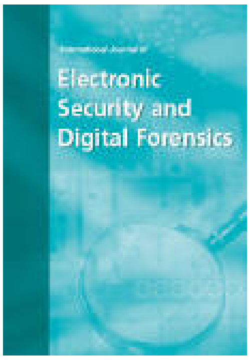


International Journal of Electronic Security and Digital Forensics

ISSN online: 1751-9128 - ISSN print: 1751-911X
https://www.inderscience.com/ijesdf

Intelligence sharing in big data forensics

Oteng Tabona, Thabiso M. Maupong, Kopo M. Ramokapane, Thabo Semong

DOI: 10.1504/IJESDF.2023.10051337

Article History:

Received:	01 October 2021
Accepted:	21 December 2021
Published online:	15 December 2022

Intelligence sharing in big data forensics

Oteng Tabona* and Thabiso M. Maupong

Department of Computer Science and Information Systems,
Botswana International University of Science and Technology,
Palapye, Botswana
Email: tabonao@biust.ac.bw
Email: maupongt@biust.ac.bw
*Corresponding author

Kopo M. Ramokapane

University of Bristol,
Bristol, UK
Email: marvin.ramokapane@bristol.ac.uk

Thabo Semong

Department of Computer Science and Information Systems,
Botswana International University of Science and Technology,
Palapye, Botswana
Email: semongt@biust.ac.bw

Abstract: With the high prevalence of digital crimes, forensic investigators rely on traditional desktop tools to conduct investigations. Most of these tools are device-specific and majority of them are desktop-based therefore they suffer from limited storage and fail to process big data. These tools also lack the analytical ability to link evidence between cases or share information between cases. Therefore, inter-links can exist between cases without being detected. The poor ability to detect links between cases may result in investigators: taking a long time to complete investigations and failing to establish organised crimes. In this paper, we propose a novel solution that can cross-link evidence between cases. Our solution is not desktop-based, nor is it restricted by the evidence source. Using real-world data for evaluation, we demonstrate that our solution is capable of uncovering evidence common between cases that could otherwise be missed.

Keywords: forensic cloud environment; FCE; big data forensic; digital forensic as a service; intelligence sharing; evidence correlation.

Reference to this paper should be made as follows: Tabona, O., Maupong, T.M., Ramokapane, K.M. and Semong, T. (2023) 'Intelligence sharing in big data forensics', *Int. J. Electronic Security and Digital Forensics*, Vol. 15, No. 1, pp.33–55.

Biographical notes: Oteng Tabona graduated with an MEng in Computer Engineering in 2012 from the Cardiff University, UK and PhD in Big Data and Digital Forensics in 2017 from the University of South Wales, UK. He is currently a Lecturer in Computer Science and Information System at the Botswana International University of Science and Technology (BIUST). His research interest includes big data forensics, and IoT forensics and security.

Thabiso M. Maupong graduated with an MEng in Computer Engineering and PhD in Electronics and Electrical Engineering in 2008 and 2017, respectively, from the University of Southampton, UK. He is currently a Senior Lecturer in Computer Science and Information System at the Botswana International University of Science and Technology (BIUST). His research interest includes data-driven control, systems theory, behavioural systems theory, application of data-driven control in power electronics, more recently he has developed an interest in data science, software-defined networks, digital forensics and cyber security.

Kopo M. Ramokapane graduated with a BEng in Computer Systems Engineering in 2011 from the University of Essex, UK, an MEng in Computer Engineering in 2012 from the Cardiff University, UK, and PhD in Computer Science in 2019 from Lancaster University, UK. He is currently a research associate at the University of Bristol, UK. His research interest lies in usable privacy and security, understanding users' challenges, and developing interventions that address their challenges.

Thabo Semong received his BSc in Computer Science from the University of Botswana in 2005, MEng and PhD in Computer Science from the Hunan University, China, in 2009 and 2018, respectively. He is currently a Senior Lecturer in Computer Science and Information System at the Botswana International University of Science and Technology (BIUST). His research interests include software-defined networking, network function virtualisation, network management and control, wireless sensor networks and IoT.

1 Introduction

Digital networks and economies have become critical enablers for the growth of cyber-crime (Goodman et al., 2007; Tropina, 2012; Yar and Steinmetz, 2019). Perceived anonymity, confidentiality, accessibility, low costs, absence of borders, ease of use and connectivity are some of the characteristics of the cyberspace that make it a more attractive environment to criminals (Goodman et al., 2007; Tropina, 2012). More often, individuals fall prey to malware, ransomware, corporate exploitation and leak of private data. Cybercriminals are developing and continue to improve techniques for committing crimes such as phishing, pharming, malware, social engineering and tools to attack commercial databases (Tropina, 2012), hence the increase in frequency and severity of attacks (Hiscox, 2019). The increase and severity are evident from Cybersecurity Ventures (2019) where it is estimated that the total cost of cyber-crimes will exceed $6 trillion annually by 2021. Furthermore, this figure will change rapidly because during the COVID-19 pandemic there has been an increase in the number of cyber-attacks (Lallie et al., 2020).

These increase, has lead to heightened demand for digital forensic investigations research. The primary purpose of digital forensics investigation research is to provide techniques and tools for discovering patterns of criminal activities from digital devices. Some of the most commonly used tools includes AccessData Forensic Toolkit (FTK) (Smith et al., 2017; Wagner et al., 2019), and Guidance EnCase (Bunting and Wei, 2006; Wagner et al., 2019; Kim et al., 2016).

The use of forensics tools like FTK and EnCase does not always yield the desired results, especially when dealing with complicated cases such as those that involve big data, organised crimes, and inter-linked multiple cases (Shalaginov et al., 2017). These tools suffer some limitations as they are workstation-based, have limited storage, processing power, device specific, and they lack the ability to detect or determine links between cases. These limitations will worsen when a case being investigated is complex and involves a large number of sources (Lillis et al., 2016). To put this into context, it is estimated that an individual currently owns about 3.96 devices and is expected to own about 9 devices by 2025 (Safaei et al., 2017). This swiftly increases the amount of data and devices that needs to be investigated.

The quantity and consequently, the complexity of digital crimes require more intelligent and proactive techniques for effective investigation in a timely manner. Furthermore, it is of paramount importance for these techniques to have the intelligence to identify potential correlation, patterns and establish links between different cases. Currently, it is very difficult to find connections between cases using traditional tools, even when using advanced solutions like triage (Montasari and Hill, 2019; Mislan et al., 2010; Hitchcock et al., 2016; Gentry and Soltys, 2019) and digital forensics as a service (DFaaS) (Van Baar et al., 2014; Van Beek et al., 2015; Stelly and Roussev, 2017). Investigators usually rely on their memory to recall similar objects (for example, name, credit card number, phone number) to link different cases. This approach is very challenging and laborious, and most importantly, it can result in omissions of important links between cases. Moreover, the problem can become more complex and demanding when dealing with big data forensic cases where size, the complexity of the data and the overall time required to investigate and conclude a case are already a challenge to forensic investigations.

Currently, there are no big data forensic tools that offers intelligence sharing as well as the link between cases or analyses. Consequently, this highlights the need to develop better ways of intelligence-sharing during forensic investigations on big data. Furthermore, prior work (Garfinkel, 2006) has shown that traditional tools examine evidence independently, and there is no opportunity to automatically 'connect the dots' on large cases involving multiple evidence sources, let alone different cases.

Intelligent sharing is an essential feature for investigating big data, and firstly, it would assuredly fill the missing pieces or gaps between cases that most investigations miss. Secondly, it would reduce the amount of time taken to investigate cases with links. Lastly, intelligent sharing would give a broader insight into cases by providing a wider scope of criminal activities, thus enabling a greater understanding of the criminal environment.

This paper makes the following contributions.

- An intelligence sharing framework for big data investigations is developed. The intelligence sharing framework allows forensic examiners to detect links between cases and provides a platform for sharing evidence between cases.

- Discussion of the intelligence sharing framework, outlining the most important features and how each component contributes towards intelligence sharing. Leveraging on the forensic cloud environment (FCE) which can already handle big data investigations (Tabona and Blyth, 2016).
- An experimental evaluation of the intelligence sharing framework by investigating three cases. Empirical results demonstrate that using this framework, it is possible to find evidence between cases which may otherwise be difficult to achieve using existing methods. Our framework design also facilitates easy and reliable intelligence sharing between investigators.

The paper is organised as follows. In Section 2, we cover some of the tools used for intelligence sharing and outline their weakness. Section 3 describes our method on intelligence sharing, and in Section 4 we demonstrate our framework with some evaluations. We discuss our results in Section 5 and conclude the paper in Section 6.

2 Related work

Prior work on big data forensics has mainly focused on adapting traditional forensics methods to enable them for big data forensic investigations. This has seen an advent of novel solutions such as digital forensics as service (DFaaS) (Van Baar et al., 2014; Van Beek et al., 2015; Tabona and Blyth, 2016; Stelly and Roussev, 2017), data reduction techniques (Scanlon, 2016; Quick and Choo, 2014; Neuner et al., 2016), triage (Mislan et al., 2010; Roussev et al., 2013; Garfinkel, 2013; Hitchcock et al., 2016; Gentry and Soltys, 2019; Montasari and Hill, 2019), artificial intelligence (Mitchell, 2010; Irons and Lallie, 2014; Mohammed et al., 2016; Costantini et al., 2019) and data mining (Beebe and Clark, 2007; Sindhu and Meshram, 2012; Tallón-Ballesteros and Riquelme, 2014; Quick and Choo, 2014) which have significantly advanced the state-of-the-art. However, these approaches are not short of limitations, particularly regarding intelligence sharing and evidence correlations. These approaches rely on the examiners to link evidence between cases which is usually time-consuming, labour-intensive, requires significant human resources, and most importantly, may result in some links between cases been missed.

In Noel and Peterson (2014), an attempt to reduce investigators' operation cost is proposed, the technique extracts hidden themes from various documents with minimum human intervention and then isolating data perceived relevant by using keywords. Other studies (Ruback et al., 2012; Rowe, 2013) have applied clustering and reduction techniques to reduce uninteresting or irrelevant data from investigations. Triage studies (Mislan et al., 2010; Roussev et al., 2013; Garfinkel, 2013; Hitchcock et al., 2016; Gentry and Soltys, 2019; Montasari and Hill, 2019) have also attempted to decrease the amount of time and effort spent investigating cases by intelligently predicting whether seized images contain relevant traces of evidence. Mohammed et al. (2016) developed a framework to handle existing issues around big data forensics such as volume, variety and heterogeneity of data. The purpose of this framework is to help investigators with understanding the nature of the evidence and correlation between artefacts. Similarly in Adedayo et al. (2016) a framework to assess the various stages of forensic examinations is introduced. In each stage, a techniques to enhance better collection, analysis, preservation and presentation while investigating big data is discussed.

As discussed above, prior studies have only attempted to solve a specific big data forensic problem which arises from the characteristic of the data but has not focused on the links between evidence or sharing of intelligence between cases. Our proposed intelligence sharing framework is different, from those discussed, because it offers the ability to analyse and link big data cases. Research closest to our framework, presented in Garfinkel (2006), focused on evidence correlation from multiple images. The authors use forensic feature extraction (FFE) and cross drive analysis (CDA) to extract, analyse and correlate evidence from multiple images. Using this technique, they analysed 750 images and were able to identify correlations between images.

Quick and Choo (2018) presented a framework for entity identification using Open Source Intelligent Tools (OSINT) which expanded the cross drive and cross case analysis used in Garfinkel (2006). They argued that criminal intelligence analysis is critical for knowledge for tactical, operational and strategic intelligence. Using M57 corpus Garfinkel et al. (2009) of approximately 498 GB from various sources (i.e., computers, portable storage, mobile phones and tablet devices) demonstrated the benefits of applying the framework to achieve a greater understanding of the evidence. With this framework, they located additional information related to the entities under study and uncovered disparate information which may add to intelligence value. A technique to associate disk drives using documents to several metrics is presented in Rowe (2018). They used a term-frequency inverse-document-frequency (TF/IDF) cosine similarity and Kullback-Leibler divergence formula to associate drives based on 18 different types of clues and developed visualisation techniques to display the associations. Another study that attempted to correlate forensics evidence from different forensic targets was conducted in Case et al. (2008). They proposed forensic automated correlation engine (FACE) which automatically discovered evidence from Linux memory dump and made some correlations between events and objects.

Compared to our work, all these approaches were only applied to multiple images and small data, in the case of FACE which is not ideal for big data where data is multi-dimensional, large in volume and without format. Prior approaches also rely on traditional tools such as relational databases which have many limitations with regard to big data. Our work contributes to this area by demonstrating intelligence sharing between cases using FCE (Tabona and Blyth, 2016) and big data sample. Our solution provides for a command-based system for determining links hence facilitating for easy investigation of organised crimes by providing links between cases, and perpetrators.

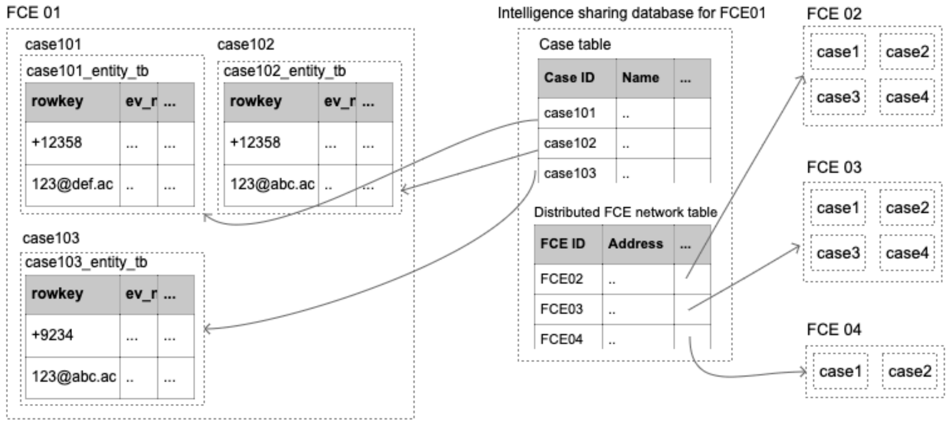
3 Proposed intelligence sharing framework

To address the lack of intelligence sharing, we develop an intelligence sharing functionality on the existing FCE proposed in Tabona and Blyth (2016). The intelligence sharing application discovers links between cases using objects such as phone numbers, e-mails, names and/or any other specific object as required by investigators. We developed two types of intelligence sharing interfaces, local and global intellishare interfaces. The local intellishare discovers links between cases within the same FCE, while the global intellishare finds connections between cases in different FCEs. In Subsection 3.1, we explore the FCE setup that makes it possible to gather data for intelligence for sharing. Then, we provide more details on the local and global intellishare features in Subsections 3.2 and 3.3, respectively. In this paper, we do not discuss the overall design, setup and implementation of FCE, for a through exposition we refer the reader to Tabona and Blyth (2016).

3.1 Setting FCE for intelligence sharing

To setup a private cloud for conducting digital forensics FCE is implemented using Apache Hadoop (White, 2012; Borthakur et al., 2011; Lam, 2010). The Hadoop framework is ideal for FCE as it offers the capabilities of investigating data from multiple sources. For example, when investigating a case with multiple sources, data is stored in the Hadoop distributed file system (HDFS), which preserves the integrity of data blocks through checksums. In HDFS, data is then duplicated and copied across multiple data nodes (three nodes by default). The data is then stored in a NoSQL database HBase (White, 2012) as it can store and process unstructured data. MapReduce programming is then used to the process data in parallel. Map codes are executed in the data blocks to avoid input/output bottlenecks associated with moving data for processing.

Figure 1 HBase tables design for local and global intelligence sharing



case101_entity_tb

rowkey	ev_r	...
+12358	...	...
123@def.ac	..	...

case102_entity_tb

rowkey	ev_r	...
+12358	...	...
123@abc.ac	..	...

case103_entity_tb

rowkey	ev_r	...
+9234	...	...
123@abc.ac	..	...

Case table

Case ID	Name	...
case101	..	
case102	..	
case103	..	

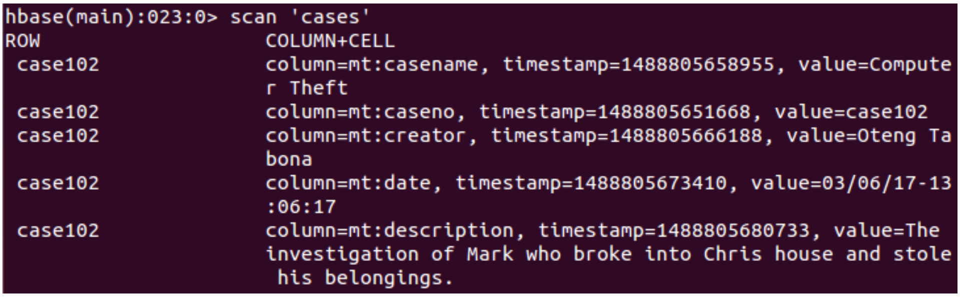
Distributed FCE network table

FCE ID	Address	...
FCE02	..	
FCE03	..	
FCE04	..	

The tables design in HBase is shown in Figure 1, these design is essential for intelligence sharing framework as it allows easy extraction of objects of interest. The *cases table* is used to store general details concerning a case such, e.g., *case number*, *case name*, *case creator*, *case description*, *date* and *time* as illustrated on the right of Figure 2. For intelligence sharing, an entity extractor program is then used to extract objects of interest from the cases. The extracted objects are then stored in an entity table, for example, case with ID case101, the entity extractor program store objects pertaining a particular case as shown in a table *case101_entity_tb* depicted on the left in Figures 1 and 3.

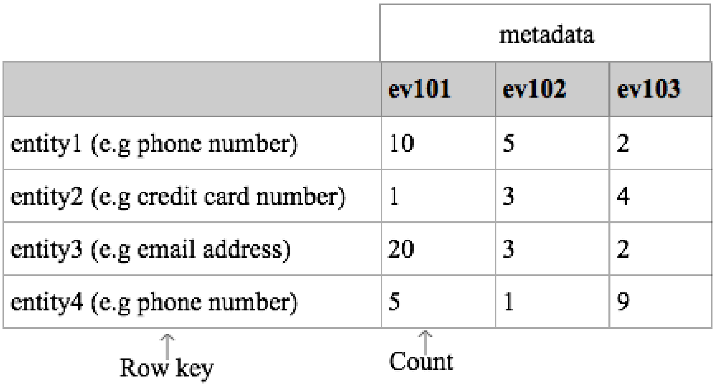
The outlined setup is crucial for both local and global intellishare and as we discuss in the following subsections.

Figure 2 Example of cases table (see online version for colours)



```
hbase(main):023:0> scan 'cases'
ROW                      COLUMN+CELL
 case102                 column=mt:casename, timestamp=1488805658955, value=Compute
                         r Theft
 case102                 column=mt:caseno, timestamp=1488805651668, value=case102
 case102                 column=mt:creator, timestamp=1488805666188, value=Oteng Ta
                         bona
 case102                 column=mt:date, timestamp=1488805673410, value=03/06/17-13
                         :06:17
 case102                 column=mt:description, timestamp=1488805680733, value=The
                         investigation of Mark who broke into Chris house and stole
                          his belongings.
```

Figure 3 Example of case entity table (where ev refers to evidence)



	metadata		
	ev101	**ev102**	**ev103**
entity1 (e.g phone number)	10	5	2
entity2 (e.g credit card number)	1	3	4
entity3 (e.g email address)	20	3	2
entity4 (e.g phone number)	5	1	9

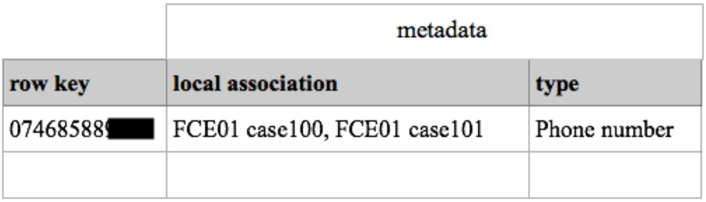
Row key

Count

3.2 Local intellishare

For local intelligence sharing purposes, we first access the *cases table* to retrieve the details of all the cases that are being investigated in the same FCE. These details, particularly the case ID, are then used to get a reference to the case entity table. Assuming that the evidence sources have been parsed and an entity identify has extracted entities from the *case entries* table. We use a MapReduce program to unify all case entity tables and the results are stored in a *local intellishare* table depicted in Figure 4.

Figure 4 Local intellishare table



	metadata	
row key	**local association**	**type**
07468588[redacted]	FCE01 case100, FCE01 case101	Phone number

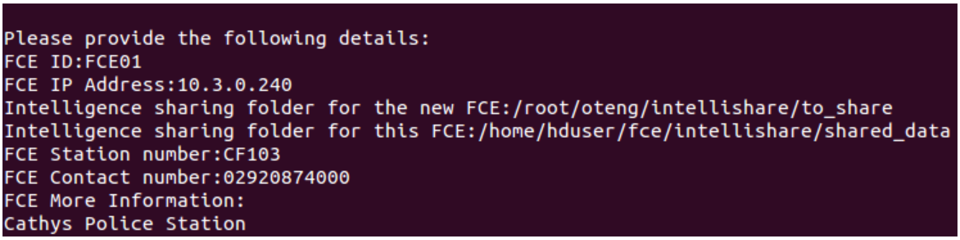
The local intellishare table consists of a *row key* column which contains a list of entities. The metadata family of columns consists of a *local association* and *type* columns – the local association list all the cases within the FCE associated with the row key while the type column provides details of the row key. For instance, considering the details of Figure 4, the row key contains a number (07468588XXX) which is present in case number ‘case100’ and ‘case101’ of FCE id of FCE01.

If an investigator is interested in knowing which evidence sources within case100 and case101 does this entity appear in, they can do a reserve-lookup. That is, in each case, they can look into their case entity table, depicted in Figure 3. The row key here is the entity and the metadata column family now consists of the evidence source ids. HBase allows the database to grow both horizontal or vertical anytime. So this particular feature allows the metadata column family columns to grow horizontally.

3.3 Global intellishare

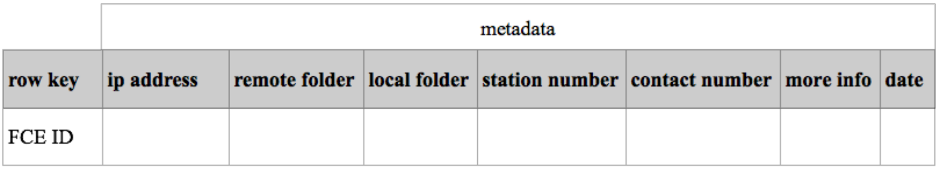
Global intellishare framework reveals links between cases in different FCEs. For this an investigator creates FCE network by adding a new FCE to current FCE. Shown in Figure 5, FCE01 is added to FCE02 and the information is added to *FCE network table* as demonstrated in Figure 6.

Figure 5 Creating a distributed FCE network (see online version for colours)



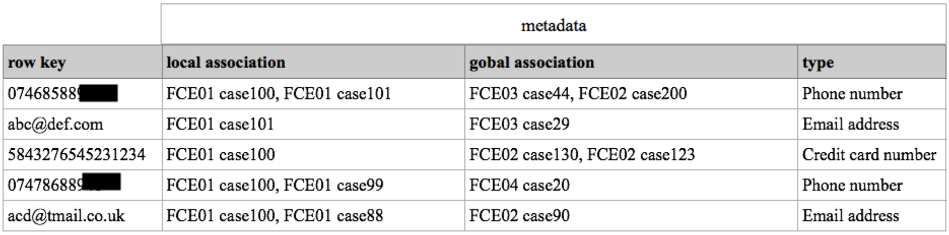
```
Please provide the following details:
FCE ID:FCE01
FCE IP Address:10.3.0.240
Intelligence sharing folder for the new FCE:/root/oteng/intellishare/to_share
Intelligence sharing folder for this FCE:/home/hduser/fce/intellishare/shared_data
FCE Station number:CF103
FCE Contact number:02920874000
FCE More Information:
Cathys Police Station
```

Figure 6 Sample of distributed FCE network table



	metadata						
row key	**ip address**	**remote folder**	**local folder**	**station number**	**contact number**	**more info**	**date**
FCE ID							

Figure 7 Local intellishare table with global association



	metadata		
row key	**local association**	**gobal association**	**type**
07468588[redacted]	FCE01 case100, FCE01 case101	FCE03 case44, FCE02 case200	Phone number
abc@def.com	FCE01 case101	FCE03 case29	Email address
5843276545231234	FCE01 case100	FCE02 case130, FCE02 case123	Credit card number
07478688[redacted]	FCE01 case100, FCE01 case99	FCE04 case20	Phone number
acd@tmail.co.uk	FCE01 case100, FCE01 case88	FCE02 case90	Email address

To find correlations between objects for global intellishare, the program first reads the FCE network table to get the location of other FCEs. As a security feature, other FCEs only connect to shared folders. The shared files from all FCEs are processed and stored in a *global intellishare* table by the FCE that initiated the process. After that, the global intellishare table is unified into a local intellishare table. A new column *global association* is added to the local intellishare table to indicate any row key that matches the global intellishare row key. The value of the new column will contain the global FCE cases that have a similar entity. A design example of the local intellishare table with global associations is shown in Figure 7.

In Figure 7, if we concentrate on the entity 07468588XXX, we find that it is linked to case100 and case101 within the initiating FCE, that is FCE01. This entity is also associated with case44 and case200 in FCE03 and FCE02, respectively.

4 Experiment

In this section, we demonstrate the intelligence sharing framework described in Section 3. We formulate three case scenarios to demonstrate how investigations are carried out on FCE and both global and local intellishare are done during the process of investigation. Firstly, in the next subsection, we provide some details on the setup of FCE.

4.1 FCE setup

We setup distributed FCEs, that is, FCE01 and FCE02 as shown in Figure 8.

The details of each are as follows. FCE01 is designed to handle big data with a total HDFS logical storage capacity of 81.5 TB. The design consists of a one master node and four slaves. The master node hosts services such as the NameNode and HBase master. The slave nodes act as data nodes and region servers. The hardware specifications of the master machine include storage capacity of 1.1 TB, 141 GB memory, 16x Intel Xenon CPU E5620 2.4 GHz, 4 cores and CentOS operating systems. The hardware specification for each slave nodes include a storage capacity of about 60 TB, memory 31 GB, 8x Intel AtomTM CPU C2750 2.41 GHz, 8 cores and also installed CentOS.

FCE02 has a total logical storage capacity of 2.6 TB and it is mainly prepared to demonstrate global intellishare. FCE02 infrastructure is made up of seven nodes including a server, master and slave nodes. The server manages the Cloudera service while the master node hosts services such as NameNode and HBase. The slave nodes are assigned data node and region server roles. Hardware specifications for each node include a storage capacity of 461 GB, 3 GB of memory, 2x AMD E-350 Processor, 2 cores and Ubuntu operating system.

We use Cloudera Distributed Hadoop (CDH) software to install Apache Hadoop and its components on both infrastructures. FCE01 we use version 5.6.0 of CDH while for FCE02 version 5.8.0 is used. Both infrastructures use the same Hadoop version 2.6.0, HBase and MapReduce. Several applications were developed to facilitate the investigation as tabulated in Table 1.

In Algorithm 1 we outline the steps for loading cases in an FCE, investigate and do intelligence sharing. Also in Algorithm 1 is the role of the applications in Table 1.

Figure 8 Distributed FCE (see online version for colours)

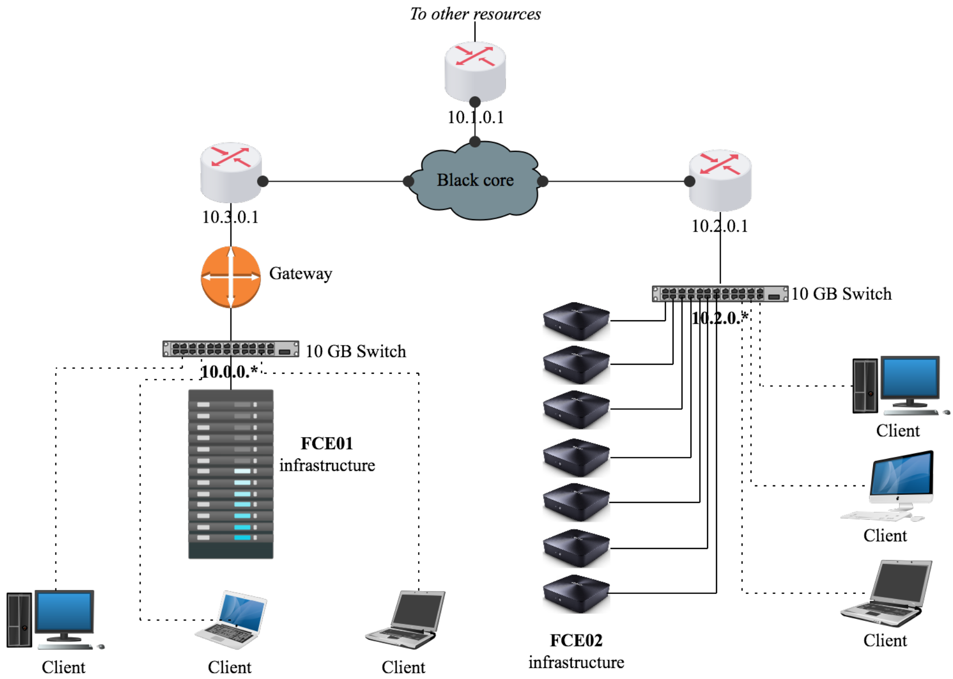


Table 1 FCE application stack

Top tier	*Middle tier*	*Bottom tier*
Network analyser	Hash calculator	Image ingester
Timeline analyser	File type determiner	File-system parsers
Communication database extractor	Known file filtre (KFF) calculator	Case management
Entity identifier		KFF loader
Local intellishare		File signature loader
Global intellishare		
Intellishare network generator		

4.2 *Case scenarios and investigation*

We start by investing case 3 (drug dealing) which will be stored in FCE02 and conducted in precinct 1, followed by case 2 (money laundering) and case 1 (crash for cash) both in FCE01 and investigated in precinct 2. For each case, we shall use communication database, i.e., e-mails and phone numbers, for timeline analysis and intelligence sharing. We remark that these are chosen for demonstration purpose and our intelligence sharing in FCE is not limited to those.

Algorithm 1 FCE intelligence share

```
Input  : Evidence sources
Output : Intellishare
1  Create a case in FCE using case management.
2  Image evidence sources using FTK Imager for HDD and USB and for mobile phones use
   Cellebrite UFED Touch.
3  Compute cryptographic hashes of images.
4  Ingesting images into an FCE case using image ingester.
5  Compute cryptographic hashes of images in the FCE using hash calculator.
6  Parse the images, using file-system parser.
7  Calculate hash values of images after parsing using hash calculator.
8  Compare the hash values from steps 3, 5 and 7 to determine if the images were corrupt
   during parsing. If the hash values are different go back to step 3. Otherwise do the next
   step.
9  Extract files and metadata.
10 Calculate hash values of the file using hash calculator.
11 Execute other optional pre-processor applications such as file type determiner, known file
   filtre, using file type determiner and known file filtre (KFF) calculator.
12 Extract necessary object for investigation, for example, using communication database
   extractor
13 Investigate using case network and timeline analyser.
14 Build entities/object tables for intelligence share using entity identifier.
15 Do either local or global intellishare using local/global intellishare application.
16 Visualise links between cases using intellishare network generator
```

For each action that is carried out on a case, an audit trail is generated as per digital forensic principles (Williams, 2012). Before generating case data we forensically erased all data sources using the US Department of Defense (DoD) wiping standard (DoD 5220.22-M) (Forte and Power, 2007).

4.2.1 Case 3: drug dealing

This case concerns drug dealing, with five suspects. A total of nine evidence sources which include mobile phones, personal computers and USB were collected for investigation. The total evidence collection for the case amounts to 676 GB. Following Algorithm 1 we generate a timeline analysis as shown in Figure 9.

The timeline graph is then analysed to find evidence pertaining to the crime starting from the 7th of March 2017. A text message was received by a suspect at 11:29:13. The content of this message is shown in Figure 10. This message shows that it was sent by a client saved as 63B client. The client was requesting for some chocolate bar to be delivered at 63B. In this instance, the chocolate bar is deemed to be a street name for drugs based on the context.

There is also a message sent to a contact saved as driver as shown in Figure 11. The content of the message reveals that the suspect is asking the driver to deliver an order to 63B. The threads within this message contain evidence that shows that the driver was frequently asked to deliver orders. The timeline graph also contains more messages from clients to the suspects concerning drugs.

At this stage, to continue the investigation it is important to know who the driver is. In an effort to identify the driver an entity identifier application on FCE02 is executed to extract all entities from the case, see Figure 12 showing the entity table. The aim of this process is to find links between the driver's number and other cases within FCE02.

Figure 9 Case 3: timeline (see online version for colours)

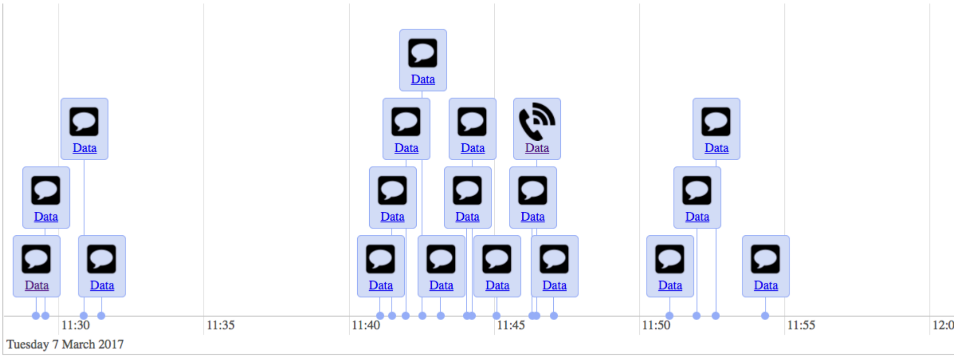
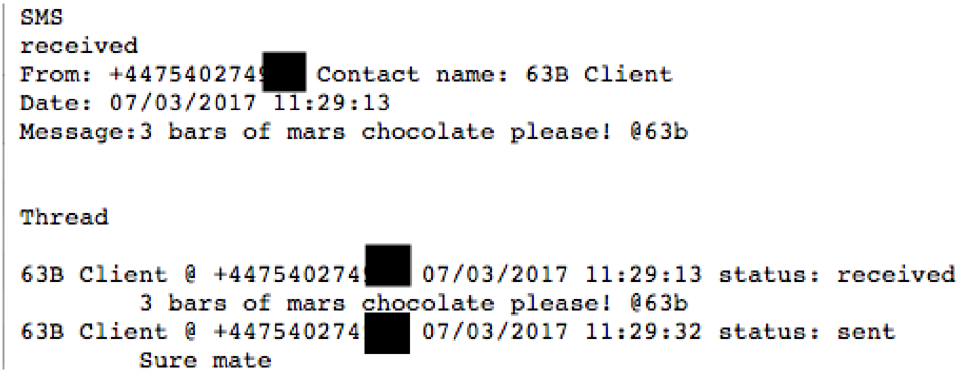


Figure 10 Case 3: drug order



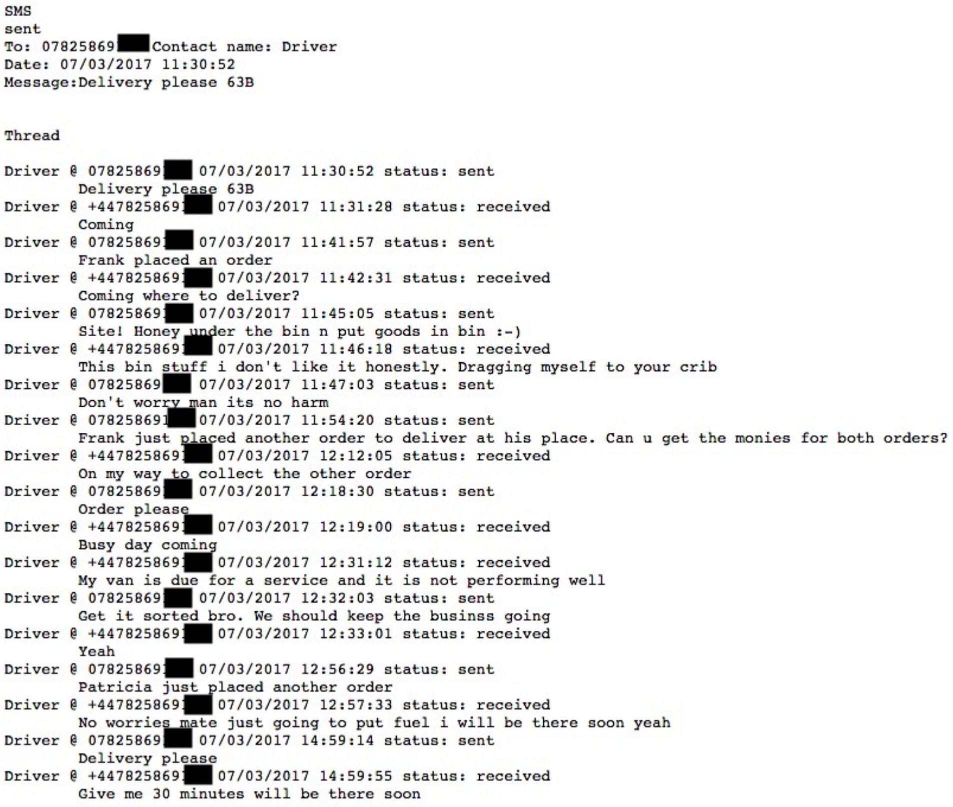
```
SMS
received
From: +447540274[redacted] Contact name: 63B Client
Date: 07/03/2017 11:29:13
Message:3 bars of mars chocolate please! @63b


Thread

63B Client @ +447540274[redacted] 07/03/2017 11:29:13 status: received
         3 bars of mars chocolate please! @63b
63B Client @ +447540274[redacted] 07/03/2017 11:29:32 status: sent
         Sure mate
```

The extracted entities are correlated with local cases using the local intellishare application. Unfortunately, there was no match that was discovered as shown in Figure 13. To elaborate further, Figure 13 only shows that there are no other cases that have the same entities that exists in FCE02 drug case or *cs-drugs-100 case*. The global intellishare application was also executed to discover any links between the driver's number and cases in other FCEs. Similarly, there was also no connection that was discovered in the local intellishare table with the global association as shown in Figure 14.

4.2.2 Case 2: money laundering

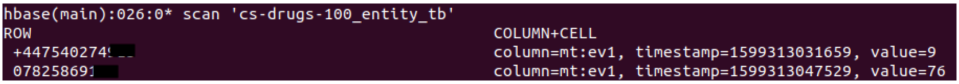
Case 2, entails money laundering and it was also investigated in precinct 2. Five suspects who all reside in one house were identified and their electronic devices confiscated. Initial investigations identified a number of suspicious transactions that were deposited into James's account. The total number of evidence sources collected is 13, including mobile phone, PCs and several external devices, with a total of 1.576 TB data.

Figure 11 Case 3: message to the driver



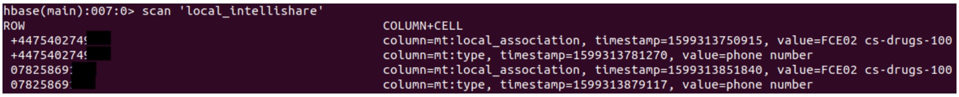
```
SMS
sent
To: 07825869[redacted]Contact name: Driver
Date: 07/03/2017 11:30:52
Message:Delivery please 63B


Thread

Driver @ 07825869[redacted] 07/03/2017 11:30:52 status: sent
        Delivery please 63B
Driver @ +447825869[redacted] 07/03/2017 11:31:28 status: received
        Coming
Driver @ 07825869[redacted] 07/03/2017 11:41:57 status: sent
        Frank placed an order
Driver @ +447825869[redacted] 07/03/2017 11:42:31 status: received
        Coming where to deliver?
Driver @ 07825869[redacted] 07/03/2017 11:45:05 status: sent
        Site! Honey under the bin n put goods in bin :-)
Driver @ +447825869[redacted] 07/03/2017 11:46:18 status: received
        This bin stuff i don't like it honestly. Dragging myself to your crib
Driver @ 07825869[redacted] 07/03/2017 11:47:03 status: sent
        Don't worry man its no harm
Driver @ 07825869[redacted]07/03/2017 11:54:20 status: sent
        Frank just placed another order to deliver at his place. Can u get the monies for both orders?
Driver @ +447825869[redacted] 07/03/2017 12:12:05 status: received
        On my way to collect the other order
Driver @ 07825869[redacted] 07/03/2017 12:18:30 status: sent
        Order please
Driver @ +447825869[redacted] 07/03/2017 12:19:00 status: received
        Busy day coming
Driver @ +447825869[redacted] 07/03/2017 12:31:12 status: received
        My van is due for a service and it is not performing well
Driver @ 07825869[redacted] 07/03/2017 12:32:03 status: sent
        Get it sorted bro. We should keep the businss going
Driver @ +447825869[redacted] 07/03/2017 12:33:01 status: received
        Yeah
Driver @ 07825869[redacted] 07/03/2017 12:56:29 status: sent
        Patricia just placed another order
Driver @ +447825869[redacted] 07/03/2017 12:57:33 status: received
        No worries mate just going to put fuel i will be there soon yeah
Driver @ 07825869[redacted] 07/03/2017 14:59:14 status: sent
        Delivery please
Driver @ +447825869[redacted] 07/03/2017 14:59:55 status: received
        Give me 30 minutes will be there soon
```

Figure 12 Case 3: case entity table (see online version for colours)



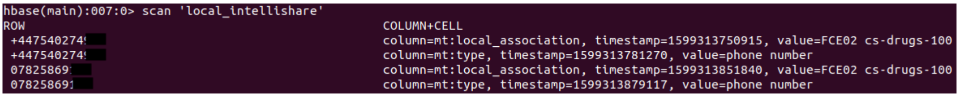
```
hbase(main):026:0* scan 'cs-drugs-100_entity_tb'
ROW                                             COLUMN+CELL
 +447540274[redacted]                           column=mt:ev1, timestamp=1599313031659, value=9
 07825869[redacted]                             column=mt:ev1, timestamp=1599313047529, value=76
```

Figure 13 Case 3: local intellishare table (see online version for colours)



```
hbase(main):007:0> scan 'local_intellishare'
ROW                                                    COLUMN+CELL
 +447540274[redacted]                                  column=mt:local_association, timestamp=1599313750915, value=FCE02 cs-drugs-100
 +447540274[redacted]                                  column=mt:type, timestamp=1599313781270, value=phone number
 07825869[redacted]                                    column=mt:local_association, timestamp=1599313851840, value=FCE02 cs-drugs-100
 07825869[redacted]                                    column=mt:type, timestamp=1599313879117, value=phone number
```

Figure 14 Case 3: local intellishare table with global association (see online version for colours)



```
hbase(main):007:0> scan 'local_intellishare'
ROW                                                    COLUMN+CELL
 +447540274[redacted]                                  column=mt:local_association, timestamp=1599313750915, value=FCE02 cs-drugs-100
 +447540274[redacted]                                  column=mt:type, timestamp=1599313781270, value=phone number
 07825869[redacted]                                    column=mt:local_association, timestamp=1599313851840, value=FCE02 cs-drugs-100
 07825869[redacted]                                    column=mt:type, timestamp=1599313879117, value=phone number
```

Initial steps of the investigation were to create a timeline graph to visualise all the activities that took place the days before the deposits were made. The timeline graph of the communication events between the 8th and 10th March as shown in Figure 15.

Figure 15 Case 2: timeline (see online version for colours)

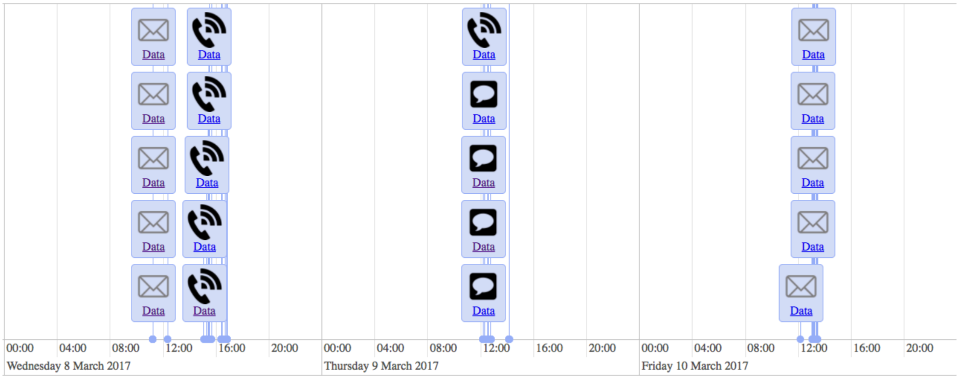
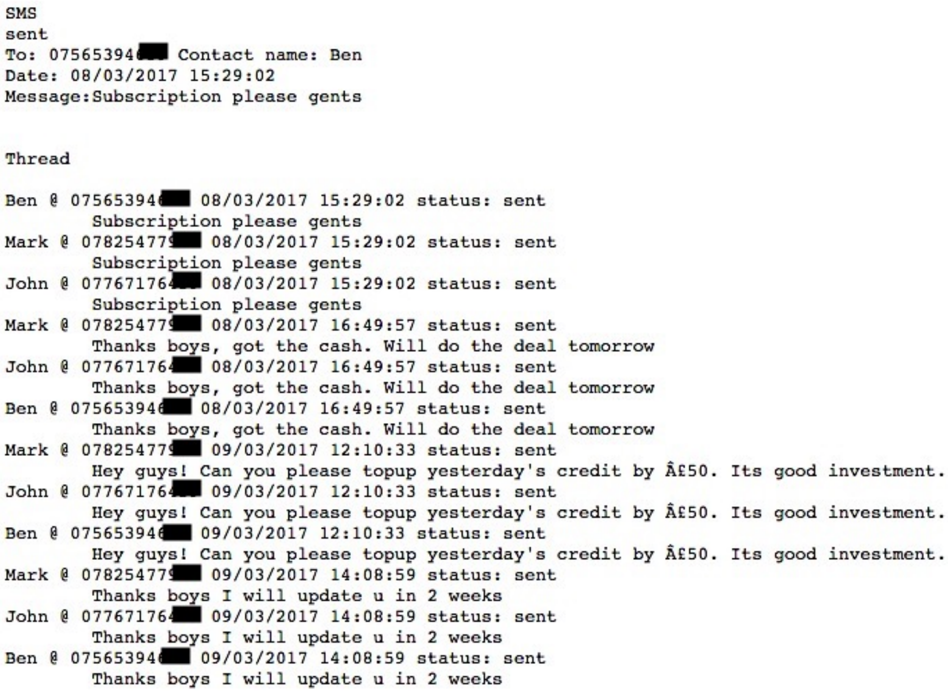


Figure 16 Case 2: request for subscription

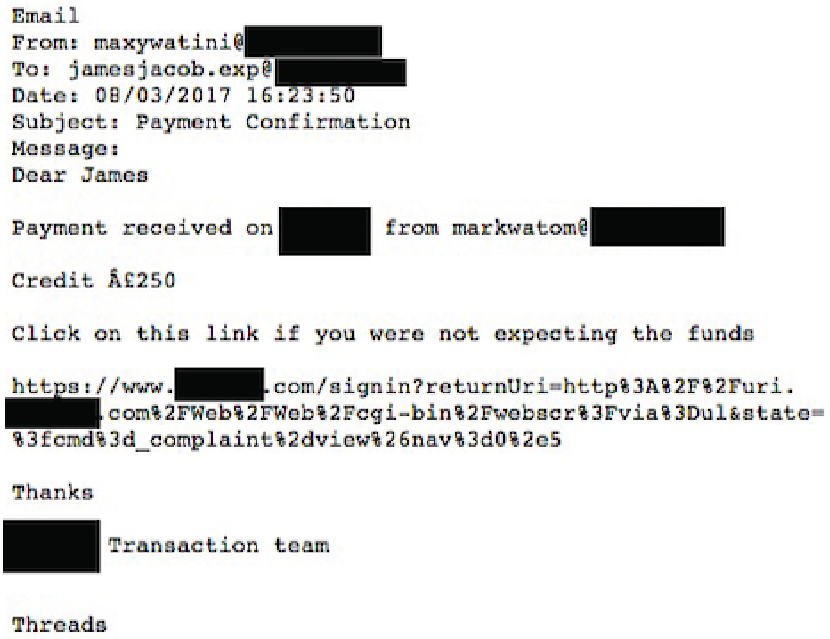


```
SMS
sent
To: 07565394[redacted] Contact name: Ben
Date: 08/03/2017 15:29:02
Message:Subscription please gents


Thread

Ben @ 07565394[redacted] 08/03/2017 15:29:02 status: sent
        Subscription please gents
Mark @ 07825477[redacted] 08/03/2017 15:29:02 status: sent
        Subscription please gents
John @ 07767176[redacted] 08/03/2017 15:29:02 status: sent
        Subscription please gents
Mark @ 07825477[redacted] 08/03/2017 16:49:57 status: sent
        Thanks boys, got the cash. Will do the deal tomorrow
John @ 07767176[redacted] 08/03/2017 16:49:57 status: sent
        Thanks boys, got the cash. Will do the deal tomorrow
Ben @ 07565394[redacted] 08/03/2017 16:49:57 status: sent
        Thanks boys, got the cash. Will do the deal tomorrow
Mark @ 07825477[redacted] 09/03/2017 12:10:33 status: sent
        Hey guys! Can you please topup yesterday's credit by Â£50. Its good investment.
John @ 07767176[redacted] 09/03/2017 12:10:33 status: sent
        Hey guys! Can you please topup yesterday's credit by Â£50. Its good investment.
Ben @ 07565394[redacted] 09/03/2017 12:10:33 status: sent
        Hey guys! Can you please topup yesterday's credit by Â£50. Its good investment.
Mark @ 07825477[redacted] 09/03/2017 14:08:59 status: sent
        Thanks boys I will update u in 2 weeks
John @ 07767176[redacted] 09/03/2017 14:08:59 status: sent
        Thanks boys I will update u in 2 weeks
Ben @ 07565394[redacted] 09/03/2017 14:08:59 status: sent
        Thanks boys I will update u in 2 weeks
```

The analysis revealed correlations between deposits that went into James's account. The transactions of interest occurred between the 8th–10th March 2017. The first suspicious transaction shows that an amount of £750 was deposited into James's account. Events leading to this transaction include three text messages sent through James mobile phone on the 8th March to contacts saved as Ben, Mark and John as shown in Figure 16.

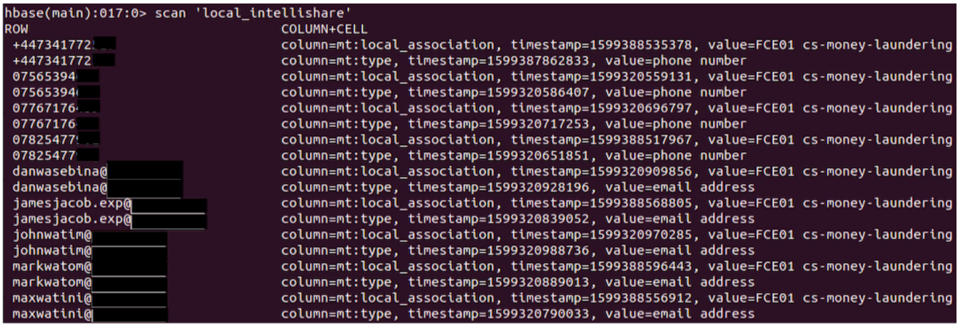
The text message indicates that James was requesting a subscription. About an hour after the text message, there are three e-mails that were sent to jamesjacob.exp@host.com. The e-mails were from online payment company (maxwatini@host.com), confirming money transfer from markwatom@host.com (see Figure 17), danwasebina@host.com and johnwatim@host.com. Each of these e-mails is approving a transfer of £250. The total amount transferred sum up to £750 which is equal to the amount of money that was deposited to James's account. This establishment led to an assumption that jamesjacob.exp@host.com is James's online payment company account.

Figure 17 Case 2: online payment company confirmation



```
Email
From: maxywatini@[redacted]
To: jamesjacob.exp@[redacted]
Date: 08/03/2017 16:23:50
Subject: Payment Confirmation
Message:
Dear James

Payment received on [redacted] from markwatom@[redacted]

Credit Â£250

Click on this link if you were not expecting the funds

https://www.[redacted].com/signin?returnUri=http%3A%2F%2Furi.
[redacted].com%2FWeb%2FWeb%2Fcgi-bin%2Fwebscr%3Fvia%3Dul&state=
%3fcmd%3d_complaint%2dview%26nav%3d0%2e5

Thanks

[redacted] Transaction team


Threads
```

Figure 18 Case 2: local intellishare table (see online version for colours)



```
hbase(main):017:0> scan 'local_intellishare'
ROW                               COLUMN+CELL
 +447341772[redacted]             column=mt:local_association, timestamp=1599388535378, value=FCE01 cs-money-laundering
 +447341772[redacted]             column=mt:type, timestamp=1599387862833, value=phone number
 07565394[redacted]               column=mt:local_association, timestamp=1599320559131, value=FCE01 cs-money-laundering
 07565394[redacted]               column=mt:type, timestamp=1599320586407, value=phone number
 07767176[redacted]               column=mt:local_association, timestamp=1599320696797, value=FCE01 cs-money-laundering
 07767176[redacted]               column=mt:type, timestamp=1599320717253, value=phone number
 07825477[redacted]               column=mt:local_association, timestamp=1599388517967, value=FCE01 cs-money-laundering
 07825477[redacted]               column=mt:type, timestamp=1599320651851, value=phone number
 danwasebina@[redacted]           column=mt:local_association, timestamp=1599320909856, value=FCE01 cs-money-laundering
 danwasebina@[redacted]           column=mt:type, timestamp=1599320928196, value=email address
 jamesjacob.exp@[redacted]        column=mt:local_association, timestamp=1599388568805, value=FCE01 cs-money-laundering
 jamesjacob.exp@[redacted]        column=mt:type, timestamp=1599320839052, value=email address
 johnwatim@[redacted]             column=mt:local_association, timestamp=1599320970285, value=FCE01 cs-money-laundering
 johnwatim@[redacted]             column=mt:type, timestamp=1599320988736, value=email address
 markwatom@[redacted]             column=mt:local_association, timestamp=1599388596443, value=FCE01 cs-money-laundering
 markwatom@[redacted]             column=mt:type, timestamp=1599320889013, value=email address
 maxwatini@[redacted]             column=mt:local_association, timestamp=1599388556912, value=FCE01 cs-money-laundering
 maxwatini@[redacted]             column=mt:type, timestamp=1599320790033, value=email address
```

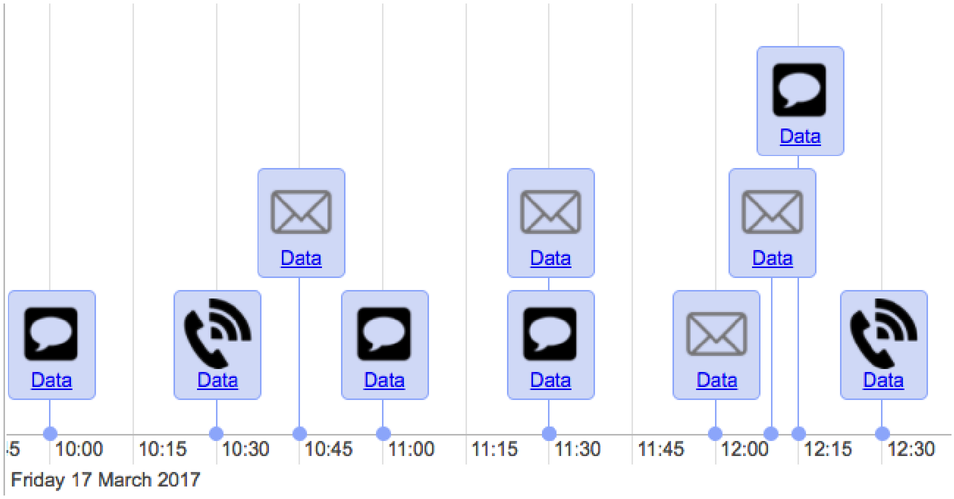
After finding the correlation between events and transaction, we were now interested in identifying the owners of markwatom@host.com, danwasebina@host.com and johnwatim@host.com. Assuming that the owners of danwasebina@host.com and johnwatim@host.com are two of the suspects who lived in the same property as James. Then the owner of markwatom@host.com is to be established. To further investigate we execute both the local and global intellishare applications to find any links with other cases, unfortunately, there were no matches established as shown in Figure 18.

4.2.3 *Case 1: crash for cash*

The case involves an insurance fraud called crash for cash incident. A perpetrator called Mark, who owns a garage is involved in deliberately crashing people's car with the aim of benefiting from the insurance. Phil, Chris and Davis are also part of the case story. Phil and Chris crashed their vehicle with the assistance of Mark while Davis was an owner of a tow truck. In this case, we collected a variety of 32 devices including mobile phones, personal computers, external hard drives and USB. The accumulative data capacity of the case was 12.704 TB.

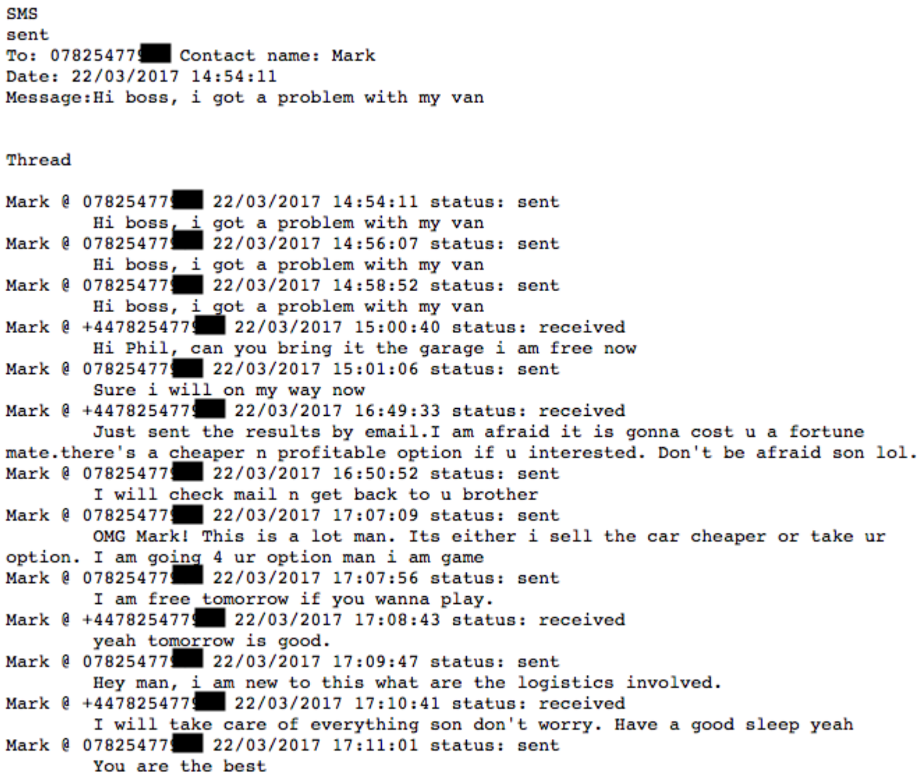
With the help of the timeline analysis graph shown in Figure 19 we are able to pick deliberate accident between Mark and Phil. Incriminating messages exchange between Mark and Phil are shown in Figure 20. Still using the timeline graph we were able to find communication between Mark and Davis as shown in Figure 21. We also looked at the events before the 22th March and we discovered more incriminating messages between Mark and Chris on the 13th March as shown in Figure 22.

Figure 19 Case 1: investigation timeline analysis (see online version for colours)



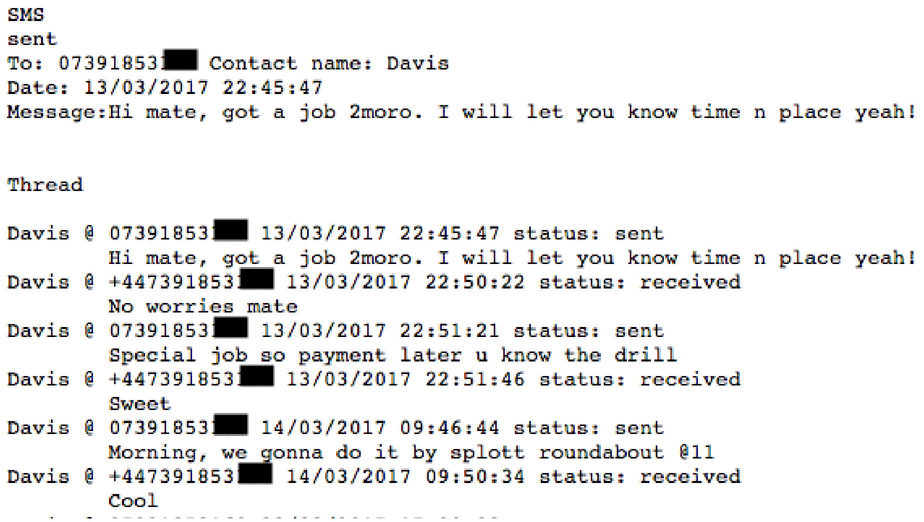
To gather intelligence between this case and other cases in the FCE01 and FCE02, we executed the local intelligence sharing framework to discover links between this case and other FCE01 cases. There is a link between cases 1 and 2 in FCE01 as captured in Figure 23. FCE01 local intellishare table was run through the intellishare network visualiser and the result is shown in Figure 24. The link is established as Mark's phone number: +447825477XXX and his e-mail address: markwatom@host.com are in both cases. Also, James's mobile number: +447341772XXX and his e-mail address: Jamesjacob.exp@host.com are in both cases. The online payment company e-mail address: maxywatini.exp@host.com are available in both cases.

Figure 20 Case 1: investigation – Mark and Phil communication



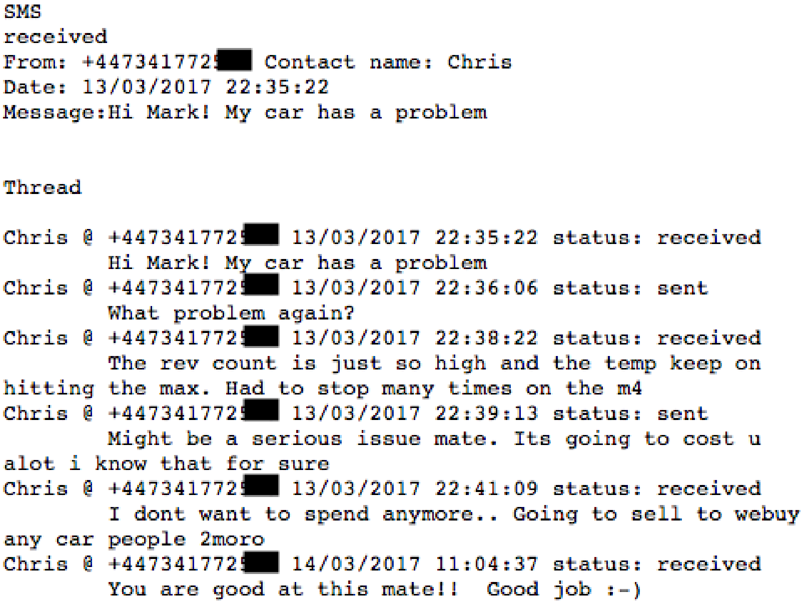
```
SMS
sent
To: 07825477[redacted] Contact name: Mark
Date: 22/03/2017 14:54:11
Message:Hi boss, i got a problem with my van


Thread

Mark @ 07825477[redacted] 22/03/2017 14:54:11 status: sent
        Hi boss, i got a problem with my van
Mark @ 07825477[redacted] 22/03/2017 14:56:07 status: sent
        Hi boss, i got a problem with my van
Mark @ 07825477[redacted] 22/03/2017 14:58:52 status: sent
        Hi boss, i got a problem with my van
Mark @ +447825477[redacted] 22/03/2017 15:00:40 status: received
        Hi Phil, can you bring it the garage i am free now
Mark @ 07825477[redacted] 22/03/2017 15:01:06 status: sent
        Sure i will on my way now
Mark @ +447825477[redacted] 22/03/2017 16:49:33 status: received
        Just sent the results by email.I am afraid it is gonna cost u a fortune
mate.there's a cheaper n profitable option if u interested. Don't be afraid son lol.
Mark @ 07825477[redacted] 22/03/2017 16:50:52 status: sent
        I will check mail n get back to u brother
Mark @ 07825477[redacted] 22/03/2017 17:07:09 status: sent
        OMG Mark! This is a lot man. Its either i sell the car cheaper or take ur
option. I am going 4 ur option man i am game
Mark @ 07825477[redacted] 22/03/2017 17:07:56 status: sent
        I am free tomorrow if you wanna play.
Mark @ +447825477[redacted] 22/03/2017 17:08:43 status: received
        yeah tomorrow is good.
Mark @ 07825477[redacted] 22/03/2017 17:09:47 status: sent
        Hey man, i am new to this what are the logistics involved.
Mark @ +447825477[redacted] 22/03/2017 17:10:41 status: received
        I will take care of everything son don't worry. Have a good sleep yeah
Mark @ 07825477[redacted] 22/03/2017 17:11:01 status: sent
        You are the best
```

Figure 21 Case 1: investigation – Mark and Davis communication



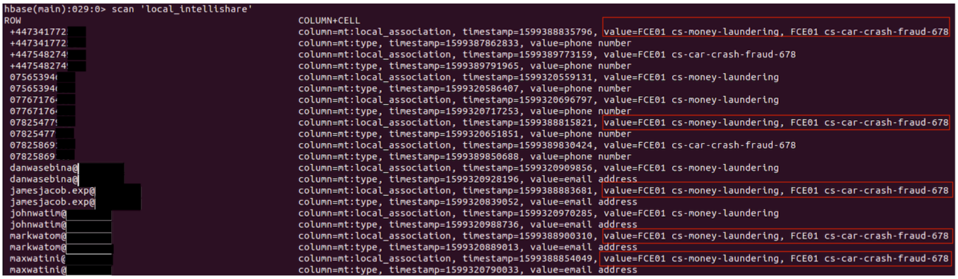
```
SMS
sent
To: 07391853[redacted] Contact name: Davis
Date: 13/03/2017 22:45:47
Message:Hi mate, got a job 2moro. I will let you know time n place yeah!


Thread

Davis @ 07391853[redacted] 13/03/2017 22:45:47 status: sent
        Hi mate, got a job 2moro. I will let you know time n place yeah!
Davis @ +447391853[redacted] 13/03/2017 22:50:22 status: received
        No worries mate
Davis @ 07391853[redacted] 13/03/2017 22:51:21 status: sent
        Special job so payment later u know the drill
Davis @ +447391853[redacted] 13/03/2017 22:51:46 status: received
        Sweet
Davis @ 07391853[redacted] 14/03/2017 09:46:44 status: sent
        Morning, we gonna do it by splott roundabout @11
Davis @ +447391853[redacted] 14/03/2017 09:50:34 status: received
        Cool
```

Figure 22 Case 1: investigation – Mark and Chris communication



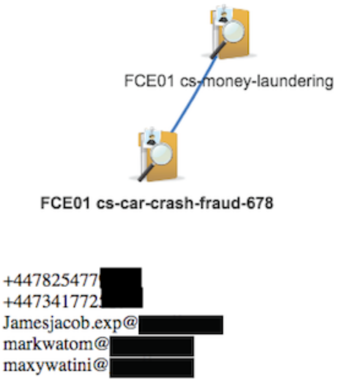
```
SMS
received
From: +447341772[redacted] Contact name: Chris
Date: 13/03/2017 22:35:22
Message:Hi Mark! My car has a problem


Thread

Chris @ +447341772[redacted] 13/03/2017 22:35:22 status: received
        Hi Mark! My car has a problem
Chris @ +447341772[redacted] 13/03/2017 22:36:06 status: sent
        What problem again?
Chris @ +447341772[redacted] 13/03/2017 22:38:22 status: received
        The rev count is just so high and the temp keep on
hitting the max. Had to stop many times on the m4
Chris @ +447341772[redacted] 13/03/2017 22:39:13 status: sent
        Might be a serious issue mate. Its going to cost u
alot i know that for sure
Chris @ +447341772[redacted] 13/03/2017 22:41:09 status: received
        I dont want to spend anymore.. Going to sell to webuy
any car people 2moro
Chris @ +447341772[redacted] 14/03/2017 11:04:37 status: received
        You are good at this mate!!  Good job :-)
```

Figure 23 Case 1: investigation – local intellishare table (see online version for colours)



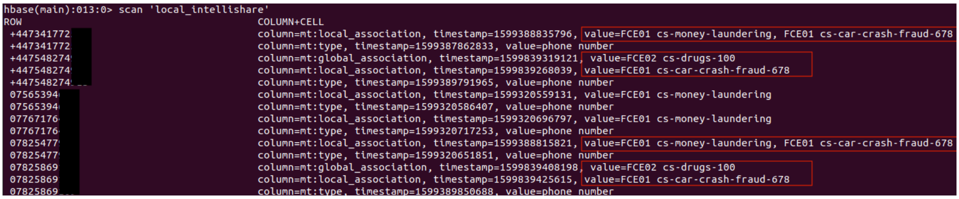
```
hbase(main):029:0> scan 'local_intellishare'
ROW                               COLUMN+CELL
 +447341772[redacted]             column=mt:local_association, timestamp=1599388835796, value=FCE01 cs-money-laundering, FCE01 cs-car-crash-fraud-678
 +447341772[redacted]             column=mt:type, timestamp=1599387862833, value=phone number
 +447548274[redacted]             column=mt:local_association, timestamp=1599389773159, value=FCE01 cs-car-crash-fraud-678
 +447548274[redacted]             column=mt:type, timestamp=1599389791965, value=phone number
 07565394[redacted]               column=mt:local_association, timestamp=1599320559131, value=FCE01 cs-money-laundering
 07565394[redacted]               column=mt:type, timestamp=1599320586407, value=phone number
 07767176[redacted]               column=mt:local_association, timestamp=1599320696797, value=FCE01 cs-money-laundering
 07767176[redacted]               column=mt:type, timestamp=1599320717253, value=phone number
 07825477[redacted]               column=mt:local_association, timestamp=1599388815821, value=FCE01 cs-money-laundering, FCE01 cs-car-crash-fraud-678
 07825477[redacted]               column=mt:type, timestamp=1599320651851, value=phone number
 07825869[redacted]               column=mt:local_association, timestamp=1599389830424, value=FCE01 cs-car-crash-fraud-678
 07825869[redacted]               column=mt:type, timestamp=1599389850688, value=phone number
 danwasebina@[redacted]           column=mt:local_association, timestamp=1599320909856, value=FCE01 cs-money-laundering
 danwasebina@[redacted]           column=mt:type, timestamp=1599320928196, value=email address
 jamesjacob.exp@[redacted]        column=mt:local_association, timestamp=1599388883681, value=FCE01 cs-money-laundering, FCE01 cs-car-crash-fraud-678
 jamesjacob.exp@[redacted]        column=mt:type, timestamp=1599320839052, value=email address
 johnwatim@[redacted]             column=mt:local_association, timestamp=1599320970285, value=FCE01 cs-money-laundering
 johnwatim@[redacted]             column=mt:type, timestamp=1599320988736, value=email address
 markwatom@[redacted]             column=mt:local_association, timestamp=1599388900310, value=FCE01 cs-money-laundering, FCE01 cs-car-crash-fraud-678
 markwatom@[redacted]             column=mt:type, timestamp=1599320889013, value=email address
 maxwatini@[redacted]             column=mt:local_association, timestamp=1599388854049, value=FCE01 cs-money-laundering, FCE01 cs-car-crash-fraud-678
 maxwatini@[redacted]             column=mt:type, timestamp=1599320790033, value=email address
```

Figure 24 Case 1: investigation – local intellishare network (see online version for colours)



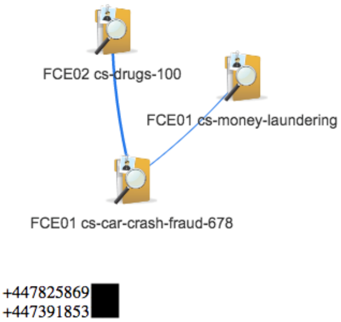
A global intellishare application was also run to find connections with FCE02 cases. The correlation between the cases is captured in Figure 25 and visualised using the intellishare network application and this is shown in Figure 26.

Figure 25 Case 1: investigation – global intellishare table (see online version for colours)



```
hbase(main):013:0> scan 'local_intellishare'
ROW                                         COLUMN+CELL
 +447341772[redacted]                       column=mt:local_association, timestamp=1599388835796, value=FCE01 cs-money-laundering, FCE01 cs-car-crash-fraud-678
 +447341772                                 column=mt:type, timestamp=1599387862833, value=phone number
 +447548274                                 column=mt:global_association, timestamp=1599839319121, value=FCE02 cs-drugs-100
 +447548274                                 column=mt:local_association, timestamp=1599839268039, value=FCE01 cs-car-crash-fraud-678
 +447548274                                 column=mt:type, timestamp=1599389791965, value=phone number
 07565394[redacted]                         column=mt:local_association, timestamp=1599320559131, value=FCE01 cs-money-laundering
 07565394                                   column=mt:type, timestamp=1599320586407, value=phone number
 07767176                                   column=mt:local_association, timestamp=1599320696797, value=FCE01 cs-money-laundering
 07767176                                   column=mt:type, timestamp=1599320717253, value=phone number
 07825477                                   column=mt:local_association, timestamp=1599388815821, value=FCE01 cs-money-laundering, FCE01 cs-car-crash-fraud-678
 07825477                                   column=mt:type, timestamp=1599320651851, value=phone number
 07825869                                   column=mt:global_association, timestamp=1599839408198, value=FCE02 cs-drugs-100
 07825869                                   column=mt:local_association, timestamp=1599839425615, value=FCE01 cs-car-crash-fraud-678
 07825869                                   column=mt:type, timestamp=1599389850688, value=phone number
```

Figure 26 Discovered criminal network (see online version for colours)



A connection between cases 1 and 3 is established. The connection between cases 1 and 3 was because of Chris and another suspect's mobile number which were in both cases.

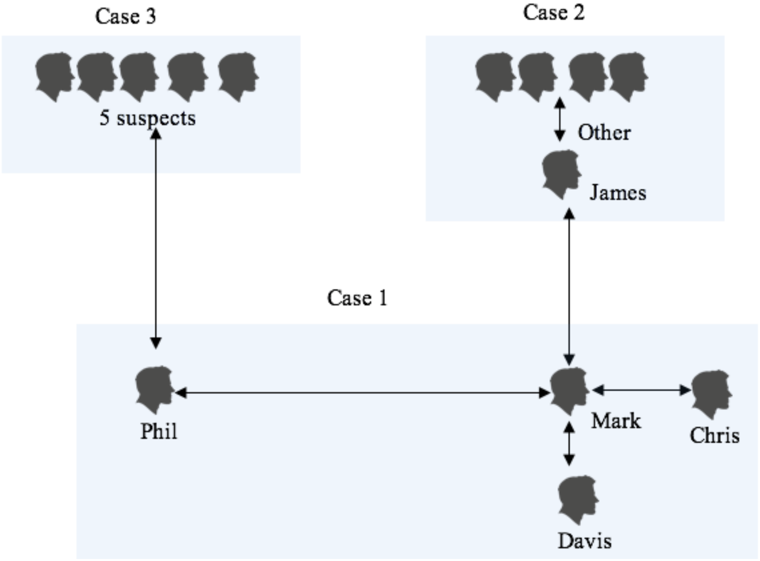
5 Discussion

It is very difficult to establish a connection between organised crime spanning multiple cases and jurisdictions. In this paper, we carried out a complex experiment to demonstrate the capability of the FCE intelligence sharing framework. We formulated three case scenarios that have artefacts that link the case stories. We did that in order to test the intelligence sharing framework.

We carried out an investigation on case 3 and extracted all the evidence except the evidence for one crucial suspect (Phil) which was missing from the collection. Similarly, we investigated case 2 and we were able to extract all the evidence except the details of one suspect (Mark) which was missing from the evidence collection. We also investigated case 1 and carried out an intelligence-gathering exercise. The intelligence sharing framework was able to find some correlations between cases 1 and 2 also, there was a link between cases 1 and 3.

The link between cases 1 and 2 was because of entities that belong to Mark and James which were in both cases. The connection between cases 1 and 3 was because of Chris and suspect's mobile number which were in both cases. A visualised criminal network is shown in Figure 27. The establishment of these correlations gave the investigators clues of who is involved in the crimes for further investigation.

Figure 27 Discovered criminal network (see online version for colours)



Without the intelligence sharing framework, the link between these cases was going to be difficult to establish if not impossible. As a result, the incorporation of an intelligence sharing framework to a forensic solution is critical as it gives the examiner the opportunity to gain more insight into a case. It helps in detecting organised crimes by providing situational awareness when criminal rings are discovered. Current forensic tools lack intelligence sharing, therefore links between crimes that extend beyond one police station cannot be identified.

It is worth noting that during the experiments, some default entities in the devices used, such as Microsoft e-mail addresses for Windows-based operating system and Google e-mail addresses for Android, were picked when building case entity tables. We filtred them out during the intelligence gathering process because they are not part of the cases.

We used the communication database to demonstrate the proposed framework, this provides a limitation as one needs to exactly know which entities to use to establish intelligence share. As such it may be possible to miss an entity and this may result in a missed link. It is necessary to fully automate the selection of the entities used for intelligence sharing based on the nature of the case.

6 Conclusions

We proposed an intelligence sharing framework on top of a FCE. Our framework will facilitate in establishing links between different case, detecting organising crimes, it is not restricted to the evidence sources, and suitable for big data forensics. The framework is based on extracting entities of interest from cases and then unifying them with entities from other cases using a MapReduce program. The use of MapReduce concept also speeds up the intelligence gathering process by parallel and distributed processing the evidence, especially when dealing with big data. We formulated three cases to demonstrate the proposed framework. The framework was used to find links between cases in the same FCE (FCE01) and between cases in a different FCE (FCE02). The framework performed accurately and was able to establish the connection between the cases. This exercise would not have been possible using traditional tools. Finally, we have also demonstrated the capability of FCE as a solution for big data forensics, as the total among of data, from the three cases, was almost 15.00 TB. With case 1 an example of a big data case with a total of 12.704 TB of data.

In future, we suggest incorporating techniques such as social network analysis (Tsai et al., 2019; Colladon and Remondi, 2017) to improve the intelligence sharing framework proposed in this paper. Other possible future improvements include using machine learning techniques to determine which objects to use for intelligence sharing. In this paper we used communication database, with machine learning, a variety of object could be used and ranked by importance depending on the type of cases investigated. For example, in a drug-dealing case, machine learning would be used to determine the importance of communication database, keywords, and GPS data and which one to use for intelligence sharing. Furthermore, issues of privacy have not been covered in this paper and remain of paramount importance in the future.

Acknowledgements

This work is part of the Botswana International University of Science and Technology (BIUST) Cyber Security Research Group (CSRG).

References

Adedayo, O.M. (2016) ‘Big data and digital forensics’, in *2016 IEEE International Conference on Cybercrime and Computer Forensic (ICCCF)*, IEEE, pp.1–7.

Beebe, N.L. and Clark, J.G. (2007) ‘Digital forensic text string searching: improving information retrieval effectiveness by thematically clustering search results’, *Digital Investigation*, Vol. 4, pp.49–54.

Borthakur, D., Gray, J., Sarma, J.S., Muthukkaruppan, K., Spiegelberg, N., Kuang, H., Ranganathan, K., Molkov, D., Menon, A., Rash, S. et al. (2011) ‘Apache Hadoop goes realtime at Facebook’, in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, pp.1071–1080.

Bunting, S. and Wei, W. (2006) *EnCase Computer Forensics: The Official EnCase Certified Examiner Study Guide*, John Wiley & Sons, Indianapolis, Indiana, USA.

Case, A., Cristina, A., Marziale, L., Richard, G.G. and Roussev, V. (2008) ‘FACE: automated digital evidence discovery and correlation’, *Digital Investigation*, Vol. 5, pp.S65–S75.

Colladon, A.F. and Remondi, E. (2017) ‘Using social network analysis to prevent money laundering’, *Expert Systems with Applications*, Vol. 67, pp.49–58.

Costantini, S., De Gasperis, G. and Olivieri, R. (2019) ‘Digital forensics and investigations meet artificial intelligence’, *Annals of Mathematics and Artificial Intelligence*, Vol. 86, Nos. 1–3, pp.193–229.

Cybersecurity Ventures (2019) *2019 Official Annual Cybercrime Report*.

Forte, D. and Power, R. (2007) ‘A tour through the realm of anti-forensics’, *Computer Fraud & Security*, Vol. 2007, No. 6, pp.18–20.

Garfinkel, S.L. (2006) 'Forensic feature extraction and cross-drive analysis', *Digital Investigation*, Vol. 3, pp.71–81.

Garfinkel, S.L. (2013) 'Digital media triage with bulk data analysis and bulk_extractor', *Computers & Security*, Vol. 32, pp.56–72.

Garfinkel, S., Farrell, P., Roussev, V. and Dinolt, G. (2009) 'Bringing science to digital forensics with standardized forensic corpora', *Digital Investigation*, Vol. 6, pp.S2–S11.

Gentry, E. and Soltys, M. (2019) 'Seaker: a mobile digital forensics triage device', *Procedia Computer Science*, Vol. 159, pp.1652–1661.

Goodman, S.E., Kirk, J.C. and Kirk, M.H. (2007) 'Cyberspace as a medium for terrorists', *Technological Forecasting and Social Change*, Vol. 74, No. 2, pp.193–210.

Hiscox (2019) *The Hiscox Cyber Readiness Report.*

Hitchcock, B., Le-Khac, N-A. and Scanlon, M. (2016) 'Tiered forensic methodology model for digital field triage by non-digital evidence specialists', *Digital Investigation*, Vol. 16, pp.S75–S85.

Irons, A. and Lallie, H.S. (2014) 'Digital forensics to intelligent forensics', *Future Internet*, Vol. 6, No. 3, pp.584–596.

Kim, H., Bruce, N., Park, S. and Lee, H. (2016) 'EnCase forensic technology for decrypting stenography algorithm applied in the powerpoint file', in *2016 18th International Conference on Advanced Communication Technology (ICACT)*, IEEE, pp.722–725.

Lallie, H.S., Shepherd, L.A., Nurse, J.R.C. Erola, A., Epiphaniou, G., Maple, C. and Bellekens, X. (2020) 'Cyber security in the age of COVID-19: a timeline and analysis of cyber-crime and cyber-attacks during the pandemic', *Computers & Security*, Vol. 105, p.102248.

Lam, C. (2010) *Hadoop in Action*, Manning Publications Co., Stanford, USA.

Lillis, D., Becker, B.A., O'Sullivan, T. and Scanlon, M. (2016) 'Current challenges and future research areas for digital forensic investigation', in *Proceedings of the Conference on Digital Forensics, Security and Law*, Association of Digital Forensics, Security and Law, p.9.

Mislan, R.P., Casey, E. and Kessler, G.C. (2010) 'The growing need for on-scene triage of mobile devices', *Digital Investigation*, Vol. 6, Nos. 3–4, pp.112–124.

Mitchell, F. (2010) 'The use of artificial intelligence in digital forensics: an introduction', *Digital Evidence & Elec. Signature L. Rev.*, Vol. 7, p.35.

Mohammed, H., Clarke, N. and Li, F. (2016) 'An automated approach for digital forensic analysis of heterogeneous big data', *Journal of Digital Forensics, Security and Law*, Vol. 11, No. 2, p.9.

Montasari, R. and Hill, R. (2019) 'Next-generation digital forensics: challenges and future paradigms', in *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, IEEE, pp.205–212.

Neuner, S., Schmiedecker, M. and Weippl, E. (2016) 'Effectiveness of file-based deduplication in digital forensics', *Security and Communication Networks*, Vol. 9, No. 15, pp.2876–2885.

Noel, G.E. and Peterson, G.L. (2014) 'Applicability of latent dirichlet allocation to multi-disk search', *Digital Investigation*, Vol. 11, No. 1, pp.43–56.

Quick, D. and Choo, K-K.R. (2014) 'Data reduction and data mining framework for digital forensic evidence: storage, intelligence, review and archive', *Trends & Issues in Crime and Criminal Justice*, Vol. 480, pp.1–11.

Quick, D. and Choo, K-K.R. (2018) 'Digital forensic intelligence: data subsets and open source intelligence (dfint + osint): a timely and cohesive mix', *Future Generation Computer Systems*, Vol. 78, pp.558–567.

Roussev, V., Quates, C. and Martell, R. (2013) 'Real-time digital forensics and triage', *Digital Investigation*, Vol. 10, No. 2, pp.158–167.

Rowe, N.C. (2013) 'Identifying forensically uninteresting files using a large corpus', in *International Conference on Digital Forensics and Cyber Crime*, Springer, pp.86–101.

Rowe, N.C. (2018) 'Associating drives based on their artifact and metadata distributions', in *International Conference on Digital Forensics and Cyber Crime*, Springer, pp.165–182.

Ruback, M., Hoelz, B. and Ralha, C. (2012) 'A new approach for creating forensic hashsets', in *IFIP International Conference on Digital Forensics*, Springer, pp.83–97.

Safaei, B., Monazzah, A.M.H., Bafroei, M.B. and Ejlali, A. (2017) 'Reliability side-effects in internet of things application layer protocols', in *2017 2nd International Conference on System Reliability and Safety (ICSRS)*, IEEE, pp.207–212.

Scanlon, M. (2016) 'Battling the digital forensic backlog through data deduplication', in *2016 Sixth International Conference on Innovative Computing Technology (INTECH)*, IEEE, pp.10–14.

Shalaginov, A., Johnsen, J.W. and Franke, K. (2017) 'Cyber crime investigations in the era of big data', in *2017 IEEE International Conference on Big Data (Big Data)*, IEEE, pp.3672–3676.

Sindhu, K.K. and Meshram, B.B. (2012) 'Digital forensics and cyber crime datamining', *Journal of Information Security*, Vol. 3, No. 3, pp.196–201.

Smith, C., Dietrich, G. and Choo, K-K.R. (2017) 'Identification of forensic artifacts in VMware virtualized computing', in *International Conference on Security and Privacy in Communication Systems*, Springer, pp.85–103.

Stelly, C. and Roussev, V. (2017) 'Scarf: a container-based approach to cloud-scale digital forensic processing', *Digital Investigation*, Vol. 22, pp.S39–S47.

Tabona, O. and Blyth, A. (2016) 'A forensic cloud environment to address the big data challenge in digital forensics', in *2016 SAI Computing Conference (SAI)*, IEEE, pp.579–584.

Tallón-Ballesteros, A.J. and Riquelme, J.C. (2014) 'Data mining methods applied to a digital forensics task for supervised machine learning', in *Computational Intelligence in Digital Forensics: Forensic Investigation and Applications*, pp.413–428, Springer, Switzerland.

Tropina, T. (2012) 'The evolving structure of online criminality: how cybercrime is getting organised', in *Eucrim-the European Criminal Law Associations' Forum*, No. 4, pp.158–165.

Tsai, F-C., Hsu, M-C., Chen, C-T. and Kao, D-Y. (2019) 'Exploring drug-related crimes with social network analysis', *Procedia Computer Science*, Vol. 159, pp.1907–1917.

Van Baar, R.B., Van Beek, H.M.A. and Van Eijk, E.J. (2014) 'Digital forensics as a service: a game changer', *Digital Investigation*, Vol. 11, pp.S54–S62.

Van Beek, H.M.A., van Eijk, E.J., van Baar, R.B., Ugen, M., Bodde, J.N.C. and Siemelink, A.J. (2015) 'Digital forensics as a service: game on', *Digital Investigation*, Vol. 15, pp.20–38.

Wagner, J., Rasin, A., Heart, K., Jacob, R. and Grier, J. (2019) 'DB3F & DF-toolkit: the database forensic file format and the database forensic toolkit', *Digital Investigation*, Vol. 29, pp.S42–S50.

White, T. (2012) *Hadoop: The Definitive Guide*, 3rd ed., O'Reilly Media, USA.

Williams, J. (2012) *ACPO Good Practice Guide for Digital Evidence*, Metropolitan Police Service, Association of Chief Police Officers, GB.

Yar, M. and Steinmetz, K.F. (2019) *Cybercrime and Society*, SAGE Publications Limited, UK.