European Commission
Directorate-General Home Affairs
Prevention of and Fight against Crime Programme

HOME/2010/ISEC/AG/INT/002
ForToo – Forensic Tools against Illegal Use of the Internet

D3: Integration Report

Work package(s):	WP3: Integration
Contractual delivery date:	30 Sep 2014
First publication date:	30 Sep 2014
Actual delivery date:	30 Sep 2014
Leading partner:	UGCS (University of South Wales)
Contributing partners:	University of Bristol
Editor:	Dr. Huw Read
Contributors:	Mr. Konstantinos Xynos, Dr. Theo Tryfonas
Internal Reviewers:	Prof. Iain Sutherland, Dr. John May
Version:	1.1

Executive Summary:

This report (a.) documents the tool integration specification with the forensics version of the DEViSE platform for security data analytics and visualization, and (b.) provides guidance to third-party developers on how to integrate their own tools with this platform.

With the support of the Prevention of and Fight against Crime Programme
European Commission - Directorate-General Home Affairs

This project has been funded with the support of the *Prevention of and Fight against Crime* Programme of the European Commission - Directorate-General Home Affairs. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Deliverable Details

Version Number of Deliverable:	Version 1.1

Contract Number:	N/A

Date of Document:	30 Sep 2014

Details of Deliverable Authors:	Dr. Huw Read Faculty of Advanced Technology, University of South Wales Pontypridd, Rhondda Cynon Taf CF37 1DL Telephone: 01443 65287 Email: huw.read@southwales.ac.uk
	Mr. Konstantinos Xynos Faculty of Advanced Technology, University of South Wales Pontypridd, Rhondda Cynon Taf CF37 1DL Telephone: 01443 654000 Email: k.xynos@southwales.ac.uk Dr. Theo Tryfonas Faculty of Engineering University of Bristol Bristol BS8 1TR
Internal Reviewers	Professor Iain Sutherland USW Commercial Services, University of South Wales Pontypridd, Rhondda Cynon Taf CF37 1DL Dr John May Faculty of Engineering University of Bristol Bristol BS8 1TR

Table of Contents

1. Introduction

This document has been created as the D3 deliverable milestone, Integration Report. It is intended to demonstrate how the tools developed through the duration of the project can be effectively integrated with the data storage, analysis and visualization platform developed by USW (UGCS). This document is separated into two parts:

- Main report – containing guidance on use of the platform and development of applications in a way that allows for data integration.
- Appendix – featuring a case study, walking through integrating mobile evidence (forensic artefacts acquired from mobile phones) to the visualization platform.

The two parts of this deliverable are closely related. In more detail, the first part presents a detailed walkthrough of typical system usage, expanding upon the concepts highlighted subsequently in the appendix. The appendix is in essence an academic paper that was written and submitted to a high profile conference with a European focus, detailing the architecture of the system and its benefits.

A worked example using the developed system is provided; this is intended to be used as a tutorial to assist understanding of the platform. It takes the form of a case study, with a forensic investigator trying to obtain a greater understanding of the data present on a set of mobile phones. The investigator uses the system to visualize data of interest and to pass subsets of relevant data between different visualization tools.

The appendix contains the paper titled *"An extensible platform for the forensic analysis of social media from multiple devices"* that has been submitted to the Digital Forensics Research Workshop (DFRWS). The paper has been submitted to the European version of the aforementioned venue (DFRWS EU[1]) on the 22nd September 2014 and the notification of acceptance is expected on 10th December 2014. As the acceptance results are unknown at the time of writing, we have not included this publication to the formal publications list of the project.

[1] http://www.dfrws.org/2015eu/

2. Using DEViSE

Starting DEViSE

The visualization process always begins with the History Manager application. It is a management front-end tool that tracks the progress from one tool to another, capturing the users progress through the data set. The History Manager is a front-end to the invocation process that happens behind the scenes when one visualisation invokes another via its context sensitive objects. This allows previous visualisations to be re-run later.

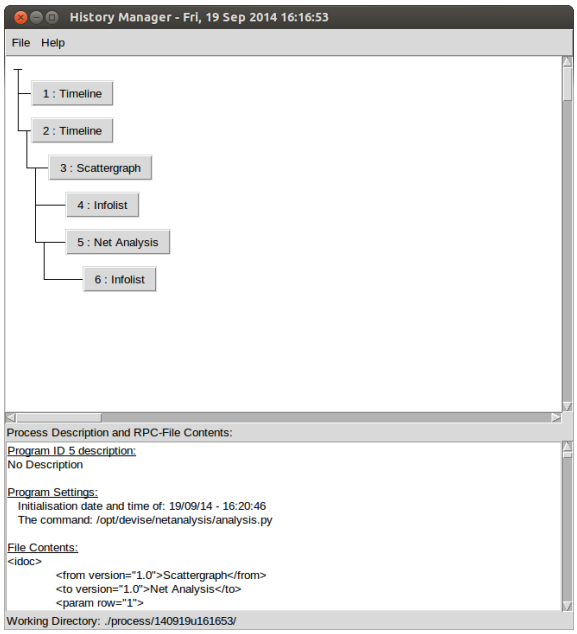


Figure 1: History Manager logs an investigators visualization activities

Starting Visualisation Tools

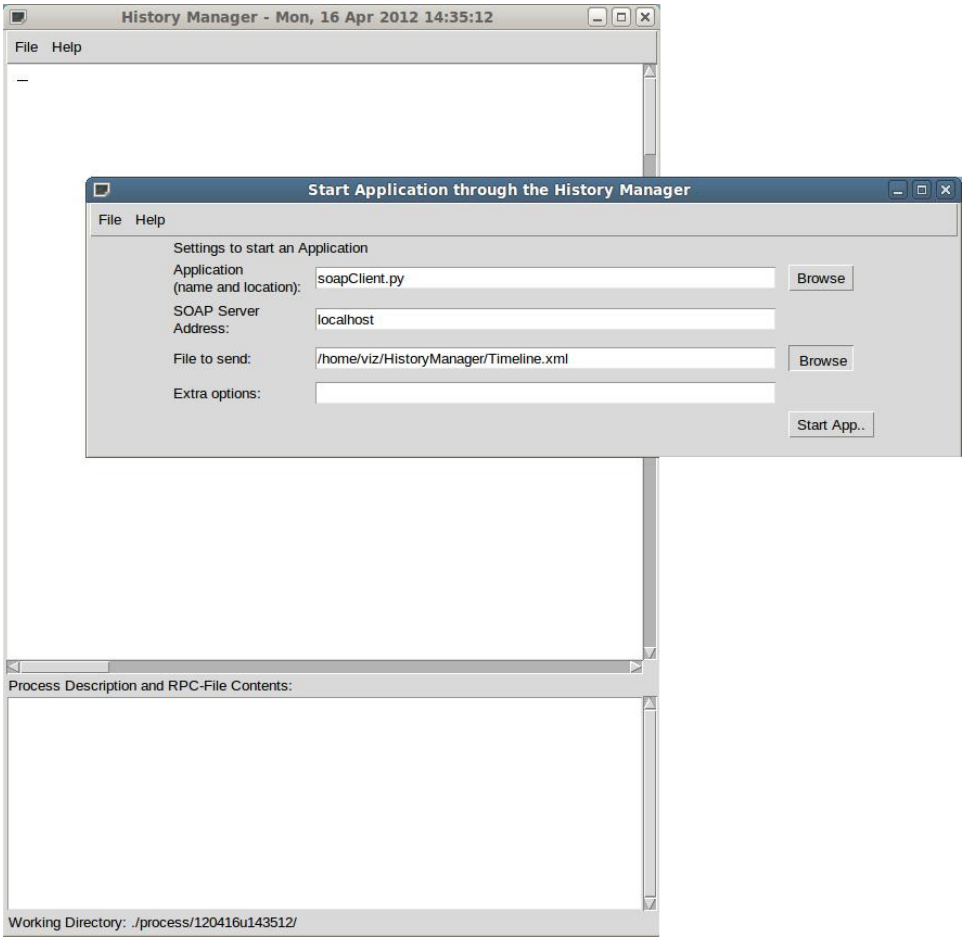


Figure 2: Selecting File > Start Application to Initiate new Visualisation Tool

Visualisation tools must initially be invoked via the History Manager. Although the tools are separate applications, the History Manager records their interactions with other tools. Selecting File > Start Application introduces a new window as seen in Figure 2. A number of options are presented:

- Application (name and location)
  This is the name of the back-end manager and should not be changed.
- SOAP Server Address
  This is the address of where the Cloud data store/query interface lies. Under normal circumstances, this is on the local machine (localhost).
- File to send
  This is where an investigator chooses an application to launch. Rather than launching an executable directly, a script is run containing details such as the path to executable and extended metadata including the tool name, description and additional default arguments to the appended at launch.
- Extra options
  Different applications may have special options that can be set as command-line arguments. If known, they may be set in the "Extra options" field and passed to the application at launch.

In this particular example, a visualization tool called Timeline has been selected.

Sending Data Between Visualisation Tools

Different areas in visualization tools act like hotspots; these are context-sensitive areas in the visualization that represent a type and quantity of data. What can or cannot be a hotspot has not been formally defined; just intuitively placed such that an investigator using the tool can interact with the application. In Figure 3, the background of the Timeline tool is just such a hotspot. The Appendix has a summary of typical hotspots found in some applications in DEViSE.

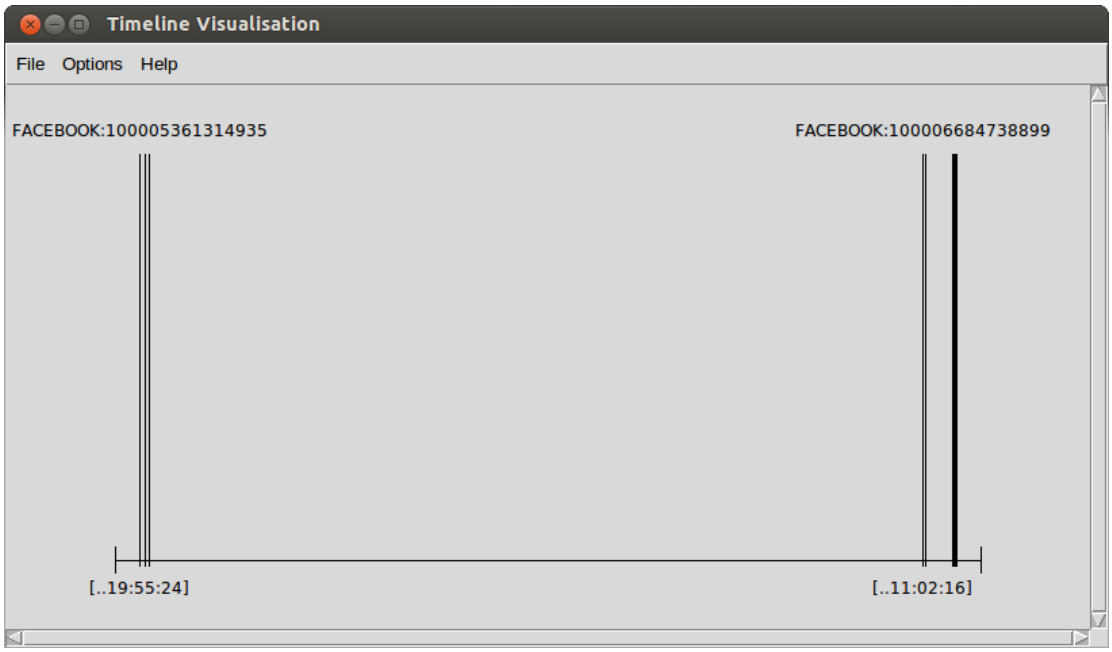


Figure 3: Timeline showing social media usernames

When an investigator interacts with a hotspot, i.e. by right-clicking or selecting a set of information, DEViSE performs a real-time scan to find other visualization applications that accept the data type and quantity. The results are presented in the form of a menu at the location of the interaction. This can be seen in Figure 4. The Scattergraph tool has been selected to visualize the selected set of data from the Timeline tool.

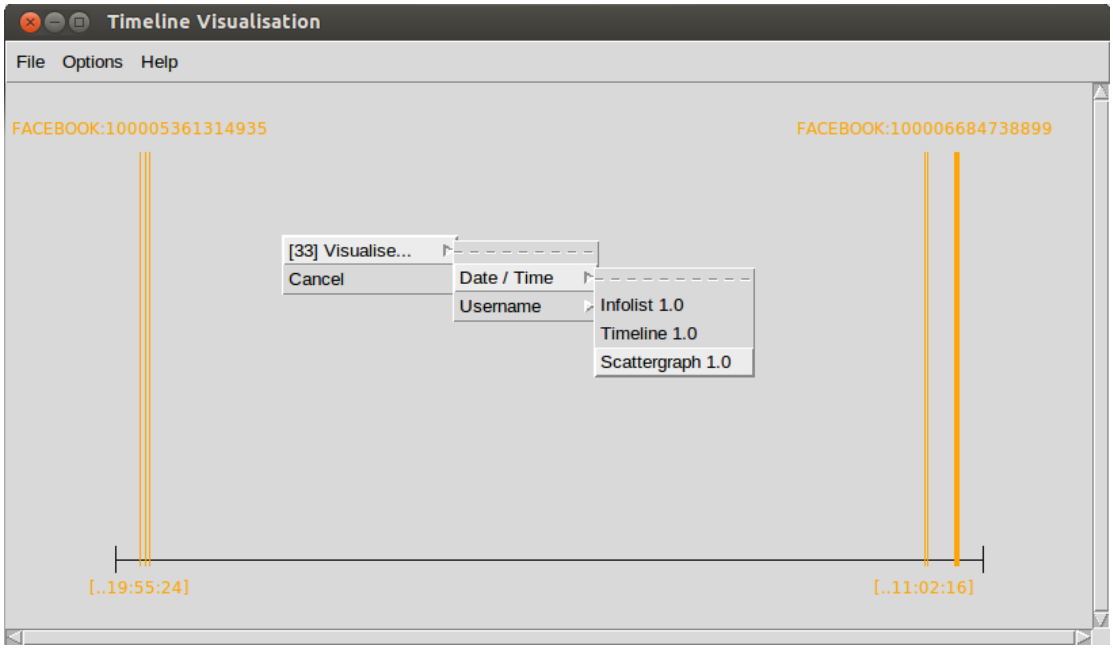


Figure 4: Selecting a compatible visualisation tool

After the Scattergraph tool has been selected, the Scattergraph application loads and visualizes the subset of data from the Timeline application (Figure 5). As all the data from Timeline has been selected, the dataset in Scattergraph remains the same. The visualization however presents quite a different image to the originating Timeline.

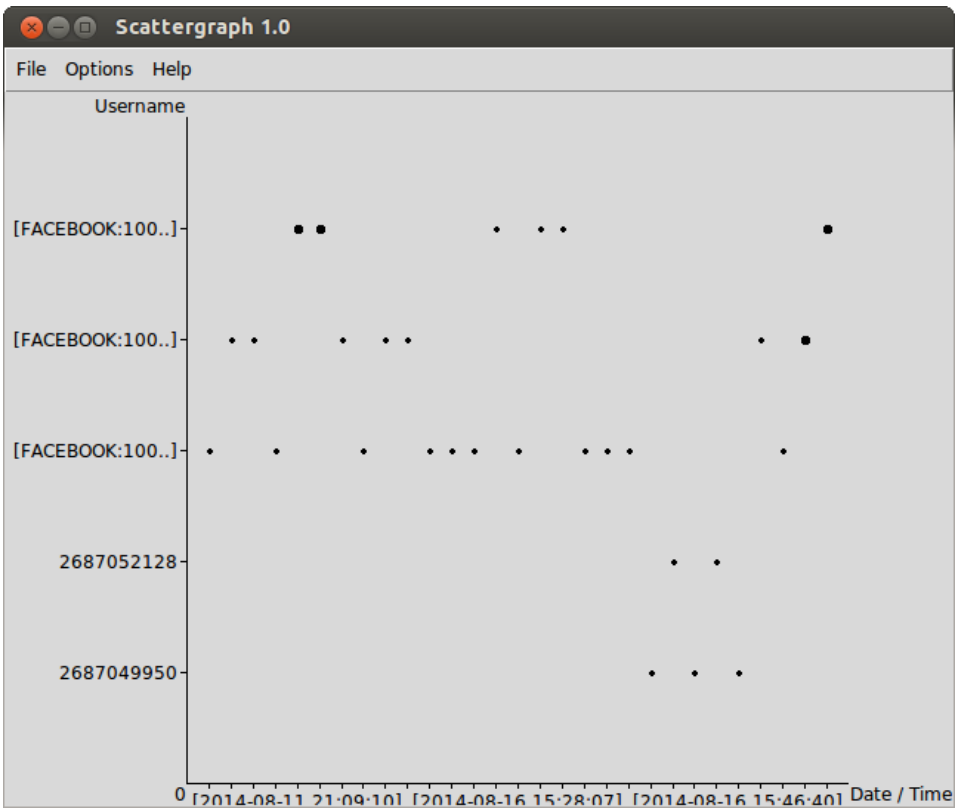


Figure 5: The Scattergraph receives data from the Timeline tool

Visualisation Tool Integration

Integrating new visualization tools into the existing visualization platform creates a number of new components:

- Configuration Document (C-DOC)

  The C-DOC holds all the configuration data about a tool, such as acceptable input, location on the hard disk and a brief description of the tools' function. There may be more than one C-DOC associated with any one tool.

- Information Document (I-DOC)

  An I-Doc is an XML file automatically generated by an application after a user clicks an object and selects a tool to pass data onto, stored with a unique name so that it is not overwritten.

These core components are described in further detail in Deliverable 2, section 2 "DEViSE Component". This should be referenced for the XML structure of the C-DOC and I-DOC formats.

During the development of the visualization tools, the structure of the configuration directories under Ubuntu have been altered from the original design deliverable D1. This decision was taken to reflect more closely the directory structure of Linux. The following structure describes the locations of DEViSE components:

`/opt/devise/tool_name`
The path to the visualization application.

`/etc/devise/toolconf`
The location of the tool Configuration Documents (C-DOC).

`/etc/devise/tool_name`
Any other settings required by the visualization tool.

`/etc/devise/namemap`
The list of all available data types visualization tools can query.

Applications must implement the following functionality in order to be compatible with the other visualization tools:

- Sending data via I-DOCs between visualization tools as described in D2.
- Identifying other tools to which the application can send data. This is performed by identifying matches in the C-DOCs.

For both of these functions, a shared python library has been developed and can either be imported directly or the source code analysed for implementation in another programming language.
`/usr/lib/python2.7/dist-packages/devise.py`

Case Study Scenario

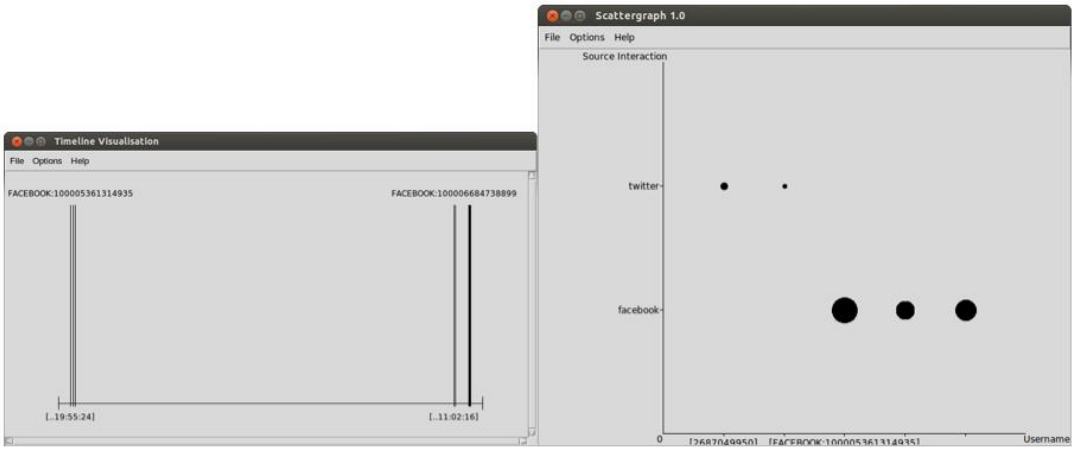
The potential sources for social media data include the user's end device, communication network and the corporate servers supporting the application. Access to corporate servers may be a lengthy legal process leading investigators to consider other possible sources. Therefore the case study uses data collected directly from smart phones.

A simple scenario was developed to test the system, with three volunteers carrying out typical social media communication across several platforms such as instant messages, voice over IP and different devices. As part of the study the individuals also shared documents using Google Drive and email. The simulated criminal activity is the selling of illegally obtained information and software. Much of the discussion between the phone owners take place over Facebook and Twitter; in particular making use of Twitter's direct messaging function and Facebook's Messenger application.

The tasking was to examine the social media communications to identify any evidence of nefarious activity. The History Manager captures the process used by the investigator. However, the thread of communication occurs over several social media types. In particular the following are found during an investigation of applications found on the cell phones:

- com.twitter.android/db/xxxxxxxxxx.db
- com.facebook.orca/db/threads_db2

These SQLite files contain private messages between the phone owner and others. The investigator begins by ingesting the files of interest into the data store. Once these have been parsed, the investigator opens the Timeline tool to obtain an appreciation of the timescales in the data.



Username	Date / Time	Message Content	Source Interaction
FACEBOOK:100006911594917	2014-08-16 15:27:29	Ok then I know how to get hold of you	facebook
FACEBOOK:100005361314935	2014-08-16 15:28:07	Good	facebook
FACEBOOK:100006911594917	2014-08-16 15:29:07	None	facebook
FACEBOOK:100006911594917	2014-08-16 15:29:22	Use this	facebook
FACEBOOK:100005361314935	2014-08-16 15:29:31	Ok i will	facebook
FACEBOOK:100005361314935	2014-08-16 15:30:21	J Tiberius Kirk sent a sticker.	facebook
FACEBOOK:100005361314935	2014-08-16 15:31:06	I will start on this and i will contact pike	facebook
2687049950	2014-08-16 15:44:35	I have got hold of some nice goods that we can work with	twitter
2687052128	2014-08-16 15:45:12	Great, lets meet up and I can get some samples from you	twitter
2687049950	2014-08-16 15:45:52	thats sounds good I will be available after 3	twitter
2687052128	2014-08-16 15:46:19	lets say 4 then	twitter
2687049950	2014-08-16 15:46:40	ok	twitter
FACEBOOK:100006684738899	2014-08-16 15:49:42	I think we are ready now	facebook
FACEBOOK:100005361314935	2014-08-16 15:50:08	Great	facebook
FACEBOOK:100006684738899	2014-08-16 15:50:40	We are ready to start	facebook

Figure 6: Rapid visual investigation. Timeline (Top Left) to Scattergraph (Top Right) to InfoList (Bottom)

The investigator then searches for correlations between the different usernames and social media formats found. To do this the investigator interacts with the Timeline tool (Figure 6, Top Left) by clicking the background, which initiates a search for other tools in the architecture that can take these visualised data-types as input. The Scattergraph tool is selected, and the dataset is transferred, as an IDOC XML file. This is stored in the History Manager and can be accessed at a later date if required.

The Scattergraph tool (Figure 6, Top Right) initially shows the two types of data it was provided (Date/Time and Usernames). The axes are altered to show the Source Interaction against the Usernames. This highlights how users sent messages via different social media formats. From this visualisation, the investigator now wants to look at the actual messages sent, across all social media types, ordered by time.

The Infolist application (Figure 6, Bottom) provides a view that lets the investigator see the messages sent in time order. It includes all social media formats from this case study and is not focused purely on one type.

3. Integration Specification for DEViSE Applications

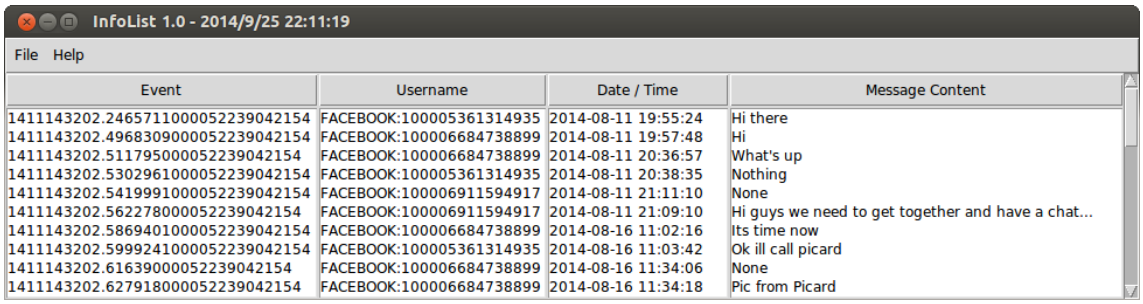


Figure 7: Infolist

Name

Infolist, version 1.0

Classification

Informational, Raw-data

Description

Infolist resembles spreadsheet applications like Microsoft Excel and OpenOffice Calc. It queries the database based on the types of data required and creates lists of each data type. It can visualise any type of data in the database.

Objects

Column Headings

Left-clicking on a heading arranges all the data by that data type, by ascending / descending order.

Data Items in Lists

Left-clicking on items of data in a list selects / deselects multiple items. After making a selection, right-click provides the user with a sub-list of tools that can accept the number of selected items and the data-type represented by the list.

Additional Features

Before Infolist is loaded a front-end application is initialised which confirms the data-types currently being sent. The user may manually add extra data-types by right-clicking in the indicated area. Upon clicking OK Infolist retrieves the new list and displays it in a window.

Figure 8: NetAnalysis

Name

Net Analysis Visualisation, version 1.0

Classification

Analytical, Logical

Description

A line is drawn between the source and destination of an event. This application is particularly effective visualising multiple attacks coming from a single source or going to a single destination.

Objects

Source IP Addresses

Right-clicking a source IP address provides a list of tools that accept this data type as input.

Destination IP Addresses

Right-clicking a destination IP address provides a list of tools that accept this data type as input.

Horizontal Lines

Right-clicking a line provides a list of tools that accept event ids, source IP's, destination IP's and the source / destination IP combo type.

Additional Features

Hovering the mouse on a line provides detailed information about the event type.

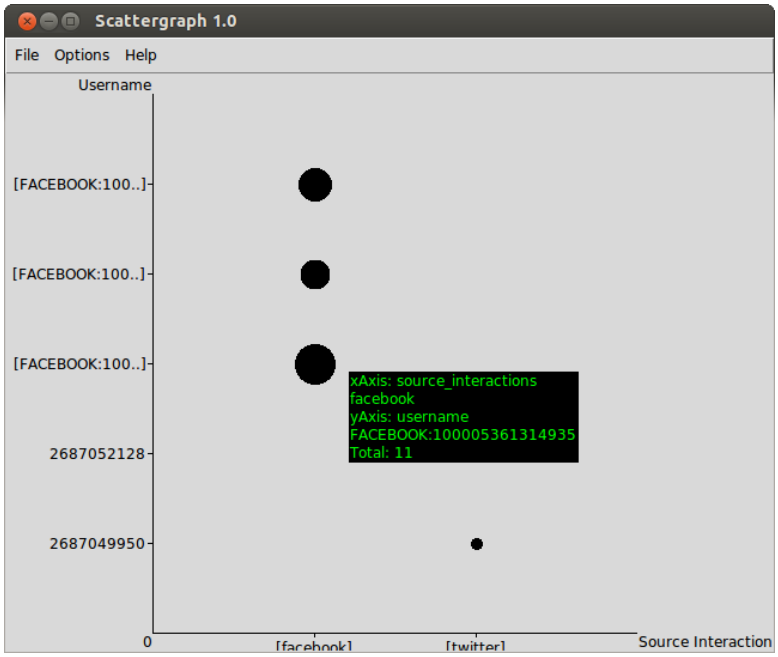


Figure 9: Scattergraph

Name

Scattergraph, version 1.0

Classification

Informational, Raw-data

Description

Scattergraph provides a horizontal and vertical axis and takes any form of data available in the database with which is put on the axis. Data on the axis is, by default, sorted alphabetically, however special consideration is given to date/time and ip address types. Date/time is converted into a single integer value before sorting to ensure correct time order is kept. IP addresses are converted into integer representations to ensure they are sorted correctly. Dots are drawn at the position between horizontal and vertical when the respective data intersects. If there are several intersections at one point, the additional events are appended to the existing dot.

Objects

Dots

Right-clicking on any dot takes the collection of events represented by it and provides a sub-menu. The sub-menu contains a selection of tools that accept this many events, and either of the data-types currently being visualised as input.

The size of the dots indicate, relatively, which x/y intersects are the most common in the dataset.

Additional Features

Hovering over a dot provides a brief pop-up that lists the horizontal and vertical data the dot represents, along with a list of all the events held within it. The options menu contains an entry called “change axis” that refreshes the Scattergraph with different horizontal / vertical data. By holding and dragging the mouse, several dots may be selected simultaneously.

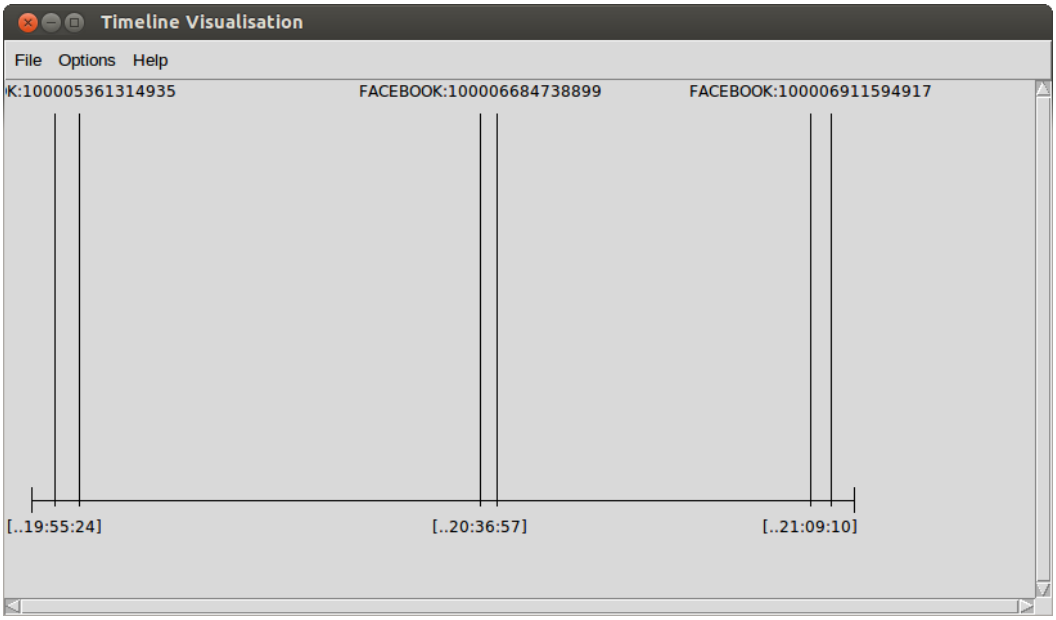


Figure 10: Timeline

Name

Timeline, version 1.0

Classification

Analytical, Temporal/Logical

Description

Timeline uses the date/time of an event to produce a diagram where the earliest event to occur appears on the left, the last event to occur appears on the right, and all other events in between are staggered accordingly. The time value appears at the bottom of the visualisation, whilst the eventid appears toward the top. The data toward the top can be changed. A slider bar near the bottom allows the timerange to be altered in real-time, stretching out the events making them easier to read, or bringing them closer together.

Objects

Text at top of lines

Right-clicking on any of the text boxes that appear at the top of lines will produce a sub-menu of tools that accept the type of data currently being shown and the event represented by that particular box.

Grey Background

Right-clicking on the grey background creates a sub-menu of the tools that accept the current data shown in the top text boxes and contains all the eventid's currently being visualised in the tool.

Additional Features

The first, “Timerange allows the user to manually enter an alternative start and end time. The database is queried and all events between these two dates are visualised. The second selection is a cascade menu entitled “Show me...” which opens to popular data types. Clicking on any of the types changes the text at the top of the lines.

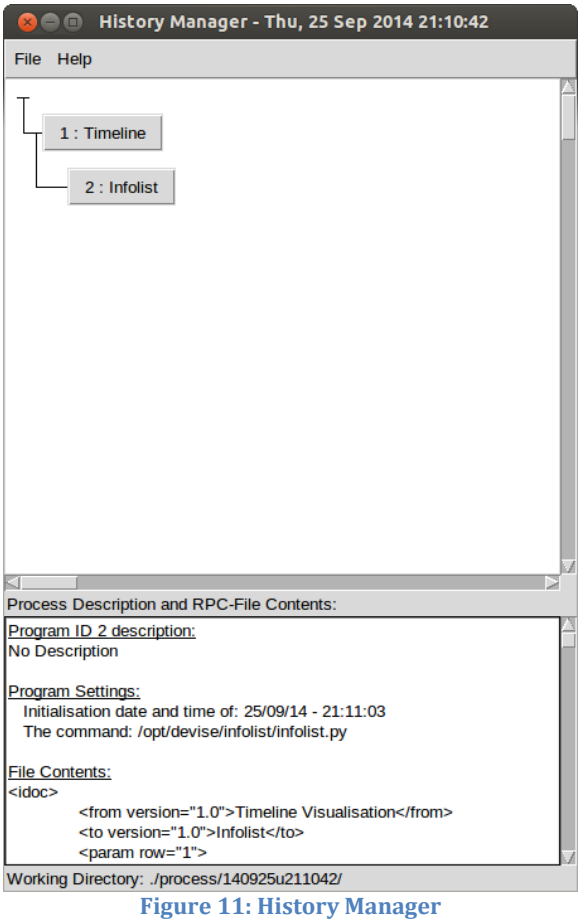


Figure 11: History Manager

Name

History Manager, version 1.0

Classification

Managerial, Logical

Description

History Manager is not a visualisation in the sense of the other tools; it is a front-end to the invocation process that happens behind the scenes when one visualisation invokes another via its context sensitive objects. It is a management front-end that tracks the progress from one tool to another and allows previous tool to be re-invoked later.

Objects

Buttons with tool names

Left-clicking a button will show a preview of the commands executed in the bottom pane. Right-clicking allows the user to re-invoke the tool with the specific settings to view the visualisation again.

The File menu provides additional functions, “Start Application”, “Load”, “Clear Window” and “Save”.

Appendix – Integration method case study

Based on "An extensible platform for the forensic analysis of social media from multiple devices", a paper submitted for review to the European version of the Digital Forensics Research Workshop (DFRWS EU, www.dfrws.org/2015eu/).

DFRWS

An extensible platform for the forensic analysis of social media data from multiple devices.

First Author[a], Second Author[b], Third Author[a,b],
First Author[a], Second Author[b], Third Author[a,b,*]

[a]*First affiliation, Address, City and Postcode, Country*
[b]*Second affiliation, Address, City and Postcode, Country*
[c]*First affiliation, Address, City and Postcode, Country*
[b]*Second affiliation, Address, City and Postcode, Country*

Abstract

Visualising data is an important part of the forensic analysis process. Many forensic tools have specialised visualisation components, but are not, as of yet, able to tackle questions concerning the broad spectrum of social media communication sources and devices. Visualisation tools tend to be stove-piped, it is difficult to take information seen in one visualisation tool and obtain a different perspective in another tool. If an interesting relationship is observed, needing to be explored in more depth, the process has to be reiterated by manually generating a subset of the data, converting it into the correct format, and invoking the new application. This paper describes a cloud-based data storage architecture and an implemented prototype consisting of interactive visualisation tools developed to allow for a more straightforward exploratory analysis. This approach developed in this tool suite is demonstrated using a case study consisting of social media data extracted from two mobile devices.

Keywords: Visualisation; social media; digital forensics; mobile device

1. Introduction

Visualisation is a subject of growing importance to the field of computer forensics. Specialist tools are required to support the analysis of large volumes of data from certain sources, one example being social media sites. In an age where personal digital communication devices that house information about our whereabouts, our conversations, or even current emotions are common [1], there are a multitude of ways that this information can be of value to a forensics investigation. Providing varied information on the different facets of an individual's activity [2]. The data sets can be extensive considering some of the new devices providing monitoring of personal fitness or health. It may be possible to query personal digital devices and social media for a wide range of data: What on-line names does an individual use? What information was sent to their contacts or received from their contacts? Even who had an elevated heart rate? The process of trying to identify a conversation traversing multiple communication mechanisms, Facebook, Twitter, WhatsApp, iMessage, email, and SMS across multiple devices further complicates the analysis and increase the volume of data.

Existing research has made some progress in addressing this issue [3], [4], [5] and [6]. However the difficulty currently faced by forensic examiners is that the process involves extensive data sets, which are time-consuming and require a degree of manual processing. There are current tools that are able to visualise aspects of the data as part of a digital forensics investigation. Current visualisation tools provide only limited interactivity and require an expert understanding of different visualisation tools, often with different input formats.

This paper presents an architecture and implemented prototype toolset that encourages direct interaction with the data in the visualisations to

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: author@institute.xxx .

compliment the exploratory and investigative mindset of an investigator. Social media application data is stored in an abstracted fashion into cloud-based storage and a defined API provides the interface for different visualisation tools to make requests for data. Visualisation tools retrieve data bound by their inherent data types and can send data directly to other visualisation tools depending on the pairing of compatible input/output data types. The paper describes the testing of the visualisation toolset using test data from smart phones.

The key contributions of this work are summarised below:

- We present a visualisation tool framework encouraging data exploration between tools.
- We demonstrate how heterogeneous social media data sources can be unified and exploited by investigators.
- We highlight a novel way of recording the visualisation trail to store the thought process of an investigator.

The rest of this paper is arranged as follows. Firstly, related work in the area of smart phone forensics visualisation is considered, secondly the architecture of the platform is described, thirdly a case study is presented to highlight the advantages of the platform, fourthly the outcome of the case study is analysed, finally the conclusions and future work are discussed.

2. Related Work

As stated by Garfinkel [7] forensic visualisations tend to serve two purposes, presentation and discovery. The former specialising in describing what has happened to a courtroom, the latter helping an investigator reach conclusions about some criminal activity. Discovery tools should be [7] data driven, have fixed, predictable, static output and be interactive for producing the visualisation. In addition to [7] we should also be seeking to record the data-mining activities the investigator uses in order to retrace through the thought/decision making process (Fig 1). The visualisations should indeed be repeatable so prosecution and defence can view the same output. In addition a means of visualising how an investigator derived their conclusions would also be beneficial when describing the process to a jury.

Social Media should be an area of considerable interest to the forensics investigator as it provides an insight into user actions and activities although there may be associated privacy, legal and ethical issues that may need to be considered [8].

Mutawa [9] has suggested when considering smartphones and social media that these devices are "...a goldmine for forensic investigators" and demonstrated that social media activity is retained on a number of smart phones, while recognising that this type of communication takes place across multiple devices [9]. A number of efforts have been made to visualize and examine the connections between mobile devices to better understand user activities [3]. Other tools have been applied to the visualization of social media including NodeXL [10] an extension of Microsoft Excel to enable a 'network graph' visualisation of data. There have also been efforts made into visualizing other aspects of social networks [4] and visualising the results. There are therefore existing tools, techniques and solutions of both commercial and open source forensics tools that can be applied to an investigation involving social media data. There are also a number of big data tools that can be used for social media analysis, including those tools used for e-discovery [11], [12], [13], [14] and [15]. The tool proposed in this paper bridges the gap between these tools providing an extensible social media analysis tool capable of visualisations supporting forensic analysis.

2.1. Cell phone forensics visualisations

The implemented prototype has been configured to analyse data from smart phones. Therefore the prototype is compared against existing tools available for forensic visualization of data from these devices. Commercial software is already looking at social-based data. However, as can be seen below, commercial applications are not yet able to visualise volumes of social networking information, instead relying on more manual query languages to obtain answers during an investigation. Social links tend to be made from more traditional digital communications, emails between individuals, SMS/MMS messages, telephone calls, etc. Open source software tends to follow a more generic approach, towards visualising any structured data rather than anything focusing on a specific niche like cell phone forensics.

The following considers the state of cell phone forensics and how recovered social media data is being incorporated into visualisations.

2.1.1. AccessData MPE+

MPE+ is the mobile phone forensics application from AccessData, producers of FTK and FTK Imager. MPE+ itself specialises in data extraction from mobile phones [16], and has its own visualisation component [17]. It uses SMS, MMS, and call histories to look for relationships, and has a social analysis that looks for relationships via email addresses. Mention is made of an SQL Builder feature that allows investigators to run custom queries on application data, but nothing specific is mentioned about correlations from social media sources.

2.1.2. Micro Systemation XRY and XAMN

Micro Systemation focuses on the data extraction with XRY and the visualisation with XAMN [18]. XAMN provides a number of different views including timeline, list, connections and geographic. Although XRY provides the means to extract application data, XAMN does not currently visualise social media communications.

2.1.3. CelleBrite UFED Link Analysis

CelleBrite provides presentation and discovery visualisations through the Link Analysis software [19]. Link Analysis provides the means to view commonalities across multiple cell phones in terms of phone calls, SMS, MMS, email, chat and a few others. Social media communications do not presently feature in Link Analysis.

2.1.4. Oxygen Forensics

Oxygen Forensics suite contains several different visualisation tools [20], timeline, web connections & locations, links & stats, but of particular interest is the Social Graph. The Social Graph looks at relationships between device owners and their contacts, i.e. how much time was spent using different communication mechanisms. Metrics include call length, number of messages sent and times of communication [20].

2.1.5. Open Source Software (OSS) Cell Phone Tools

OSS tools tend to concentrate on the extraction of data from the standard sources (call history, SMS, MMS, email [21]) and then on an application by application basis (e.g. WhatsApp [22], Skype [23]). Some of the tools generate lists as a visual aid for presentation. There are a number of OSS tools that lend themselves expertly to big-data social network visualisation [24], but are not designed for the type of focus required of the types of digital forensic investigations described in this paper.

To the best of the author's knowledge no-one has addressed the discovery side of visualisation for smart phone forensics in the open source community with a focus on the analysis and correlation of social media activity.

3. Architecture and Implementation

The architecture of the platform discussed in this paper is introduced below, highlighting the different core components of the storage, tool-to-tool data exchange and the audit process of the examiner.

3.1. Cloud storage

Sqlite is a database format found on most cell phones, as it is portable (i.e., store as files) and provides the advantages that come with relational database architectures. Mobile applications on Google's Android and Apple's iOS primarily make use of sqlite.

One of the main issues with these relational databases is the way the information is stored. This is usually in a standard format, as per the sqlite

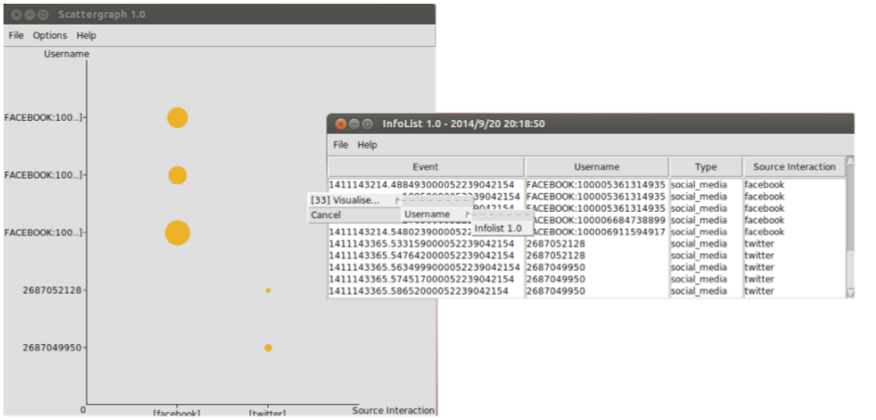


Figure 1: A scattergraph plot sends a list of usernames from gathered social media data to a listing tool

specification, although the actual table and column layouts are bespoke on the needs of the application. This poses a challenge to investigators and any tools they may rely upon. Applications are updated very frequently and it would not take much for a database design to change overnight.

In order to support the tools that have been developed, the database design proposed in [5, Fig. 2] provides a unifying database design for storing social media information. Individual parsers for Twitter and Facebook were created to convert the sqlite information into XML and then these were transformed them into the relevant social media XML format which is inserted into the cloud-based data store. Apache Cassandra was used to store the information and an XML middleware was then used to store and extract information when required by the visualisations.

3.2. Interacting with different visualisation tools

To facilitate data exchange between visualisations, the tools must incorporate the following components:

A configuration file (CDOC) that describes the quantity and type of data it can accept as input. These XML CDOCs are used whenever an investigator selects a subset of data within one visualisation tool. A list is presented to the investigator of other visualisation tools that can accept the subset as input.

An information file (IDOC) that is used to transport the data between tools. It is an XML file that describes where the relevant data is in the underlying data source that the receiving tool needs to query. The structure stores the source visualisation tool, the destination visualisation tool, the type of information to be visualised, and the identifier of the row entry that stores the information. In the following sample, we can see two tools exchanging information relating to the relationships between social media accounts (Fig. 3).

With these incorporated, the actual exchange of data becomes a straightforward task for the forensic investigator. As in Figure 1 above, an analyst selects the grouping or pattern they wish to investigate further, selects another visualisation tool that can receive such input from the library, and the new visualisation tool is invoked with the subset. The investigator can even create a new instance of the same visualisation tool (albeit with the reduced dataset) if they wish.

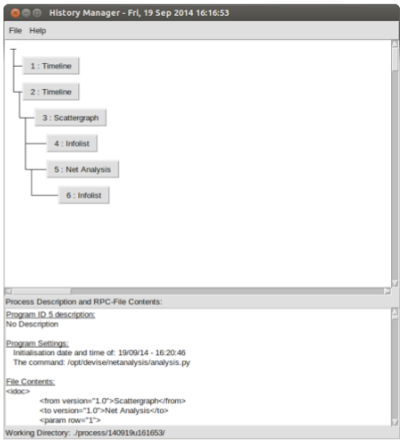


Figure 2: History Manager records an investigators progress

```
<?xml version="1.0" encoding="UTF-8" ?>
   <idoc>
       <from version="1.0">Scattergraph</from>
       <to version="1.0">Net Analysis</to>
       <param row="1">
           <object name="From Account Relationship">
               <data
type="from_account_relationships">1347865813.63255100005223
9042154</data>
           </object>
           <object name="To Account Relationship">
               <data
type="to_account_relationships">1347865813.6325510000522390
42154</data>
           </object>
       </param>
</idoc>
```

Figure 3: Sample IDOC XML

As the process of identifying compatible visualisation tools based on their data-types is performed in real-time, new visualisation tools can be integrated without closing down the entire system and disrupting the investigative process.

3.3. Recording the thought process

When a forensic practitioner is using visualisation as a mechanism to discover facts relating to an investigation, the process is typically unidirectional. Information flows from the dataset to the investigator. Large datasets are visualised, groupings of interest are investigated, subsets are extracted for further visualisation until a discovery is made (Fig. 4). With the architecture and prototype, we log the exchange that takes place between visualisation tools to let an investigator reinvoke existing visuals at a later date. As mentioned above, each visualisation tool is started with an IDOC XML file that describes the location of the data the tool needs to request data from the cloud data store.

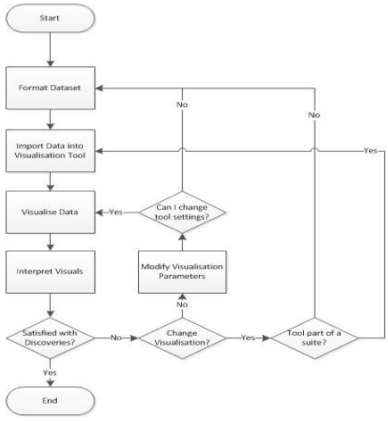
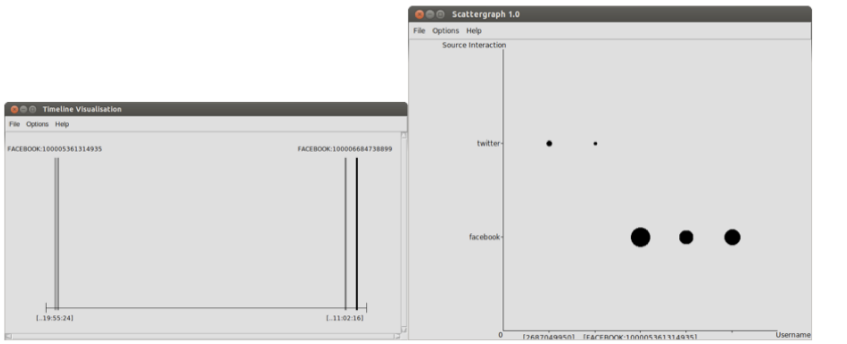


Figure 4: Recording the visualisation progression can reveal much about an investigation when expert testimony is required

As the investigator moves from one visualisation to another, we store the IDOC files and use a GUI tool called the History Manager (Fig. 2) to visualise the investigator's trail. Any views the investigator had seen previously are stored with the IDOC. The History Manager provides the facility to the investigator to select an existing IDOC and launch the visualisation tool, with the same inputs, and produce the same visual output again. The History Manager simply stores the files and other descriptive information in a hierarchical directory structure which can easily be stored and recalled for later use or archiving.

4. Case Study

There is a range of potential sources for social media data. Access to corporate servers may be a lengthy legal process leading investigators to consider other possible sources [25], [5]. Therefore the case study detailed in this paper uses data



Username	Date / Time	Message Content	Source Interaction
FACEBOOK:100006911594917	2014-08-16 15:27:29	Ok then I know how to get hold of you	facebook
FACEBOOK:100005361314935	2014-08-16 15:28:07	Good	facebook
FACEBOOK:100006911594917	2014-08-16 15:29:07	None	facebook
FACEBOOK:100006911594917	2014-08-16 15:29:22	Use this	facebook
FACEBOOK:100005361314935	2014-08-16 15:29:31	Ok i will	facebook
FACEBOOK:100005361314935	2014-08-16 15:30:21	) Tiberius Kirk sent a sticker.	facebook
FACEBOOK:100005361314935	2014-08-16 15:31:06	I will start on this and i will contact pike	facebook
2687049950	2014-08-16 15:44:35	I have got hold of some nice goods that we can work with	twitter
2687052128	2014-08-16 15:45:12	Great, lets meet up and I can get some samples from you	twitter
2687049950	2014-08-16 15:45:52	thats sounds good I will be available after 3	twitter
2687052128	2014-08-16 15:46:19	lets say 4 then	twitter
2687049950	2014-08-16 15:46:40	ok	twitter
FACEBOOK:100006684738899	2014-08-16 15:49:42	I think we are ready now	facebook
FACEBOOK:100005361314935	2014-08-16 15:50:08	Great	facebook
FACEBOOK:100006684738899	2014-08-16 15:50:40	We are ready to start	facebook

Figure 5: Rapid visual investigation: Timeline (Top Left) to Scattergraph (Top Right) to InfoList (Bottom)

collected directly from smart phones.

A simple scenario was developed to test the system, with three volunteers carrying out typical social media communication across several platforms such as instant messages, voice over IP and different devices. As part of the study the individuals also shared documents using google drive and email. The simulated criminal activity is the selling of illegally obtained information and software. Much of the discussion between the phone owners take place over Facebook and Twitter; in particular making use of Twitter's direct messaging function and Facebook's Messenger application.

The tasking was to examine the social media communications to identify any evidence of nefarious activity. The History Manager captures the process used by the investigator. However, the thread of communication occurs over several social media types. In particular the following are found during an investigation of applications found on the cell phones:

- com.twitter.android/db/xxxxxxxxxx.db
- com.facebook.orca/db/threads_db2

These SQLite files contain private messages between the phone owner and others. The investigator begins by ingesting the files of interest into the data store. Once these have been parsed, he opens the Timeline tool to obtain an appreciation of the timescales in the data.

The investigator then searches for correlations between the different usernames and social media formats found. To do this the investigator interacts with the Timeline tool (Fig. 5, Top Left) by clicking the background, which initiates a search for other tools in the architecture that can take these visualised data-types as input. The Scattergraph tool is selected, and the dataset is transferred, as an IDOC XML file.

This is stored in the HistoryManager and can be accessed at a later date if required.

The Scattergraph tool (Fig. 5, Top Right) initially shows the two types it was provided (Date/Time and Usernames). The axes are altered to show the Source Interaction against the Usernames. This highlights how users sent messages via different social media formats. From this visualisation, the investigator now wants to look at the actual messages sent, across all social media types, ordered by time.

The Infolist application (Fig. 5, Bottom) provides a view that lets the investigator see the messages sent in time order. It includes all social media formats from this case study and is not focused purely on one type.

5. Case Study Results and Evaluation

The case study highlights the advantages of the architecture. An investigator is able to easily transition from one tool to another, and rapidly data mine to gain an understanding of what has happened in their investigation. It should be clear from the case study that the emphasis in this paper is on the facilitation of interactivity within the architecture across different digital devices rather than any perceived novelty of the visualisations themselves.

The investigator has the ability to choose which visualisation tools they want to use when visually mining the data. In this fashion they are able to make discoveries more rapidly as the data mining process is effectively tailored to suit.

By looking for commonalities across social media formats we are able to process the data collectively. Using the right combination of visualisation tools conversation threads across different communication mediums can be aligned and interpreted in a more straightforward fashion than existing mechanisms.

The analysis process in the case study was kept short for brevity. However, true of both a quick triage and a full investigation, the History Manager stores all interactions between visualisation tools. The IDOCs are effectively a log or audit trail of an investigators visualisation thought process.

6. Conclusions

Visualisation of social media is of major importance to computer forensic investigators in an age where personal digital communication occurs over various applications running on multiple platforms. Existing tools and techniques for analysis include a range of commercial and open source forensics applications and Big Data visualisation tools. This paper has highlighted the lack of forensics tools that are appropriate for interactively visualising social media from multiple devices. This paper describes the testing of an open source platform and prototype tool capable of large scale forensics visualisation using social media collected from smart phones. Data is abstracted and stored in a cloud-based system the architecture includes a communication interface for multiple interactive visualisation tools to access the data. Visualisation tools retrieve data and can send data to other tools. This platform supports a forensic investigator's analysis by enabling multiple simultaneous interactive visualizations to be generated, while capturing the investigator's actions to record and log the process of the investigation. Future work will concentrate on the possibility of integrating other third-party visualisations into the platform, and the exploration of other potential evidence sources in a unified fashion to provide a more holistic view of an investigation.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement Number: HOME/2010/ISEC/AG/INT-002.

References

1. Samsung Galaxy S5 http://www.samsung.com/global/microsite/galaxys5/features.html
2. Cellebrite, *Cellebrite' s outlook for the mobile forensics industry 2014,* White Paper, 2014 http://www.cellebrite.com/collateral/OUTLOOK_FOR_THE_MOBILE_FORENSICS_INDUSTRY_2014_WP.pdf
3. Catanese S., A., Fiumara G., (2010) A visual tool for forensic analysis of mobile phone traffic, Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence Pages 71-76, ISBN: 978-1-4503-0157-2 DOI 10.1145/1877972.1877992
4. Adam Perer, Ben Shneiderman (2006) Balancing Systematic and Flexible, Exploration of Social Networks, IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 12, NO. 5, SEPTEMBER/OCTOBER 2006
5. Andriotis, P, Tzermias, Z, Mparmpaki, A, Ioannidis, S & Oikonomou, G 2013, 'Multilevel Visualization Using

Enhanced Social Network Analysis with Smartphone Data'. *International Journal of Digital Crime and Forensics*, vol 5., pp. 34-54
6. On the Development of Automated Forensic Analysis Methods for Mobile Devices, Andriotis, P., Tryfonas, T., Oikonomou, G., Li, S., Tzermias, Z., Xynos, K., Read, H. & Prevelakis, V. 2014 Trust and Trustworthy Computing - TRUST 2014. Holz, T. & Ioannidis, S. (eds.). Lecture Notes in Computer Science, Vol. 8564, p. 212-213 (Lecture Notes in Computer Science)
7. Garfinkel S.L. Forensics Visualizations with Open Source Tools, 2013. http://simson.net/ref/2013/2013-11-05_VizSec.pdf
8. Browning J.G. (2013) Keep Your "Friends" Close and Your Enemies Closer: Walking the Ethical Tightrope in the Use of Social Media. SMU Journal on Legal Malpractice and Ethics, 2013 Vol. 3, No.1.
9. Noora Al Mutawa, Ibrahim Baggili, Andrew Marrington (2012) Forensic analysis of social networking applications on mobile devices, Proceedings of the 2012 Digital Forensic Research Workshop. http://www.dfrws.org/2012/proceedings/DFRWS2012-3.pdf
10. Marc A. Smith, Ben Shneiderman, Natasa Milic-Frayling, Eduarda Mendes Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, Eric Gleave (2009) Analyzing (social media) networks with NodeXL, Proceedings of the fourth international conference on Communities and technologies, Pages 255-264, New York, NY, USA ©2009, ISBN: 978-1-60558-713-4 DOI 10.1145/1556460.1556497
11. Afentis Facebook Forensics Tool http://www.facebookforensics.com/features.html
12. Forte D. & Power (2006) Electronic discovery: digital forensics and beyond, Computer Fraud & Security, Vol. 2006(4), Pages 8-10
13. Ringtail E-discovery Tool http://www.ftitechnology.com/Products-Services/Software-and-Services/Ringtail/Ringtail.aspx
14. Attenex E-Discovery Software http://www.ftitechnology.com/Products-Services/Software-and-Services/Attenex.aspx
15. Xera I-conect http://www.iconect.com/
16. Access Data Mobile Phone Examiner http://www.accessdata.com/solutions/digital-forensics/mobile-phone-examiner
17. AccessData, Mobile Device data Visualization with MPE+, 2012. https://www.youtube.com/watch?v=bjcLDjju-kU
18. Micro Systemation, XAMN, 2014. https://www.msab.com/xry/xamn
19. Cellebrite, UFED Link Analysis, 2014. http://www.cellebrite.com/mobile-forensics/products/applications/ufed-link-analysis
20. Oxygen Forensics, Social Graph Tool, 2014. http://www.oxygen-forensic.com/en/features/analyst/social-graph
21. ViaForensics, Aflogical OSE, 2014. https://viaforensics.com/resources/tools/android-forensics-tool/#aflogical-ose
22. Ztedd, Whatsapp Xtract: Backup messages extractor, 2012. http://forum.xda-developers.com/showthread.php?t=1583021
23. Garronski, N., Skype Xtrator v.0.1.8.8, 2014. http://www.skypextractor.com/
24. Forensics WIKI, Graph and (Social) Network Visualization, 2013. http://www.forensicswiki.org/wiki/Tools:Visualization#Graph_and_.28Social.29_Network_Visualization
25. Martin Mulazzani and Markus Huber and Edgar R. Weippl, "Social Network Forensics: Tapping the Data Pool of Social Networks," in *Eighth Annual IFIP WG 11.9 International Conference on Digital Forensics*, 2012.