Assisting Digital Forensics Investigations by Identifying Social Communication Irregularities

Heloise Pieterse[a,b]
[a]Council for Scientific and Industrial Research, Pretoria, South Africa
[b]University of Pretoria, Pretoria, South Africa
hpieterse@csir.co.za

Abstract: People are creatures of habit, favouring the familiar over unpredictability, which causes them to usually follow set patterns or routines. This is especially true for social interaction or communication where people tend to prefer familiar or well-known social relationships with close family and friends. This preferred social communication is often reflected by people's usage of their smartphones. Change in social communication patterns can occur, but may be deemed exception or unusual and, therefore, such changes or irregularities must be identifiable. The identification of irregular social communication can offer insight to an examiner involved in a digital forensics investigation. Smartphone technology, however, evolves continuously, allowing for increasing amounts of social-related data to originate and be stored on the smartphone. In order to identify social communication irregularities with regards to smartphone usage, examiners are required to manually trawl through the data. This can become a very time-consuming process, leading the examiner to search for a needle in a haystack. This paper, therefore, introduces a new digital forensic analysis tool, called the Smartphone Modelling and Reconstruction Tool (SMaRT). The purpose of SMaRT is to specifically pinpoint social communication irregularities by analysing smartphone data. SMaRT achieves this functionality by combining data extraction, reconstruction and visualisation techniques to determine social patterns and locate irregularities. To determine the efficiency and evaluate the performance of SMaRT, a case study involving smartphone data is conducted. The outcome of the case study reveals that SMaRT can successfully analyse smartphone data and allows for the identification of patterns, associations and potential irregularities in an effective and timely manner.

Keywords: Smartphones, Data Analysis, Modelling, Digital Forensics, Social Communication.

1. Introduction

The rapid development of smartphone technology is leading to the design and creation of powerful devices. The powerful capabilities exhibited by smartphones that allow these devices to become integral in the daily activities of people. The reliance of users on the capabilities of smartphones, in conjunction with their ever increasing storage capacity, permits for large quantities of digital evidence to be stored such as contacts, text and instant messages, call logs, geographical data, electronic mail, web browsing history and multimedia activities (Jansen and Ayers, 2007). Such large and complex quantities of digital evidence can complicate investigations because of the time required to search for relevant digital evidence. In order to simplify investigations and quicken the analysis of the digital evidence obtained from smartphones, examiners can turn to mobile forensic tools.

Mobile forensic tools evolved along with the rapid improvements of smartphone technology. The latest available mobile forensic tools (Paraben's E3 Universal, Cellebrite UFED Pro, Oxygen Forensic Analyst and AccessData's Mobile Phone Examiner Plus (MPE+)) continue to improve and expand their provided functionality, which includes capabilities to extract and visualise data collected from smartphones running different mobile operating systems. These tools use accepted methods to recover digital evidence from smartphones (Curran et al, 2010) and provide techniques to visualise the recovered data. The focus of these forensics tools are, however, on extraction and visualisation of data and often provide limited analysis capabilities (Casey, 2009).

Manual analysis can become a very time-consuming process, causing the examiner to waste valuable time and potentially miss important digital evidence, which is not obvious from the raw data. This is especially true for the identification of regular or irregular social interaction and communication density. People are guided by known habits (Kilpinen, 2012) and social norms (Predergrast et al, 2008), which leads to a preference for familiar or well-known social relationships with close family and friends. Such relationships are reflected by people's usage of their smartphones. Change in social communication patterns can occur (such as meeting new people or contacting previously unknown individuals), but may be deemed an exception or unusual and therefore such changes or irregularities must be identifiable. The identification of irregular social communication can offer insight to an examiner involved in a digital forensics investigation.

There are currently no research studies or tools, to the best of the author's knowledge, that allows for the identification of irregularities with regards to social communication using smartphone data. This paper, therefore, introduces a new digital forensic analysis tool, called the Smartphone Modelling and Reconstruction Tool (SMaRT). SMaRT allows examiners to explore smartphone data and identify social interaction and the communication density of such interaction. The current implementation of SMaRT combines data extraction, reconstruction, and visualisation techniques to pinpoint any irregularities.

The rest of this paper is structured as follows. Section 2 describes existing research that involves the analysis of extracted mobile phone or smartphone data. Section 3 introduces the new digital forensic tool, SMaRT, and explains the design and functionality of the tool. A case study is presented in Section 4, demonstrating the efficiency and value of the new tool. Section 5 provides a short discussion on SMaRT, highlighting the strengths and limitations of the tool. The final conclusions and future work are summarised in Section 6.

2. Related Research

The continuous improvements of smartphone technology are equipping devices with various sensors that can act as witnesses for digital forensic examiners. These sensors can capture geographical locations, images, social communication, and audio (Pieterse and Olivier, 2014), all of which can easily be stored on the ever increasing storage capacity of smartphones. As smartphones become more popular, interest regarding the data stored on the devices continues to grow. Multiple studies have, therefore, focused on the mining and analysis of smartphone data.

Min et al (2013) introduced a computational model that allows for the classification of contacts according to the following life facets: family, work, and social. The computational model uses call and text message logs retrieved from mobile phones and extract from these logs features such as communication intensity, regularity, medium, and temporal tendency. Combining these features with machine learning techniques, contacts can be classified with 90% accuracy (Min et al, 2013). Weiss and Lockhart (2011) conducted a study that attempts to identify user traits, such as sex, height, and weight, by mining smartphone accelerometer data. The traits are identified by building predictive models from collected accelerometer data using supervised learning methods (Weiss and Lockhart, 2011). Altshuler et al (2012) show the possibility of predicting demographic information, such as ethnicity, age, and marital status of users, by analysing the personal features and behaviour properties of Short Message Service (SMS) messages (Altshuler et al, 2012). Chittaranjan, Blom and Gatica-Perez (2013) investigate the relationship between behavioural characteristics derived from smartphone data and the self-reported Big-Five personality traits, which are extroversion, agreeableness, conscientiousness, emotional stability, and openness to experience (Chittaranjan, Blom and Gatica-Perez, 2013). The behavioural characteristics are automatically extracted from the smartphone data, which includes call logs, SMS logs, Bluetooth scans, calling profiles and application usage. The outcome of the study presents a detailed analysis of the relationship between smartphone usage and the Big-Five personality traits (Chittaranjan, Blom and Gatica-Perez, 2013). GroupUs, proposed by Do and Gatica-Perez (2011), is a probabilistic relational model for sensing group interaction. The model uses Bluetooth data sensed by smartphones to analyse long-term dynamic social networks created by the physical proximity of people. The results produced by the model allow for the detection of different interaction types and also discover a variety of social contexts (Do and Gatica-Perez, 2011). Eagle and Pentland (2006) introduced a system for sensing complex social systems by using standard Bluetooth-enabled smartphones to measure information access. The outcome of the system shows how the collected data can be used to uncover structure and regular rules in the behaviour of individuals and organisations (Eagle and Pentland, 2006). Do and Gatica-Perez (2014) also focused on studying location characterisation of people's everyday activities by using smartphones that continuously record data. The collected data allow for the studying of human mobility, including the identification of visiting patterns and the categorisation of different places visited. The automatic labelling of locations is performed by using smartphone data only, without relying on any geographical information (Do and Gatica-Perez, 2014). Mobivis is a visual analytic tool for exploring smartphone data by presenting social and spatial information as one heterogeneous network. The tool supports the temporal and semantic filtering through an interactive timeline and can represent both individual and group behaviour (Shen and Ma, 2008). Min and Cho (2011) proposed SmartPhonebook, a tool that mines users' social network data to manage relationships by inferring social and personal contexts. SmartPhonebook uses icons and graphs to visualise social context, which allows users to understand their social situations (Min and Cho, 2011).

3. Smartphone Modelling and Reconstruction Tool (SMaRT)

Existing analysis approaches, as described in the previous section, focus mostly on the analysis of specific aspects of the data such demographic information, personality traits, and behavioural characteristics. There have been no approaches thus far that analyse smartphone data with the purpose of identifying regular or irregular social interaction and communication density. The focus of this paper is to introduce a new digital forensic analysis tool, called the Smartphone Modelling and Reconstruction Tool (SMaRT), which calculates and identifies the social interaction and related communication density for a specific set of smartphone data. The following subsections describe the design of SMaRT and the software application implementation respectively.

3.1 Design

SMaRT is created to act as a supporting tool for examiners to use during any digital investigation involving smartphone data. The tool allows for the analysis of social interaction and communication density using smartphone data with the purpose of identifying patterns, associations as well as potential irregularities. Social interaction is bi-directional communication between any two individuals, involving either telephonic calls or any form of electronic messaging (such as e-mails, text messages or instant messages). Communication density is described as the number and/or length of communication between two individuals. Using the calculated social interaction and communication density, it is possible to identify social communication irregularities. For the purpose of this research, irregular social communication is defined as the sudden increase or decrease in interaction between either known or unknown individuals.

To perform the analysis of social interaction and communication density, the current design and implementation of SMaRT consists of three components: data extraction, reconstruction, and visualisation. An illustration of SMaRT's architecture is presented in Figure 1, showing the flow of analysis between the components.

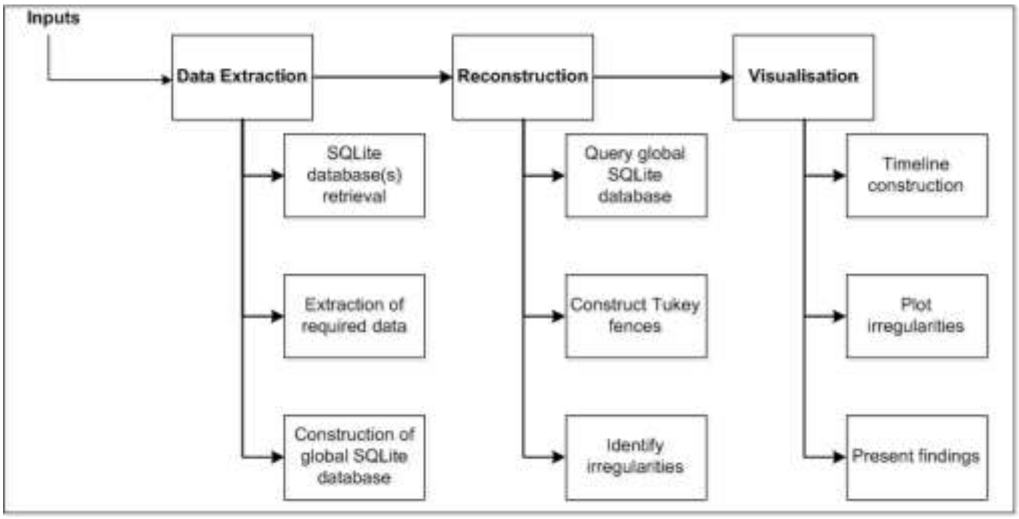

Figure 1: Architecture of the Smartphone Modelling and Reconstruction Tool

Required as input, the examiner must present SMaRT with all the available smartphone data, an investigation and visualisation periods. The investigation period is the time frame that indicates which smartphone data must be included in the analysis while the visualisation period is the time frame used to visualise the results to the examiner. Using the three supplied inputs, the data extraction component identifies and extracts the required smartphone data. Smartphone data is primarily stored in SQLite databases (Freiling, Spreitzenbarth and Schmitt, 2011) and since the current implementation of SMaRT focuses on social interaction and bi-directional communication, only social-related data is extracted from the SQLite databases. The data extraction component identifies the SQLite databases containing the social-related data and extracts from these databases the relevant data that falls within the defined investigation period. Structured Query Language (SQL), a well-known industry standard query language (Bakkum and Skadron, 2010), is used to retrieve the social-related data. SQL provides powerful statements (SELECT, WHERE and GROUP BY statements) that allow access to and the retrieval of the relevant data stored in the SQLite databases. The data is then transferred and inserted into the global SQLite database. The global SQLite database contains all the social-related data that falls within the investigation period.

The reconstruction component of SMaRT follows the extraction of the social-related data from the provided SQLite databases. The purpose of the reconstruction component is to reorganise the gathered data to give the examiner a better understanding of the transpired social interaction and communication density. The analysis and reconstruction are achieved using a collection of SQL queries and incorporating Tukey's method.

Social interaction is calculated using a collection of *SELECT*, *COUNT*, *WHERE*, *BETWEEN* and *GROUP BY* clauses. These SQL clauses analyse the social-related data in the global SQLite database and establish the social interaction between the smartphone user and other individuals for the selected investigation period. The communication density is calculated using the following collection of SQL clauses: *SELECT*, *SUM* and *GROUP BY*, between the smartphone user and other individuals for the selected investigation period. The Tukey method is then used to reconstruct the analysed social-related data to identify potential irregularities with regards to social interaction and communication density.

The Tukey method constructs a boxplot that conveys information about continuous univariate data and the identification of outliers (Tukey, 1977). The information includes the identification of the median, lower quartile, upper quartile, lower extreme, and upper extreme of a data set (Seo, 2002). The rules of Tukey's method are as follows (Seo, 2002):

- Calculate the first quartile (Q1).
- Calculate the third quartile (Q3).
- The Inter Quartile Range (IQR) is the distance between the lower (Q1) and upper (Q3) quartiles.
- Inner fences are located at a distance 1.5 x IQR below first quartile (Q1 - 1.5 x IQR) and above the third quartile (Q3 + 1.5 X IQR).
- Outer fences are located at a distance 3 x IQR below first quartile (Q1 - 3 x IQR) and above the third quartile (Q3 + 3 X IQR).
- A possible outlier lies between the inner and outer fences and is referred to as a weak outlier.
- A probable outlier is an extreme value beyond the outer fences and is referred to as a strong outlier.

Tukey's method is selected to identify potential social interaction and communication density irregularities since the method is less sensitive to data sets with extreme values and makes no distributional assumptions. Tukey's method also does not directly depend on the mean and standard deviation, which makes the method applicable to skewed or non-mound shaped data (Seo, 2002). The method is also appropriate for large data sets (Seo, 2002), making it the ideal method to detect irregularities within large collections of smartphone data. To reconstruct the analysed social-related data and identify the irregularities, the Tukey method accepts as input the data extracted using the specific SQL statements. The data is then used to construct the inner and outer fences, which allows for the detection of both weak and strong outliers. The detected outliers identify potential social interaction as well as communication density irregularities identified in the analysed data.

Visualisation is the final component of SMaRT and presents to the examiner with the outcome of the analysed and reconstructed social-related data. The presentation of the social interaction results is in the form of a timeline, which is structured according to the visualisation period selected by the examiner. The number of intervals drawn on the timeline is dynamically calculated and depends on the given visualisation period. The intervals represent a specified period of time such as an hour, day, month, or year. Table 1 shows the categories used to represent the analysed social interaction on the constructed timeline.

Table 1: Representation of Social Interaction

Social Interaction	Definition	Representation
Normal	Social interaction between two individuals is regular	Green
Weakly Irregular	Social interaction between two individuals is potentially irregular	Orange
Strongly Irregular	Social interaction between two individuals is irregular	Red

Communication density is visualised separately as a list of abbreviated phone numbers (to protect the identity of the user). Each phone number represents an individual the smartphone user communicated with and the calculated communication density is visualised in one of the following categories presented in Table 2.

Table 2: Representation of Communication Density

Communication Density	Definition	Representation
Low	Communication between two individuals occurs regularly	Green
High	Increase or decrease of communication between two individuals	Orange
Very High	Sudden increase or decrease of communication between two individuals	Red

With this new digital forensic tool, examiners can easily view the results of the analysed social interaction and communication density. SMaRT is, however, not designed to replace existing mobile forensic tools, but instead compliment these tools by overcoming limited analysis functionality.

3.2 SMaRT Application

In order for SMaRT to be accessible to examiners, a Java application was created according to the design described in the previous section. Java, which is a general-purpose computer programming language, was chosen to create and implement the application since the language allows for platform independence. The final compiled version of the SMaRT application can run on any computer that supports Java without requiring recompilation. Figure 2 shows the interface of the SMaRT application.

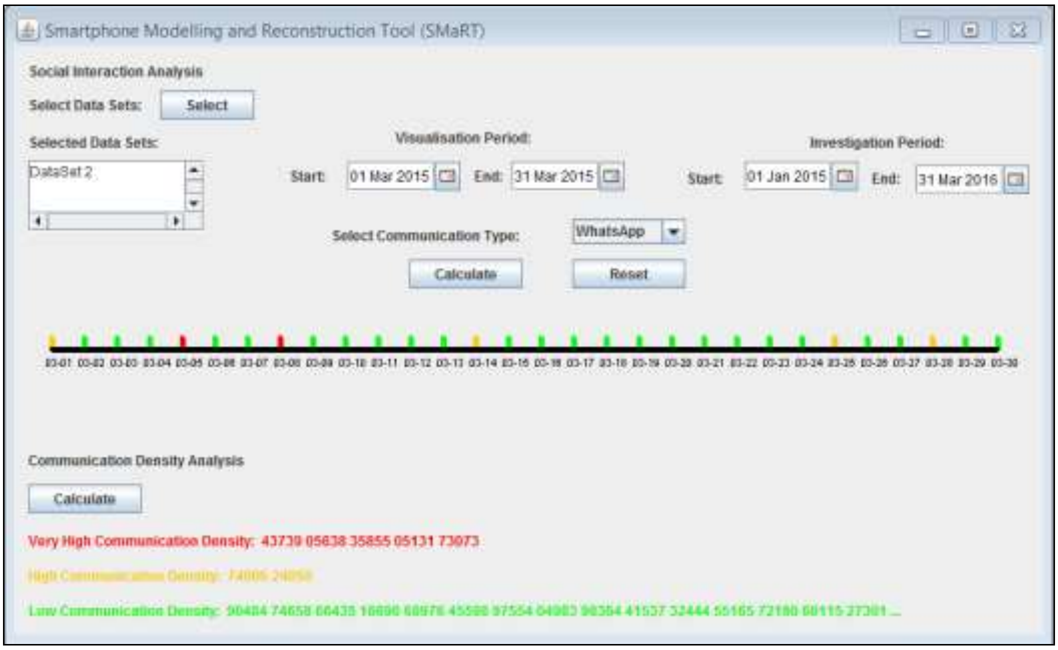

Figure 2: Interface of the SMaRT Application

The design of the interface for the SMaRT application is simple and minimalistic, allowing examiners to easily understand and comprehend the application. Using the provided functionality, the examiner can submit to the application the required input (smartphone data, an investigation and visualisation periods). The smartphone data, collected in multiple SQLite databases, is provided to the application in either a single or as multiple data sets. The examiner can then use the date dialogue boxes to select the appropriate start and end date for both the visualisation and investigation periods. In addition, the SMaRT application includes the option to select a specific communication type (text messages, multimedia messages, WhatsApp messages, Gmail e-mails, Skype messages, WeChat messages and call logs) that will dictate the focus of the analysis. By default, the SMaRT application includes all data from all communication types as defined by the selected investigation period.

To determine the efficiency and effectiveness of the SMaRT application, the performance of the application is measured by analysing various quantities of SQLite database records. The results of the performance measurements are shown in Figure 3 and identifies the time required (in seconds) to analyse a specific quantity of SQLite database records. On average, it takes the SMaRT application approximately 7 milliseconds to analyse an individual SQLite database record. The information provided by the performance measurement shows that the SMaRT application can effectively and in a timely manner analysis large quantities of SQLite database records.

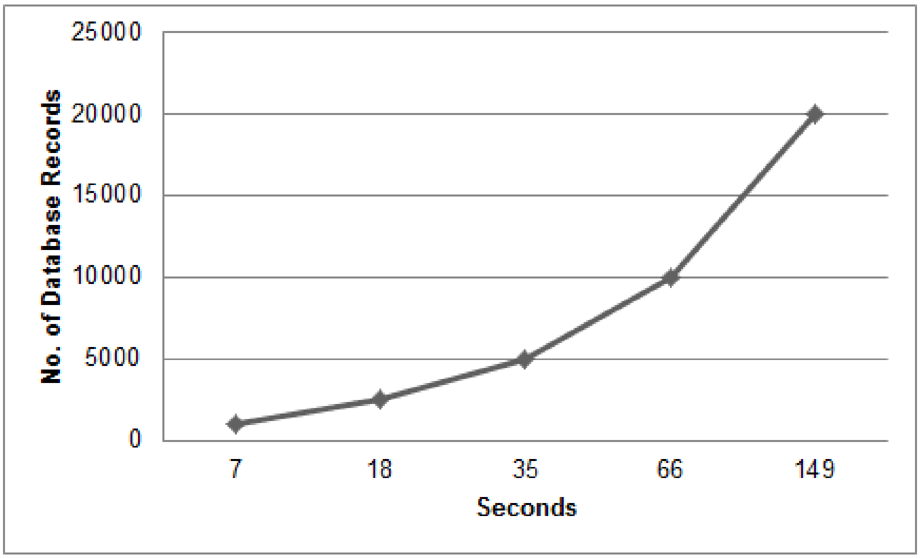

Figure 3: Performance measurement of the SMaRT application

4. Case Study

The efficiency of SMaRT and the applicability of the SMaRT application are demonstrated by conducting a case study involving smartphone data. The smartphone data is extracted from a single Android smartphone (the process followed to extract the data is beyond the scope of this paper), which belongs to an individual with a high level of smartphone usage. The smartphone data was collected during the normal operation of the smartphone and no additional data was added to the smartphone. The smartphone user is hereafter referred to as Person X to allow for the protection of the individual's identity. The smartphone data collected between August 2014 and March 2016 and includes call logs (496 records), SMS messages (445 records), WhatsApp messages (22337 records), and e-mails (500 records).

The goal of the case study is to obtain knowledge from the available data with regards to social interaction as well as the communication density of those interactions. The case study is therefore purely exploratory, following no structured methodology and allowing the examiner to learn from the available data. The role of the examiner throughout the case study is that of a participant observer. Based on the observations, the examiner can adapt the inputs and re-analyse the data. This process can be repeated continuously until the desired results are found or all options have been exhausted. The first round of the analysis for this particular case study therefore includes the entire data set and visualises the data in monthly intervals form August 2014 until March 2016. The results produced by the SMaRT application (see Figure 4) following the analysis show regular social interaction for all monthly intervals except February 2015. February 2015 is identified as period where the social interaction of Person X was weakly irregular. Even though the social interaction during the month of February 2015 is only classified as weakly irregular, further investigation of this particular period is recommended. The remainder of this analysis will, therefore, only continue to focus on February 2015.

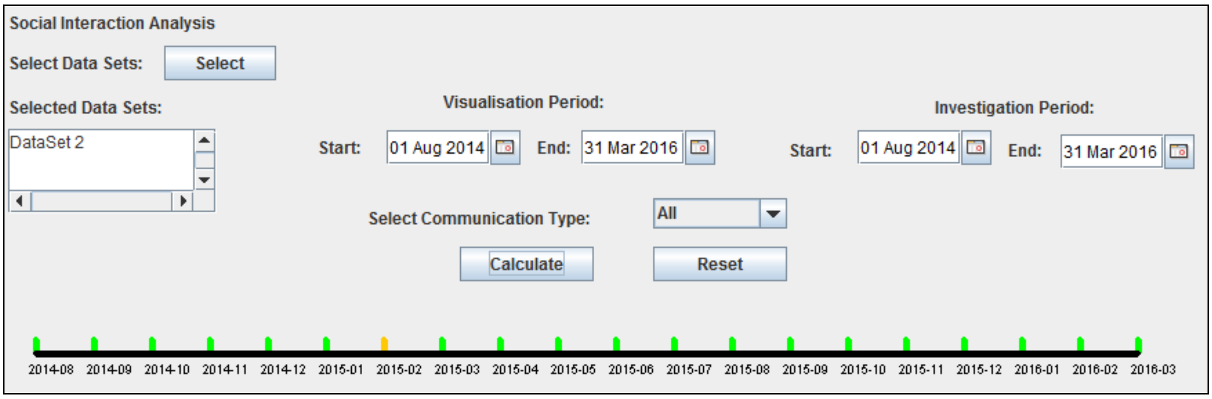

Figure 4: First round of analysis

The second round of analysis visualises the month of February 2015 but still investigates all of the data available in the data set. The visualisation period is from 1 February 2015 until 28 February 2015, allowing the results to be viewed in daily intervals. The results (see Figure 5) show eleven days with irregular social interaction, where five days are identified as weakly irregular and the other six days as strongly irregular. These results confirm that there were indeed irregular social interactions occurring during February 2015.

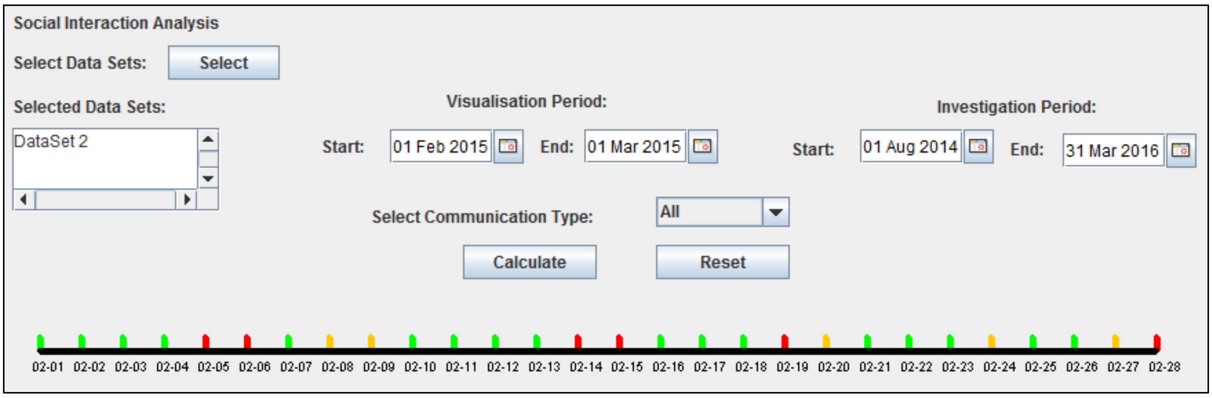

Figure 5: Second round of analysis

The third round of analysis focuses on identifying the communication type involved in the irregular social interaction of Person X. The investigation period continues to include data from August 2014 until March 2016 but now only analyse WhatsApp messages. WhatsApp messages were selected as the communication type for this round of analysis since WhatsApp seems to be the primary communication type for Person X (22337 WhatsApp messages sent during August 2014 and March 2016). The results of the analysis (see Figure 6) identify ten days with irregular social interaction, where five days were identified as weakly irregular and the remaining five days as strongly irregular. The number of days found to have irregular social interaction is similar for both the second and thirds rounds of analysis. It is, therefore, possible to confirm the source of the irregular social interaction is WhatsApp messages.

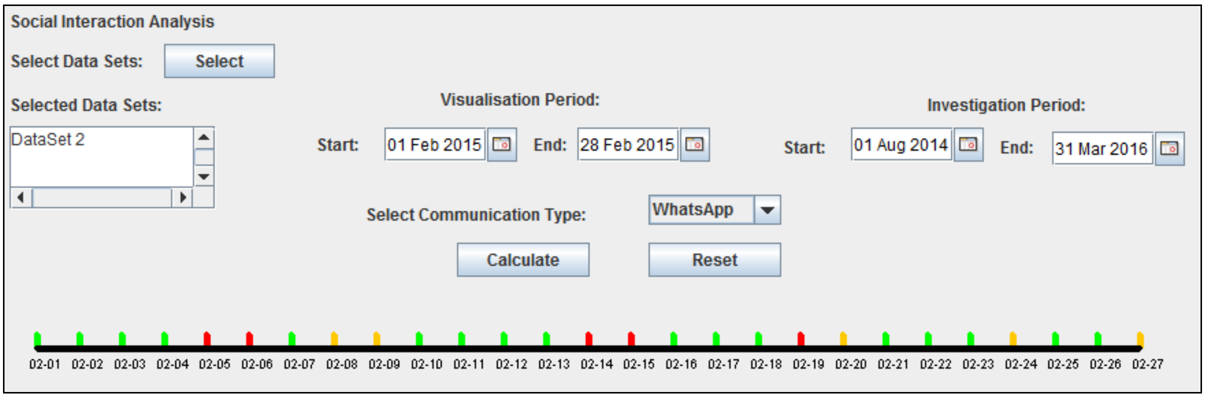

Figure 6: Third round of analysis

The final round of analysis determines the communication density for the social interaction that occurred between Person X and other individuals. The results, shown in Figure 7, reveal that Person X had sudden increased or decreased communication with the following individuals, as identified by the abbreviated phone numbers, during February 2015: *43739*, *05638*, *24050*, *35855*, *05131*, *68395* and *73073.*

Figure 7: Final round of analysis

Following the four individual rounds of analysis, the examiner can come to several conclusions regarding the case study. Firstly, the social interaction for the month of February 2015 is beyond the norm for interaction during a monthly period. Secondly, only specific days of February 2015 were identified to have irregular social interaction. Thirdly, WhatsApp is the primary choice of communication for Person X. Finally, a small collection of individuals had a sudden increase or decrease of communication with Person X, which occurred during February 2015. The examiner can now use these results to direct any further investigation and construct possible hypotheses.

5. Discussion

The evaluation of SMaRT by means of a case study and the examination of the related results emphasise the many qualities of SMaRT. SMaRT is an analysis tool designed to calculate and identify irregularities with regards to social interaction and related communication density. SMaRT supports social-related smartphone data that is collected and stored in SQLite databases. Although the current design only uses social-related data, SMaRT can easily be extended to accommodate other forms of data found in SQLite databases. Visual techniques are used to present the results to the examiner. A timeline visualises the calculated social interaction while descriptive intervals present the results as normal, weakly irregular or strongly irregular. Abbreviated phone numbers are organised into one of three descriptive categories (low, high, very high) to represent the communication density associated with the social interaction. The descriptive timeline thus highlights when the irregular social interaction occurred while the calculated communication density identifies the individual involved with the irregular communication. The visualised results, therefore, allows for the examiner to more easily pinpoint irregular social communication and identify social patterns, which can assist the examiner with the formulation of possible hypotheses and quicken the investigation.

Table 3 provides a comparison between SMaRT and other popular mobile forensic tools. The focus of the comparison is specifically on the analysis capabilities provided by these tools. Although all of the investigated mobile forensic tools support the creation of timelines and the extraction of data from SQLite databases, only SMaRT can calculate the social interaction and identify social communication irregularities.

Table 3: Comparison of SMaRT and existing mobile forensic tools

	Timeline Creation	**Extract data from SQLite Databases**	**Calculate Communication Density**	**Calculate Social Interaction**	**Identify Social Communication Irregularities**
SMaRT	✓	✓	✓	✓	✓
Paraben's E3 Universal	✓	✓			
Cellebrite UFED Pro	✓	✓			
Oxygen Forensics Analyst	✓	✓	✓	✓	
AccessData MPE+	✓	✓			

The SMaRT software application follows a simplistic and easy to use design, allowing for easy comprehension of the provided functions and capabilities. Examiners using the SMaRT application must only present the application with the necessary data, visualisation and investigation periods; the remainder of the process is completely automated. The SMaRT application can also support large collections of smartphone data and is capable of analysing 20 000 records below 149 seconds. The platform independence of the SMaRT application improves the portability of the application across many different operating systems.

The analysis capabilities of SMaRT are currently limited to the calculation and identification of social interaction and communication density. The current implementation does not support any other form of analysis of smartphone data. The SMaRT application only works with data retrieved from SQLite databases and data stored in plain text files or plist files (used in iPhones) are not supported. To obtain a realistic view of the social interaction and related communication density, and identify irregularities, SMaRT must be supplied with an adequate size data set. A very small data set may not provide an accurate view of the social interaction and can thus lead to incorrect results. Besides the current limitations of SMaRT, it still remains a valuable analysis tool that can aid examiners and provide support during investigations.

6. Conclusion

Smartphones are gold mines of information, capable of storing a wide variety of data that can become valuable evidence during investigations. The large and complex nature of the available data can, however, become difficult to analyse due to the time required to search for evidence relevant to the investigation. To calculate and identify social interaction and communication density with regards to smartphone data, this paper introduced a new digital forensic tool called the Smartphone Modelling and Reconstruction Tool or SMaRT. The current design of the tool combines data extraction, reconstruction, and visualisation techniques to analyse smartphone data and pinpoint social interaction irregularities. The output produced by the tool provides assistance to the examiners, allowing for the formulation of hypotheses relating to the investigation. To determine the efficiency and evaluate the performance of the new tool, a case study involving smartphone data is explored using the tool's provided functions. The results obtained during the analysis of the case study show the value of SMaRT as an additional analysis tool that can assist with the identification of irregular social interaction and communication density. The existing functionality of SMaRT only supports the analysis of social-related data and, therefore, future work will focus on the expansion of SMaRT to accommodate other forms of data besides the data found in SQLite databases.

References

Altshuler, Y., Fire, M., Aharony, N., Elovici, Y. and Pentland, A.S. (2012) "How many makes a crowd? On the evolution of learning as a factor of community coverage", *Social computing, behavioral-cultural modeling and prediction*, pp 43-52.

Bakkum, P. and Skadron, K. (2010) "Accelarating SQL database operations on a GPU with CUDA", *Proceedings of the 3rd workshop on general-purpose computation on graphics processing units*, pp 94-103.

Casey, E. (2009) Mobile device forensics essentials, [online], https://digital-forensics.sans.org/summit-archives/2009/37-eoghan-casey-mobile-device-forensics-essentials.pdf.

Chittaranjan, G., Blom, J. and Gatica-Perez, D. (2013) "Mining large-scale smartphone data for personality studies", *Personal and Ubiquitous Computing*, Vol. 17, No. 3, pp 433-450.

Curran, K., Robinson, S., Peacocke, S. and Cassidy, S. (2010) "Mobile phone forensics analysis", *International Journal of Digital Crime and Forensics*, Vol. 2, No. 3, pp 15-27.

Do, T.M.T. and Gatica-Perez, D. (2011) "Groupus: Smartphone proximity data and human interaction type mining", *15th Annual International Symposium on Wearable Computers*, pp 21-28.

Do, T.M.T. and Gatica-Perez, D. (2014) "The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data", *IEEE Transactions on Mobile Computing*, Vol. 13, No. 3, pp 638-648.

Eagle, N. and Pentland, A. (2006) "Reality mining: sensing complex social systems", *Personal and Ubiquitous Computing*, Vol. 10, No. 4, pp 255-268.

Freiling, F., Spreitzenbarth, M. and Schmitt, S. (2011) "Forensic analysis of smartphones: The Android Data Extractor Lite (ADEL)," *Proceedings of the Conference on Digital Forensics, Security and Law*, pp 151-160.

Jansen, W. and Ayers, R. (2007) "Guidelines on cell phone forensics", *NIST Special Publication*, 800-101.

Kilpinen, E. (2012) "Human beings as creatures of habit", *Helsinki Collegium for Advanced Studies*, University of Helsinki, pp 45-69.

Min, J.K. and Cho S.B. (2011) "Mobile human network management and recommendation by probabilistic social mining", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 41, No. 3, pp 761-771.

Min, J.K., Wiese, J., Hing, J.I. and Zimmerman, J. (2013) "Mining smartphone data to classify life-facets of social relationships", *Proceedings of the 2013 conference on computer supported cooperative work*, pp 285-294.

Pieterse, H. and Olivier, M.S. (2014) "Smartphones as distributed witnesses for digital forensics", *Advances in digital forensics X*, Vol. 433, pp 237-251.

Prendergrast, J., Foley, B., Menne, V. and Isaac, A.K. (2008) "Creatures of habit?", *The art of behaviour change*, London.

Tukey, J.W. (1977) *Exploratory data analysis*, Pearson.

Seo, S. (2002) "A review and comparison of methods for detecting outliers in univariate data", *PhD Dissertation*, University of Pittsburg.

Shen, Z. and Ma, K.L. (2008) "Mobivis: A visualization system for exploring mobile data", *Visualization Symposium, 2008*, pp 175-182.

Weiss, G.M. and Lockhart, J.W. (2011) "Identifying user traits by mining smartphone accelerometer data", *Proceedings of the fifth international workshop on knowledge discovery from sensor data*, pp 61-69.