Forensic Science International: Synergy 2 (2020) 489–520

Contents lists available at ScienceDirect

Forensic Science International: Synergy

journal homepage: https://www.journals.elsevier.com/forensic-science-international-synergy/

Interpol review of digital evidence 2016 - 2019

Paul Reedy

4th Street Global, USA

ARTICLE INFO

Article history:
Received 6 January 2020
Accepted 16 January 2020
Available online 19 March 2020

Keywords:
Digital forensics
Digital evidence
Network forensics

ABSTRACT

This review paper covers the forensic-relevant literature in digital evidence from 2016 to 2019 as a part of the 19th Interpol International Forensic Science Managers Symposium. The review papers are also available at the Interpol website at: https://www.interpol.int/content/download/14458/file/Interpol Review Papers 2019.pdf

1. Introduction

This review paper has been primarily compiled from numerous peer reviewed publications. The attempt has been made to give fair representation to the wide range of views without judgement. Consequently, many of the views expressed by various authors, although represented in this review, do not represent, and might be inconsistent with, those of this author.

In the ten years to 2017, the field of digital evidence has expanded to meet the challenges from advances in smart technology, smartphone apps, implanted medical devices, and malware. People with new skills sets in artificial intelligence and data science are joining the field, and digital investigation techniques and methods are being applied to crime analysis and intelligence. Digital forensic intelligence is becoming a priority in order to understand inter-jurisdictional criminal activity. Best practice guidelines were established over a decade ago and do not meet the challenges of smart technology, and some do not address memory forensics, database forensics, or network forensics [1].

Although important to the field to be able to demonstrate competence and provide confidence to stakeholders, best practices and automated tools are not the panacea for digital evidence. Each digital evidence case presents new challenges for which digital evidence practitioners should be problem solvers. The future digital evidence practitioner will need to be equipped with the knowledge and skills to address forensic questions in the presented case [1].

On behalf of the Organisation of Scientific Area Committees for Forensic Science, the Task Group on Digital/Multimedia Evidence prepared a document entitled A Framework for Harmonizing Forensic Science Practices and Digital/Multimedia Evidence [2]. The Task Group was commissioned to clarify how digital and multimedia evidence fits within forensic science, and to the broader question of forensic science itself. It is noted that digital and multimedia evidence is unique among forensic disciplines as it serves investigative, procedural, and scientific functions with the outcomes synthesized into expert opinions and conclusions.

Building on from the fundamental principle that every contact leaves a trace, the Task Group note that "[a] is any modification, subsequently observable, resulting from an event." Forensic science addresses questions that are, potentially, in all disciplines: authentication, identification, classification, reconstruction, and evaluation [2]. They arrived at the following definition of forensic science:

> *"The systematic and coherent study of traces to address questions of authentication, identification, classification, reconstruction, and evaluation for a legal context."*

The term *systematic* refers to empirically supported research, controlled experiments, and repeatable procedures applied to traces. The term *coherent* refers to logical reasoning and methodology. The term *legal context* refers to criminal, civil and regulatory functions, which also extends into human rights, employment, natural disasters, and security matters.

Digital and multimedia evidence includes the following subdisciplines for which descriptions are provided: speaker recognition, facial identification, video/image technology and analysis, and digital evidence.

The digital forensics market by component (hardware, software, and services), type (computer forensics, network forensics, mobile device forensics, and cloud forensics), tools, and verticals is expected to grow from USD 4.62 billion in 2017 to USD 9.68 billion by

E-mail address: paul@4thStreetGlobal.com.

https://doi.org/10.1016/j.fsisyn.2020.01.015

2022, a compound growth rate of 15.9% per year. The demand is expected to be driven by stringent government regulations and an increasing frequency of cyber attacks on organisations, and the increasing prevalence of Internet of Things devices. Advancements in technology have intensified the sophistication of attacks on devices. The majority of business and personal transactions occur electronically, including deals over email. The banking, financial services, and insurance sectors are expected to grow the fastest as they are targeted with cybercrimes and digital frauds. Mobile banking has also led to an increase in fraud incidents through access to customer information and data [3].

There are a number of common themes that appeared through the period 2016–2019. Overall, there was a widespread acknowledgement that the challenges and practice of digital forensics continues to become more complex due to the increasingly sophisticated and complex consumer and business technology environment. At a technical level and largely predictable, the impact of cloud computing and the rapidly growing prevalence of Internet of Things continues to challenge forensic analysts and their employing organisations. More surprising was the acknowledgement by several authors of the impact of human factors and human fallibility in the practice of digital forensics that, seemingly, indicates a shift (or a further shift) away from the previous general, although not universal, acceptance of digital forensics is a fact-based discipline. Moreover, human factors were shown by several authors to have an impact on quality assurance.

1.1. Statistical survey

The digital forensics market is expected to grow from $4.62B in 2017 to $9.68B by 2022, an annual compound growth rate of almost 16%. The anticipated market drivers are government regulations, increasing cyber incidents experienced by businesses, and the rapidly growing presence of Internet of Things applications and devices [4]. The report noted that business and personal transactional activities are performed electronically, deals are made over email, and confidential data is stored on personal address books and storage media. The banking, financial services, and insurance sectors are expected to contribute a substantial proportion to this growth.

As of October 3, 2019, there were 2,809,148 Android applications on Google Play [5]. The smart phone market has experienced some changes over the past decade Although some variability exists in the statistics provided by different organisations, there are shifts occurring in the market. Most notably is the increasing presence of Chinese smart phones during the period of the review.

Company	Q4 2009	Q4 2010	Q4 2011	Q4 2012	Q2 '16	Q2 '17	Q2 '18	Q2 '19
Samsung	3.3%	9.4%	22.5%	29.1%	22.7%	22.9%	21.0%	22.7%
Apple	16.1%	15.9%	23%	20.9%	11.7%	11.8%	12.1%	10.1%
Huawei			3.5%	4.6%	9.3%	11.0%	15.9%	17.6%
Xiaomi					3.9%	6.2%	9.5%	9.7%
OPPO					6.6%	8%	8.6%	8.9%
LG				3.8%				
Lenovo				4.1%				
ZTE			4%	4.4%				
vivo					4.8%			
Sony			3.9%	4.5%				
RIM	19.9%	14.3%	8.1%					
HTC			6.4%					
Nokia	38.6%	27.6%	12.2%					
Others	17.6%	24.3%	16.4%	28.6%	41.0%	40.1%	32.9%	31.0%

Global market share held by leading smartphone vendors from 4th quarter 2009 to 2nd quarter 2019 [125].

Company	2017 Q4	2018 Q1	2018 Q2	2018 Q3	2019 Q4	2019 Q1
Samsung	18.9%	23.5%	21.0%	20.3%	18.8%	23.0%
Huawei	10.7	11.8%	15.9%	14.6%	16.2%	18.9%
Apple	19.6%	15.7%	12.1%	13.2%	18.3%	11.8%
Xiaomi	7.1%	8.4%	9.5%	9.5%	6.7%	8.9%
Vivo	6.0%	5.6%	7.9%	8.3%	6.9%	7.4%
OPPO	6.9%	7.4%	8.6%	8.4%	7.9%	7.4%
Others	30.9%	27.6%	25.0%	25.7%	25.1%	22.7%

Worldwide Top 5 Smartphone Shipment Company market Share [52].

Vendor	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Samsung	4.3	10.6	18.4	25.3	32.1	31.9	32.3	33.0	30.7	31.7
Apple	29.3	27.7	24.7	24.4	23.9	20.3	19.3	19.7	20.6	22.5
Unknown	0.0	0.0	8.9	11.6	10.3	12.1	14.8	7.7	6.3	4.3
Nokia	37.0	38.2	29.9	21.4	13.7	9.2	5.5	2.5	1.5	1.1
Huawei	0.0	0.0	0.5	0.8	1.2	2.0	3.4	4.5	6.6	8.8
LG	0.2	0.6	2.0	2.8	3.5	4.2	4.1	3.7	3.2	2.7
Sony	7.7	5.4	3.2	2.5	3.2	3.4	2.8	2.0	1.4	1.0
Xiaomi	0.0	0.0	0.0	0.0	0.2	0.7	1.2	3.3	6.7	7.8
Motorola	0.6	1.0	1.5	1.4	1.8	2.2	2.2	2.2	2.7	2.7
HTC	0.1	1.0	3.8	2.9	2.3	2.3	2.0	1.4	0.9	0.6
RIM	19.4	14.7	5.5	3.6	2.0	1.3	0.8	0.3	0.1	0.0
Lenovo	0.0	0.0	0.0	0.2	0.6	1.4	2.4	3.0	2.4	1.4
Oppo	0.0	0.0	0.0	0.0	0.1	0.4	0.7	2.8	4.1	4.5
Micromax	0.0	0.0	0.3	1.1	2.0	2.5	2.2	1.4	0.9	0.4
Asus	0.0	0.0	0.0	0.0	0.1	0.7	1.0	1.2	1.0	0.8
General Mobile	0.0	0.0	0.0	0.0	0.6	1.3	1.0	0.8	0.4	0.4
Mobicel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	2.4	3.0
BBK	0.0	0.0	0.0	0.0	0.0	0.1	0.3	1.5	1.1	0.6
Google	0.2	0.2	0.4	0.5	0.6	0.7	0.6	0.4	0.4	0.6
ZTE	0.0	0.0	0.3	0.4	0.4	0.6	0.5	0.6	0.6	0.3
Alcatel	0.0	0.0	0.1	0.2	0.5	0.6	0.4	0.6	0.5	0.3
Lava	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.6	0.5	0.3
Gionee	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.4	0.4	0.2
Vodafone	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.3	0.2	0.1
OnePlus	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.3	0.3	0.6
Tecno	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.5	0.4
Turkcell	0.0	0.0	0.0	0.1	0.1	0.3	0.2	0.2	0.1	0.1
Wiko	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.3	0.3	0.1
Coolpad	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.3	0.2	0.1
Lyf	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.4	0.2	0.1
Infinix	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.3
Spice	0.0	0.0	0.0	0.0	0.3	0.2	0.1	0.0	0.0	0.0
Casper	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.1	0.1
AIS	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.3	0.1	0.0
Vestel	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.1	0.2
Hisense	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1
Itel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.2
bq	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1
Meizu	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1
QMobile	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.1
LeEco	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.1
Panasonic	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.1
TRUE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0
Acer	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.0	0.0
Kyocera	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0
Reliance Digital	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.1	0.0
Pantech	0.0	0.0	0.1	0.3	0.1	0.0	0.0	0.0	0.0	0.0
Intex	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0
Nintendo	0.2	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
InFocus	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0
Smartfren	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0
Xolo	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1
Archos	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HP	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0
Blu	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0
i-Mobile	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Avea	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
dtac	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Karbonn	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
Yu	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Condor	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0

(*continued*)

Vendor	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Symphony	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1
Lanix	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T-Mobile	0.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Sharp	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Palm	0.6	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Other	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.4	0.4	0.5

Mobile Vendor Market Share Worldwide March 2010–September 2019 [125].

2. The digital forensics environment

2.1. Changing role of forensic science

As has been understood and referenced in the previous IIFSMS digital evidence review papers, criminals are early adopters of technology. This was reinforced at the Australasian Forensic Science Summit [6] noted that criminals, acting in an ethically unconstrained environment, use new technological capabilities to redesign crime. Cybercrime will be facilitated further by encryption, alternative banking platforms and virtual currencies, while the Internet of Things will provide for new criminal opportunities as the attack surface increases. Robotics will enable the conduct of person-less crime from remote locations. Further, technology enabled globalisation has allowed for the willingness and legal capacity of multi-national organisations to oppose sovereign states seeking to apply the laws of their jurisdiction.

In parallel to the rapidly changing environment, the timeframes for dealing with crimes is shortened while the complexity of investigations has increased. Terrorism is an example where perpetrators were networked in communities and dedicated to increasing radicalisation and attack planning, whereas now the threat from previously unrecorded individuals using less sophisticated means has increased. The change has been driven by the skillful use of internet and social media by entities and groups from outside the jurisdiction. Higher degrees of cooperation between multi-agency and multidisciplinary teams of investigators and specialists will be required to conduct investigations. Further, specialist capabilities will be brought earlier into the investigation process to advise on the best approaches to solve specific issues. This necessarily provides the specialist with additional contextual information for better problem definition, and will require conceptual and methodological flexibility.

Harm reduction, prevention and disruption strategies are becoming the primary objectives in many serious crime categories. If policing is moving in this direction, it is incumbent on forensic science to also move in this direction and, therefore, to have a role in the intelligence process. This is not to replace of the usual and traditional 'after the fact' role, but it is in addition to that role. The dual role can be achieved by organising, aggregating and analysing existing data and using an approach of integration, collaboration, flexibility, and responsiveness.

It is noted that this conflicts with the issue of contextual bias discussed elsewhere in this review. Further, forensic science will necessarily become a blend of practice and theory which will require partnership between academia and forensic service agencies, with a pipeline from academia to forensic science, with specific mention of digital forensics. The partnership encompasses skills and qualifications, but also research and innovation, commercialisation and entrepreneurship.

2.2. Digital forensics strategy and process

The discipline of digital forensics is under increasing pressure to conduct forensic examinations in a more focused manner. Decision makers seek timely responses to questions regarding investigations while the volume of data continues to grow. Further, the issue of cognitive bias (addressed elsewhere in this review) has influenced the suggestion that forensic analysis is restricted to task or contextually relevant information. There is an increasing demand to link similar or related activities using distinctive digital traces, particularly for the purpose of international intelligence. Most proposed methods for speeding up digital evidence examination are based on the assumption that relevant information will be found in similar locations where it has been found in other cases. Consequently, evidence stored in previously unknown or new locations will be ignored, which disregards well known two features of the field: 1) new technology is regularly appearing in the market; and, 2) criminal behaviour constantly evolves and criminals are earlier adopters of technology.

The Task Group that developed *A Framework for Harmonizing Forensic Science Practices and Digital/Multimedia Evidence* identifies the core forensic processes that apply to all forensic disciplines including digital and multimedia evidence [2]. The core processes are:

- Authentication is the decision process that attempts to establish sufficient confidence in the truth of a claim. It is also used in the identification, classification, reconstruction, and evaluation phases to support the establishment of confidence
- Identification is the decision process that attempts to establish sufficient confidence that some identity-related information describes a specific entity in a given context, at a certain time [7]. Identification is not only applied to human beings, but also animate or inanimate entities, whether they be physical or virtual. It is also used in the authentication, classification, and evaluation phases
- Classification is the process of developing taxonomies of traces and ascribing a trace on the basis characteristics that are common among traces of the same class
- Reconstruction is the process of organising traces to disclose the most likely operational conditions or capabilities, patterns in time, and linkages between entities. It can be a sub-process within authentication, identification, classification, and evaluation, and
- Evaluation produces a value that can be fed into a decision process. It precedes every decision in the forensic lifecycle.

Further, the Task Group articulated the activities applied within forensic science including in digital and multimedia evidence [2]:

- Survey, the act of searching, founding, detecting, and recognizing traces
- Preservation of forensic traces to prevent alteration
- Examination to observe traces and their characteristics, recover information and content
- Documentation to record traces with associated contexts characteristics, forensic activities, and provenance information
- Analysis to obtain more information about their characteristics and make the results available for integration, classification, reconstruction, and evaluation or interpretation
- Integration which combines the results of multiple analysis processes to obtain a more comprehensive understanding of the traces, and

- Interpretation which explains the meaning of forensic findings in order to reach decisions.

Using the Data Reduction by Selective Imaging process, the storage demand in test cases was reduced from 339.9 GB to 207.6 MB. This has the potential to significantly improve the efficiency and timeliness of forensic analysis with the flow on potential to improve investigational outcomes [8].

Motivated by the unmet demand for trained digital evidence specialists through the world, Hitchcock et al. [9] trialed and evaluated a model in support of conducting triage in the field by non-digital evidence specialists. By training frontline personnel, ie. crime scenes investigators, in the field triage process, the time-consuming need for specialists to attend crime scenes is reduced. The frontline personnel receive basic training in digital forensic analysis, especially to ensure the integrity of the digital evidence. The enhanced capability reflects the past extension of crime scenes personnel to lift fingerprints from a crime scene, but is not expected to conduct the fingerprint comparison.

The objectives of the model were to improve investigational efficiency and to reduce the backlog of cases. Investigational efficiency significantly improved with investigators receiving actionable information in a timely manner which led to faster justice system outcomes. There remains the question of 'sufficiency of examination' for which a more in-depth forensic examination is required [9].

A multidisciplinary digital forensic investigation approach for mobile smart devices has been proposed by Lutui (2016). It is noted that data on mobile devices is easy to modify, copy and difficult to acquire. Therefore, extra precautions should be taken, and standard procedures and best practices should be carefully followed. The proposed model builds on previous work [10] that comprises three subfields in the digital forensics domain: smart device forensics, network forensics, and cloud forensics, with each subfield differing in scope, characteristics, challenges etc. Each subfield, therefore, has specific requirements to which additional attention must be provided. On evaluation of the model, Lutui (2016) found that the model's phases for network forensics were the same as for cloud forensics, yet the two types of forensics are completely different with different investigation environments and requirements. Network forensics requires additional attention to the identification phase, and that a logical acquisition of data is recommended before devices are disconnected from the network. Compare this with a cloud environment where three broad service models are employed – software as a service, platform as a service, and infrastructure as a service. In addition, the multi tenant later of cloud computing needs be considered in any forensic process together with the approach of acquiring the data rather than collecting potential evidence. In the cloud environment, investigations should implement acquisition. The objective of the model is to guide the forensic [investigator] through the execution of a digital investigation that is compliant with the applicable laws, up to date, and efficient in addressing the information technology (Lutui, 2016).

Stelly and Roussev [11] state their concern over the near-blind trust that practitioners place in commercial systems which does not allow for verification of results, and that the investigator is increasingly becoming a tool operator that is more detached form the methods used to process the evidence. They have developed a standard query interface, or domain specific language, referred to as *nugget*, that enables domain experts to use a formal specification for the computation that needs to be performed for the digital evidence process. Their architecture relieves the need to specify the computation, mapping it to the tools, and scheduling it with the resources that are available. Practitioners currently have two options comprising a selection of point and click tools; and construct an analytical strategy using a range of open source tools. *Nugget* allows practitioners to specify queries that demand responsive solutions. It is compared, in function, to SQL which is the domain specific language for the database domain.

A widely supported domain specific language, such as *nugget*, allows for unified means to specify, log, and systematically text forensic functions and integrated implementations. The authors contend that it also addresses the need to independently test the validity of tools by third party testing [11]. They note the ease with which digital forensics tools can be containerised and integrated into nugget. As it is specification driven, it allows the integration of a group of tools to accomplish a given task. It provides a potential opening to apply big data techniques to allow for the increasing volumes of data that are being encountered, and to accommodate artificial intelligence methods. The containerised tools are directed using remote procedure calls which provide for extensibility and for scaling. In summary, the domain specific language would greatly facilitate tool testing and validation, cross-tool integration, a common language for education and training, and the use of big data and artificial intelligence methods.

Consistent with the concern over the well documented pressures facing digital evidence practitioners and the organisations for whom they work, automation is one approach that might support case processing, or robotic process automation. Robotic process automation is the automation of service tasks that were previously performed by humans, technology that is based not he concept of artificial intelligence. The robot performs the instructions directed by the developer by communicating with the systems, then triggers the response to produce results. Robotic process automation is a higher level automation in which a software based task, that can be procedurally replicated, can perform the same sequence of software interactions required to complete the task. The robotic process automation core function is via element identification with an interface, and only interacts with the presentation layer of software, ie that which is visible to humans [12].

Robotic process automation has a number of benefits including lower cost and less time to implement, and no disruption to underlying systems as it operates at the human level, on top of existing software solutions, rather than integrated with those same solutions. Beneficial results of robotic process automation include: 1) accuracy as it is less prone to procedural errors; 2) improved employee morale; 3) productivity as the robot process cycle is much faster than manual processes; 4) reliability and consistency as robots can only carry pre-programmed commands and, therefore perform the same way every time; 5) non-invasive to underlying IT systems; 6) compliance with regulations and policies based on the programming of the robot; 7) low technical barrier as no programming knowledge is required to configure a software robot [12].

The authors are clear about which tasks within digital forensic examinations are suitable for the application of automated processes. Broadly, the objective investigative tasks, essentially the pre-processing tasks, are suitable for automation; whereas, subjective investigative tasks comprising analysis and interpretation of results are not suitable as these tasks are dynamic and instinctive, and are influenced by specific case circumstances [12].

An approach developed by Gladyshev and James [13] uses probabilistic sampling and priorisation in the context of file carving, an automated process for reducing the amount of data to be subjected to analysis. The approach will speed up file carving for forensics triage by processing data blocks that are more likely to contain relevant data when investigators are looking for files of a particular kind. The authors evaluate the model using: 1) decision theory, a branch of mathematics that studies decision making as a

choice between several alternative actions; 2), numeric simulation, and file carving experiments. Decision theoretic analysis allows a file carver to consider the most likely locations of relevant data based on what is known about the distribution of data on the disk. Carving times are reduced by skipping the areas on the disk that are unlikely to contain relevant data. The technique is most useful when applied in a triage situation [13].

Casey et al. [14,15] discuss the need for solutions in digital forensics that balance the multiple interests of those who have requirements for this capability. Digital forensics is used in many contexts which can be broadly described as the courtroom, the boardroom, and the war-room. Digital forensics is becoming inaccessible due to the increasing expense and complexity, which must also be balanced with privacy concerns. The editorial team make a series of recommendations:

1. Closer collaboration between industry and government
2. Centralisation of research, development, and administration of capabilities
3. Streamlined mechanisms for the exchange of digital investigation information, and
4. Improved availability of digital investigation knowledge and advanced capabilities [14,15].

The Microsoft operating system stores configuration data in the Registry which is used to run the computer. Analysis of the Registry yields very useful forensic evidence in the event of the system being attacked. Patil and Meshram [16] proposed a Registry evidence collection and analysis methodology called RegForensicTool. The tool overcomes the limitations of the pre-existing tools which they regard as time consuming to use. The RegForensicTool is portable; standalone; easy to use; has inter process communication; and presents the forensically important activity including autorun program, recent accessed documents/programs, network accessed or connected, devices connected, applications installed, login activity, and malware activity; facility for drag and drop of evidence for a user activity extracted from the Registry key; backup of individual Registry hives and entire Registry; running processes and services; and, timestamp generation [16].

Meshram and Patil [17] developed a tool for the specific analysis of free space or slack space of hard drives to obtain sensitive or malicious data that may have been stored there. Data can be hidden in an easily created Alternative Data Stream. In addition, they describe a new approach to recover deleted data. The tool performs two functions – file extraction and file analysis. A disk image is created in the file extraction phase, and it will obtain attributes that are used for the recovery of deleted files. The detailed analysis will find the evidence from deleted files, alternative data stream, and free space.

2.3. Imaging

Logical imaging is being increasingly used in digital forensics practice due to the changing computing environment. These changes include iOS devices and Mac computers where physical imaging has become increasingly impractical due to lack of access to decryption; increasing use of Software as a Service cloud-based solutions where physical imaging is not feasible; and, distribution of software directly to endpoints. Further, as logical imaging is relatively quick when compared to physical imaging, the emphasis on triage requiring rapid identification and preservation of files leads to time pressures that lessen the appeal of physical imaging [18].

2.4. Government led initiatives

The United Kingdom's Minister for Policing and Fire Services requested a collaborative review of the quality and sustainability of forensic science service provision to be conducted by the National Police Chiefs' Council, the Association of Police and Crime Commissioners, and the Home Office [19]. Concerns over quality, financial sustainability, and "… policing's failure to prioritise accreditation of its own services …" were motivation for the review. In 2019, an implementation plan was issued [20].

The United Kingdom's House of Lords Science and Technology Select Committee conducted an inquiry into the provision of forensic science services in the United Kingdom. The findings of the inquiry are covered elsewhere in this review.

2.5. Digital forensics organisational capability

With the growth and increased complexity of data and the raised recognition of its importance to an organisation as its intellectual property, the need to retain the privacy of employee and customer data, and additional compliance requirements for record keeping, the issue of organisational readiness for digital forensics is receiving additional attention. It is now recognised that organisational readiness is an active process that requires planning and expertise in execution to, for example, respond to security incidents [21].

The authors note that most organisations have data retention and disposition policies that provide a schedule for how data should be retained and how it should be disposed of. The data retention and disposal policies will be subject to the laws of the jurisdiction in which the organisation is operating. In addition, organisations should develop a digital forensics response plan in preparation for when an incident might occur and require a digital forensics response. The plan should include evidence generators that can capture the evidence of unwanted activities and correctly preserved. Further, a forensic readiness policy that details the immediate procedures so that there is a systematic, standardized and legal basis for the admissibility of digital evidence. The policy should enable the gathering of evidence relevant to the investigation without disrupting core business, conducted at a cost that is proportional to the incident and its ramifications, and the evidence has a positive impact of any legal action. Other requirements include financial support for the recruitment and ongoing training of appropriate skilled staff, and technological requirements.

Any digital forensics response investigation must comply with the data retention and disposition policies of the jurisdiction and should be consistent with the organisation's data and information governance requirements. Other factors to be considered include the impact of litigation hold requirements, releasing and disposing of court-ordered data, challenges to retention and disposal, costs associated with disposition and storage, mitigating and responding to disasters and emergencies, and dealing with organisational disciplinary issues.

As digital forensics is related to law and to technology, investigators are expected to do more than just follow known techniques. The multitude of different crimes that involve digital evidence, networks, and complexity of information and communications technology ASF to the complexity. Further, the legal processes vary from one jurisdiction to the next. This means that organisations need to adopt rigorous and flexible processes. Proper forensic examination is not just within the provenance of law enforcement agencies, but it also a responsibility for defence attorneys.

In their survey organisational preparedness to mitigate and investigate cyber threats, Ab Rahman et al. [22] found that the

needs of incident handling and digital forensics overlap. Currency in forensic awareness and capability to deal with emerging technology apps is a constant challenge as is the release of new software and computing formats. None of the organisations surveyed indicated any awareness of forensic readiness which will impact when the artifacts of any incident are being sought during an investigation.

3. Practitioners

3.1. Education and training

The Task Group that developed the framework for harmonizing forensic science practices and digital/multimedia evidence describes the foundational sciences for the various sub-disciplines as biology, physics, and mathematics, but also include computer science, computer engineering, image science, video and television engineering, acoustics, linguistics, anthropology, statistics, and data science [2].

The role of the digital forensics practitioner requires several cross disciplinary facets including an understanding pf practice, procedure, technology and law, underpinned by ethics. Due to current and predicted shortage of suitable candidates for information security jobs, the training in cyber forensics has been the subject of much attention by the governments of several countries, including the United States (NSA – National Security Agency Centre of Academic Excellence in Cyber Defense Education), and the United Kingdom (GCHQ – Government Communications Headquarters National Cyber Security Centre). These initiatives are supported by additional work such as the United States National Institute of Standards and Technology Cybersecurity Workforce Framework to ensure the consistent use of terminology [23].

The digital forensic data sets are available for training purposes including: The National Institute of Standards and Technology has a library of Computer Forensics Reference Data Sets for training purposes that cover a range of scenarios including hacking, data leakage, registry forensics, drone images, Russian tea room, memory images, mobile device images and more [24]; and, Digital Corpora that cover cell phone dumps, disk images, files, network packet dumps, and scenarios [25].

Despite the substantial teaching resources, the generation of real digital evidence is by the suspect. Carthy et al. [26] found that encouraging senior students were able to enrich their learning by generating a trail of evidence enriched their learning by providing them with a greater awareness of how evidence is formed, file provenance, and root cause analysis. Senior students, who are generating the evidence, needed to have a good understanding of best practice and procedures in the discipline. By constructing a situation where schools in two different countries (Norway and the United States) where creating and analysing digital forensic data, a richer cultural experience was had.

The American Academy of Forensic Sciences, Forensic Science Education Programs Accreditation Commission revised its accreditation standards for 2018 [27] and 2019 [28]. The 2019 version is amended to include “survey of forensic science” as a general curriculum requirement, and the option to include business statistics within the mathematics component. The 2019 undergraduate program standard still retains a requirement to complete studies in physics, chemistry and biology, but has removed the requirement for a minimum of six semester hours “ … that provide breadth in traditional forensic sciences (eg. DNA, latent prints, trace chemistry, microscopy, crime scene reconstructions, etc …”. For the postgraduate courses, the 2019 standard has removed the requirement for studies in forensic biology, but still retains pattern evidence. Additional clarity is provided for the requirements of the research project.

Verma and Bansal [29] propose taking a knowledge management approach to digital forensics education and training. In supporting this proposal, they assert that current digital forensics tools are obsolete due to the diversity and the volume of data. They describe knowledge management as the process of capturing, storing, retrieving, managing, and representing knowledge and it provides a competitive business advantage. The authors also describe several knowledge management techniques and previous attempts to map the usefulness of knowledge management techniques to digital forensics.

3.2. Ontology

The field of digital forensics comprises and encounters many technical and non-technical terminologies that can be difficult to comprehend. Several new terminologies might be encountered during the course of a single investigation which will take considerable time to comprehend and understand their role in the incident subject to the investigation. Ontologies refer to a shared understanding of a domain of interest and used as a unifying framework in solving problems. Ontologies are used for representing and reasoning about domain knowledge. Karie and Kebande [30] propose that existing tools should incorporate new approaches to assist in resolving or clarifying the meaning of new terminologies used during the investigation process. Ontologies will generate a common definition, knowledge and understanding of digital forensics domain terminologies.

The generation of an ontology comprises four main steps: 1) digital forensics terminology database; 2) develop terminology semantic annotations; 3) reasoning engine; and 4) terminology semantic repository. The critical steps focus on the meaning of digital forensic terminologies during a digital forensic investigation [30].

Building ontologies for digital forensic terminologies which will have the added benefit of providing a form of discipline knowledge, a gap in the field that has been noted by other authors in this review. It will also assist law enforcement agencies in discussing digital forensics investigations; academic institutions when teaching students; and tool developers as they develop their products in resolving the meanings of terminologies used during an investigation.

Later work by Casey et al. [124] also identified the need to harmonise how information relevant to cyber-investigations is represented and exchanged. They note that the issue is especially pressing at this time as the data sources are numerous and are derived from various tools. The proposed solution is an open community-developed specification language referred to as Cyber-investigation Analysis Standard Expression (CASE). CASE builds upon the Unified Cyber Ontology which provides a format for representing information in all cyber domains. CASE can be used in any context in which digital evidence applies including criminal, corporate and intelligence domains. It enables the fusion od information from different organisations, data sources, and forensic tools.

CASE provides a structure for capturing information for representation, sharing, interoperability, and analysis in cyber-investigations. It provides a framework for documenting how cyber-information was handled, transferred, processed, analysed, and interpreted. Without standardized approach, investigators in different jurisdictions may be unaware that they are investigating crimes committed by the same perpetrator [124].

4. Quality

4.1. Quality assurance

Quality assurance and accreditation issues are again prominent in the past three years, especially in the United Kingdom. In the report of the United Kingdom Forensic Science Regulator [32]; the regulator expressed the priority intent to work with all National Police Chiefs' Council relevant portfolios in order to comply with requirements and appropriate quality standards. Importantly, the Regulator highlighted the importance for the police to no longer procure digital forensic services from organisations that have not met compliance with accreditation standards.

The regulator is overseeing the development of several standards, including:

- Cell site analysis and communications data, but noted that there is limited published peer-reviewed research in this area [32]. The regulator further notes that areas being addressed include: the difference between technical interpretation and opinion evidence in cell site analysis, assessment of uncertainty in call data records, assessment of uncertainties of methods used within cell site analysis, and interpretation models for providing opinion in cell site analysis;
- Network forensics which covers the screening and extraction of data from a business's networked computer system;
- Open source intelligence (Internet intelligence and investigations) which includes core internet use, overt internet intelligence and investigations, and authorised covert internet intelligence and investigations.

The progress by policing organisations in meeting compliance requirements slowed in 2018 due to competing resource pressures [32]. It was also noted that commercial viability needs to be considered when procuring services from accredited providers. By November 2017, within law enforcement, 12 legal entities (of a total of 46) were accredited for imaging storage devices, three for data extraction, six for mobile phones, and two CCTV. Only four of 20–30 commercial providers to the criminal justice system have gained accreditation and smaller providers have made no progress [33]. This led to expressed concerns by smaller providers of insufficient incentive to pursue accreditation as policing continues to award contracts to non-accredited providers, and the " … [perceived] lack of commitment to quality standards in policing." Quality concerns were a motivation for a review of the provision of forensic services [19]. The implementation plan included building capacity into the system so that all providers o f digital forensic services can be accredited [20].

The issue of accreditation remains contentious. While most jurisdictions support accreditation to ISO 17025, or at least ISO 17020, for almost all of the forensic sciences, some resistance remains for the accreditation of digital evidence providers as evidenced in the United Kingdom's House of Lords [34] inquiry into forensic science. While broad support was expressed by witnesses for accreditation of digital evidence to ISO 17025, some witnesses proposed support for other standards such as ISO 27037, 27041, 27042, 27044, and 27050, albeit in the acknowledgement that the other standards are guidelines rather than expected standards of practice. There is substantial commentary on this subject with much of it ill informed, and therefore not referenced in this paper. The strength of ISO 17025 lies not just in the technical aspects of the standard, but in the requirements for the accredited organisation to demonstrate management competence, validation of the tools employed, competence of staff, ability to anticipate, detect and remediate errors, verification of results, etc. These risk that these issues impose are described elsewhere in this review [35].

Sunde and Dror [36] note the lack of formalised quality assurance procedures, such as verification or peer review, within digital forensics. Although peer review is mentioned in Scientific Working Group on Digital Evidence [37], no description as to how this should be undertaken is provided. In order to assure the elimination of bias, the peer review should be conducted independently.

A comparison for quality assurance and scrutiny between the forensic disciplines of DNA, latent fingerprints and digital evidence is made. It is noted that, as a relative newcomer, digital evidence has remained relatively unchallenged for high profile reviews of its capacity to provide reliable evidence but it has been the subject of criticism for the failure to promptly disclose evidence [38]. The authors note that "… as soon as human interaction is introduced into a process, there is the possibility of human-related error … therefore actions to prevent human error should occur". They clarify by noting that there are few formalised and enforceable peer-reviewed and quality assurance procedures enforced in digital evidence, and that implementation of a quality management system is dependent on budget or a box to be checked in order to successfully tender for work rather than establishing a framework for improving and maintaining high quality work.

Page et al. (2019) describe five hierarchical levels of review that can be undertaken in digital evidence, in descending order of resource intensity:

- Blind re-examination of the entire case,
- Verification review of the examiner's findings,
- Conceptual peer review that ensures the correct interpretation of the work, but makes assumptions that certain steps were completed correctly,
- Sense review which is just a check that it makes sense, but involves no checking of evidence, and
- Proof check which is a light administrative review.

Due to budget and time constraints, the most effective forms of review, blind re-examination and verification review, are unlikely to be conducted. Further, in smaller organisations, and due to the complexity and wide variety of types of digital evidence (devices, operating systems, applications etc), it is unlikely that a sufficient number of expert staff will be available who can adequately review the work. Further, it is speculated that fact checking and verification may be viewed as a job for the defence [38].

Page et al. [38] suggest some alternative processes to meet the requirements of accreditation and, therefore, attain better practice and to meet the intent of the accreditation standards. These include dual investigator as an extension of dual tooling whereby examiners divide a given case and collaborate in the examination, thus providing a culture of ongoing peer review; and random sampling of cases for intensive review.

Sommer [39] describes various possible approaches to assuring the quality of digital forensics for court noting that any scheme chosen needs to be viable in implementation and value for money. The approaches can be categorised into three groups: 1) individual accreditation, 2) laboratory accreditation, and 3) court procedures. There is the risk of multiple rival accrediting organisations. The argument is made that ISO 17025, which is regarded as the mainstay of accreditation in forensic science in adversarial justice jurisdictions and, therefore, digital forensics, is not the best suited for the assurance of digital forensics. This is due to certain characteristics where digital forensics differs from other evidence types,

including, but not limited to, the fast pace of development in devices, operating systems, applications etc; one-off processes; cost of compliance with accreditation requirements; and, the lack of a 'laboratory setting' for the conduct of digital forensic examinations.[1]

In October 2017, the United Kingdom's Forensic Technology Regulator [40] published the fourth version of the "Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System". The codes provide more detail on standards pertaining to occasional experts and infrequently used methods which occasionally feature in aspects of digital evidence when encountering an unusual or recently emerged technical challenge. The codes reinforce the concept that "… same level of confidence shall be required whether the method is to be used routinely or infrequently" which includes validation of methods and demonstrated competence of the staff who perform those methods. In addition, experts who testify infrequently or who are from overseas, are to fulfill certain obligations and admissibility requirements including being bound by the Code of Conduct. Specifically for digital evidence, the codes set a schedule for organisations to meet accreditation requirements.

The Task Group that developed *A Framework for Harmonizing forensic Science Practices and Digital/Multimedia Evidence* notes the importance of considering error mitigation in digital/multimedia evidence [2,41]. They note that even when operational techniques are working perfectly, there is the potential for cognitive bias, observer error, and other non-technical sources of error, some of which are discussed in other sections of this paper. Lack of competence can lead to overlooked and misinterpreted traces, as can organisational management that prioritizes speed over quality.

Horsman [42] notes the differences in opinion concerning quality assurance to ISO 17025, tool testing and validation, and their place as, effectively, a mandatory requirement in the practice of digital forensics. The fundamental tenet of ISO 17025 is as a standard to "ensure organisational competence and maintain public confidence that standards in digital forensics are maintained" [42]; page 164). It is incumbent on organisations to demonstrate the reliability of the methods they use.

4.2. Human factors

Research on miscarriages of justice has highlighted the issue of human error in forensic science with particular focus on cognitive bias in several forensic disciplines. In recent years, digital forensics has increasing taken a more scientifically sound analysis and interpretation of evidence with a growing focus on quality management, error mitigation, tool testing and verification methodologies. A number of peak organisations recognise it as a discipline of forensic science and, similarly to other disciplines, is subject to uncertainties, vulnerabilities, limitations and the potential for error.

Sunde and Dror [36] note the movement from the perception of tools and technology as the main instruments in the digital forensics process to the importance of the human role in this endeavour. In consideration of human factors and human error, the cognitive factors and their impact on decision making must also be considered. As the core processes of digital forensics are increasingly understood to be aligned to those of other disciplines, the other disciplines are an appropriate starting point. The potential for human error, that has led to miscarriages of justice or overturned convictions, has been well established in other disciplines has now been found to impact digital evidence, specifically evidence concerning CCTV recordings, SIM-cards, DVD content, and web content.

Dror [43] and Sunde [44,45] describes the taxonomy of sources of bias that may affect forensic decisions within the digital forensics process. The cognitive biases arise from the way in which the brain processes information. They are not intentional nor conscious and they are Burke the to emotions such as confidence, frustration, sorrow, and anger, personal responsibility and concern about future consequences. Much of the work of digital forensics practitioners is likely to include child sexual exploitation which will have graphic images, video and online communication which can greatly impact the emotional state of the practitioner. The taxonomy comprises seven levels: 1) the cognitive architecture and the brain; 2) training and motivation; 3) organisational factors; 4) base rate expectations; 5) irrelevant case information; 6) reference materials; and, 7) case evidence. Sunde and Dror [36] explore each level of the taxonomy in terms of brain function and normal information processing, and provide suggestions of ways in which the impacts of cognitive bias can be mitigated.

The authors note the risks to objectivity arising from situations where an organisation's digital forensics capability is integrated into the investigational teams, and base rate expectations due to prior experience [36]. Base rate expectations can lead to a bias or away from the investigational hypothesis, for example, previous experience of an inability to extract evidence from a particular type of device will possibly lessen the priority of analysis to that device when encountered in future investigations. Some of the mitigating countermeasures that can be used in digital forensics can include: 1) training of digital forensics practitioners in cognitive psychology that is practical and scenario-based that will enable practitioners to understand and experience how bias can occur; 2) testing and eliminating multiple, competing hypotheses in an investigation of the same data and information; and, 3) peer review processes that involve blind verification, which should also be applied to negative results as well as positive results.

Due to the volume of material encountered in a digital forensic investigation, searches will necessarily be customized to deal with the circumstances of that particular case. It is, therefore, important that the examiner records and reports what was searched for and the contextual information that was provided to the examiner prior to and during the examination process [44,45].

The fallibility of human reasoning provides a strong incentive for following a scientific approach when analyzing digital and multimedia evidence in a forensic context. Scientific practices cannot eliminate error, but the risks of error can be mitigated. The scientific method employs scientific reasoning, which can be described as abductive, deductive, and inductive reasoning which is, sometimes, referred to as the hypothetico-deductive model. "Abductive reasoning eliminates implausible explanations and retains the most plausible explanation for (limited) available facts and traces, drawing analogies from past experience" [2]; page 3). Deductive reasoning tests the most plausible explanation against the observable traces with a focus on contradictory facts. If contradictory traces are found, the most plausible explanation must be revised. Inductive reasoning can lead to knowledge specific to a set of circumstances and, therefore, providing trustworthy decision making. Inductive reasoning can also lead to generalized theory based on the observations of a number of circumstances, which provides ne knowledge to forensic science [2].

Scientific reasoning is applied at different stages of the justice process. During the investigative phase, practitioners develop scenarios that explain the evidence, search for contradictory and predicted facts, and interpret available information to arrive at a decision. As testimony is being prepared, practitioners consider the

[1] Note – it is not the purpose of this review paper to deconflict or contest the ideas presented in the published material that is drawn upon for the content of this review.

claims of the various parties to the litigation against the evidence of the traces, including looking for alternative explanations. The Task Group note that scientific reasoning leads to probabilistic conclusions, not absolutism. It provides a likely outcome given the available information, but that information might be limited, subject to cognitive bias, and subject to influence by external factors such as cognitive bias, fatigue among others [2].

A fundamental principle of forensic science is expert opinion should not be expressed as fact. Moreover, focusing on a single hypothesis could be an indication of bias or a failure to consider alternative possibilities. Casey [26] described a case in which the judge was concerned about the prosecution failing the meet the burden of demonstrating that the underlying science of geolocation services has gained general acceptance in the relevant scientific community. Similar challenges have also been encountered in cases involving cell site analysis. Consequently, the UK Forensic Science Regulator has included additional clauses in the Codes of Conduct for Digital Forensics – Cell site analysis that require practitioners to consider additional hypotheses; the terminology of the reports shall imply no bias so phrases such as 'in the vicinity of', and 'consistent with' can only be used with caveats. Further, limitations are placed on the use of cell site analysis as evidence used to form any hypotheses or investigative leads [32].

The previous paragraph highlights the growing expectation that digital traces are to be treated in a similar manner to that of forensic science more broadly, that is, evaluating and expressing the relative probabilities of two mutually exclusive hypotheses. In support of this approach, OSAC published "A Framework for Harmonizing Forensic Science Practices and Digital/Multimedia Evidence" to define the core forensic concepts and processes in the context of digital and multimedia evidence [2]. To put it another way, the forensic practitioner's responsibility is to focus on the digital traces, not to prove a disprove a specific claim. Subjectivity is involved in the evaluation of forensic findings, with the judge or jury responsible for considering the evidence with all other information to arrive at a verdict.

In the case of digital evidence, for example, a case of the recovery of a deleted file, forensic practitioners must consider the whether the deleted file was recovered correctly, and are the actual, original contents of the deleted file. Importantly, the deleted file recovery operations usually involve an estimation of what data was allocated to the deleted file. It is necessary for forensic practitioners to consider alternative hypotheses. Increasingly in the United States and in Europe, forensic practitioners are expected to express the probability of the evidence given one claim versus an opposing claim.

In order to assist the finders of fact to understand the results of forensic examination, the forensic expert should not advocate for a specific outcome. Bias can influence the presentation of digital evidence, especially when the stakes are high. It can result in an inappropriate conveyance to the strength of hypothesis that favors the client in an adversarial situation. Casey [26] concludes that steps must be taken to prevent forensic practitioners from acting as advocates, which can be achieved by the insistence that the practitioner's evaluation of the evidence and expression of the results is in terms of the relative probabilities of evidence given at least two alternative claims.

Collie [46] describes the impact on the quality of digital forensics in data extraction, analysis and interpretation, resulting from the pressure to reduce costs. Untrained police officers are downloading data from mobile phones and presenting very superficial interpretations as evidence which can be wrong. As the officer is often untrained, the data can be immediately misinterpreted, for example, automatically downloaded key words are mistaken for search terms; or, interpreted without context.

4.3. Tool validation

This section discusses the subject of tool validation. A number of contentious papers have been published during the period 2016–2019. It should not be inferred from this review that tool vendors are producing sub-standard tools but, rather, there are issues that need to be addressed concerning the validation of tools and processes that are used to examine digital traces.

As is the case with consumer and business software, it is understood that flaws exist in digital forensic software. Some flaws are of a severity that they can impact on an investigation with the effect and consequences of unreliable tools leading to the possibility of inaccurate evidence that, in turn, impacting the client and the practitioner [47]. In the United Kingdom, the Forensic Science Regulator requires digital forensics laboratories to obtain ISO 17025 accreditation which emphasises demonstrable development and effective implementation of adequate testing and validation methods. The regulator has developed guidelines, that embed validation into laboratory practices, by which this can be achieved.

Horseman [47] elaborates on the three error types that can be encountered in digital forensics: 1) tool error – the software misinterprets or misrepresents the data, 2) tool limitation – the confines by which the software can be expected to perform, and 3) user error – the use of software in a way for which it was not designed.

Tools errors can result from accidental errors, update errors, software rot, inadvertent and intentional bias, and flawed self test diagnostics. Detecting tools errors in digital forensics can be especially fraught as there is little opportunity for manual validation as the evidence cannot be touched or viewed. The discipline must verify and validate its tools by using the tools, therefore finding itself in an infinite loop. Consequently, the field tends to fall into an environment of recognizing certain tools as industry standards which defaults to an assumption based on wide spread by multiple practitioners [47].

Dual tooling is often used for verification and validation. This approach does not guarantee, but it does improve the chances of reliability. Tools that are used for the identification and interpretation of well documented artifacts have been subjected to long term research and scrutiny are generally accepted. However, artifacts associated with new and emerging technology are promoted as being "supported" by tool manufacturers, but the algorithm development and testing is invisible to users, so the extent of the testing for variables and reproducibility cannot not be assessed [47].

Vendors and forensic bodies advise that tools should be tested and validated by users before using it on case work, but some practitioners erroneously do not consider this to be part of the practitioner's role. Part of the practitioner's role is to engage in the court process and, therefore, adhere to the evidence admissibility and reliability governance which explicitly requires test and validation of the tools they use. It is incumbent on practitioners to know, understand and be confident in the tools that they use.

Although the distinction between each type of error is clear, categorising an error as one type or another can be more difficult. The default settings of many forensic tools are "dumbed down" to allow for a wider population of users which can lead to inadvertent misuse of the tool. If the practitioner knowledge is lacking, it can lead to misinterpretation of the evidence regarding a particular event. For example, a tool may purport to recover internet history, but what are the limitations of this recovery with variables such as browser type, version and settings; and, search engine type, version and settings among some variables that could impact on the performance of the tool. If the practitioner is unaware of any limitations, can the error be classified as a tool error, a user error, or a lack of transparency and documentation from the vendor [47].

End user license agreements set out the responsibilities and liabilities for vendors and users. In general terms, end user license agreements offer no guarantee that digital forensic software will be error free or operate without interruption, that the user assumes all risk in using the software, and that users will not disclose any results of testing or performance to any third party. Clearly, the liability lies with the practitioner to establish the reliability of the tools that they use [47].

Software updates, including bug fixes, are released from time to time along with vendors advising what the updates address. It is, therefore, reasonable to assume that a tool, when applied to a given case in certain circumstances, was operating in error. Practitioners should, with the benefit of this hindsight knowledge, should review historic cases to determine if the previous applied to those cases [47].

The restriction on publishing tool performance data negatively impacts the discipline's pursuit of reliability. As the only recourse of those who do test their tools is to report back to the vendors themselves, it prevents the timely dissemination to other users in order that they can take remedial actions. Further, testers identifying an error may be less motivated to report the error at all, and might just establish a local work around which will leave other users vulnerable to the tool error. In addition, reporting an error without reward, in the form of compensation or recognition from peers, may disincentivise testing work which would result in the testing work not being undertaken [47].

Horsman [47] provides a number of suggestions, along with their inherent challenges, to solve the issues of tool validation including a formalised error/tool limitation discovery repository, increased procedural and testing disclosures, increased functionality disclosures, test data disclosures, alerts and error handling (for example, in addition to release updates that note additional support, release updates of terminated support would also be helpful), external factors (such as practitioner competence, the prioritisation of speed over quality, and the effectiveness of organisational leadership), and implications (publication of tool errors and limitations could be exploited by those engaging in contrary conduct).

While all forensic disciplines are dependent on the tools that are used during examination to ensure valid results are produced, the level of reliance in digital forensics is greater as examiners are uanble to see what content is stored on the device without compromising data integrity [42]. If the process of interpreting digital traces is inaccurate, leading to erroneous data being presented for evaluation, the subsequent investigation could be compromised, possibly unknown to the examiner. The digital forensic practitioner often commences analysis following the acquisition and interpretation phases which are completed by the forensic software. The acquisition and interpretation phases are not manually verifiable, but are instead confirmed by signals provided by the forensic tools that are made visually accessible.

Horsman [42] is forthright in his comments regarding digital forensic tool testing, describing it as the field's "elephant in the room". The dependency on tools is acknowledged, but there remains little discussion as to whether the tools are trustworthy and how to demonstrate this. Although tool testing programs are described, he notes several significant shortcomings in the testing programs. These shortcomings include, but are not limited to: a release version of a widely used and relied upon tool was yet to be tested more than a year after its release; tests are narrowly defined and do not reflect the range of digital evidence scenarios and phenomena that are encountered in a normal digital forensic investigation; the type of image format under test is just one of multiple image formats available. Importantly, Horsman [42] notes that tool testing reached "peak academic attention" between 2007 and 2012, but the issues remain.

Critical to the discussion is the high burden of proof in criminal investigations in common law jurisdictions, that is, beyond reasonable doubt. If it cannot be guaranteed that any examination is based on a reliable representation of suspect material, then a reasonable doubt has already been introduced [42]. The inability to guarantee the required validity raises some questions: 1) why has the tool not been able to effectively acquire data; 2) what has the tool missed; and, 3) what has a tool potentially added? [37]. Horsman [42] asserts that digital forensics is a discipline that is driven by the establishment of fact, yet it is generally unable to state that the tools in use are functioning correctly or within certain limits.

The above must be considered in the context that it is impossible to test all scenarios in which a tool will be applied. Even when considering a single function of the tool, there are multiple valid outcomes with variables contained within. Further, any external factors that might affect the validity of the process need to be considered and evaluated. Further, testing and verification of tools is yet to reach the threshold of factual accuracy of their functions. This is exacerbated by the continual release of updates to existing tools and the release of new tools [42].

The practitioner survey undertaken by Horsman [42] revealed that the current state of tool testing is not yet satisfactory. He goes on to consider both centralised and federated testing approaches, noting the challenges with both approaches. It is considered that a centralised approach is unrealistic due to the cost of developing and maintaining such an organisation, but also because it would inhibit the scrutiny required to achieve a level of reliability and trustworthiness that the field requires. It does have advantages, however, in greater consistency in the testing process and protocols and greater oversight. The federated testing process, as currently implemented by the National Institute of Standards and Testing, has access to a greater number of practitioners involved in tool testing. It can potential be subject to variability in quality due to variability in oversight.

Horsman [42] concludes that the digital forensics field is under a legal and ethical obligation to improve its standards and, therefore, every opportunity for improvement must be taken. As more tool testing is undertaken, the more likely it is that tool errors will be identified and improve reliability. This will only serve to improve outcomes for those involved in the justice system and disputes. Lastly, if comprehensive validation of a tool's functionality is infeasible, then testing of those functions where the risk of error is greater in terms of frequency and severity should receive high priority and immediate attention.

4.4. Potential to compromise a write blocker

The integrity of digital evidence is of absolute importance to admissibility in court. If the data on a disk is considered to be evidence, then the whole disk should be considered to be evidence, both physically and digitally. As digital forensic tools are increasing in features such as network imaging, becoming networkable, and are being proposed as forensic cloud services, it is proposed that security testing should be integrated into the process of testing digital forensic tools. Some of the advances include the ability to remotely image a drive on a disk of interest, such as enabling the ability to browse drives that are attached to the write blocker via the Internet Small Computer System Interface (iSCSI) protocol. The iSCSI can command the SCSI to be delivered over Local Area Networks, Wide Area Networks, and the Internet. Users can be created and modified, and their settings altered, with these systems [48].

The researchers selected a popular write blocker and subjected it to a methodology comprising: 1) gaining root access; 2) constructing integrity attack scripts; and, 3) testing. They were able to

compromise the integrity of the destination drive, but were able to make it apparent to the user that there was no compromise by altering the warning message to something benign. The scenario is described in which an adversary could, relatively easily, substitute a compromised firmware update for a genuine update and convince the digital forensic practitioner to unknowingly install the compromised version. This is exacerbated by the real lack of training in cyber security and computing of many law enforcement digital forensics practitioners. Similarly, a deliberate attempt could be just as feasible and reference some examples. Hash values are the accepted authentication of a duplication, but script that the researchers constructed infers the authenticity of the generated hash value despite the alteration. The conclusion drawn by the researchers is that digital forensics practitioners should integrate security testing into the forensic tool testing process [48].

4.5. Datasets

The use of datasets can be an important aid in research, for example, in the construction of an email parser, malware analysis, or improve specific purpose algorithms. For the datasets to be useful, they must possess three features: 1) quality to ensure that results are accurate and generalizable; 2) quantity to ensure that there is sufficient data to train and validate the tools; and 3) availability for the research to be conducted and independently reproduced to ensure scientific validity. Further, funding agencies are increasingly requiring that grantees to make the results of their research available to the public (Grajeda, C., Breitinger, F. and Baggili, I., 2017). The researchers had noted from earlier work of others that: 1) many researchers produced their own datasets; 2) datasets are not released after the work has been completed; and, 3) there is a lack of labelled standardized datasets that can be used in research. These weaknesses lead to the community disadvantages of low reproducibility, comparability, and peer validated research. It is also noted that it is poor common practice to perform research and not publish the underlying dataset.

Over half the datasets found in the study were experiment generated, where researchers created specific scenarios to conduct their experiments. This was due to the lack of available real world datasets; and, datasets were created specifically to conduct experiments on new technology [49].

User generated datasets, ie real world datasets, were the second most common type of datasets. Real world datasets are crucial for developing reliable algorithms and tools. One of the inhibiting factors is copyright and privacy law which prohibit sharing. A prominent example of a real world dataset is the Enron email dataset which was posted online by the Federal Energy Regulatory Commission and later purchased by the Massachusetts Institute of Technology. Private user information and email attachments were remove to avoid violating privacy rights [49]. Some institutions collect real world information, for example, from students who have signed an agreement for researchers to capture the information. Some datasets have been generated through collaboration between law enforcement and academia; while other data is publicly available online. In addition, the National Institutes of Standards and Technology hosts collections such as the National Software reference Library, and the National Vulnerability database. Computer generated datasets are the smallest category of datasets. User generated datasets have the advantage of generated datasets is the exact knowledge of the ground truth.

Grageda et al. (2017) found 70 different datasets in their analysis of articles referring to datasets and organised them into 21 categories. The major categories are:1) Malware (computer and mobile); 2) Email; 3) File sets/collections; 4) RAM dumps; 5) Images of computer drives; 6) Images of other devices, including mobile phones, gaming systems, SIM cards, and flash drives; 7) Network traffic; and, 8) Scenarios/cases for analysis. In addition, they found 10 sources providing datasets through Google searching.

Overall, there were some gaps in the availability datasets that were summarised as: 1) a lack of variety; 2) apart from malware and network traffic datasets, no other datasets were being regularly updated; 3) lack of a single repository which has resulted in some of the most popular repositories no longer being maintained by the owners; 4) data de-identification research to remove proprietary and personal identifying information; 5) strategies to share complex data, particularly cloud data in a way that it is reproducible; and 6) publisher support for the sharing of datasets. It is noted that the US Department of Homeland Security, through its Impact Cyber Trust project, has taken some initial steps to improve the sharing and availability of forensic datasets [49].

5. Technical advances

5.1. Cloud storage forensics

Previous reviews have identified cloud computing services as an emerging issue for digital forensic examiners and investigations. The National Institute for Standards and Technology define cloud computing as "… a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources …" [50]. There are three broad categories of cloud computing services:

1. Software as a Service (SaaS): an application accesses shared infrastructure of the cloud storage provider, for example, storage as a service;
2. Platform as a Service (PaaS): user deployed applications on the cloud storage provider's infrastructure; and
3. Infrastructure as a Service (IaaS): underlying computer resources, such as the operating system or other software, are provided by the cloud storage provider.

The National Institute for Standards and Technology defines cloud computing forensic science as "… the application of scientific principles, technological practices and derived and proven methods to reconstruct past cloud computing events through identification, collection, preservation, examination, interpretation and reporting of digital evidence" [51]; page 2).

There are 65 challenges to performing forensic investigations in the cloud which are grouped as follows, although the descriptions are not comprehensive:

- Architecture – diversity, complexity, provenance, multi-tenancy, data segregation
- Data Collection – data integrity, data recovery, data location, imaging
- Analysis – correlation, reconstruction, time synchronization, logs, metadata, timelines
- Anti-forensics – obfuscation, data hiding, malware
- Incident first response – trustworthiness of cloud providers, response time, reconstruction
- Role management – data owners, identity management, users, access control
- Legal – jurisdictions, laws, service level agreements, contracts, subpoenas, international cooperation, privacy, ethics
- Standards – standard operating procedures, interoperability, testing, validation, and
- Training – forensic investigators, cloud providers, qualification, certification [51].

As of 2016, cloud forensic investigations had received little attention from researchers [52]. It is noted cloud forensic examiners are not just trying to keep up with updates to devices and software, but also to changes made to software and hardware by end users.

The traditional model of digital forensics is client-centric where the examiner works with physical evidence devices, such as storage media or mobile devices, such as smart phones. Digital forensics, therefore, was focused on the physical location of the computation and the storage of the data. The underlying assumption has been that most data is local. Gmail became the first, mass used web app. In the Software as a Service, both code and data are delivered over the network on demand. The local storage, eg hard drive, is a cache and not the data repository [53].

Roussev et al. [53] developed several tools specifically for the purpose of forensic examination of cloud storage: 1) *kumodd* which uses the service providers' API[2] to perform a complete acquisition of the drive's content; 2) *kumodocs* specifically for Google Docs to study how web apps store and work with artifacts; and 3) *kumofs* to bridge the semantic gap between cloud artifacts and legacy tools, using a filesystem interface to the cloud drive. In addition, they developed *time travel* for the ability to rewind the state of the drive as of a particular time; *(time) diff* to identify all recorded activity between two points in time; and a query interface which allows investigators to filter drive data based on the metadata provided by the cloud services.

A fundamental difference when conducting forensics in the cloud rather than client-centric analysis is that many of the required investigative functions are already present. Software development practices have changed to one where functionality can be composed from autonomous modules that communicate over APIs and distributed between clients and servers. The result is routine logging that records user input, therefore, historical information is already present and the cloud service itself can be directed to efficiently and reliably reveal it [53].

The shift from Software as a Product to Software as a Service changes the fundamental concepts of digital forensics that have been in place since its inception. The doctrine of acquiring data from physical devices does not translate well the SaaS world, and can be demonstrably incomplete and, at times, false. It is proposed that the investigative focus should be to obtain the most authoritative data source [53].

As businesses and consumers move more of their IT requirements to cloud services, forensic examiners will be increasingly called upon to examine data in cloud environments. It is noted that, while cloud might appear to be similar irrespective of the provider, there are substantial differences, particularly at the API level even if they purportedly perform similar functions. It is expected that forensic practitioners will need to be able to write case-specific solutions that can perform acquisition using APIs. Further, it is likely that, due to the rapidly increasing volume of cloud stored data and the associated logistical problems with moving/copying it to an examiner managed environment, a solution will be to forward deploy forensic tools to the cloud in order to conduct forensic analysis [53].

Mohtasebi, Dehghantanha and Choo [54] researched the forensic implications of cloud storage of three providers (SpiderOak, JustCloud, and pCloud). Users of the three cloud services can download, upload, and access their data using a web-browser and a client application, such as an app. Other functionalities that might also available, depending on the provider and the means of access (browser or app) include, but are not necessarily limited too: creation, schedule, and restoration of backups; sharing files with or without password protected links; syncing across devices; encryption of all cloud stored data; upload by other users who have account access; and, backups from other services including social media. They experiment with a Windows environment running on a virtual machine and on an iOS environment via respective apps, for each of the cloud providers under investigation, running on an iPhone 5S device.

Detailed observations and findings were made for each of the three cloud providers, including: 1) account creation; 2) cloud application program; 3) uninstalling the cloud application program; 4) downloading from the cloud using the browser; and, 5) browsing and downloading from the respective iOS app [54]. Various forensic artifacts were located when using Internet Explorer, Firefox, and Google Chrome browsers, the client application on Windows machines and iOS devices. The artifacts included email addresses, the identity and the name of the created account, and the names of the uploaded and downloaded files. User credentials could be recovered from memory. When downloaded from the cloud service, the files were identical to those that were uploaded as verified by the checksum values, however, the timestamp and the ADS were subject to change. The metadata of the doc file was not altered.

pCloud is a free online cloud storage service that users are able to store, sync, and share files in addition to backing up from other cloud services. It provides client-side encryption meaning that, as data leaves the client's system, it is encrypted. Dargahi, Dehghantanha and Conti [55] conducted a forensic study of pCloud to determine what data can be found on Windows, Ubuntu, Android, and iOS operating systems when using pCloud; what data is leaked when using Google Chrme and Internet Explorer browsers on Windows operating systems; what data of forensic interest can be discovered in live memory' and, what data can be captured in network traffic?

In the Windows based browser experiments, uploaded file names and user names could be revealed; passwords, email addresses, file names and directories were discoverable in physical memory as Internet Explorer saves pCloud credentials in the registry. Similarly, username and passwords could be found when Google Chrome was used as the browser [55].

In the Android experiments, pCloud specific folders were created and a database containing usernames, email quota, and tables related to pCloud communications could be found. Following deinstallation of the pCloud app, website information and cookies related to pCloud could be found in the memory [55].

Locating pCloud artifacts in the iOS experiments was more difficult than in the other experiments and pCloud login details could not be found. Some other useful information could be found including "session ID", "API key", the pCloud installation directory location, and uploaded file names. Following uninstallation, some deleted files could be recovered [55].

In the Ubuntu experiments, many artifacts could be found in the memory including, importantly, the username and password. In addition, the uploaded file names and file path could be obtained. After deletion fo the files from the app, the username could be recovered from memory [55].

CloudMe is a European cloud service that offers secure cloud storage, syncing of files, and client software for managing cloud data across various devices. 360Yunpan is a Chinese cloud service notable for its huge (36 terabytes) free storage space for users [56]. Experiments were conducted in Internet Explorer, Google Chrome, and Mozilla Firefox browsers in a Windows environment (client application and browser); Android client application; and in Apple iOS client application. Three file operations were conducted:

[2] API is the set of functions or procedures that allow the creation of applications that access the features or data of an operating system, application, or other service.

upload, download, and delete. Valuable forensic evidence could be found related to CloudMe and 360Yunpan storage accounts on various platforms. Digital traces included information related to user credentials, device names, and filenames. The data could be found on hard drives, live memory, internal phone memory, backup files, network traffic and more [56].

Amine Chelihi, Elutilo, Ahmed, Papadopoulos and Dehghantanha [57] create a taxonomy to aid in the investigation of cloud storage applications. Artifacts of 31 free cloud apps (of a total 240 that were considered) that appear on an Android mobile device are assessed. Of the 31 cloud apps investigated, 15 generated database files in memory. Artifacts are usually found in the internal storage for some apps and comprise pictures, documents, audio files, and web files. The authors categorised the apps into three groups based on the retrieved files: 1) no recovered data; 2) database files only generated in internal storage but without file recovery; and, 3) database files and cloud-based data recovered.

To successfully meet the challenge of a malicious cyber attack, a teamwork model comprising a group of people with diverse skills is proposed. The team members would include the: 1) cloud customer; 2) trusted third party who can assure identification and validate the integrity of service providers; 3) cloud service provider; and, 4) cloud forensics investigation team. If suspicious activity is suspected on the network, the cloud forensic investigation team can capture the digital evidence, perform analysis to provide a narrative for the events and identify the perpetrator(s), and present evidence in court if necessary [58].

Cloud exploits are major risks to cloud consumers which are difficult to mitigate. Digital forensic readiness is a proactive process that precedes incident detection. It could be achieved by deployment a botnet, acting as a distributed agent-based solution, to capture potential digital evidence. The captured information is preserved for digital forensic readiness. A botnet describes a set of scripts written to perform systematic predefined functions, which are usually associated with malicious intent, but could be used in non-malicious for the purpose of digital forensic readiness [59].

In outlining their proposed approach, Kebande and Venter [59] identify and comment on the challenges to be met. The challenges include the very use of an agent based solution in a cloud environment, a phenomenon that cloud service providers mitigate with disinfection strategies. Challenges would also be incurred through the increase in distributed computing devices, for example mobile devices, and across networks; the rapidly changing cloud environment as operational demand and on-demand solutions result in changes; trustworthiness of the chain of custody of evidence; large scale data management; and, monitoring of forensic evidence. Additional technical challenges include live evidence acquisition; virtualisation; data integrity; data volatility; anti-forensics; potential digital evidence handling; malicious activity; privacy; multitenancy; big data; and, encrypted data. On top is the afore mentioned challenges, there are operational challenges including legal authority; "[c]olossal forensic evidence analysis in the cloud; contractual and service level agreement obligations; and, standard operating procedures [59].

Imran et al. [60] identify a weakness with the cloud provenance information used in digital forensic investigations as that information itself is susceptible to tampering. They propose a scheme that ensures software security and cloud provenance security using a series of steps. The first step binds the provenance information with user data, then merging the provenance information with unstructured web data for improved security intelligence.

Virtualisation technology has become increasingly prevalent in information systems environments. Chaus et al. [61] reviewed some existing tools and their suitability for the conduct of forensic investigations in a virtual environment. In order for forensic analysts to conduct forensic investigations in virtual environments, the analyst should have a thorough understanding of the virtual environment and the storage details of log files. In addition, to meet the requirements of a forensic investigation in the virtual environment, Chaus et al. [61] created a new tool specific for this purpose.

5.2. Mobile phones

As the use of mobile phones continues to evolve, so do the forensic challenges. Emerging challenges for practitioners engaged in the examination of mobile phones include cloud applications, malware, mobile phones used as part of botnets, and SCADA systems [62]. It has been well established that no one tool or technique recovers all data, and therefore information of potential forensic interest, from a device.

The importance of mobile forensics continues to grow as it is a more affordable means of accessing the internet for a significant proportion of users. In addition, there is a proliferation of mobile malware with users less likely to be able to recognise the threats, and poor cyber hygiene as users do not seem to manage their mobile security. These factors increase the attack surface for mobile devices. This is exacerbated by the proliferation of devices, systems and apps with the need for digital forensics practitioners to adhere to the principles of sound collection of evidence [63].

Understanding the behaviour of mobile device users can be useful to digital forensic practitioners when conducting an examination of mobile devices. Petraitye et al. [63] conducted a social engineering experiment using QR codes which, while they have a useful and legitimate purpose, attackers have realised that they can also serve as a tool for redirection to fake websites and for the installation of malware onto a user's device. It was found that the secure use of mobile phone is largely influenced by cognitive impulsivity of the user. The authors propose a mobile forensics investigation guideline based on exploiting possible remnants of user activities that resulted from user impulsivity and lack of knowledge.

The digital forensic examination of local storage on mobile devices sets to achieve three objectives: 1) what information is stored; 2) where the information is stored; and, 3) how the information is stored [64]. Dynamic analysis is the most common method for data acquisition, but it has several drawbacks including: 1) it is hard to trigger all interesting programs paths, which could result in criminal behaviour remaining undetected, or content that is encoded or of unknown format can be very difficult to analyse; 2) manual reverse engineering, which is arduous and time consuming and, therefore, problematic if producing results are subject to time pressures. Consequently, several researchers have been exploring the potential for automated mobile application forensic analysis.

Ali et al. [65] developed a Mobile Forensic Metamodel for mobile forensics based on a metamodel that identifies common concepts. It simplifies the investigation process and enables investigation teams to capture and reuse specialised forensic knowledge that, in turn, supports training and knowledge management. The authors noted that previous publications discussed mobile forensic evidence as a subset of computer forensics which, therefore, did not focus on case domain information from investigations. Existing mobile forensics models are based on proprietary solutions.

The Mobile Forensic Metamodel clarifies all of the activities conducted in the course of an examination of mobile forensic evidence. Further, it creates a unified view of the domain and a consistent lexicon as the field includes multiple words and descriptions for similar processes, and single words can have multiple meanings. The metamodel defines the relationships between the concepts that form the metamodel into three groupings –

Association, Specialisation, and Aggregation. Association indicates the functional relationship between the concepts; Specialisation represents the hierarchies between concepts; and, aggregation represents relationships between concepts that comprised of other concepts. The model underwent two rounds of validation: 1) comparison to ensure that all concepts of other mobile forensic models are represented in the metamodel, and 2) frequency to determine the importance of each concept to the metamodel [65]. The metamodel provides a guideline to domain users through the various concepts and who can then find decision solutions from semantic models.

SQLite is accepted as the most popular storage engine for messaging applications on mobile devices. Therefore, digital evidence requires forensic analysis of SQLite databases and mobile forensic commercial tools are targeted to performing and presenting this function. However, little in known about the ability of tools to reliably perform this function, a fundamental principle of forensic science and a requirement for admissibility in court. Nemetz, Schmitt and Freiling, (2018) note the absence of the ability to objectively compare the relative strengths and weaknesses of different tools due to the lack of a standardized test data set. In response, they construct a publicly available test data set, a forensic corpus specific to the SQLite database management system, that aims to assist mobile phone forensic tools become more robust, reliable and trustworthy. The corpus comprises 77 databases grouped into five categories based on their peculiarities, which is then used to evaluate strengths and weaknesses of existing tools. Importantly, they note that none of the tested tools handle all of the analyses reliably.

Drawing on earlier work, Nemetz, Schmidt and Freiling [66] constructed a corpus that meets the following criteria:

- Representative of data encountered in the normal course of forensic examinations, that is, variation in settings, internal structures and contents
- Complex with intertwined information of varying sizes from 2048 bytes to 286720 bytes; and, in human languages with all SQLite encodings represented in the corpus
- Heterogeneous derived from a range of computer systems and usage patterns
- Annotated so that new algorithms can be validated against earlier versions with extensive documentation regarding the generation of each database
- Available and unrestricted without files that are restricted in any way. All data included are test data
- Distributed in open file formats with accompanying metadata
- Maintained with versioning and augmented to reflect contemporary and new information that is major and confounding feature of the digital evidence space and mobile forensics in particular. All SQL statements used to produce the entire corpus are included.

The SHA256 hashsum of all files has been included to verify the integrity of all databases and their metadata [66].

The corpus includes potential pitfalls and unusual structures and values as can be encountered in real data. Each database file of the 77 databases includes at least one peculiarity in its contents and/or internal structure. To test whether or not a tool correctly handles SQL statements, weird table names, encapsulated column definitions, and specific SQL keywords and constraints are included. These can be special characters can be included in column definitions [66].

There exist three encodings supported by the SQLite file format: UTF-8, UTF16le (little endian), and UTF-16be (big endian) and used by mobile phone manufacturers. They are, therefore, represented in the corpus with Unicode, Latin and non-Latin (Chinese) characters. The corpus is designed to test the ability of tools to handle different codings. To test the ability of a tool to handle database elements other than regular tables, some databases include different types of elements, such as virtual and temporary tables [66].

When database contents exceed the length of a page, the record is split and stored on overflow pages. To test the ability of tools to handle tree and page structures, including fragmented contents, different scenarios regarding internal tree and page layouts are included. This can include hidden data and pages that do not belong to a database element. As analysis of deleted data being an important aspect of forensic analysis, particular attention is paid to different settings that can impact deletion actions. To test the ability of a tool to correctly recover deleted contents, databases include deleted and (partially) overwritten data [66].

There are two broad categories of mobile forensic tools – those that do not recover deleted artifacts and those that do recover deleted artifacts. Those tools that do not recover deleted artifacts extract data that is logically present in the database. A tool conducting a physical extraction of data purports tp recover deleted contents provided that they are still present [66].

Nemetz et al. [66] tested the performance of six commercial and open source tools against the corpus. The tools were *Undark*, *SQLite Deleted Records Parser*, *SQLiteDoctor*, *Stellar Phoenix Repair*, *SQLite Database Recovery*, and *Forensic Browser for SQLite*. Note that none of the names will be familiar as commercial providers of mobile forensic tools. In general, none of the tools performed perfectly and there was variation on the severity of impact on the failure. Of particular concern was the performance of undeletion of entries containing numeric values by recovery tools.

The authors conclude that a forensic tool used for the analysis of SQLite: 1) should not destroy underlying evidence when converted or transferred to the output of a forensic analysis; 2) should not the elimination or (silent) omission of other evidence when erroneously analysed; and, 3) should not degrade the analysis of existing, logically present data when activating or using the data recovery function [66].

Noting that commercial mobile phone forensic vendors continue to use physical acquisition techniques, Guido et al. [67], introduced an automated differential forensic acquisition technique. The new technique and algorithm use baseline datasets and hash comparisons to limit the amount of data acquired from a mobile device. The acquired data was forensically valid bit-for-bit copies of the original and obtained in a shortened time of 7 min compared with one to 3 h by traditional methods. Notably, the final product is a physical image and is the equivalent of that obtained by a traditional method.

Saleem, Popov and Baggili [68] note the diversity of devices, the types of evidence, and the range of tools that are available. Failure to select the correct tool may lead to incomplete and/or improper extraction and, therefore, compromising the integrity of the evidence and diminishing its probative value. For example, one tool might be better for recovering text messages while another might be superior for recovering standalone files. This could result in erroneous analysis, incorrect interpretation and wrongful conclusions.

The authors propose a decision-making framework for the selection of the most suitable tool to conduct an examination of a mobile phone and other small devices for a given investigation. In constructing the framework, the authors applied theories of decision analysis: 1) probability theory, noting that, in the past, examiners, selected a tool based on previous experience and without measuring the performance of the tool; 2) utility theory based on a survey of experts in the field regarding their degree satisfaction for the relevance of all types of digital evidence; and, 3) multi-criteria

decision analysis where the cornerstones of the problem are uncertainties and utilities associated with different criteria (types of digital evidence) and alternatives (forensic tools) [68].

The framework is based on a multi-criteria decision-making process with 19 criteria evaluated and balanced against performance and relevance as the two main factors. The process is tested against seven different types of cases, namely drug trafficking, sexual assault, homicide, credit card fraud, harassment, espionage/eavesdropping, and child exploitation. The model was able to determine a clear difference in performance between two tools for a particular device. They state their intention to conduct further work with additional devices and tools to aid in the selection of the most appropriate forensic tool for a given scenario [68].

An important application of multimedia forensics is to be able to identify the device used for producing a recorded file. Jin et al. [69] present a novel method for source smartphone identification by using encoding characteristics derived from MP3 codec identification as the intrinsic fingerprint of recording devices. Through an analysis of several makes and model of smartphones, they are able to achieve high identification rates of over 97%. The smartphone to which a given recorded speech file belongs can be recognised. This is restricted to specific formats, MP3, AAC, and M4A, which are the default format of speech recording by most of the popular smartphones.

iOS and Android operating systems are prevalent having grown from an average of 16 GB in 2007 to 512 GB in 2017. MicroSD card storage has grown from 512 MB to 512 GB, and SD cards from 1 GB to 1 TB over the same time period [8].

5.2.1. Analysis of android phones

Lin et al. [64] seek to automate forensic analysis on Android devices by static analysis which can be scaled to a large number of applications without human intervention. It does not need to set up a test environment and can cover all application codes. The model uses *Fordroid* which uses an Android APK. It builds control flow and data dependency graphs are decompiling the APK; identifies the types of sensitive information written in local storage through taint analysis; then, reveals the file path where the information is stored. Finally, *Fordroid* identifies the structure of the database tables.

In testing 100 Android applications, *Fordroid* took approximately 64 h and found that approximately one third write sensitive information to local storage, and successfully located the places where sensitive information was written for 98% of paths, and identified the structure of all database tables [64]. Android applications typically have three modes to store information in local storage: 1) SharedPreferences; 2) database; or, 3) file. *Fordroid* handles all three modes differently because each mode requires different APIs and code patterns. The information revealed included: 1) category; 2) number of APKs; 3) number of components; 4) time for analysis; 5) number of paths discovered by taint analysis, including leaked formation; 6), 7) and 8) number or paths writing sensitive information to SharedPreferences, database, and file respectively; 9) and 10) number of paths *Fordroid* succeeds and fails to find sensitive information respectively; 11) number of APKs leaking sensitive information; and 12) number of APKs writing sensitive information [64] The researchers found that more than half of applications leak sensitive information, and more than one third write sensitive information to local storage. Importantly, the researchers note that information leakage is prevalent, even for applications that are not malware, and the sensitive information is more likely to be written into SharedPreferences.

Lin et al. [64], do note some limitations for their work. The proposed model may take infeasible paths into consideration; it cannot easily analyse highly obfuscated paths; some features of Java language will increase the difficulty of status analysis.

Scrivens and Lin (2018) contend that an alternative approach when conducting a digital forensics investigation may be to extract and examine particular mobile applications, rather than the whole device. This approach can apply in situations where the digital evidence pertaining to that specific application is generated and stored on the device. An automated forensic analysis is developed that can be scaled to a large number of applications as no human intervention is required. The tool was tested on 100 applications where 36 applications were found to have written information to local storage. Further, noting that Android applications typically have three storage modes – Shared Preferences, database, or file - the application was able to handle all three modes and to identify the structure of the databases where the information was stored. The authors provide the technical detail of their work, including the algorithms, so that it can be reproduced.

In an experiment conducted by Ogazi-Onyemaechi, Dehghantanha and Choo [62] to investigate the recovery of deleted data, a known dataset was loaded into a Samsung model mobile phone with a 16 GB internal memory and 1 GB RAM. The phone was then factory reset to simulate deletion of the pre-loaded data. The phone was then imaged using *AccessData FTK* and *Backtrack dd*, and the images examined using *Photo Image* Carver, *AccessData FTK*, *Foremost*, *Recover My files*, and *DiskDigger*. Examination of the subsequent logical acquisition did not contain any files. Analysis of physical images revealed less than 100% of the phone memory when acquired by multiple tools. Different images from a different acquisition tools yield differences in the volume of the evidence recovered when analysed using the same tool, and there were significant differences in the yields of various file types. It was found that the.dd images compared more favourably than *Phone image Carver AccessData FTK* under the experimental conditions. On analysis, *Foremost* recovered more file formats and a large number of data files. *Recover My File* had the best recovery function under the conditions of the experiment. It demonstrated the deepest search penetration, recovered more file formats, and recovered a high number of large sized files. It is noted that it was not the best performing tool in all measurements. Importantly, it is noted that most of the tools used recovered major file formats that other tools did not recover, reaffirming that no single forensic tool recovers all evidence on a phone.

With many phone manufacturers using Android operating systems, there are many Android applications on the market [5]. Associated with this growth, there has been an increase in security threats attributed to Android applications. An Android application is a single file in the Android Application Package format which might comprise: 1) a file containing essential data about the application which the phone must read before it can run the code; and 2) at least one Android Virtual Machine Dalvik EXecutable (DEX) file which is the application itself [70]. The authors outline four common procedures for analysing DEX files with their inherent disadvantages and, instead, present Rapid Android Parser for Investigating DEX files (RAPID). RAPID is an efficient, open source tool that is easy to use for examiners and can handle large amounts of data. It also proved to be more reliable than traditional methods and it can support dynamic analysis. For example, of 11,711 Android applications tested, 16 were unable to be analysed with existing tools, whereas RAPID was able to. The efficiency was demonstrated by a reduction in total query time (for 11,695 applications tested) from 1368 min to 88 min [70].

With the introduction of HTML5's web storage feature, the five major web browsers have rapidly increased their web storage capability. The data held in the web storage feature is an area of interest for forensic investigators. Sariboz and Varol [71] examine the web storage feature on the Android platform for the five major browsers (Google Chrome, Samsung, Firefox, Opera, and Web

Explorer). It was shown that the implementation of web storage on the Android platform is substantially similar to that on desktop platforms. Further, the information is beyond that presented by the previous web stored browser information that used cookie technology. The improvement provided by HTML5, therefore, means the browser is now a potentially richer source of forensic evidence than was previously available.

5.2.2. Huawei smartphones

The increasing presence of Huawei smartphones in the consumer market means that the ability to examine Huawei phones is becoming of increasing importance. Smartphones are usually backed up locally on the device's internal storage and as well as on PC. However, some of the backup data is encrypted to protect privacy, which the examiner must decrypt in order to analyse the data. If the backup data has been encrypted with a user-centered value, such as a password or personal identification number (PIN), recovering the value should take presence [72].

The authors reverse engineered the Huawei smartphone backup application, KoBackup, and its PC backup program, HiSuite, to reveal the local and PC backup processes, including the password-based encryption. Local backup is performed by the phone itself and the data is stored in the internal memory, an SD card, or a USB drive. The local backup requires a password and the encryption only applies to database files. The PC backup is synchronised between the phone and the HiSuite on the PC via a USB connection. Unlike the local backup, the PC encrypts both database and media files, and will do so even in the absence of a password [72].

The researchers found that it is impossible to decrypt password based encrypted data on Haowei smartphones without a user-entered password. It is, therefore, necessary to recover the password, of which, they found four password recovery methods, ie four different password authenticators. Two of the password recovery authenticators are created during the backup process. The third password authenticator is in a "backupinfo.ini" file created after backup on the PC. The fourth method is a plaintext attack media file based on the user-entered password. For each method, estimates of the time to recover passwords is provided, with estimates ranging from less than a minute to multiple years. The fastest method for an eight digital password is up to seven years [72].

5.3. Apps

Instant messaging has become an essential means of communication exceeding that of voice calls and SMS. Instant Messaging applications have pervaded beyond personal use and are now increasingly used for business and professional communications. But, they are also used for criminal activities.

5.3.1. LINE

Instant messenger is an internet based category of applications that has become a popular medium for the conduct of cyber crime. LINE has increased in popularity as a communications app growing growth from 170 million users from the second quarter of 2014 to 217 million users by the fourth quarter of 2016, and is particularly popular in Asian where it is ranked as the second most popular instant messaging app. LINE uses unencrypted messages. Riadi et al. [73,74] test the ability of two mobile forensic tools, Oxygen and MOBILedit, to examine digital evidence from the LINE messenger app.

Oxygen could generate LINE text message artifacts using physical acquisition. Oxygen was able to perform timeline analysis for calls, messages, calendar events, geolocation data and applications activities. MOBILedit was able to obtain contact information, text messages, deleted data, and pictures, but video artifacts could not be obtained. The picture artifact includes metadata such as file path, size, and dates created and modified [73,74].

5.3.2. Blackberry Messenger

Blackberry Messenger is one of the world's most popular smartphone instant messaging apps with high uptake in Britain, India, South Africa, and Indonesia. It was originally designed only for smartphones using the Blackberry operating system, but is now available on Android, iOS, and Windows platforms. Riadi, Unar and Firdonsyah (2017) conducted experiments following the NIST Mobile Forensic method using *Andrilla* on a Sony Xperia Z running Android Lollipop. *Andrilla* was able to acquire "several' messages to reconstruct the conversation, but images could not be displayed. Reports and logs could be generated in HTML format and text files and contained: email accounts, Wifi passwords, applications, SMS, and call logs. The text file report included the date of data acquisition, Android version, IMEI and other data.

5.3.3. iPhone health app

The iPhone health app automatically collects activity data for health purposes, including the number of steps taken and distance travelled, which are recorded with timestamps. In addition to the Health App that is shipped with the iPhone, users can access other apps and wearable sensors that can be synced with the health app where the data, or a copy, can be stored. The information could be very useful in forensic investigation in a number of scenarios including, but not limited to, assessing probability statements, in the form of a likelihood ratio, about scenarios or routes; or, the analysis of physical user activity over time. It is important to note that the reliability of Health App information cannot be assumed [75].

In a study of five subjects using iPhone 6, iPhone 7 and iPhone 8, the accuracy of steps and distances was assessed under a range of conditions, and against manual measurements. Variables that were tested included carrying locations trouser pockets, jacket pockets, backpack, and hand); walking and running; and, a range of distances travelled. The data for the number of steps taken was found to correlate well the manual measurements, part from a few outliers. The distances registered by the iPhones was found to be dependent on the carrying location, the walking speed, and the walking style of the subjects. For example, a walking (or running) style with vigorous arm movements led to higher registered distances travelled. Although little information is available as to how the app functions, the researchers determined that the geolocation APIs are not utilized by the Health App during locomotion, which means that it is reliant on accelerometer and gyroscope sensor data [75].

5.3.4. Snapchat

Snapchat is a popular social network app that is available for Android and iOS devices. It allows users to send messages, photos and videos with a predetermined time to view. Once the time has expired, the contents are automatically deleted and the recipient can no longer view it. An examination for potential Snapchat artifacts on an Android platform was conducted using two forensic tools – Autopsy and AXIOM Examine [76].

Autopsy was able to view ~10% of Snapchat images and videos and some basic information. But, it was not able to indicate deleted snaps, chat messages, user, and friends. AXIOM Examine presented event logs, sent snaps, 100% of friends, 100% user, 58% chat messages and 6% of delivered video with detailed information such as sender, receiver, time, and status. But, it was not able to indicate deleted, story, and delivered photo snaps.Using both tools manually, more artifacts could be found [76].

5.3.5. Kik

Kik is a relatively new messaging app that has grown popular quickly among young users with 300 million users. The marketing appeal was the promise of anonymity as users were not required to provide personal details, a phone number, verify an email address, nor, importantly, verify the individual's age. Verifying the identity of the Kik user can be difficult for the forensic examiner. The app consequently gained a reputation as a preferred app for child abusers and bullying [77]; Ovens and Morison, 2016). Although the company was on the verge of closing down the app due to a dispute with regulators, it was acquired by a holding company, MediaLab, which will invest in its future [77]. Kik do not store and, therefore, cannot retrieve any sent or received message, meaning any forensic evidence is the responsibility of the forensic examiner.

Ovens and Morison (2016) studied forensic artifacts produced by the use of Kik on iOS devices. They used iTunes to perform a logical acquisition (not primary purpose of this app) of the target device. Apart from message attachments, Kik related files on the iOS device have names and suffixes suggestive of their content. However, the filenames are more obscure on the iTunes computer back up files. The study reveals not only contact information can be retrieved, but also other Kik users suggested by the search engine when the *Find People* feature is used, and bots run by Kim's administrators and marketing companies. Additional information is available that suggests the frequency of communication between the user and the group (Ovens and Morrish, 2016).

Messages from blocked users are delivered to the device, but are invisible to the user, unless the user unblocks the corresponding party. Message data includes message content, sender/receiver, time stamps, and chronology. Also, data specified if the messages were direct between two users or part of a group chat. The date and time of blocking and unblocking was not apparent. Deleted contacts and chats could be recovered by the examiner in the kik.sqlite database. Entire conversations could be retrieved even when the conversation had been deleted.

When. Kik user sends a video or image, it is uploaded to the Kik servers and a copy is stored on the device, along with a preview version of the attachment. The recipient is notified of a new message (if permitted). On opening the Kik app, all chats are automatically updated and attachments downloaded. Attachments can also be retrieved from the Kik servers via a web browser using the URL that can be found on the device. Attachments that have been deleted from the Kik app can still be retrieved from the iOS device and the Kik server for eight weeks and four weeks respectively. Moreover, preview versions are still recoverable from the device and backed up on iTunes three months after deletion.

5.3.6. WeChat

WeChat is one of the world's most popular instant-messaging smartphone apps in the world. The app has multimedia capabilities including text, images, voice, and video, in addition to services such as WeChat Moments (where users share their lives with friends) and Official Accounts. To protect the privacy of users, WeChat encrypts the database of messages, and data acquisition through the backup functionality is prohibited. By end 2015, there were 697 million active users in over 200 countries. Importantly, WeChat is the instant messaging mobile application with the highest number of Chinese users. The app is used widely by criminals for communication, and for the organisation and coordination of criminal acts such as selling illegal items, fraud, and child exploitation. The ability to retrieve and interpret data from WeChat is, therefore, an essential source of evidence for investigation [78].

Wu et al. [78] studied the retrieval and interpretation of several versions of WeChat (version 5.0 through to version 6.3.27) on six different Android smartphones. Notably, the authors cite other studies that demonstrate that each app requires its own forensic method and that the literature regarding one app cannot simply be applied to WeChat. One of the solutions the authors used was to downgrade the version of WeChat to version 6.0 as later versions cannot backup up the data using the backup command.

The SQLite database of the user's chat messages is encrypted and the decryption key can be calculated from data stored on the phone, ie the identity of the phone itself, and user specific information. The authors describe the specific details of the retrieval concerning all the various types of messages, as the different types (text, images, audio, video) have different storage schemes. 'Moments' are stored unencrypted. The multimedia resources can be acquired from the WeChat server after extracting the URL of the multimedia file. The thumbnails can also be extracted from the device. In all, the researchers were able to perform decryption; and, extract text, image, voice, and video messages, and moments [78].

5.3.7. Telegram

No single commercial tool always interprets the all of the information from artifacts correctly, and may produce false results, or not manage the application or version under examination. No single forensic tool supports all instant messaging applications or all of their features. Consequently, several tools are required in order to cover the full range of applications. Tool vendors often base their support on the number of downloads of a given app, or on client requests. In their study, Gregorio et al. [79] noted that none of the three tools tested offered satisfactory support for Telegram Messenger on Windows Phones. Since they published their study, Microsoft has decided to discontinue development of the Windows Phone [80].

Notwithstanding the above, the approach taken by Gregorio et al. [79] is relevant to understanding the process of forensic analysis for an unfamiliar app and phone combination. They use a methodology of a combination of open knowledge, analysis of artifacts, and analysis of source code.

5.3.8. WhatsApp

WhatsApp is a smartphone communications app with over 1 billion users in over 180 countries. It can be used on several platforms including Android, BlackBerry, iOS, and Symbian, and can be used for secure calls, text, video, images, and audio messages. One approach to forensic analysis of the WhatsApp content is to use text mining to process the evidence. The text mining process employs word weighting to obtain a value comparison of a conversation between two actors; and cosine similarity to calculate the similarity between two objects [81].

5.3.9. Skype, Viber and WhatsApp on android

The three most popular mobile voice over internet protocol (mVoIP) apps available from the Google Play (Android) store are Skype, Viber and WhatsApp messenger. Onovakpuri [82] conducted experiments using both logical and physical extractions from an Android device with a rootkit installed.[3] The examination tools used included Access Data FTK Imager, SQLite Database Browser, Internet Evidence Finder, and Epoch & Unix Timestamp Converter.

For WhatsApp Messenger, unique directories could be found that include information records and logs related to the sent and received activities of the user: contacts; and chat messages, pictures, audio and video. For Viber, two unique directories were found and included contacts information, calls made and received, and GPS coordinates. Similarly, comprehensive information was

[3] A rootkit allows privileged access to the device's Android operating system and can be used in forensic examination of Android devices.

found for the Skype experiments, in addition to the IP address of the device which provides further information concerning the location of the user.

5.4. Internet of Things (IoT)

In their review, Quick and Choo [8] note the increasing prevalence of connected devices providing societal benefits and benefits. Devices such as connected cars, refrigerators, smart homes, fitness bands, early warning tide measuring buoys, air monitoring balloons. Large amounts of data for example, from wearable technology, can be transmitted to a mobile device and sent to the cloud. The data can then be accessed to using web-based applications to interpret and represent the data to users and decision makers, such as health care professionals.

The data from devices can be in many, often proprietary, forms that can impact on the digital forensic processes. When the data from multiple devices is merged and combined with data from other sources, and considered together with other information concerning the circumstances of an investigation, a chronology of events in time and place can become rich in detail.

One of the major challenges with digital forensics is to be able to place the person at the keyboard. Many IoT devices have biometric information and personal identity built in. Information and logs from IoT devices can, therefore, lead to the identification of a person of interest. Smart homes with security systems can have biometric data stored within the cloud.

Accessing the data for an investigation can be an issue if the cloud stored data is in another jurisdiction, privacy issues are not carefully considered, and maybe subject to security measures [8].

Internet of things devices communicate with each other directly or via Application Programming Interface (API) over the internet, and they can be controlled by learned devices with high computing capabilities. The growth in the prevalence of IoT devices now presents a much great attack surface including virus, mass surveillance, denial of service, and disruption of IoT networks. Digital forensics is key to Investigations these attacks. Notably, current digital forensics tools and standard procedures, as the community currently understands them, are not ideal for investigations of IoT devices. For example, IoT devices generation large volumes of diverse data in formats that can be confusing to digital forensics practitioners, and the lack of real-time log analysis solutions. Notably, the key evidence on the devices must be extracted from the firmware or flash memory. Further, the data is mostly stored and processed in the cloud which presents access issues for the investigators. This is exacerbated by the two tier processing and storage of data, where commutation is mostly performed at the edge of the network, and metadata is stored in the cloud; and the proprietary nature of hardware and software used in IoT devices [83].

In their study, Yaqoob et al. [83] consider the various broad groupings of IoT devices with a view to constructing an IoT digital forensics taxonomy. Their groupings are smart home, smart vehicles, smartphones, drones, BitTorrent Sync peer-to-peer cloud storage service, and general IoT systems. They then go onto elucidate the taxonomy: 1) forensics phases, 2) enablers, 3) networks, sources of evidence, 5) investigation modes, 6) forensics models, 7) forensics layers, 8) forensics tools, and 9) forensics data processing. In discussing the requirements, the Yaqoob et al. [83] define the following requirements: 1) managing the IoT data volume in a structure specified by a framework that can store and manage diverse types of data that has been generated by various IoT devices; 2) mitigation of privacy risks including the awareness of data owners to monitor and control how their data is being accessed and used; 3) data integration across the spectrum of all data sources including IoT, social media, and other communications generated data; 4) guidelines for the IoT deployment approaches including suggested user managed, smart home forensics system; and, 5) dealing with system identification and human behaviors to form a predictive model to locate relevant evidence. Finally they outline the challenges for which more research is required.

Most smarthomes lack any forensic preparedness and therefore is not well placed if it became the scene of a crime. In a study of various home devices (multifunctional surveillance camera, an alarm system with a base station; motion sensor, and contact sensor; another surveillance camera; and a smoke and CO detector) digital traces were extracted from the devices and the associated smartphone applications. Traces generated by the devices were found on the physical devices themselves, but also on the smartphones and the cloud. The traces could provide information such as when a door was opened or when an alarm was disabled. Digital traces that were available on the smartphone included cached image thumbnails and fragments of camera streams, cached events triggered by the sensors, and event logs. The traces provide investigators provide information concerning what happened, when, which user account sent commands to the device, and recorded images and video. In addition, cloud account credentials can also be recovered from the smartphone applications [84].

Significant challenges in the conduct of the forensic examination were encountered. An increasing amount of network traffic is encrypted; and, communication protocols between the device and the base station are not limited to WiFi and ethernet, some devices use ZigBee, Z-Wave, Bluetooth or custom radio frequencies. The traces on the devices themselves might be limited to configuration settings; were limited in the time period for which the data was retained due to limited memory or until a reboot; or, could only be accessed by non-automated techniques such as JTAG or chip-off [84].

IoT forensics presents additional challenges beyond the technical ones. Traditional digital forensics has generally not required the voluntary participation of citizens and relatively little regard has been paid to privacy. IoT devices, however, function more as a digital witness for which voluntary participation is citizens is required. This can only be achieved if privacy of individuals is guaranteed. Nieto et al. [85] propose that the digital witness solution is adapted to comply with the PRoFIT (Privacy-aware IoT Forensics) model, which allows citizens to retain control of their sensitive information stored in their personal IoT devices.

Cardiac implantable medical devices are increasingly being used to treat patients to manage health conditions. The devices include defibrillators and pacemakers. The devices are surgically implanted and wirelessly configured by healthcare professionals. Due to insecure wireless communication, the devices are vulnerable to attack. Ellouze et al. [86] propose a digital investigation system for the postmortem analysis of lethal attack scenarios on the devices. The postmortem analysis would seek to establish: 1) what functions of the implanted device were impacted, ie either did not execute or executed incorrectly; 2) the role of the malfunctioning of the device in the health event; 3) the malfunctioning was due malicious intent or improper deployment; 4) the attack scenario; and, 5) the vulnerabilities that were exploited.

Interpretation of digital evidence obtained from implanted devices is unique to that of other sources of digital evidence: 1) the consequences of an action of an implanted device will vary from one patient to the next; 2) implanted devices are resource-constrained; and, 3) the evidence is technical and medical and,

therefore, it should be interpreted by different experts. The researchers developed techniques that allow the secure storage of digital evidence logs that track the executed sensitive events, and they implement a security solution allowing for the protection of the devices. Further, they construct a library of medical rules that infer potential medical scenarios that might have led to the death of the patient, or that created cardiac emergency situations. The examination is a three step process:

1. The cause of death is identified based on the observations collected and stored in memory by the device and the log of actions performed by the device;
2. Based on the access and system logs, reconstruct potential attack scenarios that would generate the similar content of the collected evidence; and,
3. Correlate the technical and medical evidence to arrive at a conclusion supported by the evidence [86].

An efficient investigation of attacks on cardiac implantable medical devices should reflect the following requirements:

1. The postmortem investigation should be capable of differentiating between a natural death and a criminal death caused by an inappropriate response of a previously attacked device;
2. The digital traces should be reliable and accurate, and encompass three types - collected prior to death; when arrythmias have been detected; and, collected related to sensitive activities;
3. Protection against alteration;
4. Secure access even when the battery has been exhausted; and
5. Reconciliation of the evidence interpreted by technical investigator, and the forensic pathologist and other medical experts [86].

The authors also outline several attack scenarios:

- Simple attacks – including eavesdropping (unauthorised interception of communication between a device and an authorised programmer); unauthorised access to execute remote attacks (for example, repetitive electrical shock generation, data in the data log, clock alteration, and therapy modification); attacking the device availability (jamming, replay, repeated access attempts, or exploiting software vulnerabilities (for example, remotely update the device's software);
- Advanced attacks – including a combination of simple attacks described above; and,
- Advanced complex attacks - following an attack on a device, an adversary might perform some anti-forensic techniques, such as deleting all logged events relating to the attack; or, prior to the attack, create a drift in the device clock so that the time logs of the event do not correlate with the time of death [86].

As of 2016, there were no accepted digital forensics frameworks for the conduct of digital forensic investigations in an Internet of Things environment. Kebande and Ray [87] propose a framework a generic digital investigative framework for IoT that can support IoT investigative capabilities with a degree of certainty. It complies with ISO/IEC 27043:2015, the standard for information technology, security techniques, incident investigation principles, and process.

The framework comprises three processes:

1. The proactive process involves planning and preparation before an incident occurs and includes the IoT scenario definition, evidence source identification, planning incident detection, potential digital evidence collection, digital preservation, and storage of potential evidence, which are all defined in ISO/IEC 27043:2015;
2. The IoT forensics, including cloud forensics, network forensics, and device level forensics, which have the potential for being investigated using forensically sound methods; and,
3. The reactive process which is the actual investigation and includes initialisation, acquisitive, and investigative components.

The authors suggest that the proposed framework should be incorporated into future digital forensic tool development [87].

5.5. Network forensics

The validity and integrity of data can be compromised by failures in system security, of which intrusion detection systems are an integral part. Intrusion detection systems generally include a sniffing process, observing data traffic, and traffic log analysis. SQL injection is a technique used to exploit web applications that store data in a database. An attacker can take advantage of SQL syntax and capabilities by influencing what is forwarded to the database. Detection of SQL injection attacks is identified by forensic evidence that is collected, checked, analysed, and reported. The evidence can be collected from various sources depending on the given situation, and can include the WebServer, network switch, router, cloud, email, and the suspect source device. Caesarano and Riadi [88] conducted experiments with Snort, an open source intrusion detection system using the NIST 800-30 standard. They found that the implementation of the Snort Intrusion Detection System on the web server can provide information concerning SQL injection attacks. Analysis of the log files produced by Snort identify unauthorised actions that occur on the web server.

Rizal et al. [89] note the expanding domain of security attacks on IoT devices due to the multiple vulnerabilities. The vulnerabilities can include attacks on the physical device (micro probing, reverse engineering), side channels (timing, power, fault, electromagnetic), environmental, crypto (cyphertext, known plain, chosen plain, man in the middle); software (virus, Trojan, logic bomb, worms, denial of service), and network (monitor and eaves dropping, traffic, camouflage, denial of service, node subversion or malfunction or capture or outage, message comption, false node, replication, and routing).

The researchers experimented with a flooding attack using an infected Bluetooth device on an IoT device to perform network forensic testing on the device and identify the attack packets. Noting the large amount of data that will be produced in such a scenario, it will be difficult to locate the evidence that identifies the source of the attack. They describe a nine step process for the forensic model which identified three IP addresses that committed the unauthorised actions and led to the traffic overload [89].

Network forensics is dealing with dynamic and volatile data instead of static and stored data, ie the crime is constantly changing. Network forensics is the scientific process that ensures investigation of attacks that are performed in a network or network devices. In their review of network forensics, Jayakrishnan and Vasanthi [90] note that current network forensics processes do not address the forensic challenges presented by new networks such as Internet of Things. Further research needs to be conducted to meet the network forensics challenges of IoT and 5G.

5.6. New devices and apps

There has been much discussion in the media and within

forensic science (and cyber security) of the Internet of Things. Perhaps the most pervasive of these devices are the digital virtual assistants[4] such as Amazon's Alexa, Google Assistant, and Apple's Siri, but others are also appearing on the market. All three of the main ones have voice matching technology, 'delete recording' options, instant translation technology, are compatible with a range of Internet of Things brands, and support multiple languages [91]. As can be imagined, as each device is in 'always on' mode, they will be rich in the data that it has captured and will present challenges for digital forensics examiners to retrieve and interpret the data.

The digital virtual assistants are designed to act in an ecosystem where they can access cloud services (such as Alexa cloud services and other clouds), use companion devices (such as personal computers, mobile devices and smart devices), access third party applications (such as pizza delivery and ride sharing), communicate with other IoT devices (such as smart lighting and smart smoke alarms). The Amazon Echo family of devices, including the Dot and Tap, connect to the intelligent cloud-based voice service known as Alexa. There is a convergence of Alexa with connected cars, smart refrigerators, and robots [92,93].

Chung et al. [92] propose a new digital forensic approach that combines cloud-side and client-side forensics. The device operations are based on Alexa, therefore the artifacts are located in the cloud. In order to access these artifacts, valid user accounts are required; and, it is difficult to recover deleted data from the cloud. The authors propose a multi-level strategy that analyses the data from the hardware (the Amazon Echo device), the network to understand the communications between each component, the client(s) (mobile apps and web browsers) which are used to set up and manage Alexa enabled devices, and the cloud.

In addition to the well-known AI virtual digital assistant offerings from Google, Samsung, Apple and Amazon, more recent offerings are now available from large Chinese companies including Xiaomi and Alibaba. Despite their recent appearance in the consumer market, evidence from digital virtual assistants have already been used in several homicide investigations. Jo et al. [93] conducted digital forensic analysis of the four major providers of digital virtual assistants, referred to as 'AI speakers'[5] that are available in the Republic of Korea – *Clova* of NAVER, *Kajao I* of KAKAO, *NUGU* of SKT, and *GiGA Genie* of KT. Five forensic analysis methods employing, both static and dynamic analyses, are proposed, with a focus and in-depth examination of the *Clova* system. Multiple analytical approaches are very useful for validating results.

As the digital virtual assistant functions as part of an ecosystem, there are five analysis techniques that can be applied to the system:

1. Packet analysis via the AI speaker studies the communication process between the AI speaker and the cloud, and are collected in real time
2. Packet analysis via the Android mobile app studies the communication between an application and the cloud, and user information data is collected in real time
3. Android directory analysis of the data that is stored by the Android mobile app which communicates with the cloud while using applications such as AI speaker configuration and voice commands. Artifacts available here include personal information, connected speaker information, and voice command information
4. Android Application Package (APK) decompilation analysis which looks at the communication between an Android mobile app with the cloud to process the user's voice input. This data can reveal the API address and the data transmitted to the server, and other data stored on the server of the device
5. AI speaker chip-off analysis studies the identity information of the device required for the cloud to recognise the user, the user's personal information, and device history information. The information can include, for example, the user's name and address [93].

The authors provide a significant amount of instructive detail of the kind of data that can be obtained from the digital virtual assistant, and what conclusions can be interpreted from that data. In addition, they describe the Clova Digital Forensic Investigation Tool. The forensic strategy for a given device and ecosystem will depend on the way in which the vendor has designed the ecosystem and how it has been configured on installation. For example, in situations where no mobile apps linked to the assistant, some of the methods cannot be used. Some assistants reinstall all applications and check for updates every time they run, which will result in overwriting of previous data. But, if the metadata of the file system can be identified, the deleted file system can possibly be restored [93].

Notably, the authors provide a word of warning that, when using the directory analysis method, the integrity of the resultant data is compromised during the process of acquiring administrator privileges. The same caveat does not hold the situation when the tools are used to collect data from the service provider's cloud as a legitimate communication protocol has been employed. Further, the identification information obtained in most analyses employed by the authors carries a high degree of surety, as does the chip-off analysis [93].

5.7. Apps – non-phone

5.7.1. Database forensics

The database is at the heart of any digital application and, with the growth in available applications, databases are becoming increasingly important for the storage of important and sensitive information. Database forensics, a sub field of digital forensics, focuses on the detailed analysis of a database including its contents, log files, metadata, and data files. The principles of digital forensics apply to database forensics. Chopade and Pachghare [94] review the state of database forensics for various relational databases including MySGL, Oracle, SQLite, PostgreSQL, DB2, and SQL Server; and, NoSQL databases like MongoDB and Redis. The rising popularity of NoSQL databases is due to their ability to handle even larger amounts of data. The authors review several database forensic investigation models, artifacts (including metadata, application schema, triggers, data structure, storage engine, and logs), tools for SQLite (including Undark, SQLite Parser, SQLite Doctor, Phoenix Repair, and Forensic Browser), tools for database extraction (including Oxygen Forensic Detective, Xplico, Digital Detective Blade, Kernal Data Recovery, SysTools Analyzer, WinHex, NetCat, Windows Forensic Toolchest, SQLCMD, and Forensic Toolkit) [94].

5.7.2. Spotlight

Apple's Spotlight allows a user to search files, mail archives, address books, contacts and other digital assets embedded in a file. Spotlight organises and accesses information using metadata, and collects additional data about files such as Last Opened timestamp, number of times used, and dates and times of usage. The Apple operating system (*macOS*) maintains extended attributes in the file

[4] The term 'digital virtual assistants' has been used for consistency in this review. As it is a new field, there are alternative naming conventions employed by some authors.

[5] The authors refer to the devices as 'AI speakers'.

system which Spotlight also collects and indexes this data [95]. The researchers wrote a python script to read the relevant database and parse all of the metadata contains within. By reading the data directly, instead of using macOS utilities, it is possible to recreate the directory structure and ascertain the last time that the record for a particular file or folder was updated. The author has made the script freely available.

The author presented a case where the method was used to investigate a theft of intellectual property. The files relevant to the intellect property had been removed from a 500 GB disk at some prior time with no visible remnants present. The disk had a Spotlight index which indicated that it had been attached to a Mac system, yet the office environment did not have any Mac systems. On examination of the Spotlight metadata, it was found that there was complete metadata present that referenced the files in question. The disk was in heavy use in the office Windows environment, but Windows does not interact with Spotlight, so the Spotlight database was preserved despite heavy use in the approximately three month interval between the theft and the investigation [95].

5.7.3. America online instant messaging

While digital forensic practitioners need to maintain proficiency in techniques, they also need to maintain current understanding of the artifacts that could be recovered from different types of instant messaging products. One such product is America Online Instant Messenger desktop version (AIM). Yang et al. [96] sought to identify the digital traces from AIM version 7 running on a Windows 8.1 environment. Their results were inconsistent with the published results of previous studies. They found that the caches are no longer a source of potential data for AIM 7, with recent conversations and login credentials not evident. The digital forensics practitioner may potentially retrieve the most recent user image and personal messages from the server using the corresponding links. Time stamp and file path information can be recovered from the system files (short cuts, event logs, thumb cache etc and registry keys) of the Windows client application. Artifacts of the contact lists and conversations can only be recovered from the memory dump. Additional data such as portions of conversations and transferred files can potentially be recovered from the swap files and unallocated space. Although most network traffic is encrypted, the IP addresses and URLs may assist in understanding the activity of a suspect. Notably, the trend of users storing their data in the cloud was consistent with users of AIM 7 [96].

5.8. Drones

Drones, or unmanned aerial vehicles, have grown in popularity among hobbyists and for commercial use alike such as package delivery. They are also being used for law enforcement surveillance; agricultural maintenance; monitoring of poaching of wild life in Africa; and, acquiring specialist movie and sports event footage. There are also reports of the technology being used for nefarious purposes such as physical assaults; intrusions into protected places such as the UK Parliament, Royal residences, the White House, and prisons; and, interference with civil aviation. There is a requirement for forensic analysis of these devices [97]. The recent attack of the Saudi Arabian oil production facilities is believed to have been conducted using drones.

There are four challenges to be addressed by the digital forensics practitioner during the course of investigating the use of a drone: 1) acquisition of data as it can be difficult to directly access the physical disk for imaging; 2) establishing location and flight path, which is key to establishing any offences, when the recording of data will differ between manufacturers and may not be recorded at all by the device and recorded only on the controller; 3) metadata of media captured by the device might provide geo-location data; and 4) establishing ownership which can be difficult if the vehicle has been abandoned. These challenges can be made even more difficult with the availability of components, users can now build their own vehicles specifying their own customisations. Directions for the acquisition and analysis of the device's internal storage are provided, including the interpretation of in-flight data, captured media and operating system. As the drone can be controlled via Android (Samsung Galaxy S3) and iOS (Apple iPhone 6) devices, analysis of these devices is also available. There's are limitations in identifying the owner of the vehicle [97].

A sound forensic investigation will include consideration of all evidence including DNA and fingerprints which could assist in establishing ownership of the device. It is recommended that, on seizing the device, that it is powered down to prevent the data being compromised. As drones continue to grow in popularity, it is expected that their use for illegal activities will also increase, as will the range of drone manufacturers and models. Examination of the range of drones is likely to present, and be analogous to, the challenges faced in mobile forensics. Other methods of data acquisition, including JTAG and chip-off are likely to be appropriate for the analysis of drones [97].

5.9. Volatile memory forensics

Over the past decade, the subfield of volatile memory forensics has evolved to become a reliable and effective technique for recovering forensically sound information from computer systems [98]. Once memory has been acquired, the challenge is to interpret the raw memory into higher level artifacts. This is complicated by the absence of publicly available documentation of the internal structure of software, therefore requiring reverse engineering. As has been referenced elsewhere in this review, reverse engineering is time consuming, difficult and not scalable.

User space malware utilizes code injection techniques to manipulate other processes or hide its existence. Current tools and plugins are unreliable when attempting to reveal existing malware. Attackers can use a variety of methods to evade detection, for example, by creating an executable file that does not appear to be executable; or, by exploiting the paging mechanism. A novel approach that reveals all executable memory pages that are of potential interest to an investigator, despite the use of hiding techniques [99]. The approach involves examining the Page Table Entries, for the executable state of a page, which are enumerated via the paging structures which is faster and more reliable than alternative, predecessor approaches. The approach was tested on Windows 7 and Windows 10 environments.

Memory smear is a common problem when acquiring forensic memory from an active system, particularly when the system is under heavy load. Smear can result in corruption of a memory sample. Further, malware targeting memory can tamper with in-memory data. To address the issues of memory smear and tampering, strenuous testing of the memory parsing components of analysis frameworks must be conducted. Due to the large volumes and complexity of memory data, the testing must be conducted automatically. *Volatility* is one of the most widely used frameworks with a total functionalist comprising over 60,000 lines of code. It cannot be reliably testing by manual means [31].

An automated testing method is 'fuzzing' which are programs that generate input to cause programs to crash or to behave incorrectly. Case et al. [31] describe *Gaslight*, an automated fuzzing architecture which they tested against *Volatility* and *rekall*. *Gaslight* supports seamless testing of the memory forensics frameworks. Although the testing was not exhaustive, *Gaslight* was able to find crashes in numerous core *Volatility* plugins for linux, and OS X, but

not in Windows; and for some plugins for *rekall* [31].

Block and Dewald [99] provide detailed descriptions of the fundamentals of code injection techniques including: 1) Remote Shellcode Injection; 2) Reflective DLL injection; 3) Atom Bombing; 4) Process Hollowing; and, 5) Gargoyle. They also describe fundamentals pf: 1) private and shared memory; 2) Page Table Entries and the Page Frame Number database; 3) the different states of Page Table Entries (including hardware state, transition state, Proto-pointer PTE, Pagefile state, Unaccessed and state); and, 4) large and huge pages. The authors also list complementary work and resources made available by others.

Block and Dewald [99] demonstrate that it is possible for to prevent injected code from being reported by current code injection detection plugins. In their novel approach, executable pages are detected despite any intentional or unintentional hiding techniques. Two injection techniques were successful in hiding from the new approach. The authors' algorithm will report a huge amount of data that will require investigation. It is provided as a plugin that should be integrated with code injection detection plugins in order to strip out benign data. Finally, the authors note limitations of their approach of which investigators should be mindful and take mitigating actions.

Recognizing that malicious software (malware) is the enabling technology for most forms of cybercrime, Palutke and Freiling [100] note the demand for methods to detect, acquire and analyse the software in a forensically sound manner. Existing methods have improved in their ability, but the emerging challenges of malware in hidden memory and hypervisor-based malware can potentially impact their reliability. Memory is divided into reserved and unreserved memory in order to perform different functions, with reserved memory generally avoided by acquisition tools. Data can actually be hidden in the reserved areas and are often referred to as hidden memory. The hypervisor-based malware takes advantage of processor virtualisation that migrate a running system onto a virtual machine. But, hypervisor-based rootkits are detectable. They researchers were able to combine both approaches, which they refer to as *Styx*, and locate it in hidden memory. *Styx* was not detectable with any current forensic memory acquisition software.

Albabtain and Yang [101] used computer forensics processes to perform graphics recovery from GPUs, particularly focused on last visited web pages and last opened images from GPU global memory. They found that recovery of the artifacts from GPUs is possible, but subject to three major challenges: 1) the elusive global memory allocation scheme of GPUs; 2) varying levels of support for different GPU drivers; and, 3) the prerequisite od using certain types of operating system and applications.

5.10. Dark net

A lack of privacy offered by digital communication became a global debate following the revelations of Edward Snowden concerning mass surveillance. Subsequently, use of the Tor Browser and network became mainstream in 2013 for members of the public and criminals alike. Tor is intended to protect the user from both network and local adversaries which is achieved through design that obfuscates network activity and employs anti-forensic techniques. The Tor Browser Bundle is an extended support release of Mozilla's Firefox browser. Firefox, without the Tor extension, stores history, download and cookie information, which are very useful to the forensic investigator.

The increased popularity of the Tor browser has led to an increase in research concerning the effectiveness in protecting users. Muir, Leimich and Buchanan [102] conducted a forensic analysis of Tor software and the host operating system. The experiment used VirtualBox to expert the contents of RAM, which were then analysed using Volatility. RAM collected at four moments: 1) with the browser open after browsing had been performed; 2) after closing the browser window; 3) after simulated uninstalling; and, after the user logs out.

They found that: 1) artifacts proving the installation and use of the browser are generated in memory and on disk in the form of default bookmarks. Said artifacts are attributable to a particular user, uniquely identify the Tor browser, and persist though uninstallation and logout: 2) user activity is written to the Windows registry as a consequence of recent updates to Windows 10, therefore revealing the titles of pages visited using the browser; and, 3) a forensic methodology can be devised. The information that can be revealed under static analysis includes HTTP header information, web page titles, and a URL. Under live analysis, traces of Tor processes after the browser had been closed and the user logged out. The path to browser executable was visible in RAM and included the username and the device from which it was run.

After uninstalling the Tor Browser Bundle and logging out, the Tor related processes had ended, but the outputs parsing the volatile memory showed Tor related artifacts, including the absolute path to the Tor install directory. Other artifacts, were also located including such as the page title of visited websites suffixed with *Tor Browser* as well as the absolute path to the Tor install directory, the user name, and referencing Firefox.exe within the browser directory. Of note, one of the title pages contained the German word for search, Suche, which suggests that the Tor exit node was located in a German speaking country. Several artifacts were recovered from unallocated space and demonstrated considerable browsing data leakage. References to the Tor installation directory were found [102].

The most surprising finding is that it is evident that, under Windows 10, browsing data from user sessions is written to non-volatile storage. This occurred when Tor was used in Firefox's Extended Support Release or in Firefox's Private Browsing mode, and also when a portable browser is used. It was concluded that Tor is easily identified, cannot be securely deleted, and activity from the browsing session is determinable. The persisting Firefox.exe process could not be fully terminated by closing the browser window and exists in a traceable but inactive state [102]. Tor can be easily detected using live forensics, particularly when the browsing session is still active.

The conclusion that Tor writes browsing data to disk means that the use of static forensics by forensic investigators is potentially more worthwhile than examining the contents of RAM. The vast majority of the browsing protocol was found in the NTUSER.DAT Windows Registry file, making it possible for the activities of the user to be reconstructed.

The proposed forensic methodology includes the following recommendations:

- Analyse a RAM dump using Volatility to establish the use of Tor and to find the username. This will also reveal timestamps even after the user has uninstalled Tor and logged out
- Extract the registry hive of the previously identified user, or all users, from non-volatile storage
- Search Tor and/or Firefox for titles of web pages visited from the contents of the shellactivities key
- A keyword search for 'obfs4' in unallocated space can reveal bridging IP addresses that may have been used by Tor [102].

A common investigative and intelligence method is to monitor the forum 'Reddit' to identify emerging trends in what people (Reddit users) are thinking about. It is useful to focus on specific subreddits that attract specific users. Researchers conducted an analysis of all posts on the subreddit 'DarkNetMarkets' for a period

of 12 months, specifically to examine the impact of a compromise to, or take down of, multiple international darknet markets in July 2017 [103].[i]

It was noted that, following the actions of July 2017, the disposition of DarkNetMarkets subreddit users went from casual and relaxed to a state of concern, uncertainty, and security-mindedness. Words associated with law enforcement became highly relevant in many topics, and the void left by the previously most popular markets was filled by a multitude of newer and smaller markets [103].

Users appeared to be concerned about trust of the new markets and hackers evidenced by discussions concerning secure transactions between untrustworthy markets. Many discussions featured words referring to Bitcoin, drugs, and delivery logistics. Cryptocurrency and security tools were consistent topics of conversation with the popular cryptocurrencies being Monero and Bitcoin. There was also interest in VPN services. After the July takedown in which Alphabay and Hansa were removed, the most relevant market name became Dream, with additional markets named Aero, Agora, Traderoute, Sourcery, and Trishula. In addition, decentralised market concepts such as OpenBazaar [103].

Discussion topics regarding cryptocurrency are useful intelligence gathering for law enforcement as they are not just restricted to security. To enhance anonymity, darknet market users often use additional services such as 'mixing' or 'tumbling' where users exchange cryptocoins with each other to increase the difficulty in tracing transactions. Mixing services include Dash, Helix, Bitmixer (now taken down), Coinbase, Seraphim, Localbitcoinds, Bitbay, Shapeshifter, and Viabtc [103].

Users have now gone beyond just using Tor for anonymity. 'Tails' is the most recommended operating system to enhance operational security as it automatically configures software to connect to the internet via Tor. Other operating systems include Whonix and Qubes. There was an increased interest in virtual private networks (VPNs) with PureVPN the most relevant; and, authenticated and confidential communication with the subject of PGP encryption being discussed more frequently [103].

Topic modeling is a useful intelligence gathering technique from darknet markets and forums. Although not reviewed in this paper, several references to topic modeling are provided are provided and include the types of items being sold on Alphabay and the top vendors [104]; dragnet hacker forums for source code, attachments, hacking tutorials [105], and malware [106]; and, identifying topics on Chinese hacker forums which revealed new communication methods, specific security mechanisms, and caution over faulty transaction (Hang et al., 2016). It has also been used to detect anxiety related posts from multiple subreddits. As Reddit posts include usernames against posts, users exhibiting a behaviour of interest can be identified [126].

5.11. Anti-forensics

Anti-forensics relates to the impeding of forensic processes by various means, some of which are subject to research. Anti-forensics can be defined as "any attempts to alter, disrupt, negate, or in any way interfere with scientifically valid forensic investigations" [107].

Research in anti-forensics represents just 2% of total digital forensics research by number of articles published, with very little research having been conducted of hardware write blockers [48].

The taxonomy of anti-forensics tools comprises:

- Data hiding
  - o Encryption
  - o Steganography
  - o Other forms of data hiding
- Artifact wiping
  - o Disk cleaning utilities
  - o File wiping
  - o Disk degaussing/destruction techniques
- Trail obfuscation
- Attacks against computer forensic tools and processes (in Ref. [48].

It is noted that should root access be gained to a forensic writing blocking or duplicating device, then many elements of the anti-forensics elements can be compromised as the integrity of the collected evidence is tainted.

Conlan et al. [107] collect and categorise 308 anti-forensic tools and include variables for each of the tools such as anti-forensic capability, developing party, country of origin etc. Building on earlier work, they then devise an extended, comprehensive anti-forensic taxonomy that facilitates a linguistic standardization with deeper, more granular specifications. The expansion to a more granular level was necessary due to the growth in volume and complexity of the anti-forensics domain. Importantly they include tools that were not designed for anti-forensic purposes, but can be used with malicious intent. The taxonomy was designed to capture as many possible situations that a forensics practitioner might encounter in the course of their work.

The resultant extended taxonomy is as follows [107]:

- *Data hiding* including encryption, steganography, data contraception, file system manipulation, hard disk manipulation, memory hiding, and network-based hiding. Each of these categories are further broken down into sub-categories that provide considerable granularity. Most of the anti-forensics tools fell into this category;
- *Artifact wiping* was extended to include, but not limited to, subcategories such as wiping of files, disk, removeable disk, generic, registry, and disk degaussing/destruction techniques;
- *Trail obfuscation*, the deliberate activity to disorient and divert a forensic investigation on a digital system or network includes P2P networking, IP address spoofing, data fabrication, data misdirection/misinformation, and proxy server among others; P2P networking software was found to be very prevalent; and
- *Attacks against forensic tools and methods* includes alerts to forensic tool usage, anti-reverse engineering, hash value integrity attacks among others. These tools have the potential to be the most devastating anti-digital forensic activity in an investigation.

The researchers share their data, including the categorical data on the anti-forensic tools plus the unique hash values related to the installation files of 191 publicly available anti-forensic tools. The 2780 unique anti-forensics installation related files are analysed for their presence in the National Software reference library. Of these, 423 distinct hashes were found to be in the 2016 Reference Data Set. When considering the identifiable country of origin of anti-forensics tools, the three most prevalent source countries were the United States, Germany and Finland [107].

Ext 4 is a popular file system used by Android and many Linux distributions. Within the data structure is the inode table which contains all of the metadata of a file or directory. Gobel and Baier [108] examine the feasibility of using ext4 timestamps to hide data, by using the data structure in the inode table. Data that matches the normal internal structures of the inode table will not be recognised by a digital forensics analysis tool. The authors use a steganographic approach as it raises no suspicion of an information exchange unlike a cryptographic approach. Cryptographic approaches can be

identified but the contents cannot necessarily be read by unintended audiences. As fall back security, in their experiment and before they embedded the information into the timestamps, the authors also encrypted the information to be hidden.

For each file or directory in an ext4 file system, the following timestamps are provided: 1) last modification time; 2) last access time; 3) last metadata change, eg. change of ownership, permissions, or file size; 4) deletion time; and, 5) creation time. The creation time timestamp was not available in ext 2 and ext 3, but was added to ext4. Notably, the timestamps support nanosecond timestamps although end users do not have visibility to that level of detail. This provides a capacity of a few megabytes in which to hide information. Of the five timestamps, the creation time is the only one that is not subject to change and is, therefore, suitable for hiding data [108].

A bitmap file of 357,574 bytes was able to be hidden. The steganographically hidden file was found to be indistinguishable from normal system usage as the timestamp distribution did not significantly deviate from a uniform distribution; and, the timestamps containing hidden information are indistinguishable from that of a normal file system operation. Using SHA-256 hashing, the integrity of the recovered data was found to be assured. The proposed hiding technique has capacity limits and is only suitable for small text files, and not for image or video files [108].

It is recommended that the forensic investigator, in the absence of encryption, use statistical analysis for pattern recognition. Other artifacts which might suggest an anti-forensics technique has been used might be found in the log files; a non-sensical sequence of timestamps, such as access or modification states occurring before a file was created, or just a few nanoseconds after creation; backups containing different timestamp information to the original [108].

5.12. Deleted and fragmented files

The concept of dating and time in computing is an important consideration in digital forensics. As files are created, modified, deleted, and overwritten, date/time events are important in the reconstruction of events that have taken place. Some deleted and fragmented files provide useful evidence in the consideration of criminal activity. Although some attributes can be modified, the dates in the $FILE_NAME attribute can only be modified by the system kernel and are, therefore, immune from any known anti-forensics tools [109].

A digital fragment is a remnant of a deleted file that resides in one or more contiguous sectors of a hard drive. A single file might leave several remnants which can be found in several ways. Slack spaces occur in various forms of which there are two main types: 1) volume slack is the unallocated space left after creating a hard drive partition; and, 2) file slack is occurs in files that do not fully align with a multiple of a cluster size [109].

The physical allocation of files by the file system follows the rules of the applicable file system. When a file in the NTFS file system is deleted, the file record in $MFT table is marked deleted and the corresponding clusters are marked available in the system $Bitmap. The deletion event is recorded in the transaction journals but none of the dates change in the $MFT drive. At this point, no dating is required and the file can be fully restored, but without any guarantee for how long it will remain intact. The file record can be overwritten in two ways: 1) the record in the $MFT is allocated to a different file, but the file can be recovered by creating a new pointer to the file which will also create new system dates; and 2) the available clusters are later allocated for a different file which results in overwriting of the file content. Many files also have a date contained within the file which can be used for fragment dating [109].

Dating file fragments is an important step for event reconstruction when deleted files form part of the evidence. Using their model, Bahjat and Jones [109] were able to determine the date of deleted files and file fragments with a high degree of accuracy, although the accuracy is subject to certain conditions that they specify. They observed that if the file created date is similar to the file modified date, then the file is intact and has not been modified. It is noted that the creation time and the modified time do not always refer to the actual creation time on the file system.

The research by Bahjat and Jones [109] is a foundation for building a dating framework for file fragments. The dates of neighboring files can be used to infer a minimum boundary for when a deleted file was created. Further, the maximum date from the currently allocated file can be used to define the upper-bound period for when the file was deleted. Together, the minimum and the upper boundaries create a time window for a deleted file for which a fragment was found. The dating accuracy is affected by heavy usage of the hard drive, the frequency of defragmentation, and the type of the file system that is in use.

5.13. Images

One of the major trends impacting the practice of digital forensics is increasing rate of growth in the number and size of digital images and video encountered in seized data. This is, in part, due to cameras being a standard feature of smart phones, the increased penetration of CCTV, and now featuring in connected (and unconnected) motor vehicles and drones. The forensic examination of images is within the province of the other digital evidence review paper and will not be considered here in depth, but some issues that directly impact digital forensic practitioners will be introduced.

Two questions that feature frequently in investigations are: 1) source identification – were images made with the same camera? and, 2) common source identification – were different images made with the same camera? Common source identification is much more computationally intensive and, therefore, more expensive. The method used is referred to as Photo Response Non Uniformity (PRNU) which measures the imperfection of an image sensor. It is common practice to compromise accuracy for performance, by a reduction in the size and/or number of sample images, in order to reduce the cost. A solution is proposed using the use of high performance computing systems with an, importantly, variety of many-core processors. Such a complex system can improve application performance, but also apply different algorithms that can provide higher accuracy [110].

The PRNU technique, as applied in most approaches, is sensitive to random noise within systems, and susceptible when simple manipulations are applied to the images. A feature-based PRNU approach is proposed, for source camera, identification that chooses the features that are robust fo image manipulations. The PRNU noise is extracted from the images, with the source camera identified through vector machine classifiers. The proposed algorithm can identify the source camera of a given image with ‘good’ accuracy. Images could be differentiated even when captured from cameras of similar make and model. The technique was robust even when challenged with simple image manipulations or geometric variations [111].

5.14. Chip-off forensics

Chip-off is a technique used to extract data from memory in some circumstances, for example, when the tools available at the investigator's disposable do not support the device, or the device is damaged and cannot be accessed by the tool. The chip-off process involves the removal of the NAND flash memory chip from the

device, and the chip is then accessed directly to extract the raw data. The chip-off process for older devices is quite reliable as the number of raw bit errors was quite low. Advances in technology have increased the storage capacity of NAND flash memory resulting in the number of raw bit errors increasing by several orders of magnitude. In normal use, modern NAND flash memory controllers employ sophisticated error-correcting codes which can correct raw bit errors. Consequently, the standard chip-off method often cannot recover the data in modern NAND flash memory. The forensic process must also extract the error correcting information, in addition to the raw data, that is stored within the chip controller and use this information to correct the errors [112].

In the interval between when the device is seized and the time that the investigator extracts the data, errors can be introduced as a result of charge leakage from the cells of the NAND flash memory (referred to as data retention errors). Further, when thermal based chip removal is employed, the high temperature can result increase the number of introduced errors within the NAND flash memory by two to three orders of magnitude. The number of errors following thermal chip-off procedures may exceed the ability of the error-correcting codes to correct. The chip-off procedure is quite destructive and can corrupt a large proportion of the data, therefore the technique is becoming less reliable [112].

Fukami et al. [112] develop a new hardware-based approach to reduce the number of errors resulting from the chip-off process. Flash memory manufacturers incorporate a *read-retry* mechanism in modern flash memory chips which significantly reduces the raw bit error rate. By incorporating the *read-retry* based error mitigation into the forensic data recovery procedure, the errors are mitigated when the thermal-based chip removal and read procedure is used in certain circumstances.

5.15. Cryptocurrency

Previous reviews of digital evidence for the Interpol International Forensic Sciences Managers Symposium have included material on cryptocurrencies and their use for nefarious purposes. Cryptocurrencies appeal to those undertaking criminal conduct due to three features: 1) ensuring limited anonymity, however users may reveal their identity either negligently or knowingly, or might be revealed by other parties who use external data, independence from a central authority with rules made by consensus, and Connor be abolished or regulated by force, and double spending attack protection where the owner of cryptocurrency cannot use the same units to pay two different recipients. As of January 2016, there were over 600 cryptocurrencies [113] which has since grown to 1596 as of April 1, 2018, and 9914 markets available to trade the currencies [114]. With the number of cryptocurrencies and the number of markets, this represents an impossible task for law enforcement and regulating authorities to monitor.

There is a growing acceptance of cryptocurrency in conventional transactions. Cryptocurrency is distinct from electronic money, which is not discussed in this review. According to Lansky [113]; cryptocurrency systems:

1. Do not require a central authority
2. Retain an overview of cryptocurrency units and their ownership
3. Defines whether new units can be created, the circumstances in which they are created, their origin, and how to determine their ownership
4. Exclusively and cryptographically prove ownership of the units
5. Allows transactions in which ownership of the units changes, and
6. Perform one transaction at the most when simultaneous instructions for changing ownership are received.

Cryptocurrency uses a peer-to-peer system to store transactions within a Blockchain database. The Blockchain is a public ledger that keeps a track of every transaction and is available to anyone within the network. Cryptocurrencies can be owned by through cryptocurrency accounts that comprises a combination of a private key and a cryptocurrency address. A weakness in the Bitcoin system is that the account address can be calculated from the private key. There is no limit on the number of attempts at guessing the password [113].

Lansky [113] describes four levels of anonymity for cryptocurrency accounts:

- Transparent account owner has revealed their identity in a credible manner
- Semi-transparent account is traceable by the government administration
- Pseudo-anonymous account owner can only be known to the owner's business partners which might not include knowing the owner's name, but also being in possession of information that can lead to ascertaining the owner's identity, and
- Anonymous account owner is unknown to anyone.

Cryptocurrencies are only anonymous as the owner is determined by a random set of alpha-numeric characters with no known association to the legal entity. When used in conjunction with Tor and a Virtual Private Network, the entity's identity is protected. Transactions can be further obscured by 'Mixers' who take coins form different sources and redistribute them to hide the original owner of the of the coin and the transactions with which they are involved. This can be taken a step further by breaking coins up into smaller bits before distribution. These features are what has made it so attractive for criminal transactions and used for sex trafficking, drugs, guns, fake identity, assassination, financing terrorism, tax evasion, identity theft, money laundering, malware (such as ransomware), child abuse [114].

Each country has chosen how to, or not, regulate the trading of cryptocurrencies within its borders. Some countries have banned cryptocurrencies from operating or trading within its borders, but often with little impact. It is noted that, at this point in time, transactions occurring outside of conventional systems will generally result in a loss of revenue from transaction fees to the state. Cryptocurrencies are not subject top the usual financial levers that governments can use to control the economy. Conversely, the lack of control and transparency allows legitimate users to purchase goods and services electronically and protects them from criminal actors who may seek to control the local economy [114].

As criminal organisations change their approach to one of exploiting the characteristics of cryptocurrency, an understanding of the digital forensics that is indicative of transactions in the blockchain is essential. This is especially so when cryptocurrency is used to transact between criminal groups [114].

Investigating global currencies, specifically cryptocurrency, has specific requirements beyond those that have traditionally been part of the investigator's tool kit. The tools are not restricted to technical tools, but will also require legislative permission to make enquiries of other jurisdictions. But, transactions are public, so special permissions of financial institutions is not required. In addition, there are a range of applications available to users to assist in the management of cryptocurrency. Knowledge of the applications and where they store the data is important, especially if the applications encrypts or hides the data [114].

When conducting a digital forensic investigation, the usual digital forensic steps should be taken to ensure that all evidence is collected. The steps include:

- Acquire the Random Access Memory (RAM) using the usual tools for this purpose and with which the investigator is familiar. The RAM will help to determine if the data is encrypted; which programs are running; applications that might contain necessary artifacts; indication of additional connected devices
- Locate any wallets which contain artifacts of cryptocurrency. The wallets might contain transactional information with time stamps. They can be tracked and used to identify people or groups, and disclosed during litigation
- Artifacts are stored on the drive in different locations according the file system and depend on the purpose of the device in the currency exchange. For example, the device might be unknowingly used for currency mining; or, it might be encrypted to hide transactions. Logs of internet searching can also identify other entities in the actor's network
- Network traffic can be captured which can reveal transactional data, the IP addresses of collaborators, and online shopping sites for illegal goods and services.

Tools are emerging that assist to identify illegal activity using digital currencies. The tools use public blockchain data with known addresses of threat actors to track the usages of currency.

6. New applications for digital forensics

6.1. Behaviour

Current digital evidence practice places more emphasis on the principles of computer science and engineerings than it does on traditional investigative approaches. Behavioural evidence analysis within digital forensics investigations has become increasingly recognised as a viable practice, but it has not been widely adopted. No model is in existence that investigators would be able to adopt and incorporate into their investigation process. Al Mutawa et al. [115] describe a multidisciplinary approach to a behavioural digital forensics model which incorporates behavioural evidence analysis into the laboratory examination of seized devices. The model integrates behavioural evidence analysis into the digital forensics examination, analysis, and interpretation of the data contained within the digital devices.

The model follows the standard digital forensics of: 1) review; 2) recognition and collection; 3) examination and analysis; and, 4) interpretation and reporting. While the process is usually linear, in practice when following the behavioural digital forensics model, the phases are dynamic and iterative where new evidence about the suspect, victim, and the events can be introduced into the investigation. The new evidence can prompt the re-investigation of previous stages [115].

The model specifies behavioural evidence analysis in the review, and the examination and analysis phases. During the review phase, the currently available evidence and the established facts are considered, in addition to potential offender motivations, behaviour and characteristics. During this phase, the context, classification and prioritisation can be subjected to behavioural evidence analysis.

This process will assist the investigator to fill gaps in the evidence, and to which the conflicting and changing accounts of the incident. During the examination and analysis phase, information is produced concerning the case that will confirm or refute the associated hypotheses. At this stage, content analysis, and timeline analysis and mapping are conducted using the quantitative and qualitative techniques of frequency analysis and language analysis [115].

The authors test the model against 35 inter-personal cases concerning cyberstalking and the possession and dissemination of child exploitative material; and evaluate it against five cases of online impersonation and defamation. The model provided for a more effective focusing of the investigation and direction towards the location of additional evidence. In the example investigation, the time spent on the examination and analysis of devices was reduced from 13 dats to five days. The approach also provided for a better understanding and interpretation of victim and offender behaviours, leading to a better overall understanding of the dynamics of the specific crime. The consequent information enabled the identification of the suspect's collaborators in some cases [115].

6.2. Digital forensic intelligence

Quick and Choo [8] advance the idea for the potential of intelligence to be gained from digital evidence obtained from the increasing sources of data, noting that, in order to do so, the volume of data must be reduced to manage the storage and review. They promote the Data Reduction by Selective Imaging process, which produces a subset of the seized data to a smaller volume to that which has greater potential for evidentiary and intelligence purposes, and removing data and files with low potential. This reduces the issues of collection, storage, analysis, archiving, extracting and creating intelligence products associated with dealing with big digital forensic data.

They found that digital evidence used for intelligence purposes has the potential to identifying to review and gather intelligence from a wide range of case data and to provide insight into the into emerging trends in technology. It can also be subjected to other intelligence applications to provide information on entity information and extraction, keyword filters, entity relationships, emerging crime types and criminal craft, common websites, communication applications, etc. This can be enriched by merging with additional intelligence arising from call charge records, intelligence reports, arrest reports, traffic stops, social media etc. The enhanced intelligence capability derived from digital forensics can be applied to tactical, operational, and strategic intelligence.

When considering digital traces from mobile phone data, Quick and Choo [8] used several mobile forensic and other forensic software from MSAB, Oxygen, Cellebrite, EnCase, Paraben, and Internet Evidence Finder, to extract data from the phones. The volume of exported data varied significantly between each software package. Using intelligence tools, there were able to demonstrate relationships between the entities. They applied their methods to real world devices obtained from a law enforcement agency.

The data reduction strategy employed can vary depending on the nature of the investigation, but there is some predictability based on the crime type. For example, a drug investigation is usually primarily concerned with communications, whereas, a child exploitation investigation will usually be concerned with images. However, these are not exclusive requirements. The methods have applicability to a wide range of investigations or intelligence probes including terrorism, homicide, child exploitation, drug trafficking, fraud, computer crime in addition to others.

Porter [103] refers to the use of the topic modeling approach for other intelligence gathering purposes such as to determine the types of items being sold on Alphabay and the top vendors; source code, attachments, and hacking tutorials from darkness hacker forums to better understand hacker assets; malware; Chinese hacker forums; specific security mechanisms; and, Noel communications methods. There is also potential for Author-Topic modeling on Reddit data to identify users with specific behaviours, such as anxiety. These additional topics are reviewed here, but the references are available in additional references.

6.3. Open source intelligence

The analysis of seized data can be enhanced for intelligence and investigational purposes by drawing on open source intelligence [8].

Intelligence is especially useful for combating organised crime and terrorist organisations, with organised crime featuring in the perpetration of human trafficking, drug trafficking, extortion murder, and high technology crime [8]. The report on Australian Organised Crime reports the enablers of organised crime as money laundering, technology and infrastructure, professional facilitators, identity crime, corruption with the public sector, and violence and intimidation. It is estimated to cost Australia $36 billion, or $1561 per capita, each year (Australian Criminal Intelligence Commission, 2018). Importantly, the report states that the majority of serious and organised crime are enabled, to some extent, by technology. Technology provides criminals with anonymity, obfuscates activities and locations, and increases their global reach by connecting them to potential victims and information around the world. Using technology to commit crime is also significantly more efficient and less resource intensive than traditional methods of perpetrating crime.

Two key technologies enabling organised crime are crypto currencies and identity crime. In addition, organised crime groups use high end encrypted communication devices and applications such as Phantom Secure Blackberry and Wickr (Australian Criminal Intelligence Commission, 2018).

Open source intelligence involves the extraction of information from publicly available sources, with the internet now a major source of that information which is expected to double in size every two years and to reach 44 zeta bytes by 2020 [8]. One of the challenges with open source intelligence is the prevalence of multiple languages and the need for translation capabilities. The authors provide a framework for digital forensic intelligence and open source intelligence, and emphasise the importance of maintaining identity protection measures and network security when using open internet connections.

The digital forensic intelligence analysis cycle is a merger of digital forensics and intelligence analysis methodologies. The process is one of: 1) commence, 2) prepare, 3) evaluate and identify, 4) collect, 5) preserve, 6) collate, 7) analyse, 8) inference development, 9) presentation, and 10) completion or further tasks. The process of combined digital forensic intelligence analysis and open source intelligence is a sub-cycle and described as follows:

1. Commence (scope/tasking)
2. Prepare
3. Identify and collect
4. Data reduction by selective imaging
5. Quick analysis and entity extraction
6. Open source intelligence
7. Entity chart
8. Inference development
9. Presentation
10. Complete

In their study, Quick and Choo [8] used deep web resources such as electoral roles, telephone, and business databases, LinkedIn, Facebook, Twitter, YouTube, Flikr, Instagram, PhotoBucket, web blogs, Tripod, and online sales sites such as eBay, Gumtree, Craigslist and Whirlpool. The information was combined with digital forensic information and charted in a relationship map.

6.4. Other applications for digital forensics methods

Digital forensics techniques and processes are now being referenced for other purposes. One such purpose is archiving, the disposition of records, and maintaining collections of historical records for which government agencies and other organisations are required to comply with laws and regulations governing the management of records. The growing volume of data does not impact law enforcement alone. Vinh-Doyle [116] notes the growing rates at which emails are sent across the world. He also notes the repurposing of email from a means of communication to now also being used for task management and personal archiving. Collecting institutions, currently managing their collections by manual processes, need to improve their methods of discovery, identification, and redaction or they will lose the trust of donors and accumulate a backlog of unprocessed material. This is particularly fraught when the managed information contains personal and personally identifying information. Further, employing digital forensics methods in archives can assist archivists in discovering valuable information for clients, such as credit card numbers, phone numbers, email addresses, social security numbers and other private information. The author notes that, by employing digital forensics methods, light has been shed on the misuse of organisational resources, including illegal and politically sensitive records, such as pornography and misogynistic content. The digital forensic processes assisted in the organisation gaining an understanding equal employment opportunity culture of the organisation by identifying toxic language that might be used in communications between employees [116].

7. Crime and law

7.1. Crime types

Although digital evidence is almost ubiquitous in any criminal investigation, some crimes and especially impacted through the use of technology and, therefore are more likely dependent on the use of technology to a good investigational outcome.

Identity theft is conducted through phishing activities, hacking online accounts, retrieving personal information on social media accounts, and the illegal access to personal information held on databases (Australian Criminal Intelligence Commission, 2018). In Australia, the incidence of identity theft exceeds that of other personal and household thefts.

Grivna and Drapal [117] analysed cybercrime cases brought before court in the Czech Republic between 2008 and 2016. They describe the provisions of the Czech criminal code that pertain to cybercrime inlcuding illegal access, illegal interception, data interference, system interference and misuse of devices. In their analysis, they grouped the crimes into the categories of:

- Password misuse – in order to gain access to private data for a range of purposes, or for financial gain. Common motives were to discover something about a partner or ex-partner, or to seek revenge on ex-partners, or to impact custody arrangements for children. Other motives were to seek financial gain, particularly through internet backing accounts,
- Abuse of position – this occurred in two realms, the public and the private. In the private realm, the motivation was usually for financial gain with bank employees featuring frequently, but also accountants. Public employees featured, notably, police officers with various motives including personal gain and when there was no personal gain at all;
- Hacking – to make a material gain, for example, by obtaining account details or to insert fraudulent payment details; to cause

damage such as taking routers out of circulation or disconnecting a company's services; or, to paid job offer portals;
- Database misuse – for material gain where, most frequently, perpetrators obtained client data and information and offered them to competitors. The database misuse offences were often perpetrated after the offender had ceased working at the company concerned;
- Misuse of information found on flash drives – where perpetrators found information by chance and sought to exploit it;
- Information deletion – from databases especially frequent from after the termination of employment, particularly, but not exclusively, by "computer experts"; and
- Gambling machines and roulette wheels – alterations to the way gambling machines and roulette wheels operated for financial gain.

Ransomware has become prominent in recent years. A ransomware payload will encrypt a system and demand payment, invariably in the form of a cryptocurrency. When payment is made, the encryption key is released and the data can be restored [114]. Organised crime has found that ransomware is a very useful way to meet their goals.

Cryptocurrencies themselves can be the subject of criminal activity rather than just the spoils. It can be the theft of the wallet, containing all of the ownership information needed to access the coins, either physically or via malware. For example, the URL website for an Initial Coin Offering can be altered to capture owner information and currencies; or, a payment gateway could be hacked to intercept cash flows [114].

7.2. Notable cases

City of London Police investigated the cryptocurrency OneCoin, promoted as an investment opportunity and a rival to Bitcoin, for fraud believing that it was a pyramid scheme. Companies associated with the scheme were investigated in the United Kingdom, the United States, Ireland, Italy, Canada, and Ukraine. The company's servers were located in Bulgaria. Investors had been enticed to part with up to 28,000 pounds sterling with a promise of 10% commission from others who they encouraged to invest [118]: [119].

Phone charge records is a commonly used method to establish links between entities in investigations, or even between investigations. They are especially useful in the investigation of organised crime. Investigators in the United Kingdom became aware that phone charge records included information that directed investigators to innocent third parties, or other nonsensical unissued numbers that could not be linked to any real subscriber. It was found that the suspects were using standard feature mobile phones with deliberately limited functionality, and with customised SIMs, referred to as 'stealth', 'spy', or 'spoofer' SIMs. The customised SIMs had been sold in a country from where it is difficult to obtain data. Customised SIMs make use of network features that allow for call costs to be managed by allowing for redirection to an alternative provider, or using a callback process. The SIMs have created their own Mobile Virtual Network Operators and make use of reprogrammed SIMs [120]. The researchers have developed a manual process that requires a minimum of: 1) start date and time of call; 2) end date and time of call; 3) type of call; 4) calling number; and, 5) called number. The process correlates call time points from the outgoing and incoming records and requires both phones to be available.

7.3. Digital music consumption on the internet

The transformational impact on the music industry caused by digitisation is not news. Apart from improving the efficiency in production and distribution, there was also real concern within the industry for a negative impact on revenues, especially with respect to piracy. Piracy, through digitisation, has the potential to weaken copyright protection and, therefore, devalue creative works. For this to hold true, the possession and distribution of pirated works would displace those of purchased sales. Most studies have supported the notion that piracy causes harm to revenues [121].

The authors followed the online behaviour of 16,500 internet users in five European Union countries, through their clickstream activity, identifying specific visits to websites related to music consumption, both licensed and unlicensed. The authors found no negative affect of unlicensed music downloading on music purchasing behaviour. This is despite controlling for individual unobserved heterogeneity. It was observed that there was, in fact, a positive relationship between licensed and unlicensed acquisition although there were significant cross country differences in these affects. Further, there was a positive relationship between the use of licensed streaming websites and licensed websites selling digital music as consumers review licensed and unlicensed acquisition as complementary sources of music. Consumers will place a valuation on the price for the music. If the price exceeds the retail price of the music, the consumer will not purchase the music. It then follows that, if the consumer decides to download an unlicensed copy of the music, then it complements rather than displaces the purchased music. It is also posited that downloading unlicensed music can increase sales of licensed music as it allows the consumer to sample, for example, an artist with a view to making additional purchases [121].

It was concluded that, despite the breach of copyright, music piracy does not negatively impact digital music purchasing behaviour. This research was conducted in 2011 and the authors note that music purchased in the physical format represented the larger proportion of purchased music. At the time of publication, the authors noted that, if piracy continues to grow, it will have a negative impact on overall music industry revenues [121].

7.4. Law and jurisprudence

Jordanian researchers conducted a comparative study of the legal provisions of unauthorised access crime as prescribed in Jordan with other Arabic legislation and French law, and clarifying the position on international conventions regarding this crime type [122]. They make several recommendations for amendments to theJordanian Electronic Crimes Law:

- Aggravate the penalty for the crime of unauthorised access as the current penalty is insufficient to achieve deterrence
- Link the aggravating circumstance to the consequences of access rather than the objectives of the actor as proving the perpetrator's purpose is difficult
- Include an explicit provision to criminalize remaining within the information system illegally
- Access to a state specific information system should be aggravated, and
- Oblige companies to protect their systems.

8. The future

There is the potential for portable storage to grow to between 512 TB and 1 PB over the next 10 years. Future research in digital evidence intelligence suggests the inclusion of data from forensic analysis of a range of devices and locations including phones, computers, portable storage, GPS, CCTV, cloud storage, biomedical data, and Internet of Things. Research into the potential to use XML

data and the development of software to automatically merge the output of various tools into a common format would be useful [8].

8.1. 5G mobile phone networks

The fifth generation of mobile phone networks is now becoming a reality in many countries.It will bring user speeds of 10 Gigabits per second (currently up to 35 Megabits per second), 1000 fold increase in system capacity, and 100 fold increase in connection density over current LTE and LTE Advanced networks. Sharevski [123] notes that mobile network forensics is a cross discipline of digital forensics and cellular networks with the objective to "… investigate cellular network-facilitated crimes …". Key technologies that accompany the introduction of 5G include Control and User Plane Separation (CUPS), Network Functional Virtualisation (NFV), network slicing, and Cellular Internet of Things (CIoT). 5G will support the deployment of new devices and functions including high-speed vehicles and trains, Internet of Things, commercial air to ground service, and service for light aircraft and helicopters, which will be facilitated by the new and/or enhanced 5G network technologies. These technologies (CUPS, NFV, network slicing, and CioT) provide new opportunities for lawful interception and lawful access location services. The NFV will cause a significant reconfiguration of processes and law enforcement agencies cannot assume regulated forensic readiness and pre-established points of interception and localization. Network slicing allows network operators to create customized network partitions based on their preferred business models which can include sharing portions of the network with other operators. This allows for multi-tenancy of the network with multiple options for management of the network [123].

Laws governing mobile network forensics differ between jurisdictions but, in general, require a warrant and privacy protections for safe storage and analysis of acquired evidence. With the anticipated increase in Internet of Things devices, another avenue for warrantless acquisition of mobile network evidence might be available. An Internet of Things device can be a digital witness that can identify, collect, safeguard, and communicate mobile network evidence. It might be necessary for evidence obtained from the IoT device to be correlated with evidence collected from the IoT network operator [123].

8.2. The risks for digital forensics

The quality of digital forensic results is decreasing and the comprehension of cybercrime is diminishing. As has been identified by a number of authors, the consequences of errors and omissions result in miscarriages of justice and dangerous criminals at large to perpetrate further crimes against persons and organisations [35].

The increasing quantity, diversity, diffusion, structural intricacy, and complexity of use of these data make it difficult for the digital forensic to find the most investigatively useful information. Attorneys and judges are struggling to learn how to evaluate and interpret digital forensic results. The intimate and detailed nature of digital traces raises privacy concerns that must be considered in all stages of the data preservation, examination, and reporting.

The situation is further complicated by the competing demands to follow methodical scientific practices and to respond in shorter timeframes, yet deal with dual challenges of growth in cybercrime and big data. Further, there is an increasing demand for decentralised forensic capabilities (for example, at the crime scene) and for correlation capabilities to identify emerging trends and seriality.

A framework is needed to facilitate forensic science and digital forensics to reinforce each other. In its early history, digital forensics practitioners considered the data from devices as fact-based evidence with little consideration given to evaluation or alternative interpretations. This approach still persists today to a significant degree with the effect of denying the scientific basis to the field. To this day, there is still debate about what aspects of digital forensics are or are not science, and some forensic science publications still do not recognise digital forensics as a forensic discipline.

Casey [35] infers that the risks in digital forensics are currently inadequately addressed as technical and interpretive errors continue to be ongoing challenges. There is an inadequate understanding of the operation of hardware and software, and flawed interpretation of the analysis of data with practitioners heavily relying on tools to process data without due regard to limitations and bugs in the tools. This is exacerbated by the highly dynamic technical and operational environments. Casey [35] draws attention to a number of cases where incorrect conclusion, false accusations, and misinterpretation have led to poor investigational and court outcomes.

Non-technical errors, such as insufficient practitioner knowledge, laboratory management, and cognitive bias can also influence digital forensics results. In particular, forensic laboratory management that emphasises speed over quality of results can contribute to errors. Inadequate case management and training can lead to sub-optimal practices, documentation not being properly maintained, and forensic tools not being used properly.

Treating the field as fact based, rather than a scientific discipline, is useful when the data is to be used as information to assist in investigations, including develop and fact check scenarios, locate additional data sources, or to find potential suspects or victims. But, given that digital traces can be altered or parsed incorrectly by the tools, and digital forensics results can be open to interpretation and, therefore, misinterpretation, the assumption that digital forensics is based in fact is dangerous. Some courts have questioned the validity of digital forensics reports due to the absence of demonstrable scientific validity in the analytical process.

The future risks to digital forensics arise in a number of areas including, but not limited to:

- It can be applied in many contexts including investigations, military, critical infrastructure protection, and intelligence operations, with each environment treating it differently and developing its own standard procedures.
- Decentralisation, including the deployment of advanced digital forensic techniques by persons with limited knowledge, can result in the errors described above, and the lost opportunity for broader visibility across the crime environment and to compare multiple crimes.
- The dynamism of the field with new technology and devices, such as the Internet of Things, outpaces the scientists' ability to understand the new technology that they are likely to encounter.
- The volume continues to grow at massive rates.
- Weak knowledge management and information sharing between groups within the justice system.
- Poor quality management with many of the processes used in digital forensics occurring outside of a quality framework that increases the risk of errors and omissions.
- Privacy where governments and business can access huge amounts of personal data, but the tension between privacy and digital forensics is complex. Ignoring privacy concerns may result in the limitation of utility of digital evidence by means of regulation and legislation.

Some steps are being taken to address the risks. The SWGDE

(Scientific Working Group on Digital Evidence) has developed an error mitigation approach that will identify each potential source of error in both technology and human factors. It has some overlap with ISO 17020 and ISO 17025. Error mitigation analysis involves testing and validation of digital forensic tools, but it does not deal with evaluation of evidence and mitigation of bias.

Work is being undertaken to harmonise forensic science and digital forensics. The Digital Media Scientific Area Committee (of the Organisation of Scientific Area Committees) has developed a framework for digital traces, but with a view to it being applied to other disciplines. It includes a framework of scientific reasoning to address defined questions of authentication, identification, classification, reconstruction, and evaluation in a broad range of legal contexts.

Casey [35] describes several knowledge management strategies to address the challenges in digital forensics. These include the definition of three tiers of forensic examination (triage, preliminary examination, and in-depth examination); codifying digital forensic knowledge in automated solutions; collaborative knowledge exchange including multi-disciplinary conferences, structured knowledge management systems (such as instructional documents and videos); forensic advisors who specialise in digital forensics; forensic intelligence that specialises in digital forensics; interoperability and automation, for example, the ability to combine the results of multiple tools that are used to extract information from all data sources will significantly improve the efficiency and effectiveness of an investigation, facilitate verification, and the sharing of information. Several initiatives are under development including the support of forensic intelligence capabilities. Some of the developments in digital forensic capabilities are progressing in excess of the pace at which forensic science can adapt.

Disclaimer

This is a republication in journal form of a conference proceeding that was produced for the 19th Interpol Forensic Science Managers Symposium in 2019 and was originally published online at the Interpol website: https://www.interpol.int/content/download/14458/file/Interpol Review Papers 2019.pdf. The publication process was coordinated for the Symposium by the Interpol Organizing Committee and the proceeding was not individually commissioned or externally reviewed by the journal. The article provides a summation of published literature from the previous 3 years (2016-2019) in the field of digital evidence and does not contain any experimental data. Any opinions expressed are solely those of the authors and do not necessarily represent those of their agencies, institutions, governments, Interpol, or the journal.

Declaration of competing interests

The author has no competing interests to declare

References

[1] E. Casey, Editorial: the broadening horizons of digital investigation, Digit. Invest. 21 (2017) 1–2.

[2] Nist, A Framework for Harmonizing Forensic Science Practices and Digital/multimedia Evidence, National Institutes of Standards and Testing: Organization of Scientific Area Committees for Forensic Science, 2018. Retrieved from, https://www.nist.gov/sites/default/files/documents/2018/01/10/osac_ts_0002.pdf.

[3] Pr Newswire, Digital Forensics Market Worth 9.68 Billion USD by 2022, 06 March 2018.

[4] Market Insider, Digital Forensics Market - Global Forecast to 2022, 16 March 2018. Retrieved from, https://markets.businessinsider.com/news/stocks/digital-forensics-market-global-forecast-to-2022-1018885400.

[5] AppBrain, Android and Google Play Statistics, 2019. Retrieved from, https://www.appbrain.com/stats.

[6] S. Walsh, Australasian forensic science summit 2016: the external future context and the case for change, Aust. J. Forensic Sci. 50 (3) (2018) 245–258.

[7] E. Casey, D.-O. Jaquet-Chiffelle, Do Identities Matter? *Policing: a Journal Of Policy And Practice, Special Issue*, 2017.

[8] R. Quick, K.-K.R. Choo, Big Digital Forensic Data, Volume 2: Quick Analysis for Evidence and Intelligence. Springer Briefs on Cyber Security Systems and Networks, Springer, 2018a.

[9] B. Hitchcock, N. Le-Khac, M. Scanlon, Tiered forensic methodology for digital field triage by non-digital evidence specialists, in: Digital Investigation, DFRWS 2016 Europe – Proceedings of the Third Annual DFRWS Europe, 2016.

[10] B. Cusack, R. Lutui, Updating investigation models for smart phone procedures, in: Proceedings of the 12th Australian Digital Forensics Conference, 2014, pp. 53–63.

[11] C. Stelly, V. Roussev, Nuggest: a digital forensics language, Digit. Invest.: DFRWS 2018 Europe - Proceedings of the Fifth Annual DFRWS Europe 24 (2018) S38–S47.

[12] A. Asquith, G. Horsman, Let the robots do it! Taking a look at robotic process automation and its potential application in digital forensics (unedited manuscript as accepted for publication), Forensic Sci. Int.: Report 19 (2019).

[13] P. Gladyshev, J. James, Decision-theoretic file carving, Digit. Invest. 22 (2017) 46–61.

[14] E. Casey, S. Barnum, R. Griffith, J. Snyder, H. van Bek, A. Nelson, Advancing coordinated cyber-investigations and tool interoperability using a community developed specification language, Digit. Invest. 22 (2019a) 14–45.

[15] E. Casey, Z. Geradts, B. Nikkel, Editorial: transdisciplinary strategies for digital investigation challenges, Digit. Invest. 25 (2019b) 104.

[16] D. Patil, B. Meshram, RegForensicTool: evidence collection and analysis of Windows registry, Int. J. Cyber-Secur. Digital Forensics 5 (2) (2016) 94–105.

[17] B. Meshram, Patil, Digital forensic analysis of hard disk for evidence collection, Int. J. Cyber-Secur. Digital Forensics 7 (2) (2018) 100–110.

[18] B. Schatz, AFF4-L: a scalable open logical evidence container, in: Digital Investigation: DFRWS 2019 USA – Proceedings of the Nineteenth Annual DFRWS USA, vol. 29, 2019, pp. S143–S149.

[19] Home Office, Apcc, & Npcc, Forensics Review: Review of the Provision of Forensic Science to the Criminal Justice System in England and Wales, 2018. Retrieved from, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/800447/Joint-review-of-forensics-provision-July-2018.pdf.

[20] Home Office, Apcc, & Npcc, Implementation Plan: for the Joint Review of Forensics Provision, 2019. Retrieved from, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/796812/Forensics-implementation-plan-April-2019.pdf.

[21] N.M. Karie, S.M. Karume, Digital forensic readiness in organisations: issues and challenges, Journal of Digital Forensics, Security & Law 12 (4) (2017) 43–53.

[22] N. Ab Rahman, G. Kessler, R. Choo, Implications of emerging technologies to incident handling and digital forensic strategies: a routine activity theory, in: Contemporary Digital Forensic Investigations of Cloud and Mobile Applications, 2017, pp. 131–146.

[23] L. Carthy, E. Ovensen, R. Little, I. Sutherland, H. Read, Committing the Perfect Crime: a Teaching Perspective, European Conference on Cyber Warfare and Security, 2018.

[24] Nist, Computer Forensic Reference Data Sets, The National Institute of Standards and Technology, 2019. Retrieved from, https://www.cfreds.nist.gov.

[25] Digital Corpora, Retrieved from, https://digitalcorpora.org, 2019.

[26] E. Casey, Editorial: clearly conveying digital forensic results, Digit. Invest. 24 (2018) 1–3.

[27] Fepac, Forensic science education programs accreditation commission: accreditation standards, American Academy of Forensic Sciences (2017). Retrieved from, http://www.fepac-edu.org/sites/default/files/FEPAC%20Standards%2002122017%20v3.pdf.

[28] Fepac, Forensic science education programs accreditation commission: accreditation standards, American Academy of Forensic Sciences (2019). http://www.fepac-edu.org/sites/default/files/FEPAC%20Standards%2007252019.pdf.

[29] R. Verma, P. Bansal, Scope of managing knowledge in digital forensics, in: Proceedings of the International Conference on Sustainable Computing in Science, Technology and Management ((SUSCOM-2019), 2019. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3363040. Retrieved from.

[30] N. Karie, V. Kebande, Building ontologies for digital forensic terminologies, Int. J. Cyber-Secur. Digital Forensics 5 (2) (2016) 75–82 (The Society of Digital Information and Wireless Communications).

[31] A. Case, A. Das, S.-J. Park, J. Ramanujam, G. Richard, Gaslight: a comprehensive fuzzing architecture for memory forensics, in: Digital Investigation: DFRWS 2017 USA – Proceedings of the Seventeenth Annual DFRWS USA, vol. 22, 2017, pp. S86–S93.

[32] United Kingdom Forensic Science Regulator, Annual Report 17 November 2017 to 16 November 2018, 2019. Retrieved from, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/786137/FSRAnnual_Report_2018_v1.0.pdf?_ga=2.127790285.1384007655.1566857069-1676743936.1458493850.

[33] United Kingdom Forensic Science Regulator, Annual Report: November 2016-November 2017, 2018. Retrieved from, https://assets.publishing.

service.gov.uk/government/uploads/system/uploads/attachment_data/file/786137/FSRAnnual_Report_2018_v1.0.pdf.
[34] House of Lords, Forensic Science and the Criminal Justice System: a Blueprint for Change, House of Lords: Science and Technology Select Committee, May 1, 2019. Retrieved from, https://publications.parliament.uk/pa/ld201719/ldselect/ldsctech/333/33302.htm.
[35] E. Casey, The checquered past and risky future of digital forensics, Aust. J. Forensic Sci. 51 (6) (2019).
[36] N. Sunde, I. Dror, Cognitive and human factors in digital forensics: problems, challenges, and the way forward, Digit. Invest. 29 (2019) 101–108.
[37] Swgde, Establishing Confidence in Digital Forensic Results by Error Mitigation Analysis. Version 2.0, 2018. November 20, 2018.
[38] H. Page, G. Horseman, A. Saran, J. Foster, A review of quality procedures in the UK forensic sciences: what can the field of digital forensics learn? Sci. Justice 59 (1) (2018) 83–92.
[39] P. Sommer, Accrediting digital forensics: what are the choices? Digit. Invest. 25 (2018) 116–120.
[40] United Kingdom Forensic Science Regulator, Codes of Practice and Conduct for Forensic Science Providers and Practitioners in the Criminal Justice System. Issue 4, 2017. Retrieved from, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/651966/100_-_2017_10_09_-_The_Codes_of_Practice_and_Conduct_-_Issue_4_final_web_web_pdf__2_.pdf.
[41] Swgde, Establishing confidence in digital forensic results by error mitigation analysis, Scientific Working Group on Digital Evidence (2017) version 1.7. Retrieved from, https://www.swgde.org/documents/Current%20Documents/SWGDE%20Establishing%20Confidence%20in%20Digital%20Forensic%20Results%20by%20Error%20Mitigation%20Analysis.
[42] G. Horsman, Tool testing and reliability issues in the field of digital forensics, Digit. Invest. 28 (2019) 163–175.
[43] I. Dror, Human expert performance in forensic decision making: seven different sources of bias, Aust. J. Forensic Sci. 49 (5) (2017) 541–547.
[44] N. Sunde, Non-technical Sources of Errors when Handling Digital Evidence within a Criminal Investigation, Master's Thesis, Norwegian University of Science and Technology, 2017a.
[45] N. Sunde, Non-technical Sources of Errors when Handling Digital Evidence within a Criminal Investigation, Masters Thesis, The Faculty of Technology and Electrical Engineering, Norwegian University of Science and Technology, Sunde N. Gjovik, 2017b.
[46] J. Collie, Commentary: digital forensic evidence – flaws in the criminal justice system, Forensic Sci. Int. 289 (2018) 154–155.
[47] G. Horsman, "I couldn't find it your honour, it mustn't be there!" – tool errors, tool limitations and user error in digital forensics, Sci. Justice 58 (2018) 433–440.
[48] C. Meffert, I. Baggili, F. Breitinger, Deleting collected digital evidence by exploiting a widely adopted hardware write blocker, Digit. Invest. 18 (2016) S87–S96.
[49] C. Grajeda, F. Breitinger, I. Baggili, Availability of datasets for digital forensics – and what is missing, Digit. Invest. 11 (2017) S94–S105.
[50] Nist, The NIST Definition of Cloud Computing, National Institute of Standards and Technology: Computer Security Resource Center, 2011. Retrieved from, https://csrc.nist.gov/publications/detail/sp/800-145/final.
[51] Nist, Draft NISTIR 8006: NIST Cloud Computing Forensic Challenges, National Institute of Standards and Technology: NIST Cloud Computing Forensic Science Working Group, Information Technology Laboratory, 2014. Retrieved from, https://csrc.nist.gov/csrc/media/publications/nistir/8006/draft/documents/draft_nistir_8006.pdf.
[52] R. Choo, M. Herman, M. Iorga, B. Martini, Editorial: cloud forensics: state-of-the-art and future directions, Digit. Invest. 18 (2016) 77–79.
[53] V. Roussev, I. Ahmed, A. Barreto, S. McCulley, V. Shanmughan, Cloud forensics – tool development studies & future outlook, Digit. Invest. 18 (2016) 79–95.
[54] S. Mohtasebi, A. Dehghantanha, R. Choo, Cloud storage forensics: analysis of data remnants on SpiderOak, JustCloud, and pCloud, in: Contemporary Digital Forensic Investigations of Cloud and Mobile Applications, vol. 13, 2017, pp. 205–246.
[55] T. Dargahi, A. Dehghantanha, M. Conti, Investigating Storage as a Service cloud platform: pCloud as a case study, in: Contemporary Digital Forensic Investigations of Cloud and Mobile Applications, Chapter 12, 2017, pp. 185–204.
[56] A. Dehghantanha, T. Dargahi, Residual cloud forensics: CloudMe and 360Yunpan as case studies, in: Contemporary Digital Forensic Investigations of Cloud and Mobile Applications, Chapter 14, 2017, pp. 247–283.
[57] A. Amine Chelihi, A. Elutilo, I. Ahmed, C. Papadopoulos, A. Dehghantanha, An Android cloud storage apps forensic taxonomy, in: Contemporary Digital Forensic Investigations of Cloud and Mobile Applications, vol. 12, 2017, pp. 285–305.
[58] S. Manoj, D. Bhaskari, Cloud-forensics – a framework for investigating cyber attacks in cloud environment, Procedia Computer Science: International Conference on Computational Modelling and Security 85 (2016) 149–154, 2016.
[59] V. Kebande, H. Venter, On digital forensic readiness in the cloud using a distributed agent-based solution: issues and challenges, Aust. J. Forensic Sci. 50 (2) (2018) 209–238.
[60] A. Imran, S. Aljawarneh, K. Sakib, Web data amalgamation for security engineering: digital forensic investigation opf open source cloud, J. Univers. Comput. Sci. 22 (4) (2016) 494–520.
[61] D. Chaus, A. Pathak, A. Boramani, S. Rajguru, R. Kalantri, A virtual environment forensic tool, International Journal fo Cyber-Security and Digital Forensics 7 (1) (2018) 63–70.
[62] B. Ogazi-Onyemaechi, A. Dehghantanha, R. Choo, Performance of Android forensics data recovery tools, in: Contemporary Digital Forensic Investigations of Cloud and Mobile Applications, vol. 7, 2017, pp. 91–110.
[63] M. Petraitye, A. Dehghantanha, G. Epiphaniou, Mobile phone forensics: an investigative framework based on user impulsivity and secure collaboration errors, in: Contemporary Digital Forensic Investigations of Cloud and Mobile Applications, vol. 6, 2017, pp. 79–89.
[64] X. Lin, T. Chen, T. Zhu, T. Yang, F. Wei, Automated forensic analysis of mobile applications on Adroid devices, Digit. Invest.: DFRWS 2018 USA – Proceedings of the Eighteenth Annual DFRWS USA 26 (2018) S59–S66.
[65] A. Ali, S.A. Razak, s.H. Othman, A. Mohammed, F. Saeed, A metamodel for mobile forensics investigation domain, PloS One 12 (4) (2017).
[66] S. Nemetz, S. Schmitt, F. Freiling, A standardized corpus for SQLite database forensics, Digit. Invest.: DFRWS 2018 Europe – Proceedings of the Fifth Annual DFRWS Europe 24 (2018) S121–S130.
[67] M. Guido, J. Buttner, J. Grover, Rapid differential forensic imaging of mobile devices, in: Digital Investigation, DFRWS USA 2016 – Proceedings Of the 16th Annual USA Digital Forensics Research Conference, vol. 18, 2016, pp. S46–S54.
[68] S. Saleem, O. Popov, I. Baggili, A method and a case study for the selection of the best available tool for mobile device forensics using decision analysis, Digit. Invest. 16 (2016) S55–S64.
[69] C. Jin, R. Wang, D. Yan, Source smartphone identification by exploiting encoding characteristics or recorded speech, Digit. Invest. 29 (2019) 129–146.
[70] X. Zhang, F. Breitinger, I. Baggili, Rapid Abdroid parser for investigating DEX files (RAPID), Digit. Invest. 17 (2016) 28–39.
[71] D. Sariboz, C. Varol, Acquisition of browser artifacts from Android devices, Int. J. Cyber-Secur. Digital Forensics 7 (2) (2018) 175–182.
[72] M. Park, G. Kim, Y. Park, I. Lee, J. Kim, Decrypting password-based encrypted backup data for Huawei smartphones, Digit. Invest. 28 (2019) 119–125.
[73] I. Riadi, Sunardi, A. Fauzan, Examination of digital evidence on android-based LINE messenger, Int. J. Cyber-Secur. Digital Forensics 7 (3) (2018a) 336343 (The Society of Digital Information and Wireless Communications).
[74] I. Riadi, R. Umar, A. Firdonsyah, Identification of digital evidence on Android's Blackberry Messenger using NIST mobile forensic method, Int. J. Comput. Sci. Inf. Secur. 15 (5) (2018b) 155–160.
[75] J. Van Zandwijk, A. Boztas, The iPhone Health App from a forensic perspective: can steps and distances registered duriung walking and running be used as digital evidence? Digit. Invest. 28 (2019) S126–S133.
[76] T. Alyahya, F. Kausar, Snapchat analysis to discover forensic artifacts on Android smartphone, in: The 7th International Symposium on Frontiers in Ambient and Mobile Systems. in: Procedia Computer Science, 109C, 2017, pp. 1035–1040.
[77] S. Liao, Kik Ap Won't Shut Down after Acquisition by MediaLab, CNN Business, 19 October 2019. Retrieved from, https://www.cnn.com/2019/10/19/tech/kik-messenger-saved/index.html.
[78] S. Wu, Y. Zhang, X. Wang, X. Xiong, L. Du, Forensic analysis of WeChat on android smartphones, Digit. Invest. 21 (2017) 3–10.
[79] J. Gregorio, Gardel, B. Alarcos, Forensic analsysis of tepegram messenger for windows phone, Digit. Invest. 22 (2017) 88–106.
[80] C. Reilly, Windows 10 Mobile Gets its Final Death Sentence, CNET, 18 October 2017. Retrieved from, https://www.cnet.com/news/windows-10-mobile-features-hardware-death-sentence-microsoft/.
[81] A. Marfianto, I. Riadi, WhatsApp messenger forensic analysis based on Android using text mining method, Int. J. Cyber-Secur. Digital Forensics: The Society of Digital Information and Wireless Communications 7 (3) (2018) 319–327.
[82] P. Onovakpuri, Forensic analysis of Skype, viber and WhatsApp messenger on android platform, Int. J. Cyber-Secur. Digital Forensics 7 (2) (2018) 119–131.
[83] I. Yaqoob, I. Hashemite, T. Ahmed, A. Kazmi, C. Hong, Internet of things forensics: recent advances, taxonomy, requirements, and open challenges, Future Generat. Comput. Syst. 92 (2019) 265–275.
[84] F. Servida, E. Casey, IoT forensic challenges and opportunities for digital tracers, Digit. Invest. 28 (2019) S22–S29.
[85] A. Nieto, R. Rios, J. Lopez, IoT-forensics meets privacy: towards cooperative digital investigation, Sensors 18 (2) (2018) 492.
[86] N. Ellouze, S. Rekhis, N. Boudriga, M. Allouche, Cardiac implantable medical devices forensics: postmortem analysis of lethal attackers scenarios, Digit. Invest. 21 (2017) 11–30.
[87] V. Kebande, I. Ray, A generic digital forensic investigation framework for Internet of Things (IoT), in: 2016 IEEE International Conference on Future Internet of Things and Cloud, 2016.
[88] A.R. Caesar, I. Riadi, Network forensics for detecting SQL injection attacks using NIST method, Int. J. Cyber-Secur. Digital Forensics 7 (4) (2018) 436–443.
[89] R. Rizal, I. Riadi, Y. Prayudi, Network forensics for detecting flooding attack on Internet of Things (IoT) device, Int. J. Cyber-Secur. Digital Forensics 7 (4) (2018) 382–390.

[90] A. Jayakrishnan, V. Vasanthi, Empirical survey on advances of network fornsics in the emerging networks, Int. J. Cyber-Secur. Digital Forensics 7 (1) (2018) 38–46.
[91] A. Dennon, The Best Voice Assistants, 16 July 2019. Reviews.com Retrieved from, https://www.reviews.com/voice-assistant/.
[92] H. Chung, J. Park, S. Lee, Digital forensic approaches for Amazon Alexa ecosystem, in: DFRWS 2017 USA - Proceedings of the Seventeenth Annual DFRWS USA vol. 22, Digital Investigation, 2017, pp. S15–S25.
[93] W. Jo, Y. Shin, Kim, D. Yoo, D. Kim, C. Kang, J. Jin, J. Oh, B. Na, T. Shon, Digital forensic practices and methodologies for AI speaker ecosystems, Digit. Invest.: DFRWS 2019 USA – Proceedings of the Nineteenth Annual DFRWS USA 29 (2019) S80–S93.
[94] R. Chopade, V. Pachghare, Ten years of critical review on database forensics research, Digit. Invest. 29 (2019) 180–197.
[95] Y. Khatri, Investigating spotlight internals to extract metadata, Digit. Invest. 28 (2019) 96–103.
[96] T. Yang, A. Dehghantanha, R. Choo, Z. Muda, Investigating America Online instant messaging application: data remnants on Windows 8.1 client machine, in: Contemporary Digital Forensic Investigations of Cloud and Mobile Applications, vol. 3, 2017, pp. 21–39.
[97] G. Horsman, Unmanned aerial vehicles: a preliminary analysis of forensic challenges, Digit. Invest. 16 (2016) 1–11.
[98] B. Schatz, M. Cohen, Editorial: advances in volatile memory forensics, Digit. Invest. 20 (2017) p1.
[99] F. Block, A. Dewald, Windows memory forensics: detecting (in)intentionally hidden injected code by examining page table entries, in: Digital Investigation: DFRWS 2019 USA – Proceedings of the Nineteenth Annual DFRWS USA, vol. 29, 2019, pp. S3–S12.
[100] R. Palutke, F. Freiling, Styx: countering robust memory acquisition. *Digital investigation: DFRWS 2018 Europe –*, Proceedings of the Fifth Annual DFRWS Europe 24 (2018) S18–S28.
[101] Y. Albabtain, B. Yang, The process of recovering image and web page artifacts from the GPU, Int. J. Cyber-Secur. Digital Forensics 7 (2) (2018) 132–141.
[102] M. Muir, P. Leimich, W. Buchanan, A forensic audit of the tor browser Bundle, Digit. Invest. 29 (2019) 118–128.
[103] K. Porter, Analyzing the DarkNetMarkets subreddit for evolutions of tools and trends using LDA topic modeling, Digit. Invest. 26 (2017) S87–S97.
[104] J. Grisham, C. Barreras, C. Afarin, M. Patton, H. Chen, Identifyin top listers in Alphabay using latent dirichlet allocation, in: 2016 IEEE International Conference Intelligence and Security Informatics (ISI), 219-219, 2016.
[105] I.S. Samtani, R. China, H. Chen, Exploring hacker assets in underground forums, in: 2015 IEEE International Conference Intelligence and Security Informatics (ISI), 2015, pp. 31–36.
[106] I. Deliu, Extracting Cyber Threat Intelligence from Hacker Forums (Master's Thesis), NTNU, 2017.
[107] K. Conlan, I. Baggili, F. Breitinger, Anti-forensics: furthering digital forensic science through a new extended, granular taxonomy, Digit. Invest. 18 (2016) S66–S75.
[108] T. Gobel, H. Baier, Anti-forensics in ext4: On secrecy and usability of timestamp-based data hiding, Digit. Invest. 24 (2018) S111–S120.
[109] A. Bahjat, J. Jones, Deleted file fragment dating by analysis of allocated neighbors, Digit. Invest. 28 (2019) S60–S67.
[110] B. Van Werkhoven, P. Hijra, C. Jacobs, J. Maassen, Z. Geradts, H. Bal, A jungle computing approach to common image source identification in large collections of images, Digit. Invest. 27 (2018) 3–16.
[111] K. Akshatha, A. Karunaka, H. Anitha, U. Raghavendra, D. Shetty, Digital camera identification using PRNU: a feature based approach, Digit. Invest. 19 (2016) 69–77.
[112] A. Fukami, S. Ghose, Y. Luo, Y. Cai, O. Mutlu, Improving the reliability of chip-off forensic analysis of NAND flash memory devices, Digit. Invest. 20 (2017) S1–S11.
[113] J. Lansky, Possible state approaches to Cryptocurrencies, J. Syst. Integrat. 9 (1) (2018) 19–31.
[114] D. Orr, D. Lancaster, Cryptocurrency and Blockchain: a discussion of forensic needs, Int. J. Cyber-Secur. Digital Forensics 7 (4) (2018) 420–435, and The Society of Digital Information and Wireless Communications).
[115] N. Al Mutawa, J. Bryce, V. Franqueira, A. Marrington, J. Read, Behavioural digital forensics model: embedding behavioural evidence analysis into the investigation of digital crimes, Digit. Invest. 28 (2019) 70–82.
[116] W. Vinh-Doyle, Appraising email (using digital forensics): techniques and challenges, Arch. Manuscripts 45 (1) (2017) 18–30.
[117] T. Grivna, J. Drapal, Attacks on confidentiality, integrity and availability of data and computer systems in the criminal case law of the Czech Republic, Digit. Invest. 28 (2019) pp1–13.
[118] S. Higgins, London Police Investigate OneCoin Cryptocurrency Scheme. *Coindesk*, September 27, 2016. Retrieved from, https://www.coindesk.com/london-police-investigate-onecoin-cryptocurrency-scheme.
[119] A. Penman, Alleged Pyramid Scheme OneCoin Faces Collapse as Police Move in. *Mirror*, January 25, 2018. Retrieved from, https://www.mirror.co.uk/news/uk-news/alleged-pyramid-scheme-onecoin-faces-11911858.
[120] A. Marshall, P. Miller, CaseNote: mobile phone call data obfuscation & techniques for call correlation, Digitial Investigation 29 (2019) 82–90.
[121] L. Aguiar, B. Martens, Digital music consumption on the Internet: evidence from clickstream data, Inf. Econ. Pol. 34 (2016) 27–43. January 22, 2016.
[122] H. abu issa, M. Ismail, O. Amar, Unauthorized access crime in Jordanian law (comparative study), Digit. Invest. 28 (2019) 104–111.
[123] F. Sharevski, Towards 5G cellular network forensics, Eurasia Journal on Information Security (2018) 8, 2018.
[124] E. Casey, M. Biasiotti, F. Turchi, Using Standardization and Ontology to Enhance Data Protection and Intelligent Analysis of Electronic Evidence, University of Lausanne, 2017. Retrieved from, https://serval.unil.ch/resource/serval:BIB_EFAFD05944CB.P001/REF.pdf.
[125] Statcounter GlobalStats, Mobile Vendor Market Share Worldwide March 2010 – September 2019, StatCounter, 19 October 2019. Retrieved from, https://gs.statcounter.com/vendor-market-share/mobile/worldwide/#monthly-201003-201909.
[126] J. Shen, F. Rudzicz, Detecting anxiety through reddit, in: Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology – from Linguistic Signal to Clinical Reality, 2017, pp. 55–65.