-	-	-	1	0.19
PM14	-	-	-	-	-	-	-	-	-	-	-	-	-	1

A visual representation of the association measures, shown in Table 4, is presented in Fig. 1(a) as an undirected graph $G_1 = (V, E_1)$, where V is the set of all process models, and $E_1 \subseteq V \times V$ is the set of edges connecting the pairs of process models. The thickness of an edge is proportional to the association score between the connected nodes. As stated earlier, the more common activities are available between two models, the association score between the models is higher, and thereby the edge thickness. In this graph, it can be observed that, the edges (PM_5, PM_6) , (PM_5, PM_7) , (PM_6, PM_7) , and (PM_2, PM_4) have the maximum association scores, which implies that the numbers of mutual activities between the corresponding process models (nodes) are more than any other combination of the nodes. Fig. 1(b) presents a refined version of the graph G_1 (named as G_2), where an edge between a pair of nodes is created, provided the respective association measure is greater than or equal to 0.2. These two graphs can be used to assess the similarity between different forensics process models at different levels of granularity. After comparing both graphs, it can be noticed that the maximum weak edges were associated with the node PM_{12} , whereas node PM_2 has the strongest bond with its neighbour nodes.

Though from Figs. 1(a) and 1(b) the highly connected process models can be identified, one cannot identify which process model is served as most influential among the other models. In a social graph, an influential node is a node, which inspires its association with other nodes. From the descriptions of the individual process models, it can be perceived that most of the process models are inspired by some of the former process models. Therefore, in order to identify the basic and influential models, a directed graph, G_3 is constructed from graph G_2 (Fig. 1(c)), wherein the edges between the models are drawn in such a way that the head of an edge is associated with the inspirer node and that the tail with the selected node. Other edges that do not satisfying any such relation are discarded from G_3 .

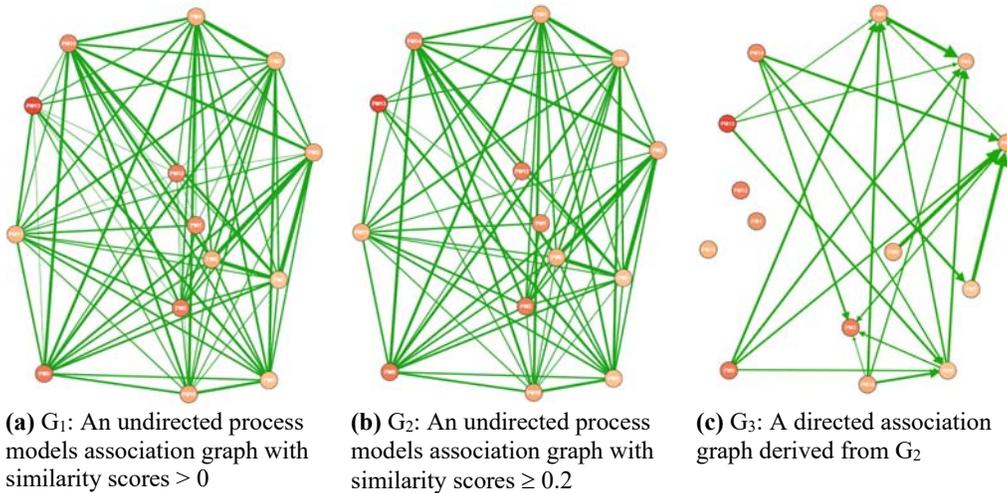


Fig. 1. Graphical representation of process models and their associations at different levels of granularity

In graph theory, the relative importance of a vertex within a graph is determined using the centrality measures. The node-degree and the edge-weight are the most enlightening characteristics of a graph to determine the advantage of a node with respect to its neighbours [33]. Though there are various centrality measures available, the appropriate measures are dependent on graph properties and its structure. Conceptually the simplest centrality measure is degree centrality, which is defined as the number of links incident upon a node. In case of a directed graph, like G_3 , two separate measures of degree centrality, namely *in-degree* and *out-degree* are calculated. The *in-degree* of a node is a count of the number of ties directed to the node, and it is often interpreted as a form of popularity, whereas the *out-degree* is the number

of ties that the node directs to others, and it is generally interpreted as a form of gregariousness [46]. The concept behind measuring degree centrality is that vertices with larger degrees exert greater response on a network. It measures the immediate influence of nodes in a network. Hence identifying the most connected vertices is a useful way to determine the important vertices. The degree centrality of a node v of a graph G , $C_d(v)$, is calculated using equation 3.

PageRank, a variant of eigenvector centrality measure, is another ranking method, which measures the influence of a node in a network-based graph. It works by counting the number and quality of edges associated to a node, and determines a rough estimation about the importance of the selected node. The underlying assumption of such measurement is that more important nodes are likely to receive more inbound edges from its neighbours. The PageRank of a node v of a graph G is calculated using equation 4, where, $M(v)$ is the set of pages which links to node v and $L(u)$ is the number of outbound edges on u .

The ranking of the process models based on both degree centrality measure and PageRank is presented in Table 5. The top two process models based on in-degree rank are PM_2 and PM_4 proposed by Palmer [17] and Reith et al. [20], respectively. However, on the basis of PageRank score, the process models PM_2 and PM_3 , proposed by Palmer [17] and Lee et al. [19], respectively have higher ranks than other models. After analysing the top-ranked four models in Table 5, it can be seen that the process models PM_2 , PM_3 , and PM_4 are competing with each other and can be considered as more basic and popular forensics process models.

$$C_d(v) = \frac{\text{deg}(v)}{|G| - 1} \quad (3)$$

$$PR(v) = (1 - d) + d * \sum_{u \in M(v)} \frac{PR(u)}{L(u)} \quad (4)$$

Table 5. Ranking of various process models

Rank	Process Models Ranking Criteria				
	Degree centrality w.r.t. G_1	Degree centrality w.r.t. G_2	Degree centrality w.r.t. G_3		Page rank
			In degree	Out Degree	
1	PM2	PM2	PM2	PM14	PM2
2	PM3	PM4	PM4	PM9	PM3
3	PM 4	PM11	PM3	PM10	PM4
4	PM11	PM3	PM8	PM8	PM5
5	PM14	PM14	PM5	PM13	PM8
6	PM8	PM8	PM7	PM5	PM7
7	PM9	PM9	PM11	PM4	PM6
8	PM10	PM10	PM14	PM7	PM13
9	PM5	PM7	PM9	PM6	PM9
10	PM6	PM5	PM10	PM2	PM10
11	PM7	PM6	PM6	PM3	PM14
12	PM12	PM13	PM13	PM11	PM11
13	PM13	PM12	PM12	PM12	PM12
14	PM1	PM1	PM1	PM1	PM1

4. Privacy-Preserving Digital Forensics Framework

Privacy issue is a major concern in computer forensics and security. In case of digital forensics investigation, an investigator needs to collect and examine user data from digital devices. The data for investigation can be associated with suspects or an accused. A suspect is

a person believed to be the one who committed a crime being investigated. In further proceeding of investigation, it may be found that the suspect don't have any role in the occurred crime. On the other hand, an accused is a person who has been charged of committing a crime. It is also important to mention that an accused does not mean a convicted person unless and until he/she is condemned in final verdict. Hence, it is unethical for an investigator to extract and use the user data which is private to a suspect or accused. Consequently, there is a need to consider only those data that are related to a crime and should handle them in a confidential manner [32]. However, it is quite impossible to provide full protection to such private data during data acquisition, storage and transfer. Forensic investigators may face challenges in determining the balance between key evidence and user privacy. Srinivasan [9] identified the importance of privacy policies in order to protect the confidential data of individual. Such privacy-protecting policies restrict an investigator to analyze any private data. The author defined ten important privacy-protecting policies both from the user as well as investigator perspectives. Out of total 10 policies, 5 policies were made with reference to the investigator, 2 policies were made with reference to the users, and 3 were made with reference to both investigator and users. Table 6, shows the list of policies and their association with investigators and users. However the privacy policies mentioned by Srinivasan [9] cover data acquisition phase only. Importance of defining policies in case of other investigation steps (i.e., preservation, presentation, etc.) is not mentioned.

Table 6. Privacy-preservation policies and their general criteria proposed by Srinivasan [9]

Criteria	Policies
Made with reference to investigator	(i) Keep one identical hard disk copy with user, (ii) Limit search for evidence to the goal of investigation, (iii) Treat time-stamped events and keep confidential, (iv) Transaction log store safely, and (v) Organizational policy should examine privacy violation
Made with reference to user	(i) Wipe out any irrelevant data, and (ii) Acknowledge packets via token number
Made with both user and investigator	(i) Preserve event logs in external nodes, (ii) Safeguard backed up relevant data, and (iii) Handle disposal of data in secure manner

The state of privacy and digital forensics in USA was studied by Adams [34], in which the author discussed about legal issues that may arise during the use of a forensic tool. The developer should be aware of some (or all) of the legal issues that may arise during the use of the forensics tools. According to the author, the major legal constraints for the forensic tools are reliability and privacy preservation, and both should be kept in mind while designing the tools. This study also suggests that the investigation scope and search warrant should be specific and clear, and it should be ensured that the digital investigator shouldn't exceed the investigation scope and goal as well to allow the court of law to verify the digital evidence reliability by tracing investigator's activities.

In 2009, Reddy et al. [35] presented a forensics framework to provide enterprises with a generic forensic readiness capability for information privacy incidents. Their framework has been consisted as a series of business processes and forensics approaches, and follows a hierarchical tree-like structure with four levels and multiple blocks. The blocks contain technical and non-technical readiness components with various privacy and security policies. Using such structure, an enterprise can conduct quality privacy-related forensics investigations on information privacy incidents. In a specific way, the framework provides guidance for determining high-level policies, business processes, and organizational functions [35].

In general, once private information come out from an expected flow and disclosed, it can not be restored as private again. Thus, there is a need to balance the efficacy of the investigation against privacy violation [36]. Keeping in mind, the importance of balancing privacy and investigation process, Croft et al. [36] defined a way of forensic investigation

through the release of private information in a sequential manner. The proposed mechanism in this paper allows data to be packaged in cryptographic way, such that it benefits both suspect (whose privacy is protected until incriminating evidence has been shown) and investigator (no need for repeated permission). The user data in the proposed framework are classified into four hierarchical levels that act as a guiding channel for the investigator to cross-check the information he/she is accessing. After classifying data into levels, they are encrypted using a searchable encryption scheme. The number of encryption process for each data type depends upon the level to which the data are assigned. For example, the data in level 4, are encrypted four times and the investigator must run four-pass decryption process. The searchable decryption scheme is used to allow decryption if some parts of encrypted data are possessed. However, in this model, a prior knowledge as well as proof of hypothesis are necessary in granting access to information.

As discussed earlier, irrelevant personal data should be excluded during computer forensic examination to enable data privacy. Law et al. [37] designed a cryptographic model to be incorporated into the current digital framework to protect users' private data. According to their model, the forensics investigators have to allow the data owner to encrypt storage media with a key and perform indexing over it. After performing the encryption, the investigators could perform a keyword search (or list of keyword relevant to investigation) with the help of the encryption key. Such process is targeted to add a possible way to protect data privacy.

The use of third party storage (which is shared by many users; i.e., shared drive, cloud data, etc.) creates problem for forensic analyst to acquire data and privacy preserving issue due to huge volume and access right concerns. In order to solve such problems, Hou et al. [38] presented a homomorphic and commutative encryption (HCE) scheme where remote server and investigators encrypt data and queries. The investigator first encrypts the necessary queries and server administrator results relevant encrypted data based on investigator key. Finally, the investigator decrypts data with the administrator's key. The commutative encryption introduces a Trusted Third Party (TTP) that supervises the administrator to prevent any unfair play. The TTP also inspects whether the server administrator returns all searched results or not.

However, there is a limitation in the above schemes where administrator finds subset of documents that match a certain keyword rather than simultaneous multiple keywords. The subset of documents in case of single keyword search is generally huge and it may decrease the analysis efficiency. To obtain a better search results and to improve the investigation efficiency, it is essential to perform multiple keyword searches. Sequence of keyword search in encrypted documents is introduced in [39]. Such framework limits data disclosure during forensics investigation. However, most of the privacy preserving forensic frameworks employ a binary privacy level (user privacy is either fully protected or not at all) to user privacy. In contrary, Halboob et al. [40] introduced a four-level privacy protection mechanism, based on data relevancy and investigator authorization. Such privacy model is more flexible and more acceptable as it restricts unauthorized investigator from accessing relevant data.

In addition to the privacy concern, a framework called "Privacy Preserving Efficient Digital Forensic Investigation" (PPEDFI) proposed by Gupta [41] in 2013. The PPEDFI framework contains three modules – *expert system*, *evidence extraction*, and *ranking*. The *expert system* module assists investigator by providing the basic information to start the investigation. The *evidence extraction* module extracts evidences based on search query, whereas the *ranking* module assigns rank to files obtained from evidence extraction module on the basis of their relevancy to containing evidences. Such ranking scenario can help an investigator to analyze a crime efficiently.

In 2014, Dehghantanha et al. [42] established a foundation towards a privacy-respecting digital investigation model, which targets a cross-disciplinary field of research favoring both the legal requirement and data privacy in-line. The authors reviewed and

elaborated the main research efforts in this research discipline, followed by a promising conclusion. Potential privacy concerns during digital investigation in light of the European Union (EU), United States (US), and Asia-Pacific Economic Cooperation (APEC) jurisdictions have also been discussed in this paper.

An ontology-based forensics framework is proposed by Wan et al. [43], which focuses on various issues related to personal privacy protection. The privacy protection is realized through allocating rights at different levels of hierarchy. As claimed by the authors, the suggested framework can protect privacy through means of identifying authentication and the scenario as well as protecting sensitive information and images. In order to reduce data encryption cost and investigation time, Halboob et al. [44] proposed a forensic framework which introduces privacy levels that are useful for collecting only relevant data, rather than taking bit-by-bit image of the physical storage, resulting in reduce investigation cost and time. The forensic data is marked as four possible groups based on its relevance and privacy. In order to access relevant but non-private data, no particular privacy concern is needed. However, for relevant and private data, a well-defined privacy preserving techniques are needed during data imaging and analysis process, which can reduce the computational cost to 50% in comparison to the frameworks that collects all data blindly.

5. Review of the State-of-the-Art Digital Forensics Tools

Digital forensics tools are predefined software or list of integrated methods, which are engaged in accomplishing a digital investigation process. Computer forensics was developed as an independent field in early 2000, when computer-based crime started growing with increasing popularity [48]. However, in recent years, the exponential growth of technologies has brought some serious challenges for digital forensics research. As a result, the tools and techniques for digital investigation have been changed due to the advances in forensics. Usage of tools in digital forensics has much more advantages. A tool can yield a better analysis and visualization by minimizing the investigation time and efforts. During evidence examination, digital evidence sources are interpreted using one or more forensic tools that can provide a file system abstraction to the digital evidence source in such a way that their contents may be examined for evidence tracing [48].

Table 7. A summarized view of the well-known digital forensics tools

S. No.	Tool/ Application name	Major tasks	Supported platform	Software license type	Provider/ Developer	Reference forensics model
1	EnCase	Data identification, Acquisition, Analysis, Documentation, Reporting	Linux, Mac, Windows, Solaris	Commercial	Guidance Software	PM5, PM8, PM14
2	Forensics Toolkit (FTK)	Identification, Imaging, Analysis, Reporting	Windows	Commercial	AccessData	PM5, PM14
3	X-Ways	Imaging, File carving, Data recovery, Analysis	Windows	Personal, Commercial	X-Ways Software Technology AG	NA
4	WinHex	Imaging, Analysis, Privacy-protection, Wipe/ Erase confidential files	Windows	Personal, Commercial	X-Ways Software Technology AG	NA
5	The SleuthKit and Autopsy	Data acquisition, Analysis, Data carving	Windows, Linux, Unix	Freeware	Brian Carrier [50, 51]	PM5

6	Passware Recovery Kit Forensic	Password recovery, Memory acquisition, Live memory analysis, Data decryption, Evidence discovery, Encryption analysis	Windows	Commercial	Passware	NA
---	--------------------------------	---	---------	------------	----------	----

As suggested by the domain experts, process-specific tools should be used for different stages of digital forensics investigation process. The tools selection should be based on both the performance of the tools and relevance of the digital evidence towards solving a specific case [49]. A number of tools are available for specific tasks including disk imaging, data recovery and carving, file analysis, document metadata extraction, memory imaging, memory analysis, network forensics, log file analysis, and mobile device forensics. On the basis of the underlying operating system and supported environment, forensics tools can be categorized as Linux-based tools, Macintosh-based tools, Windows-based tools, Android-based tools, and so on. On the other hand, based on the license and usage permission, forensics tools can be classified as open-source, freeware, and commercial tools. In addition to these specific tools, there exists few general-purpose forensics toolkits and software like EnCase, FTK, X-Ways, etc. that can be used to perform the majority of investigation steps under one umbrella. A survey conducted by James [47] reveals that EnCase is the most preferred primary software chosen by 80% of the digital investigation organizations. The other famous tools such as FTK, X-Ways forensic, and other miscellaneous tools are also used, but not nearly as often. The average percentage of cases in which only the chosen primary softwares (EnCase, FTK and X-Ways) are used is 77.9% cases [47]. In this survey, it is also mentioned that majority of the cases focused on Windows-based forensics (86.99%), rather than Linux (7.3%) or Mac (5.7%). In another survey by Hibshi et al. [53] it is reported that FTK and EnCase are the most famous among the forensics practitioners. This survey was conducted among 114 participants where 43% were forensics experts, 39% were intermediates, and the rest were just beginners. The most preferred open-source forensics tools were voted as Autopsy and the Sleuth Kit. Table 7 presents a list of the popular forensics tools along with other related details including license type, supporting platform, developer, and reference forensics models. A brief description of these tools is presented in the following sub-sections.

5.1. EnCase

EnCase is the most famous digital forensics tool which conducts efficient, forensically-sound data collection, identification, analysis, and reporting in a repeatable and defensible manner. EnCase technology is developed by Guidance Software [45] and it is available within a number of products, currently including: EnCase Forensic, EnCase Endpoint Security, EnCase eDiscovery, and EnCase Portable. Each of the products has their own strength and limitations. EnCase is also acceptable by court of law around the globe for digital media analysis. The Encase uses the de-facto standard file format for preserving crime-related digital evidence and allows an examiner to acquire data from a wide range of devices that are either volatile or non-volatile in nature, unearth potential evidence, and summarizes the findings as detailed reports. EnScript scripting allows the examiners to manipulate data according to their own needs. Its enterprise edition supports Unix, Linux, Mac, Solaris, and Windows platform for the investigation purpose. Examiner privileges are also defined to prevent investigator from performing any unauthorized data alteration. A customizable report assists decision makers to summarize case-relevant information and audit logs are used to generate detailed reports to obtain information about the steps performed during an investigation process. EnCase is one of the complete tools which can be used in most of the digital forensics phases.

5.2. Forensics Toolkit (FTK)

FTK is another famous and most widely used digital forensics investigation solution developed by AccessData [54]. It provides comprehensive processing and indexing up front, which caters faster filtering and searching. FTK provides innovative and integrated features to support data processing integrity, speed, imaging, indexing, and analysis depth. It also takes control of Big Data. The mature database-driven and enterprise-class architecture of FTK allows to handle and make sense of exponentially grown datasets through processing stability and data visualization. Apart from processing a wide range of data, an FTK user can also analyze the registry, decrypt files, crack passwords, and build reports within a single platform. It also includes a standalone disk imaging program called FTK Imager, which is a simple tool used for concise graphical user interface. The images of a hard disk created by FTK imager can be saved in one file or in a segment, and such files or segment of files can be reconstructed later on, if required. It checks MD5 hash values and confirms the integrity of data before closing the files.

5.3. X –Ways

X-Ways Forensics [56] is a general-purpose data recovery and forensics tools mostly available for Windows platform. It performs safe recoveries on digital storage devices like hard disks, memory card, flash disks, CDs, DVDs etc. It incorporates several automated file recovery mechanisms and allows to conveniently recover data manually. It has multi-purpose functionalities like disk imaging, RAM content analysis, data analysis, file carving, text search etc. In X-Ways Forensics, disks, interpreted image files, virtual memory, and physical RAM are opened in strictly read-only mode. Such write protection of X-Ways Forensics ensures that no original evidence can be altered accidentally, which can be a crucial aspect in court proceedings [55].

5.4. WinHex

WinHex [57] is an advanced binary editor which provides access to all files and hard disk sectors. Though WinHex and X-Ways Forensics share the same code base, WinHex is the base tool with lesser features than the X-Ways. In case, investigator needs to edit disk sectors, free disk space, slack space, and remove irrelevant private data, WinHex can be chosen; otherwise, X-Ways can be used to preserve the original evidences with more options.

5.5. Sleuth Kit and Autopsy

The Sleuth Kit is a collection of command line tools and C library which can be used to perform in-depth analysis of various file systems. It is an open-source tool developed by Brian Carrier [50], mainly focusing on data acquisition and disk image analysis. Nevertheless, Autopsy [51] is a graphical interface that sits on top of the Sleuth Kit. The features associated with Sleuth Kit and Autopsy are timeline analysis, hash filtering, file system forensics analysis, keyword searching, Web artifact, and email analysis. Both Sleuth Kit and Autopsy are the most suitable free tools for disk image analysis in Linux as well as Windows environment.

5.6. Passware Recovery Kit Forensic

Passware Kit Forensic is the digital evidence discovery solution that reports all password-protected items on a computer and decrypts them efficiently. Though Passware software supports Windows platforms only, it can recover passwords for some files created on Macintosh and it can be run on virtual PC or parallel desktop as well. Another important feature of Passware is that, it can be easily integrated with EnCase.

6. Conclusion and Future Challenges

The digital forensics practices have been broadened significantly because of their importance. Such practices were first measured in early 2000. It is a rapidly growing field in information technology where a crime is traced using digital media linked with computer-assisted crimes. In this paper, we have presented an in-depth comparative analysis of the well-known digital forensics process models. In addition, the evolutionary development of forensics models is discussed in a proper way and their similarity and reference inter-relations are modeled as an undirected and directed graph, respectively. The advantages and limitations of the forensics models are discussed and various ranking approaches are used to rank them from different perspectives, which can assist in choosing most suitable model for a digital investigation task. Due to increasing concerns about users privacy in investigation process, a number of privacy-preserving process models are proposed by various researchers. A brief review of such models and their mapping with the current technologies are also presented in this paper. We have also presented a review of the available digital forensics tools and their mapping to the different phases/sub-phases of the forensics models.

In current age, the care for users' privacy while investigating their digital media is necessary. But preserving users' privacy is itself a challenging task. Though, a number of forensics policies are developed in this regard, they are not scientifically strong enough to protect users privacy. Moreover, most of the existing forensics policies and frameworks are organization-based, and country-based investigation frameworks that still need more research. On the other hand, due to significant increase in the volume of digital data, investigation task is getting tedious and annoying. From the existing literatures, it is noticeable that, evidence mining and preservation from enormous and varied digital data (aka Big Data) is yet an open issue to be solved in near future. Moreover, the available forensics tools are mostly dedicated to analyze digital data in static manner. To re-inspect an investigation process, the changes have to be analyzed again from the beginning, even though there is minor change in the underlying policy. Hence, a real-time analysis of investigation process needs to be devised for efficiently handling the dynamism of data and investigation policies.

References

- [1] Kohn, M. D., Eloff, M. M., and Eloff, J. H., *Integrated Digital Forensic Process Model*, Computers & Security, Vol. 38, 2013, pp. 103-115.
- [2] Richter, J., Kuntze, N., and Rudolph, C., *Security Digital Evidence*, In Proceedings of the IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE), Oakland, USA, IEEE Computer Society, May, 2010, pp. 119-130.
- [3] Garfinkel, S. L., *Digital Forensics*. American Scientist, Vol. 101, No. 5, 2013, pp. 370.
- [4] Chen, B. Y., *Computer Forensics in Criminal Investigations*, Dartmouth Undergraduate Journal of Science Online, Vol. 13, 2013.
- [5] Rogers, M. K., and Seigfried, K., *The Future of Computer Forensics: A Needs Analysis Survey*, Computers & Security, Vol. 23, No. 1, 2004, pp. 12-16.
- [6] Marcella Jr, A., and Greenfield, R. S., *Cyber Forensics: A Field Manual for Collecting, Examining, and Preserving Evidence of Computer Crimes*. CRC Press, Second Edition, 2002.
- [7] Reilly, D., Wren, C., and Berry, T., *Cloud Computing: Forensic Challenges for Law Enforcement*. In Proceedings of the International Conference on Internet Technology and Secured Transactions (ICITST), London, UK, IEEE, November 2010, pp. 1-7.
- [8] Mohay, G., *Technical Challenges and Directions for Digital Forensics*, In Proceedings of the International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE), Taipei, Taiwan, IEEE, November 2005, pp. 155-161.

- [9] Srinivasan, S., *Security and Privacy in the Computer Forensics Context*. In Proceedings of the International Conference on Communication Technology (ICCT), Guilin, China, IEEE, November 2006, pp. 1-3.
- [10] Noblett, M. G., Pollitt, M. M., and Presley, L. A., *Recovering and Examining Computer Forensic Evidence*, Forensic Science Communications, Vol. 2, No. 4, 2000, pp. 1-13.
- [11] Köhn, M., Olivier, M. S., and Eloff, J. H., *Framework for a Digital Forensic Investigation*, In Proceedings of Information Security South Africa (ISSA), Sandton, South Africa, IEEE, July 2006, pp. 1-7.
- [12] Garfinkel, S. L., *Digital Forensics Research: The Next 10 Years*. Digital Investigation, Vol. 7, 2010, pp. S64-S73.
- [13] Irons, A., and Lallie, H. S., *Digital Forensics to Intelligent Forensics*. Future Internet, Vol. 6, No. 3, 2014, pp. 584-596.
- [14] Quick, D., and Choo, K. K. R., *Data Reduction and Data Mining Framework for Digital Forensic Evidence: Storage, Intelligence, Review and Archive*, Trends & Issues in Crime and Criminal Justice, Vol. 480, 2014, pp. 1-11.
- [15] Guarino, A., *Digital Forensics as a Big Data Challenge*. Securing Electronic Business Processes, Springer, Chapter 17, 2013, pp. 197-203.
- [16] Pollitt, M., *Computer Forensics: An Approach to Evidence in Cyberspace*. In Proceedings of the National Information Systems Security Conference (NISSC), Maryland, USA, October 1995, pp. 487-491.
- [17] Palmer, G., *A Road Map for Digital Forensic Research*, In proceedings of the Digital Forensic Research Workshop (DFRWS), Utica, New York, August 2001, pp. 27-30.
- [18] Willassen, S. Y., and Mjølunes, S. F., *Digital Forensics Research*. Teletronikk, Vol. 30, 2005, pp. 92-97.
- [19] Lee, H. C., Palmbach, T., and Miller, M. T., *Henry Lee's Crime Scene Handbook*, Academic Press, First Edition, 2001.
- [20] Reith, M., Carr, C., and Gunsch, G., *An Examination of Digital Forensic Models*, International Journal of Digital Evidence, Vol. 1, No. 3, 2002, pp. 1-12.
- [21] Carrier, B. D., and Spafford, E. H., *Getting Physical with the Digital Investigation Process*, International Journal of Digital Evidence, Vol. 2, No. 2, 2003, pp. 1-20.
- [22] Carrier B. D., and Spafford E. H., *An Event-Based Digital Forensic Investigation Framework*, In Proceedings of the Digital Forensic Research Workshop (DFRWS), Baltimore, USA, August 2004, pp. 1-12.
- [23] Baryamureeba, V., and Tushabe, F., *The Enhanced Digital Investigation Process Model*, In Proceedings of the Digital Forensic Research Workshop (DFRWS), Baltimore, USA, August 2004.
- [24] Ciardhuáin, S. Ó., *An Extended Model of Cybercrime Investigations*, International Journal of Digital Evidence, Vol. 3, No. 1, 2004, pp. 1-22.
- [25] Beebe N. L., and Clark J. G., *A Hierarchical, Objectives-Based Framework for the Digital Investigations Process*, Digital Investigation, Vol. 2, No. 2, 2005, pp. 147-167.
- [26] Pilli, E. S., Joshi, R. C., and Niyogi, R., *A Generic Framework for Network Forensics*, International Journal of Computer Applications, Vol. 1, No. 11, 2010, pp. 1-6.
- [27] Zainudin, N. M., Merabti, M., and Llewellyn-Jones, D., *Online Social Networks as Supporting Evidence: A Digital Forensic Investigation Model and Its Application Design*, In Proceedings of the International Conference on Research and Innovation in Information Systems (ICRIIS), Kuala Lumpur, Malaysia, November 2011, pp. 1-6.
- [28] Jeong R. S. C., *FORZA – Digital Forensics Investigation Framework that Incorporates Legal Issues*, Digital Investigation, Vol. 3, 2006, pp. 29-36.
- [29] Freiling, F. C., and Schwittay, B., *A Common Process Model for Incident Response and Computer Forensics*, In Proceedings of the International Conference on IT-Incident

- Management & IT-Forensics (IMF), Stuttgart, Germany, SIDAR, September 2007, pp. 19-40.
- [30] Perumal S., *Digital Forensic Model Based on Malaysian Investigation Process*, International Journal of Computer Science and Network Security, Vol. 9, No. 8, 2009, pp. 38-44.
- [31] Halboob, W., and Mahmud, R., *State of the Art in Trusted Computing Forensics*, In Future Information Technology, Application, and Service, LNEE, Springer, 2012, pp. 249-258.
- [32] Bui, S., Enyeart, M., and Luong, J., *Issues in Computer Forensics*, Santa Clara University Computer Engineering, USA, 2003, pp. 1-35.
- [33] Noori, A., *On the Relation Between Centrality Measures and Consensus Algorithms*. In Proceedings of the International Conference on High Performance Computing and Simulation (HPCS), Istanbul, Turkey, IEEE, July 2011, pp. 225-232.
- [34] Adams, C. W., *Legal Issues Pertaining to the Development of Digital Forensic Tools*, In Proceedings of the International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE), California, USA, May 2008, pp. 123-132.
- [35] Reddy, K., and Venter, H., *A Forensic Framework for Handling Information Privacy Incidents*, In Proceedings of the IFIP International Conference on Digital Forensics, Florida, USA, Springer, January 2009, pp. 143-155.
- [36] Croft, N. J., and Olivier, M. S., *Sequenced Release of Privacy-accurate Information in a Forensic Investigation*, Digital Investigation, Vol. 7, No. 1, 2010, pp. 95-101.
- [37] Law, F. Y., Chan, P. P., Yiu, S. M., Chow, K. P., Kwan, M. Y., Tse, H. K., and Lai, P. K., *Protecting Digital Data Privacy in Computer Forensic Examination*, In Proceedings of the International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE), California, USA, May 2011, pp. 1-6.
- [38] Hou, S., Uehara, T., Yiu, S. M., Hui, L. C., and Chow, K. P., *Privacy Preserving Confidential Forensic Investigation for Shared or Remote Servers*, In Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), Dalian, China, October 2011, pp. 378-383.
- [39] Hou, S., Uehara, T., Yiu, S. M., Hui, L. C., and Chow, K. P., *Privacy Preserving Multiple Keyword Search for Confidential Investigation of Remote Forensics*, In Proceedings of the International Conference on Multimedia Information Networking and Security (MINES), Shanghai, China, November 2011, pp. 595-599.
- [40] Halboob, W., Abulaish, M., and Alghathbar, K. S., *Quaternary Privacy-levels Preservation in Computer Forensics Investigation Process*, In Proceedings of the International Conference for Internet Technology and Secured Transactions (ICITST), Abu Dhabi, UAE, December 2011, pp. 777-782.
- [41] Gupta, A., *Privacy Preserving Efficient Digital Forensic Investigation Framework*, In Proceedings of the International Conference on Contemporary Computing (IC3), Noida, India, August 2013, pp. 387-392.
- [42] Dehghantanha, A., and Franke, K., *Privacy-Respecting Digital Investigation*, In Proceedings of the International Conference on Privacy, Security and Trust (PST), Toronto, Canada, July 2014, pp. 129-138.
- [43] Wan, X., He, J., Huang, N., and Mai, Y., *Ontology-Based Privacy Preserving Digital Forensics Framework*, International Journal of Security & Its Applications, Vol. 9, No. 4, 2015, pp. 53-62.
- [44] Halboob, W., Mahmud, R., Udzir, N. I., and Abdullah, M. T., *Privacy Levels for Computer Forensics: Toward a More Efficient Privacy-Preserving Investigation*, Procedia Computer Science, Vol. 56, 2015, pp. 370-375.
- [45] Guidance Software: EnCase®. Available at: <http://www.guidancesoftware.com/> (Last accessed: 15/11/2016).

- [46] Panda, M., Dehuri, S., and Wang, G. N., *Social Networking: Mining, Visualization, and Security*, First Edition, Springer, 2014.
- [47] James J. I., *Survey of Evidence and Forensic Tool Usage in Digital Investigation*, AER Specification Result, University College Dublin, 2009.
- [48] Raghavan, S., *Digital Forensic Research: Current State of the Art*, CSI Transactions on ICT, Vol. 1, No. 1, 2013, pp. 91-114.
- [49] Saleem, S., Popov, O., and Baggili, I., *A Method and a Case Study for the Selection of the Best Available Tool for Mobile Device Forensics Using Decision Analysis*, Digital Investigation, Vol. 16, 2016, pp. S55-S64.
- [50] Carrier, B., The Sleuth Kit (TSK). Available at: <http://www.sleuthkit.org/sleuthkit/>, (Last accessed: 15/11/2016).
- [51] Carrier, B., Autopsy Forensic Browser. Available at: <http://www.sleuthkit.org/autopsy/> (Last accessed: 15/11/2016).
- [52] Cohen, M. I., *Pyflag – An Advanced Network Forensic Framework*, Digital investigation, Vol. 5, 2008, pp. S112-S120.
- [53] Hibshi, H., Vidas, T., and Cranor, L., *Usability of Forensics Tools: A User Study*, In Proceedings of the International Conference on IT Security Incident Management and IT Forensics (IMF), Stuttgart, Germany, May 2011, pp. 81-91.
- [54] Access Data: Forensic Toolkit. Available at: <http://www.accessdata.com> (Last accessed: 15/11/2016).
- [55] Fleischmann, S., X-Ways Forensics/ Winhex Manual, Available at: <http://www.x-ways.net/winhex/manual.pdf> (Last accessed: 15/11/2016).
- [56] X-Ways Forensics, Available at: <http://x-ways.net/forensics/index-m.html> (Last accessed: 15/11/2016).
- [57] WinHex, Available at: <http://x-ways.net/winhex/index-m.html> (Last accessed: 15/11/2016).