Future Generation Computer Systems 78 (2018) 558–567

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Digital forensic intelligence: Data subsets and Open Source Intelligence (DFINT+OSINT): A timely and cohesive mix

Darren Quick [a], Kim-Kwang Raymond Choo [b,a,*]

[a] Information Assurance Research Group, University of South Australia, Adelaide, SA 5001, Australia

[b] Department of Information Systems and Cyber Security, The University of Texas at San Antonio, San Antonio, TX 78249-0631, USA

HIGHLIGHTS

- Digital forensic intelligence.
- Big forensic data reduction and management.
- Forensic data subsets and open source intelligence.

ARTICLE INFO

Article history:
Received 16 October 2016
Received in revised form
16 November 2016
Accepted 29 December 2016
Available online 31 December 2016

Keywords:
Big data forensics
Forensic data reduction
Intelligence analysis
Open Source Intelligence
Criminal intelligence
Digital forensic intelligence

ABSTRACT

Advances in technologies and changing trends in consumer behaviour have led to an increase in the volume, variety, velocity, and veracity of data available for digital forensic analysis. A benefit of analysis of big digital forensic data is that there may be case-related information contained within disparate data sources. This paper presents a framework for entity identification and open source information cohesion to add value to data holdings from digital forensic data subsets. Application of the framework to test data resulted in locating additional information relating to the entities contained within the digital forensic data subsets, which led to adding intelligence value relating to the entities. Analysis of real-world data confirmed the potential to add value to big digital forensic data to uncover disparate information and open source information. The results demonstrate the benefits of applying the process to achieve greater understanding of digital forensic data in a timely manner.

1. Introduction

The increasing volume of digital forensic (DF) data has been highlighted over many years [1–3]. Whilst this raises issues for government agencies and private sector organisations (e.g. organisations involved in fraud and misconduct investigations, and electronic discovery) that have not met the demand for service, such as increasing backlogs of work [4], there are positive aspects to the pervasive nature of technology. Big Data is defined as high volume, variety, velocity, and veracity information holdings [5]. One positive is that with the increasing volume, variety, velocity, and veracity of evidence and information available, this can enable investigators, managers, and decision makers to have a greater understanding of the criminal environment, and assist with informed decisions, with a process of adding value to big data.

The role of intelligence is to independently and impartially inform decision makers by providing a clear understanding of issues [6]. Intelligence-led policing (ILP) is a philosophy which includes data analysis and criminal intelligence as pivotal to decision making. ILP originated from developments in information and communications technologies (ICT) and a need for greater management of crime, with a view to disrupt and prevent prolific and serious offenders.

Whilst there is a large volume of data and information available within digital forensic data holdings, the sheer volume, variety, velocity, and veracity of data can be overwhelming to intelligence analysts. This highlights a need for a way to manage or reduce the volume of data to that which is necessary to the task or scope of analysis. Quick and Choo [7,8] developed a method to considerably reduce the volume of digital forensic data. Using both test and real-world datasets, the authors demonstrated that their data reduction method, Data Reduction by Selective Imaging (DRbSI), can reduce the volume of data to 0.2% of the source volume, retaining key evidence and intelligence data [8]. This data reduction process has also enabled much faster processing and analysis of digital forensic

* Corresponding author at: Department of Information Systems and Cyber Security, The University of Texas at San Antonio, San Antonio, TX 78249-0631, USA.
E-mail address: Raymond.choo@fulbrightmail.org (K.-K.R. Choo).

http://dx.doi.org/10.1016/j.future.2016.12.032
0167-739X/

data, i.e. from many hours to minutes to process large volumes of data, using a Quick Analysis process [9].

The DRbSI and Quick Analysis process have potential applications for criminal intelligence purposes. In this research, we explore the ability to use DRbSI subsets and Quick Analysis with Bulk Extractor software [10] to gain an understanding of digital forensic data holdings. From the extracted data, we explore a process of value-adding to the data by drawing on Open Source Intelligence (OSINT) resources. This allows us to add value to the data and information; thus, enabling greater understanding of the criminal environment.

The contributions of this paper are:

- a process of semi-automated scanning of multiple digital forensic data subsets from a variety of devices, including computers, hard drives, mobile phones, portable storage, cloud storage, and Internet-of-Things (IoT) data, with a view to extract entity information; and
- value-add to the entity information by drawing on the resources of OSINT with a view to expanding cross-device and cross-case analysis, leading to greater overall knowledge relating to disparate cases.

In the next section, we outline the background and related work for digital forensic intelligence (DFINT) and open source information. The process of using DRbSI subsets in conjunction with semi-automated analysis to enable entity extraction and open source information searching is then outlined. Following this, we apply the process of DFINT+OSINT analysis to M57 test data to enable an understanding of its application, and then explore the application of the methodology to real-world data. The final sections discuss the research findings, and we then conclude the paper and outline future research opportunities.

2. Background

The role of intelligence is to provide decision makers with independent and impartial information which is timely, accurate, relevant, verifiable, answers a question, and enables proactive decision-making [6]. Criminal intelligence analysis is a term used to describe how information and intelligence can be used in the investigation of crime and persons involved or suspected of being involved in crime [11]. Criminal intelligence encompasses a range of techniques, including analysis of data and information to provide intelligence and understanding of issues and the criminal environment.

2.1. Intelligence-led Policing (ILP)

The process of intelligence analysis is well entrenched within enforcement and other agencies, and includes a concept of "Intelligence led policing", which is defined as; "*The application of criminal intelligence analysis as an objective decision-making tool in order to facilitate crime reduction and prevention through effective policing strategies and external partnership projects drawn from an evidential base.*" [12].

ILP focuses on four elements, namely;

- targeting offenders,
- management of crime hotspots,
- linked crimes and incidents, and
- preventive measures.

Organised crime groups and terrorist organisations often consist of members with antecedents for the predilection of crime, and members often associate with persons with similar criminal background, but also draw on new recruits with limited criminal background. Organised crime was previously associated with the Cosa Nostra, but is nowadays quite different [11]. Organised criminal groups now can have well-developed organisational structures, mainly established for obtaining power and/or wealth. Such groups include outlaw motorcycle gangs, Russian organised crime, Asian organised crime, African organised crime, drug cartels, and street gangs, such as; Asian, Korean, Hispanic, black, and white supremacy groups [11]. It is reported that the complexity of these groups is increasing, with fluid structure-less networks, and increasing cooperation between different organised crime groups and networks [11,13,14].

Organised crime involvement now extends to; trafficking in human beings, drug trafficking, extortion, fraud, murder, and high-technology crime, facilitated with the growth in Internet resources, opening new opportunities for profit. The National Organised Crime Response Plan 2015–18 of the Australian Government estimates that organised crime costs Australia $15 billion per annum in transnational crime, money laundering, identity crime, and the growth in technology facilitates this [15]. The United Nations Office on Drugs and Crime (UNODC) has reported an "escalation of high-technology crime is a challenging and relatively new arena for law enforcement" [11, p. 8], and that organised crime groups are more sophisticated and dynamic than ever before [11]. The UNODC clearly states that "the challenge for law enforcement is to be prepared for this increasing sophistication in order to reduce the impact of criminal activities on our communities" [11, p. 8]. With the increasing volume, variety, velocity, and veracity of digital forensic data, there is an opportunity to focus and learn from the information contained within this data, with an appropriate method to process the data in a timely manner.

The aim of a criminal intelligence analyst is to gather information in relation to criminals and criminal enterprises, and prevent or disrupt crime and criminal activity. The focus should, therefore, be to develop useful sources of information, including details of associates and relationships between individuals and their role in a criminal enterprise [11]. The pervasive nature of ICT has resulted in a vast amount of information on devices, and data transiting from devices via the Internet to be stored in cloud storage environments. The role of digital forensics is to identify, collect, and preserve digital data which may provide assistance in legal enquiries. This is not limited to evidence alone, and there is a role for digital forensics in criminal intelligence analysis, supported with open source and external source intelligence and information.

2.2. Digital forensic intelligence

Digital Forensic Intelligence is the intelligence and information which is available to be distilled from digital forensic data holdings. Weiser, Biros [16] proposed a National Repository of Digital Forensic Intelligence, comprising four aspects; Information Knowledge Base, Best Practice, Tools Index, and a Case Index, with the aim of sharing ideas and methodologies leading to efficiency gains. Their proposal related to a corpus of knowledge for practitioners, but did not address the intelligence information available from within digital forensic data holdings. We propose that digital forensic intelligence should encompass the information potential contained within digital forensic data holdings.

Forensic Intelligence is; "the accurate, timely and useful product of logically processing (analysis of) forensic case data (information) for investigation and/or intelligence purposes" [17]. By applying criminal intelligence analysis methodologies to digital forensic data extracts, there is a potential to gain a greater understanding of computer and mobile phone data, and contribute to a greater knowledge for tactical, operational, and strategic intelligence purposes. This is different to the current main focus of digital forensic practitioners which currently relates to a focus on digital evidence in relation to a crime or as focus of an enquiry.

It can be seen that "digital evidence" is a reactionary analysis process examining what occurred at a previous point in time, whereas "digital forensic intelligence" is a process of forward-looking analysis with a view to predicting what may occur in the future based on the information from previous events.

2.3. Open source intelligence

One of the tools of criminal intelligence is Open Source Intelligence (OSINT). OSINT originates from security and law enforcement agencies, and refers to intelligence derived from publicly available information sources [18]. These sources include global media, web blogs, government reports, satellite pictures, academic papers, Wikipedia, YouTube, and Facebook, and a range of other information via the Internet and other media resources. Open source intelligence is utilised by a range of agencies, including the United States Federal Bureau of Investigation (https://www.fbi.gov/about-us/intelligence/disciplines), United States Central Intelligence Agency (https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html), Royal Canadian Mounted Police (http://www.rcmp-grc.gc.ca/pubs/cc-strategy-strategie-cc-eng.htm), and Europol (https://www.europol.europa.eu/ec3/cyber-intelligence-products).

OSINT has been utilised by many agencies and contributes to strategic, operational, tactical, and technical intelligence needs [6]. OSINT is potentially a cost effective and rapid source of information, and the information and intelligence derived can potentially be shared [18]. OSINT involves information extraction from publicly available sources (e.g. social networking sites). The Internet is now a major source of information, with estimates that data volume will grow from 4.4 zettabytes (ZB) in 2013 to 44 ZB by 2020, doubling in size every two years [19]. The digitalised society we are witnessing has led to the globalisation of commerce, and opportunities for the globalisation of crime. This is enabled by the ability to easily travel across borders and transit international distances with relative ease.

This growth in data requires software which can provide for rapid content discovery, search, and retrieval. The Internet itself is not a source, but the means to access information sources, which can include media monitoring services, specialist information sources, 'grey literature' such as academic papers, and satellite imagery [6]. Care must be taken with OSINT as information in the public domain is not necessarily verified and may be biased or inaccurate [11]. Identification of sources of information is an ongoing process, as different means of communication go through a cycle of popularity. Search engines such as Google, which enable searching web sources by crawling and indexing hosted content, are slow, but effective. Web mining tools that focus on specific sources can provide alerts to keyword terms when changes are made, or matches are located.

OSINT can be fast, flexible, dynamic, communicable, shareable, partner forming, can encompass rapid evaluation or in-depth analysis at strategic, operational, tactical, and technical levels, and identify and mitigate risk, i.e. from 'horizon scanning' to sophisticated targeting [6]. One challenge with OSINT is the data source language and the need to translate the language, which may necessitate the use of a translation service, training for analysts, or the use of translation software [11].

Today's unparallelled access to global satellite data, coupled with Google street view pictures of a large percentage of populated environments, has provided for vast amounts of data and information. As this information is not classified, there are fewer barriers to information sharing, although the intelligence derived from OSINT may need to be classified. It is quite clear that data available via the Internet is potentially and currently the greatest data source ever available, and with the rapid rate that data is doubling, this is anticipated to only increase in future. The vast amount of data and information can assist with providing actionable intelligence to decision makers.

It is apparent that in today's connected world, with a vast volume of data a mere click of a mouse away, that criminals and organised crime groups will utilise new and sophisticated methods of communication. Hence, agencies tasked to investigate these groups have a need to gather information about activities undertaken in the digital realm, or with a digital footprint. Furthermore, with the growing number and storage capacity of mobile phones, computers, and portable devices increasing dramatically, there is a vast pool of information in relation to criminal activity.

The FBI has applied the process of criminal profiling to the digital offending realm, developing cyber-criminal profiles [20]. However, this may not necessarily take into account the information available from victims' computer and devices, or the potential to develop psychological profiles of a wider range of crimes using the digital intelligence available from computers and devices, which when analysed can reveal a lot of information about the user.

A process of forensic feature extraction and cross drive analysis was first proposed in [21] and has been refined over the years to develop software to assist with processing and extracting entity information from digital media storage. The next step of this is to build on the extracted entity information, and again, an automated process is necessary, given the sheer volume of data and entities resulting from even small hard drive storage. Research in relation to the development of a method to value-add to the entity and information extracted is necessary, as is a method which is applicable in the current environment involving real-world cases and real-world data volume.

Beebe [22] emphasised the need for research in crime network identification. The author also highlighted the importance of applying analysis techniques to digital investigations, to which we undertook in [9] and this paper continues this research focus by applying link analysis methods to digital forensic data and expanding on this with open source and external source information.

In Weiser, Biros [16] a repository of information is proposed to store information relating to digital forensic cases, to build a knowledge base of cases including a case tracking system that stores forensic discoveries related to a case, an expert system record of best practice, and certified tools index. This aspect of digital forensic intelligence appears to relate to a higher level focus to build a knowledge base, rather than a focus on the use of intelligence analysis techniques to assist with the process of digital forensic analysis. In our research, we focus on the merging of digital forensic analysis and intelligence analysis techniques to add value to the process of both digital forensics and intelligence analysis. A national repository knowledge base for digital forensic practitioners is an admirable goal, and with the appropriate security and controls, could be expanded to include intelligence extracted from digital forensic data, with the adoption of a suitable method of data reduction, such as that proposed in [7,8].

2.4. Digital forensic intelligence + open source intelligence

With the vast quantity of data available from digital forensic analysis, there is a wealth of information which can be potentially improved with open source intelligence data to enable a greater understanding of events or persons, and greater decision making opportunities. One issue affecting rapid and timely analysis of large volumes of digital forensic data is the growth in media volume and the associated increase in the time to undertake searches and review information. The recent development of a method of data reduction has enabled faster collection and processing of digital forensic data. In our previous research, the time from collection to analysis was reduced from many hours to minutes

using both research and real world datasets [7,8]. This process of rapid extraction and analysis enables the processing of large volumes of digital forensic data in a timely manner.

In the next section, we outline our proposed framework to enable the inclusion of open source information with digital forensic analysis to encompass building on the data that is extracted through digital forensic analysis, such as computer hard drives, mobile phones, tablets, media storage devices, IoT devices, and cloud stored data. Using our previously published data reduction process [8], Data Reduction by Selective Imaging (DRbSI), and semi-automated entity information extraction (Bulk Extractor), we explore a process of using OSINT to expand knowledge and intelligence from large volumes of data by a process of

- digital forensic data reduction [8],
- semi-automated entity extraction [9], and
- external source OSINT searches.

In the proposed process for digital forensic and open source analysis, we seek to enhance the information from a wide scope of devices and information storage. This is to enable those involved in an investigation or intelligence probe to make decisions with as much relevant knowledge as is available, in a timely manner using our proposed digital forensic intelligence and open source information (DFINT+OSINT) framework.

3. Methods

The proposed framework for the process of digital forensic intelligence and open source intelligence (DFINT+OSINT) is based on the digital forensic intelligence analysis cycle [23] and the digital forensic data reduction methodology [8]. The proposed DFINT+OSINT framework (Fig. 1) is outlined in this section.

When working with open source data for intelligence purposes, consideration should be made regarding the following points for each source: authority, accuracy, objectivity, currency, and coverage [6]. The use of a rating system for intelligence and sources is also appropriate, such as the 4 × 4 system or 6 × 6 system (admiralty scale) [11].

As with any investigation or intelligence probe, timely and accurate notes of the steps undertaken should be maintained, and any information and evidence secured according to agency directions or industry best-practice [24–26] Legislation and legal authority must be checked and appropriate prior to commencing, and ensure compliance at all stages.

When working online, practitioners are advised to ensure identity protection measures are appropriate, and network security is paramount when dealing with source access via open Internet connections. Covert operations need to ensure the appropriate approvals and security is in place prior to commencing, as it potentially only takes one miss-step to jeopardise an operation or personnel involved, which in some cases may be life-threatening consequences.

Digital forensic analysis follows a well-established framework, namely: identification, preservation, analysis, presentation [27]. Intelligence analysis involves a similar process of collection, collation, analysis, and presentation. The Digital Forensic Intelligence Analysis Cycle (DFIAC) is a merger of the methodologies of digital forensics and intelligence analysis to form a process of Commence, Prepare, Evaluate and Identify, Collect, Preserve, Collate, Analyse, Inference Development, Presentation, Completion or Further Tasks Identified [9].

Using the DFIAC to frame our proposed process (see Fig. 1), we outline DFINT+OSINT as follows:

Commence (**Scope/Tasking**): the focus, aims, and scope of analysis are outlined to enable preparation and guidance for the overall examination. This could be done with a written tasking request, or verbal request, which should be written down in agency specific record keeping format. As all agencies differ in the methods of recording taskings, we will not elaborate on this, but practitioners should adhere to their agencies intelligence and/or digital forensic record keeping requirements.

Prepare: gather the anticipated equipment required and expertise, including confirmation of legal authority, network security, and covert considerations.

Identify and collect: identify electronic devices with potential evidence and/or intelligence, and undertake appropriate physical examination and documentation, according to agency policy and procedures.

Data Reduction by Selective Imaging (DRbSI): a digital forensic subset of data is collected from the identified device or media. This involves utilising the DRbSI process outlined by Quick and Choo [8], which is summarised as; connect the identified media through a write-blocked mechanism to a PC, use forensic software such as EnCase or X-Ways to run a filter for pre-identified files and data, preserve the selected data in a logical forensic container (L01 or CTR). As necessary, reduce the dimension of pictures, and thumbnail video data and include these in the logical container. Export a list of all files on the media, and include this in the logical forensic container (e.g. L01, AD1, or CTR). See Quick and Choo [9] for further information on the DRbSI process.

As an example, we previously used test data from the M57 corpus [28] for digital forensic data reduction research [8]. We used Encase 6.19.7 forensic software to access each of the forensic image files (E01) and ran a filter process for a range of files using conditions to filter on file extension and file type. We used XnView software to reduce the dimensions of the picture files, and "mtn" software was used to create thumbnail images of the video files. Using EnCase 6.19.7 spreadsheet listing of all files and data was created, and these were stored in logical forensic containers (L01). This was done for the various forensic image (E01) files within the M57 corpus, resulting in a smaller DRbSI subset corpus of relevant data able to be searched in a timely and rapid manner. The original source data volume was 498 GB and this was reduced to 4.25 GB in DRbSI data subsets, representing a reduction to 0.85% of the source volume.

Quick Analysis and Entity Extraction: A process of Quick Analysis and Entity Extraction is outlined in Quick and Choo [9]. The entity extraction process involves using software (e.g. Internet Evidence Finder, RegRipper, Bulk Extractor, NetAnalysis, and others as required) to process the various data types retained within the DRbSI subsets and merge the output into a single source of information. The use of software such as Bulk Extractor, assists with extracting a range of key entity information from the DRbSI subsets.

In our previous research [9], we used the DRbSI subsets of test data [8] from the M57 corpus [28] to demonstrate a process of quick analysis and entity extraction. A range of forensic software was used to undertake analysis of the data subsets and determine individuals and communication of relevance to an investigation or intelligence probe (including Internet Evidence Finder, RegRipper, Bulk Extractor, NetAnalysis, and others). We also used Bulk Extractor software to scan and locate entity information from within the DRbSI forensic subsets for the various devices in the M57 corpus. The extracted data and information was loaded into Pajek64 software to create an entity link chart of the persons and data identified, which assisted to highlight the interlinked nature of the data (Fig. 2).

In this we could see that there were a number of linkages between the various entities, including Jasper, Pat, Charlie, Terry, Jo, Jean, Cod, and Aaron. To enable an understanding of the Pajek64 chart, we undertook analysis of the Pajek64 linkages to determine

Digital Forensic Intelligence and Open Source Intelligence (DFINT+OSINT) Framework

Commence	Outline focus and scope of investigation.
Prepare	Ensure correct equipment and expertise.
Identify and Collect	Identify devices with potential evidence/intelligence, photograph and document
Data Reduction by Selective Imaging (DRbSI)	Load forensic image, or connect write-blocked physical device. • Filter and select pre-identified DRbSI files; File and OS, Software, Internet History, User created Documents, Email, Pictures, Video, etc. • Deselect overwritten data and non-relevant files [if necessary, mount forensic image as logical drive/s] • Convert Video files to thumbnail/s (e.g. using mtn batch file) • Shrink Picture files [1024 or 800 pixels wide/high] and place into L01 • Export a File List to CSV and convert to XLSX or compress to ZIP Export selected files to L01 or other forensic container
Quick Analysis and Entity Extraction	Undertake analysis of the DRbSI Subset, processing the data holdings with various tools to extract data and information, such as; RegRipper for Registry Information, File Metadata, Bulk_Extractor, Internet Evidence Finder, NetAnalysis, etc.
Open Source Intelligence (OSINT)	Process and load the entity information into software, such as Maltego and undertake web searches using Zotero to capture, collate, and store data. Search other OSINT and External Source resources to gain a greater understanding of the Entity and extracted Digital Forensic data
Entity Chart	Process and Load the Entity information into entity charting software, such as Maltego or Pajek64
Inference Development	Develop hypotheses and inferences based on the information findings, producing intelligence product
Presentation	Produce written and/or verbal report of hypotheses, inferences, and findings
Complete	Finalise the matter, or identify further avenues of enquiry

Fig. 1. Proposed digital forensic and open source intelligence (DFINT+OSINT) framework.

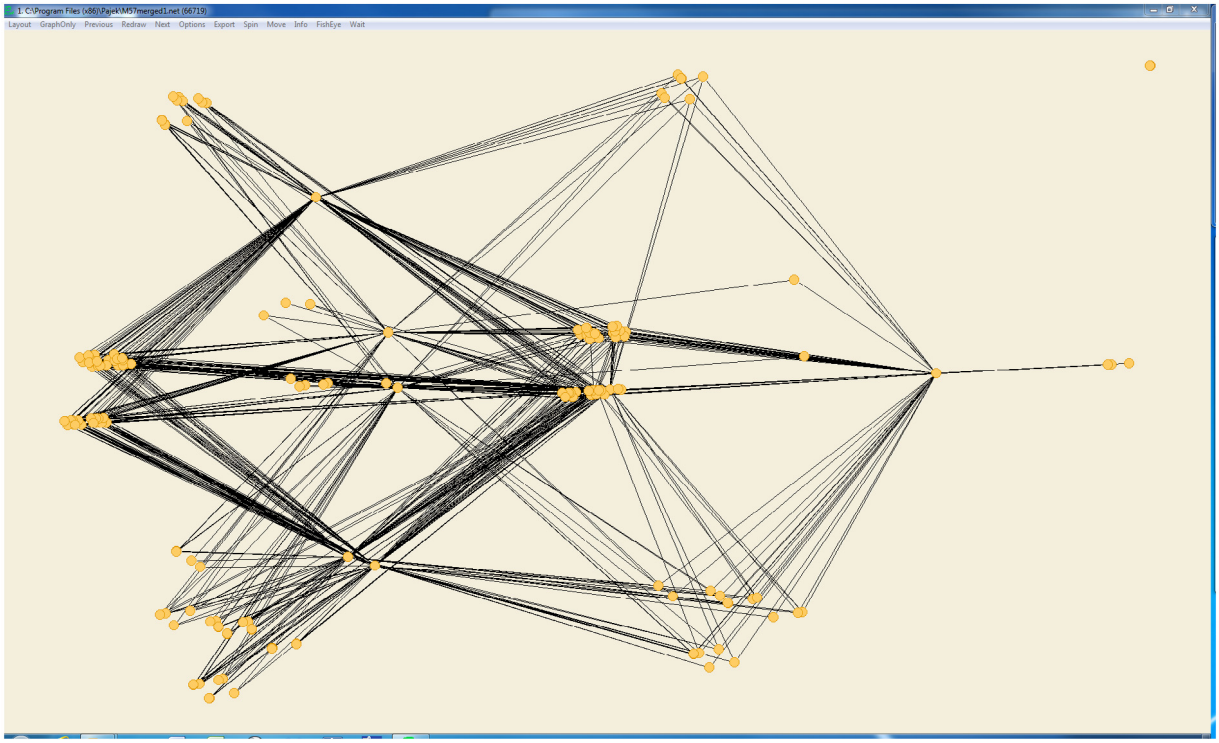

Fig. 2. Pajek64 entity link chart from M57 corpus.

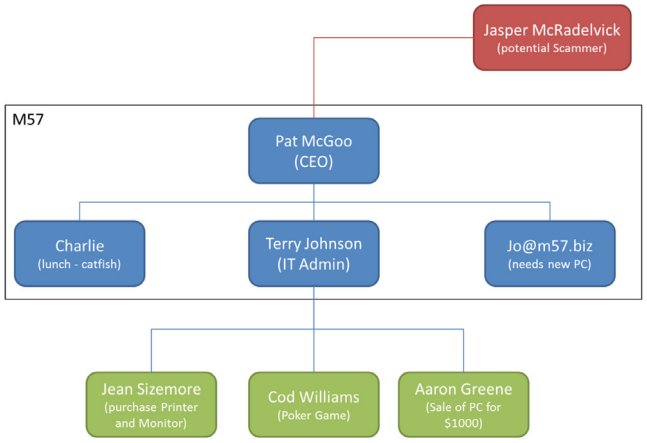

Fig. 3. Relationship chart for main entities in m57.biz [9].

the nature of the information, and produced an entity link chart to highlight the relationships (Fig. 3).

OSINT: using the extracted entity information, identify further sources of data and potential evidence and intelligence, such as: data which may be on social media websites, global or local media reports, 'grey' literature, satellite images, Google Street view information, or other open source media. Once data or information source is identified, conduct searches for known relevant entities; names, addresses, email addresses, phone numbers, vehicle details, and other information pointers.

To summarise a guide to OSINT analysis from [29] as follows:

- Choose an effective search tool, (in our research we use Maltego CE to automate the searches)
- Use extended search capabilities (GREP), (using Maltego CE or Google, etc.)
- Search deep web resources, such as; databases, electoral roles, telephone, business databases, (using Maltego CE or Google, etc.)
- Review linkedin, facebook, twitter, youtube, flickr, Instagram, photobucket, blogs, tripod, online sales sites, such as; ebay, gumtree, craigslist, whirlpool, (using Maltego CE or Google, etc.)
- Run WHOIS searches for domain names, (using Maltego CE or Google, etc.)
- IP addresses (extracted from Registry and Internet History), genealogy sites, maps, traceroute, wayback machine, review source HTML, trace email headers, (using Maltego CE or Google, etc.)
- Search for names, usernames, account names, email, phone numbers, addresses, family members, friends, associates, image EXIF data, GEO DATA, etc., (using Maltego CE)
- Use zotero to collect, collate, and store data as you go (we use Maltego CE for this purpose)
- Use snipping tool or print to pdf (e.g. the in-built Windows Snipping Tool or CutePDF Writer).

Contemporaneous notes should be made of the searches conducted, including; keyword search terms used, which websites were examined, any email references, personal information, associates, online images, chat, social network, further avenues of enquiry, time and date, and importantly, legal authority to undertake the analysis.

If relevant data is located, then it should be forensically handled where possible, i.e. saved, printed to PDF, use screen capture (Microsoft Windows Snipping Tool), or screen capture software, to preserve the data. The data preservation can align with a data collection and reduction process, such as that of [8] to collate a subset of relevant data. Further in-depth analysis of websites may be necessary, which may locate additional information not normally presented within a browser window, i.e. within HTML source code for websites. The use of software, such as Maltego, may assist with bulk data analysis and retrieval.

If during the process of analysis, additional sources of data are identified (e.g. other social media websites, media reports, or cloud stored data), the process forks to the 'Preparation' stage to collect the new data, whilst the analysis progresses.

In the next section, we will outline the process of using the extracted entity information from the M57 corpus (as outlined above in Figs. 2 and 3) to undertake OSINT searches to build on the information and intelligence in relation to the persons associated with the investigation or intelligence probe. We utilise Maltego CE software to store and chart the extracted data and undertake open source information searches using Maltego, with the available in-built search functions.

Entity Chart: the identified and collected OSINT data is then included with other data subsets and data extracted from computers, mobile phones, Internet of Things (IoT) devices, including data from cloud stored providers, for examination for evidence or intelligence, with a focus of that of the scope of the task [30]. The use of charting software, such as Maltego CE, IBM I2, or Pajek64 is recommended to automate the process of linking the extracted entities, which when considered in relation to cross case and cross device analysis can result in many hundreds or thousands of entities and linkages, which would be too time consuming to manually process and chart.

In the following section, we will outline the process of using the extracted entity information from the M57 corpus (as outlined above in Figs. 2 and 3) and Maltego CE software to create entity interlink charts of the information and intelligence in relation to the persons associated with the investigation or intelligence probe.

Inference Development: with the knowledge gained during the process of collation and analysis, ideas are formed in relation to the questions of who, how, what, when, why, and where. The gained knowledge is used to form inferences about the investigation or intelligence probe to answer questions or outline findings. The intelligence and source data can also be rated using a rating system, such as the 4×4 system or 6×6 systems [11].

Presentation: the findings of the overall process of analysis are formed into a report (written and/or verbal) which is communicated to the requesting persons involved in the investigation, legal process or probe.

Complete: the matter is finalised, and data archived according to agency practice and procedures. If further tasks are identified, the process continues in the cycle until complete.

4. Results

4.1. Digital intelligence and OSINT from M57 test data

The growth in the volume and number of devices encountered in investigations has resulted in a need to collect and analyse growing volumes of digital forensic data in a variety of formats. We previously used test data from the M57 corpus [28] for digital forensic data reduction research, where we explored a data reduction method and successfully reduced the volume of data to 0.2% using our DRbSI process [8]. We then outlined a process of Quick Analysis [30] to distil relevant information from the digital forensic data subsets. Subsequent to this, we merged the digital forensic data from the test data (M57) computers, portable storage, mobile phones, and tablet devices, with source data comprising approximately 498 GB. This was successfully reduced to 4.25 Gb of DRbSI subsets and logical container files encompassing potentially relevant information.

We applied a bulk data analysis process to the data subsets, using semi-automated entity extraction from the forensic subsets using Bulk Extractor 1.5.5 software [10]. This scanned the DRbSI data subsets, and the output encompassed 2.02 GB, comprising 23,496 email features, and 22,962 picture files, in approximately 30 min. Mobile phone extracts from 41 mobile devices, comprising 207 MB, were merged into a single source file for analysis [9].

The data output from the Quick Analysis process [30] which included parsing the Windows Registry files, Internet Evidence Finder, NetAnalysis software, and information extracted from a variety of data sources within the DRbSI subsets, were merged with the output from Bulk Extractor, along with the previously merged mobile phone extracted data, resulting in a very large file of extracted entity information with associated source and relationship links. This was loaded into Pajek64 software, with the resulting entity chart shown in Fig. 2.

This process has encompassed the first five stages of the DFINT+OSINT framework, and we now move to the next stage, by using the extracted single-source entity information with Maltego CE to explore the process of expanding our knowledge of the persons and entities contained within the data by locating available OSINT relative to the M57 test data. We loaded the extracted entity information from the test data into Maltego CE, and the output is displayed in Fig. 4. This link analysis chart shows the interlinked nature of the data and entities identified within the M57 corpus, which is consistent with the Pajek64 chart in Fig. 2 which was summarised into the entity chart in Fig. 3.

We then conducted Maltego Transform searches of the entities within the link chart. This resulted in additional data matches, expanding our knowledge in relation to the entities in the test data, including associated URL locations with references to the email addresses and entities contained within the M57 corpus. The resulting data is displayed in Fig. 5, highlighting the URL references.

A selection of the URL matches located using open source intelligence analysis is listed in Table 1. This highlights the additional information we were able to locate in relation to the email and entity information contained within the DRbSI subsets and extracted using the Quick Analysis and Bulk Extractor processes. Without undertaking open source intelligence techniques we would not be aware of this information. In Table 1, we highlighted the entity associated with the URL which led us to the additional information.

In a rapid and timely manner, we were able to add-value to the information in relation to the M57 test data, expanding our knowledge-base with information available from open source information. Using this process, we have expanded out digital forensic intelligence with open source information, resulting in DFINT+OSINT. Whilst this increased information and knowledge-gain is of benefit in research, more importantly, this type of information and intelligence building can greatly assist in real world investigations.

4.2. Applying DFINT+OSINT to real world data

From our prior research [7,8], the entity information extracted from the M57 test data is similar to that which is extracted from real world data. The volume in real world data holdings is often much larger, which can result in longer search times. However, this has greater information potential.

There is also an opportunity to develop a method of refining the data and entities to that which relates to a case, and exclude or filter out generic data that exists in many operating systems and software installations, such as Microsoft Windows URL and email links for help and assistance. There is potentially a large volume of entities which can be excluded as these are unlikely to be related to an investigation. Undertaking a bulk extraction of a newly installed operating system and software and using this as a source of 'known-good' entities, which can then be used to remove these 'known-good' entities from real world data, would enable a further reduction in the number of entities for OSINT research. The use of the National Software Reference Library (NSRL) hash databases is also of potential benefit to further reduce the volume of entities extracted.

With real world data, we can also consider data from IoT devices such as fitness bands which record GPS locations of the wearer and upload this to cloud storage, security systems such as smart door locks which record biometric information as a person enters or exits a smart home, or wireless internet connected doorbell systems with video recording capability which record movement of persons near the device. Data from these systems can be extracted and merged with the digital forensic data for analysis, and provides for even more information for investigators and analysts to examine. Hence, a process of data reduction, quick analysis, and external source intelligence can be beneficial to those involved in an investigation to understand the context of the information available.

By adding value to the information contained within digital forensic data, there is an opportunity to explore cross-case intelligence analysis, which in real-world cases may highlight cross-case linkages which were previously unknown to investigators. Indeed, the first author has experience where case linkages were unknown to disparate investigations due to the different focus of the various investigations, i.e. drug importation investigations and local service area property crime offending, and when the cross-case linkages have been brought to the attention of the separated investigation teams, this has enabled a greater understanding of the volume of disparate but actually connected offending.

By implementing a method of building a knowledge base of cases, such as that proposed by Weiser et al. [16], there is an opportunity to assist with disparate cross-case linkages being discovered early enough in an investigation to ensure appropriate resourcing of investigations, with a potential for a more timely resolution of cases. This must be balanced with appropriate security of the information, and legal authority to access and review the data.

In our previous research, we examined the metadata contained within archived cases when we were afforded limited access to South Australia Police Electronic Evidence Section data backup archives [7,8]. We did not examine the contents of case data, and only reviewed the times and data from processing of limited archived meta-data.

One archived case we examined comprised 18 computers, laptops, portable storage, mobile phones, and tablet devices totalling 2.7 TB of source data. We reduced the volume of data according to the process outlined in [7,8], resulting in 46.1 GB of DRbSI subsets. Full imaging took approximately 42 h, and the DRbSi process took less than 4 h. To test the use of a semi-automated analysis process, the subsets were batch processed with Bulk Extractor 1.5.5 software, processing in 1 h 19 min. In comparison, processing the original source data took 43 h and 11 min. As per the process undertaken with the M57 data, it would be possible to merge the output entity information for further analysis with Maltego and adding value to the information with OSINT data analysis (this process was not undertaken with this data).

In addition, we loaded the DRbSI subsets from 544 archived devices into NUIX 6.2.3, EnCase 6.19.7, and EnCase 7.10.5. Again, the data was not viewed, rather, the times for processing was noted. EnCase 6.19.7 took approximately three [3] minutes to load and open the 544 L01 files. File signature analysis was run, and took 2 h and 8 min. Over 10 million files were presented for analysis, including 907,015 documents, 52,742 emails, 2,221,521 picture files, and 2333 container files. Within this data was potentially relevant entity information which could be improved with OSINT data.

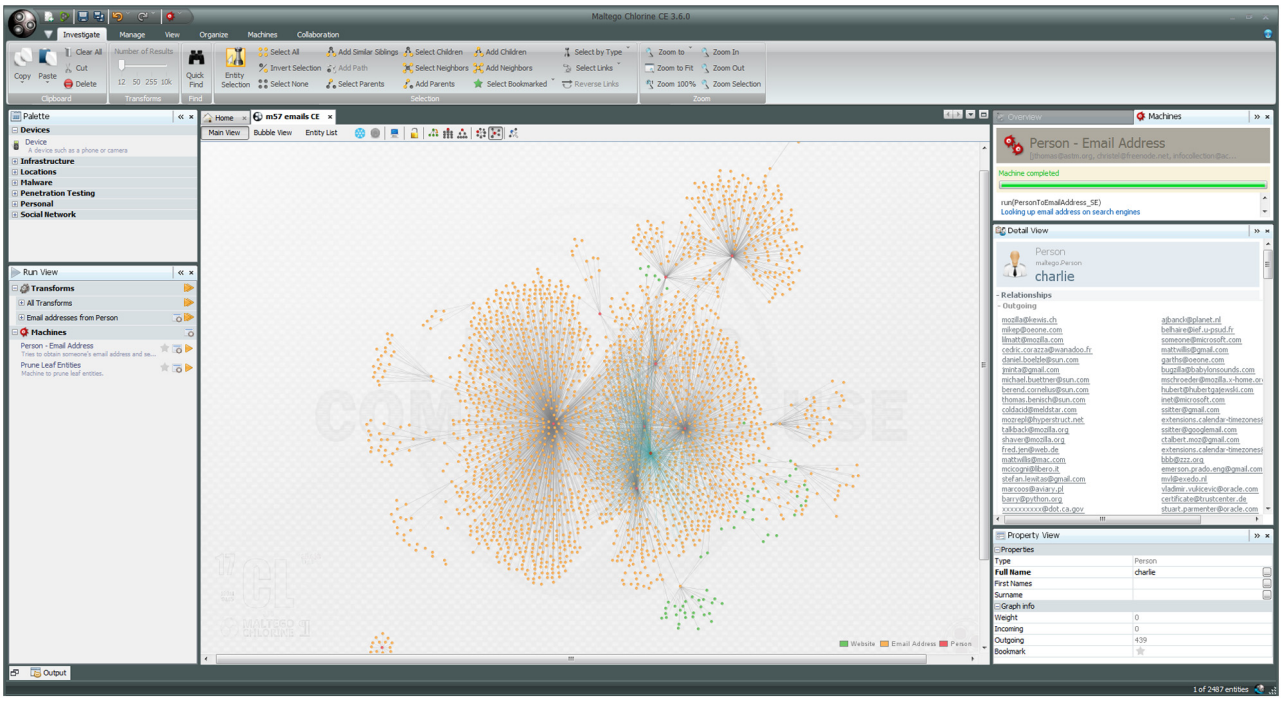

Fig. 4. Test data loaded into Maltego CE.

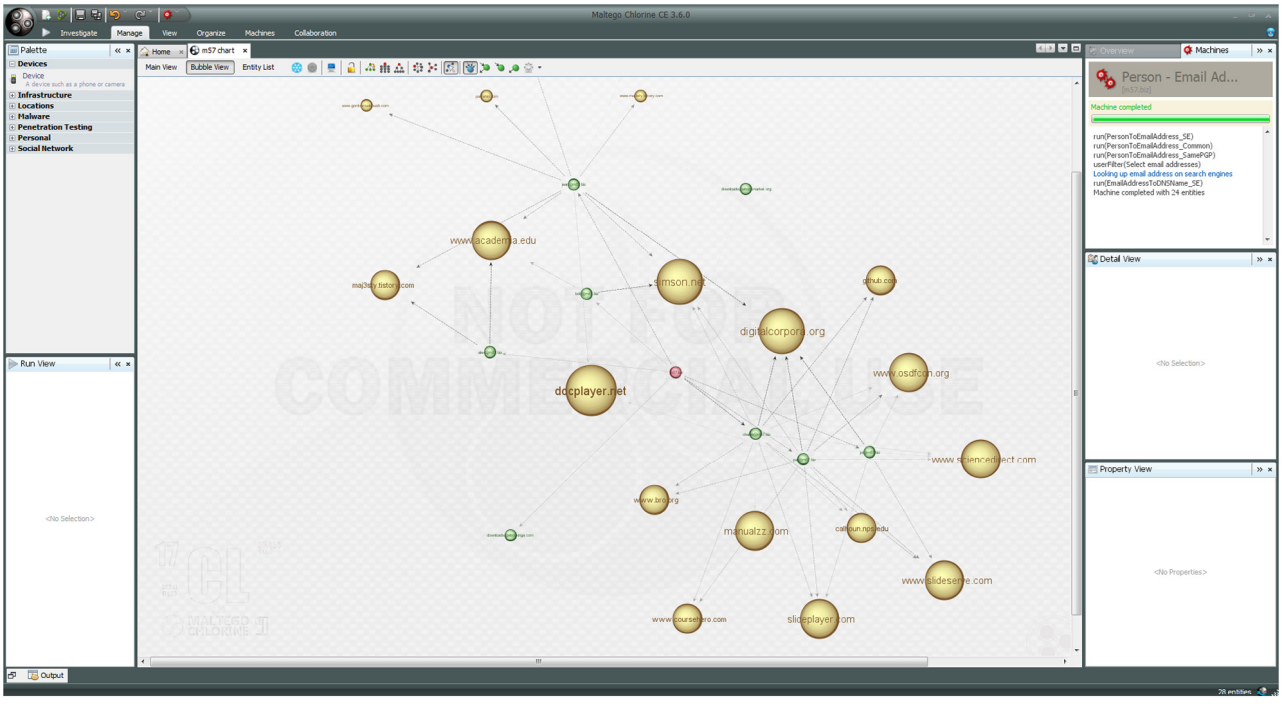

Fig. 5. OSINT URL references displayed in Maltego CE.

Table 1
OSINT URL references located with Maltego CE.

Name	URL
Jean	http://maj3sty.tistory.com/1034
Jean	http://maj3sty.tistory.com/category/%5B+%5D%20Forensic?page=8
Jean	http://www.doc88.com/p-1836912450568.html
Charlie	http://simson.net/ref/2012/2012-08-08%20bulk_extractor%20Tutorial.pdf
Alison	http://www.tuicool.com/articles/eiYNzuU
Jo	http://www.osdfcon.org/presentations/2015/McCarrin-Allen_osdfcon.pdf
Pat	http://digitalcorpora.org/downloads/bulk_extractor/BEUsersManual.pdf

The subsets when loaded into NUIX 6.2.3, provided for metadata analysis, which included strategic intelligence analysis to identify the types of phones presented for analysis, identifying that mainly iPhone and Samsung mobile phones were present. This highlighted that research into the device storage of these devices is warranted as they appear to be quite popular in relation to other devices. Further device specific information was not viewed. In addition, a focus on the JPEG EXIF metadata highlighted that Panasonic, Nikon, and Canon camera identifiers were present (further analysis was not undertaken on this data).

This research demonstrated a capability to process DRbSI subset L01 files from real-world devices ranging from USB storage to multi-terabyte hard drives, and it was possible to load and process these with EnCase 6.19.7, EnCase 7.10.5 and NUIX 6.2.3, and a potential to conduct further analysis of the data using Bulk Extractor across the subsets which could be further enhanced with Maltego to locate any open source information relating to the entities. Further analysis of the data was not undertaken due to privacy considerations.

5. Discussion

As outlined, our literature review highlighted a need for further research in the use of intelligence analysis techniques with digital forensic data. We collected a corpus of test data to undertake experiments, and were able to determine a method to undertake in-depth analysis to value-add to the entity information present in digital forensic case data. We described this in our proposed DFINT+OSINT framework. We applied aspects of this framework to real world data, which indicated the potential application to real world cases. In this manner, we have enhanced intelligence from DFINT+OSINT data.

Criminal intelligence is described as a national asset, which should be; "collected once and used often for the benefit of many and therefore adds value to the decision-making process" [31, p. 62]. This principle is also applicable to digital forensic data, and the analysis of disparate case data can have benefits to society in solving and progressing cases in a timely manner. As IoT devices, computers, portable storage, mobile phones, and tablet devices become more pervasive throughout society, there will be a growing need for forensic analysis of these devices. With the growing volume of disparate data, there is a need to be able to undertake analysis on growing volumes of structured and unstructured data. The method outlined in the previous sections as applied to test data (M57) and real world data, demonstrated an ability to undertake analysis of a large volume of disparate data, and locate potential evidence and intelligence.

In our experiments, it was possible to scan data-reduced subsets in a semi-automated manner, and then merge the output to enable the examination of a large volume of data in a timely manner for linkages across devices and cases. As more and more devices are seized and presented for digital forensic analysis, there will be a larger source of data for criminal intelligence analysis, potentially locating evidence and intelligence to enable investigators and decision makers a greater understanding of events from large volumes of data. Strategic and management level information can be drawn from digital forensic data; operational knowledge can be located and provided to investigators and managers, including information relating to crime trends. Tactical, target specific information can also be located and communicated in a timely manner.

By enhancing the information contained within digital forensic data by undertaking semi-automated analysis of entity information with open source data sources and information, there is an opportunity to fast-track investigations, and locate disparate linkages, which may otherwise remain unknown.

The DFINT+OSINT process utilises the Digital Forensic Data Reduction Framework [7], which comprises a process of Data Reduction by Selective Imaging [8], Quick Analysis and Entity Extraction [9] merging the output from Mobile Device Analysis [23] and Cross Case and Cross Device processing [32].

6. Concluding remarks

As technology continues to become more pervasive and prevalent throughout society, there will be an associated growth in the use of these devices by criminals and for criminal purposes. This has already impacted on demand for digital forensic analysis of these devices, and the data they generate. This has highlighted a need for improved analysis techniques, and by drawing on methods used in criminal intelligence analysis, we have proposed a framework for enhanced analysis of digital forensic data. This includes reducing the volume of data to that which is necessary to achieve the goal of analysis, semi-automated processing of the data to locate entity and associated information, and automated searching of the entity information with other source data, including open source resources, to improve the value of the data. This is further enhanced by utilising cross-case link analysis methods, as outlined in the framework.

In experiments with our test data corpus, we were able to reduce the volume of data, process the data to extract entities, and then search the entity information with open source information to add value to the data in a timely manner. We applied aspects of our framework to real-world data which demonstrated the potential to apply the framework to real-world data, using currently available software and tools to achieve the goals of value-adding to the data, and improve and assist with timely analysis of digital forensic data.

Future research opportunities include building and deploying a list of 'known-good' entities from standard operating system and software installations to further refine the entities extracted to those that are more specifically case-related. This refining process can also draw on the NSRL reference libraries.

By implementing a digital forensic intelligence process at an organisation level, in conjunction with value-adding to the entity data, there is an opportunity to have the data to potentially answer the questions of strategic, tactical, and operational intelligence and investigational needs. Applying the Data Reduction by Selective Imaging (DRbSI) [8] process, in conjunction with Quick Analysis [9] and the DFINT+OSINT framework can assist with achieving the scope or goals of investigation and intelligence in a timely manner.

Acknowledgements

The views and opinions expressed in this article are those of the authors alone and not the organisations with whom the authors are or have been associated or supported. The authors also thank the editor and the anonymous reviewers for their constructive feedback.

References

[1] S. Garfinkel, Digital forensics research: The next 10 years, Digit. Investig. 7 (Supplement(0)) (2010) S64–S73.
[2] S. Raghavan, Digital forensic research: current state of the art, CSI Trans. ICT 1 (1) (2013) 91–114.
[3] D. Quick, K-KR. Choo, Impacts of increasing volume of digital forensic data: A survey and future research challenges, Digit. Investig. 11 (4) (2014) 273–294.
[4] H. Parsonage, Computer Forensics Case Assessment and Triage - some ideas for discussion. 2009 [updated 2009; cited 2013 4 August]; Available from: http://computerforensics.parsonage.co.uk/triage/triage.htm.
[5] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, P. Tufano, Analytics: The real-world use of big data (IBM institute for business value-executive report), IBM Inst. Bus. Value (2012).
[6] S. Gibson, Open source intelligence: An intelligence lifeline, RUSI J. 149 (1) (2004) 16–22.

[7] D. Quick, K-KR. Choo, Data reduction and data mining framework for digital forensic evidence: Storage, intelligence, review and archive, Trends Issues Crime Criminal Justice 480 (2014) 1–11.
[8] D. Quick, K.-K.R. Choo, Big forensic data reduction: digital forensic images and electronic evidence, Cluster Comput. 19 (2) (2016) 723–740.
[9] D. Quick, K.-K.R. Choo, Big forensic data management in heterogeneous distributed systems in smart cities: Quick analysis of multimedia forensic data, Softw.: Pract. Exp. (2017) in press. http://dx.doi.org/10.1002/spe.2429.
[10] S. Garfinkel, Digital media triage with bulk data analysis and bulk_extractor, Comput. Secur. 32 (2013) 56–72.
[11] UNODC. United Nations Office on Drugs and Crime - Criminal Intelligence Manual for Analysts. Vienna, Austria: United Nations, New York, Vol. 8, 2011.
[12] J. Ratcliffe, Intelligence-led policing, in: Trends Issues Crime Criminal Justice, Australian Institute of Criminology, 2008.
[13] K.-K.R. Choo, Organised crime groups in cyberspace: a typology, Trends Organized Crime 11 (3) (2008) 270–295.
[14] K.-K.R. Choo, R.G. Smith, Criminal exploitation of online systems by organised crime groups, Asian J. Criminol. 3 (1) (2008) 37–59.
[15] A. Government, National Organised Crime Response Plan 2015-2018. Australia; 2015 [updated 2015; cited]; Available from: https://www.ag.gov.au/CrimeAndCorruption/OrganisedCrime/Documents/NationalOrganisedCrimeResponsePlan2015-18.pdf.
[16] M. Weiser, D.P. Biros, and G. Mosier (Eds.), Development of a national repository of digital forensic intelligence, in: Proceedings of the Conference on Digital Forensics, Security and Law. Association of Digital Forensics, Security and Law, 2006.
[17] O. Ribaux, S. Walsh, P. Margot, The contribution of forensic science to crime analysis and investigation: Forensic intelligence, Forensic Sci. Int. 156 (2–3) (2006) 171–181.
[18] C. Best, Open source intelligence, Min. Massive Data Sets Secur.: Adv. Data Min. Search Soc. Netw. Text Min. Appl. Secur. 19 (2008) 331–344.
[19] IDC. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. EMC Corporation; 2014 [updated 2014; cited 2016 1 June]; Available from: http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm.
[20] M. Rogers, The role of criminal profiling in the computer forensics process, Comput. Secur. 22 (4) (2003) 292–298.
[21] S. Garfinkel, Forensic feature extraction and cross-drive analysis, Digit. Investig. 3 (Supplement(0)) (2006) 71–81.
[22] N. Beebe, Digital Forensic Research: The Good, the Bad and the Unaddressed, in: (Advances in Digital Forensics), Springer, 2009, pp. 17–36.
[23] D. Quick, K.-K.R. Choo, Pervasive social networking forensic intelligence and evidence from mobile device extracts, J. Netw. Comput. Appl. (2017) in press. http://dx.doi.org/10.1016/j.jnca.2016.11.018.
[24] ACPO. Good Practice Guidelines for Computer Based Evidence v4.0. Association of Chief Police Officers 2006 [updated 2006; cited 5 March 2014]; Available from: www.7safe.com/electronic_evidence.
[25] NIJ. Electronic Crime Scene Investigation: A Guide for First Responders, Second Edition. 2008 [updated 2008; cited]; Available from: http://www.nij.gov/pubs-sum/219941.htm.
[26] NIJ. Forensic Examination of Digital Evidence: A Guide for Law Enforcement. 2004 [updated 2004; cited]; Available from: http://nij.gov/nij/pubs-sum/199408.htm.
[27] R. McKemmish, What is forensic computing? in: Trends Issues Crime Criminal Justice, Australian Institute of Criminology, 1999, pp. 1–6.
[28] S. Garfinkel, Roussev Farrell, Dinolt, Bringing Science to Digital Forensics with Standardized Forensic Corpora, DFRWS 2009, Montreal, Canada. Montreal, Canada: DFRWS 2009; 2009 [updated 2009; cited 2013 9 September]; Available from: http://digitalcorpora.org/corpora/disk-images.
[29] Toddington_International. Online Investigator's Checklist. Toddington International Inc.; 2016 [updated 2016; cited 2016 7 July]; Available from: https://1x7meb3bmahktmrx39tuiync-wpengine.netdna-ssl.com/wp-content/uploads/TII_Online-Investigators-Checklist_v2-1.pdf.
[30] O. Ribaux, A. Baylon, E. Lock, O. Delémont, C. Roux, C. Zingg, et al., Intelligence-led crime scene processing. Part II: Intelligence and crime scene examination, Forensic Sci. Int. 199 (1) (2010) 63–71.
[31] Australia Co. Parliamentary Joint Committee on Law Enforcement Inquiry into the gathering and use of criminal intelligence; Available from: http://www.aph.gov.au/~/media/wopapub/senate/committee/le_ctte/completed_inquiries/2010-13/criminal_intelligence/report/report.ashx.
[32] D. Quick, K.-K.R. Choo, Internet-of-Things Forensic Data Reduction: It is not just Computers and Phones Anymore! Manuscript (submitted for publication), 2016.

Darren Quick received a M.Sc. in Cyber Security and Forensic Computing from University of South Australia, and is currently undertaking a Doctor of Philosophy (Ph.D. by research) with the University of South Australia. He is a Digital Forensic Investigator with the Australian Border Force, and previously an Electronic Evidence Specialist with the South Australia Police, has undertaken over 650 digital forensic investigations involving many thousands of digital evidence items, and has given evidence in Court in relation to a range of criminal matters. In 2012, Darren was awarded membership of the Golden Key International Honour Society, in 2014 received a Highly Commended award from the Australian New Zealand Policing Advisory Agency, in 2015 received the Publication of the Year award from the Australian Institute of Professional Intelligence Officers, and in 2016 was awarded Best Author in the Consensus IT Writers Awards.

Kim-Kwang Raymond Choo received the Ph.D. in Information Security from Queensland University of Technology, Australia. He currently holds the cloud technology endowed professorship at The University of Texas at San Antonio, and is an associate professor at University of South Australia. He was named one of 10 Emerging Leaders in the Innovation category of The Weekend Australian Magazine/Microsoft's Next 100 series in 2009, and is the recipient of various awards including ESORICS 2015 Best Research Paper Award, Highly Commended Award from Australia New Zealand Policing Advisory Agency, British Computer Society's Wilkes Award, Fulbright Scholarship, and 2008 Australia Day Achievement Medallion. He is a Fellow of the Australian Computer Society, and a Senior Member of IEEE.