Mobile Social Network Forensic Analysis Based on Visualization Method

Jiating Li[1, 2], Chunguang Ma[1*], Min Yu[2, 3*], Chao Liu[2]

[1]College of Computer Science and Technology, Harbin Engineering University, Harbin, China
[2]Institution of Information Engineering, Chinese Academy of Sciences, Beijing, China
[3]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[a] machunguang@hrbeu.edu.cn, [b] yumin@iie.ac.cn
*corresponding author

Keywords: Forensic Analysis, Visualization, Mobile Social Network

Abstract. Nowadays, so much personal information is stored in the mobile phone, especially in mobile social network. Due to the large amount of data in the mobile phone, it is very difficult to extract and analyze the evidence of the mobile phone. The paper presents a visualization forensic analysis method based on mobile social network, and a forensic analysis of mobile phone messages and call records was conducted, the results shows which makes the evidence more intuitive and improves the efficiency of phone forensics. The method can be used as reference for forensic analysts.

1. Introduction

In 2016, smart phone users accounted for 73.2% of all mobile phones. Due to the portability of mobile phones, which makes it better reflect a person's various types of information. The current smart phone features more powerful, comparable to a microcomputer, the user use the more features, leaving the greater the possibility of digital evidence. Which includes some privacy information, these privacy information is a proof of the behaviors of the owner of the phone, which criminal investigation and evidence collection is a great help, which the rise of domestic and foreign research on smart phone crime evidence. Smart phone forensics analysis technology has a very urgent application needs, automated forensic analysis tasks are handled by the system, can greatly reduce the time of evidence and personnel to pay, check the work of the efficient development of evidence has important significance and development prospects.

Most of the existing research on mobile phone forensics is the data extraction and recovery, A large amount of data was extracted; there are many unrelated information, do not see the correlation between the characteristics of the data. Resulting in forensic personnel often have a lot of evidence data, but failed to effectively get from the data to the event information. The COPLINK system[15] constructs the conceptual space of an entity and object using data mining techniques to help find the relationships between entities. And provides a visual support, including hyperchromatic tree views and spring embedded graphics layouts for related entities. TRIST[16] can represent, refine, organize, and execute queries on large collections of documents. TRIST is optimized for querying large

databases and analyzing comparisons. Based on the data latency model of vector clock[17], a vector clock algorithm is proposed, which can produce partial order relations between events in distributed systems and detect causal conflicts. At the same time, vector clocks can be used to find causal relationships between different events basis to visualize the formation of a two-dimensional map. Graphical representation adopted in social network analysis since the origin[12] is a natural and fast method to highlight links among individuals. Emilio Ferrara[19] proposes a method which finds the configuration of crime organizations through using call records. He uses the theory of network centrality in the process and proposes a program which visualizes a network. Cosimo Anglano[20] discusses all process smartphone forensic process model. Jisung Choi and Sangjin Lee[18] propose a method which shows connectivity, between a user and another as a numerical value, by using recorded data of SMS/MMS, call applications, contact information and stored time information.

Here, the contributions of this article are summarized as follows, we release the visualization forensics method based on mobile social network, which highlights the different aspects and characteristics of the network under consideration, and allows the elements of the network themselves to be checked, the visualization results can be more intuitive and facilitate the work of forensic work. Based on the key data of mobile information, the social network model of mobile information is constructed. Based on the characteristics of mobile social information network model, the algorithm of graph layout based on social network is realized. In the following of the paper, mobile social network and the phone social network are the same meaning.

The rest of the paper is organized as follows, Chapter 2 introduces the research status of mobile phone forensics technology and evidence visualization method. Chapter 3 introduces the social network based visualization method and implementation method proposed in this paper. Chapter 4 introduces the mobile data visualization scheme and the related experimental results. Chapter 5 summarizes the work of the article and explains the next step of the study.

2. Visualization Method

We take full advantages of the large amount of metadata in the mobile phone information, extract the key information, the use of social network analysis technology to build a criminal suspect user-centric star interpersonal network diagram, mobile phone information in the core of the communication behavior model, potential information associated with the suspect.

2.1. Mobile Social Network

Mobile information network refers to the user through the mobile platform with other people to interact with information generated by the information stored in the mobile phone set of information network, because the information set includes relationship between communications and its behaviors, which can be expressed using social networks.

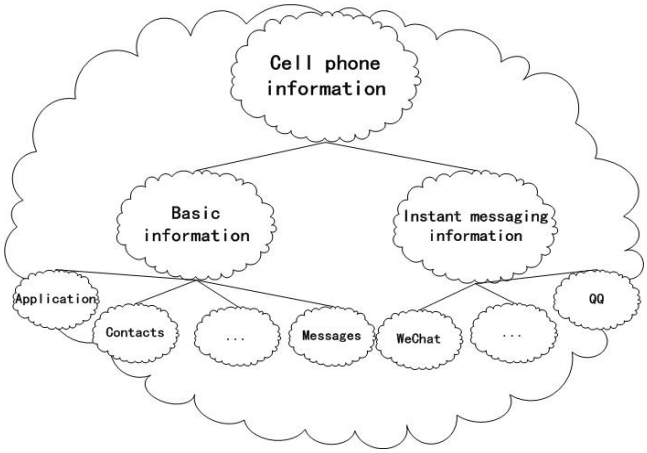

Figure 1 Mobile Information Network

Graph is a more complex than the linear table of a data structure, display form more intuitive, is widely used in social, chemical and other different areas. In the graph G = (D, R), D represents the set of data elements in the graph, and R represents the set of relationships that exist between these data elements. If the data elements and relationships are abstracted as nodes and edges, respectively, you can use G = (V, E) to represent the graph, where V is the set of nodes and E is the set of edges.

Constructing mobile information network with directed graph. Its formal definitions are as follows:

Definition 1: The mobile information network is represented by G = (V, E), where V represents the set of user accounts on both sides of the communication behavior relationship, ie V = $\{v_i \mid v_i$ represents the user account$\}$; E is the edge (communication behavior relationship) of the set, that is, E = $\{(v_i, v_j) \mid v_i$ and v_j successful communication$\}$.

Definition 2: Mobile information in the network communication between the two sides of the degree of C = *Wij*, where $W_{ij} = \sum_{k=1}^{n} 1$ user account *i* and user account *j* between the side weights, *n* that user *i* and user *j* to send communication between the total number of interactions, The greater the value of *Wij*, the closer the relationship between users is the more frequent the interaction. The value of intimacy C can be used to measure the affinity of the user.

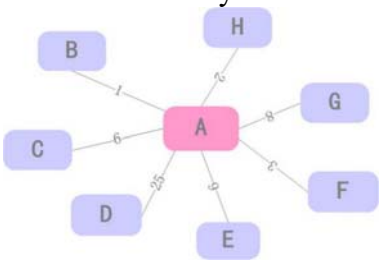

Figure 2 Mobile Social Network in Mobile Phone

Figure 2 is a simple social network diagram with A as the center. We can see that user A communicates with users B, C, D, E, F, G, and H respectively. The number on the connection line indicates the number of communication records. Forensic personnel through the social network map can be intuitive to find a regular contact with the user A, through the intimacy to determine their intimate relationship, the following figure, the user A and user D the highest degree of intimacy.

2.2. Layout Algorithm

Force-directed algorithms are typical of spring theory algorithms and are widely used to describe relational information graphs such as social networks. The Fruchterman-Reingold algorithm [5] imagines the entire network as a virtual physical system. Each node in the system can be seen as a discharge of particles with a certain amount of energy, between the particles and particles there is a Coulomb repulsion, so that they are mutually exclusive. At the same time, some of the particles are implicated by some "edges", which produce a spring-like Huke gravitational force, and tightly contain the "edge" at both ends of the particles. Under the constant action of particle repulsion and gravitational force, the particles are constantly displaced from the random and disorderly initial state, and gradually tend to balance the orderly final state [6, 7]. While the energy of the entire physical system is also constantly consumed, after several iterations, the particles are almost no relative displacement between the whole system to achieve a stable and balanced state, the energy tends to zero. At this point, the social network drawing is done. Mobile phone forensics data with large amount of data, data types and more features, a large number of complex information is not easy to evidence for evidence analysis, the use of force-oriented layout algorithm generated social network map, reasonable layout, clear and easy to understand. Can bring great convenience to the work of forensic personnel, improve the quality and efficiency of forensic work.

In the algorithm for nodes i and j in the graph, the Euclidean distance of two points is denoted by d (i, j), s(i, j) represents the natural length of the spring, k is the elastic coefficient, r represents the electrostatic force constant between two points, W is the weight between two points. The following are two models in force-oriented algorithms:

Spring Model: $$E_s = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}k(d(i,j)-s(i,j))^2 \quad (1)$$

Energy Model: $$E = E_s + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{rw_i w_j}{d(i,j)^2} \quad (2)$$

Here is the pseudo-code for the force-directed algorithm:

```
Set the initial speed of the node to (0,0)
  Set the initial position of the node to an arbitrary but non-overlapping position
    Total kinetic energy: = 0 // The total kinetic energy of all particles is zero
    For each node i
      Net force f: = (0,0)
    For each node j out of the node
      Net force f: = net force f + j node corresponding to the i node of the Coulomb repulsion
  The next node j + 1
  For each spring on the node s
      Net force f: = net force f + spring on the node of the Hooks elasticity
    Next spring s + 1
    // If there is no damping, the whole system will keep going
    The node speed: = (the node speed + step size * net force) * damping
    The node position: = the node position + step size * the node speed
    Total kinetic energy: = total kinetic energy + node quality * (the node speed) ^ 2
    The next node i + 1
```

The force-directed layout can be used for most network datasets, achieving better symmetry and local aggregation, which is easy to understand and easy to implement. Based on the combination of statistical analysis and social network visualization technology, the interpersonal network composed of the association between user communication behaviors in mobile information is visualized from the point of view of evidence. The mobile information network not only allows the forensic staff to understand the relationship between the network structure data more intuitively, but also can assist the forensic person to excavate and obtain the hidden characteristic information between the data in the most natural way in a short time, so as to take the fast and effective Strategy. Therefore, the mobile information network in the mobile phone forensics data analysis process has a very important significance.

3. Case Analysis

3.1. Mobile forensics analysis system

The architecture of the system is shown in Figure 3, which is made up of an extensible level: the data of the authenticated phone is imported (usually a flat file); the data is cleaned by data cleansing, and the redundant edges and nodes are removed to normalize the data. Convert to GraphML format [9], which is a structured XML format that is more suitable for graphical discovery and graphical rendering of applications between interchange, visualization and dynamic exploration of linked networks. Finally, through the layout algorithm output forensic results, that is, mobile information network map.

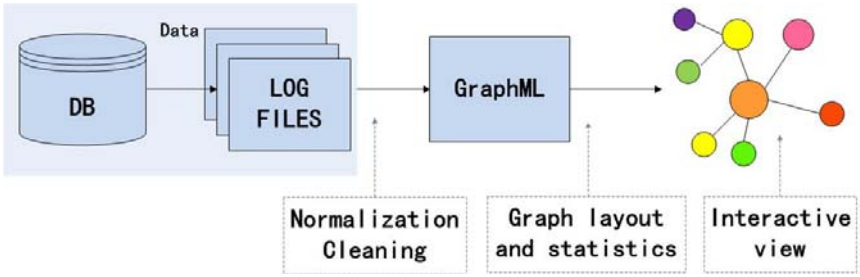

Figure 3 Architecture of Forensic System

3.2. Experiment

The main contents of this experiment are call records, short messages and WeChat chat records, and the related information of the chat records is mainly composed of chat records, group chat records and system information of users and friends. The main contents are: sender ID, sender name, information content, receiver ID, recipient name, information creation time, information type, information status, notes and other information.

Table 1 Data Types

Field	Data Type	Data Length	NULL
ID	VARCHAR2	64	NOT
Profile_ID	VARCHAR2	64	NOT
TYPE	NUMBER	1	NULL
Sender_ID	VARCHAR2	32	NULL
Receiver_ID	VARCHAR2	32	NULL
CreateTime	VARCHAR2	32	NULL
State	NUMBER	1	NULL

The data source used in the experiment comes from the test cases published on the web. The data source is a Samsung Galaxy series of smart phone system image file, the specific model for the Sam Sung Galaxy Mini GTS5570, the operating system for Andrews 2.2.1, recorded in the mobile phone users Patrick Payge. Figure 4-6 are the results of the evidence collection of the short massage records, call records and the WeChat records stored in the Sam Sung Galaxy Mini GTS5570.

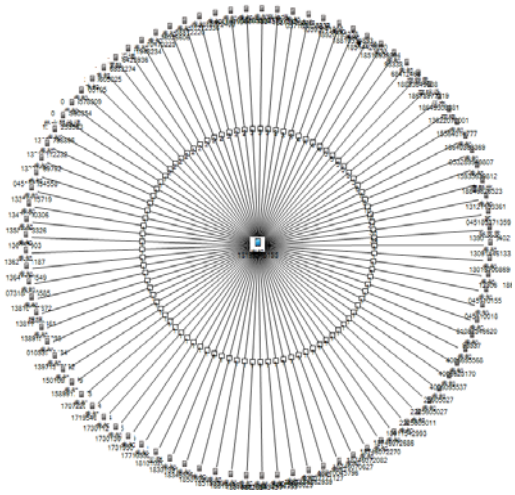

Figure 4 Result of Short Massage Records

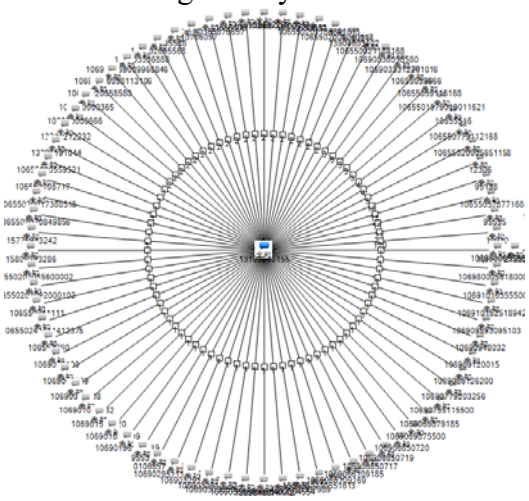

Figure 5 Result of Call Records

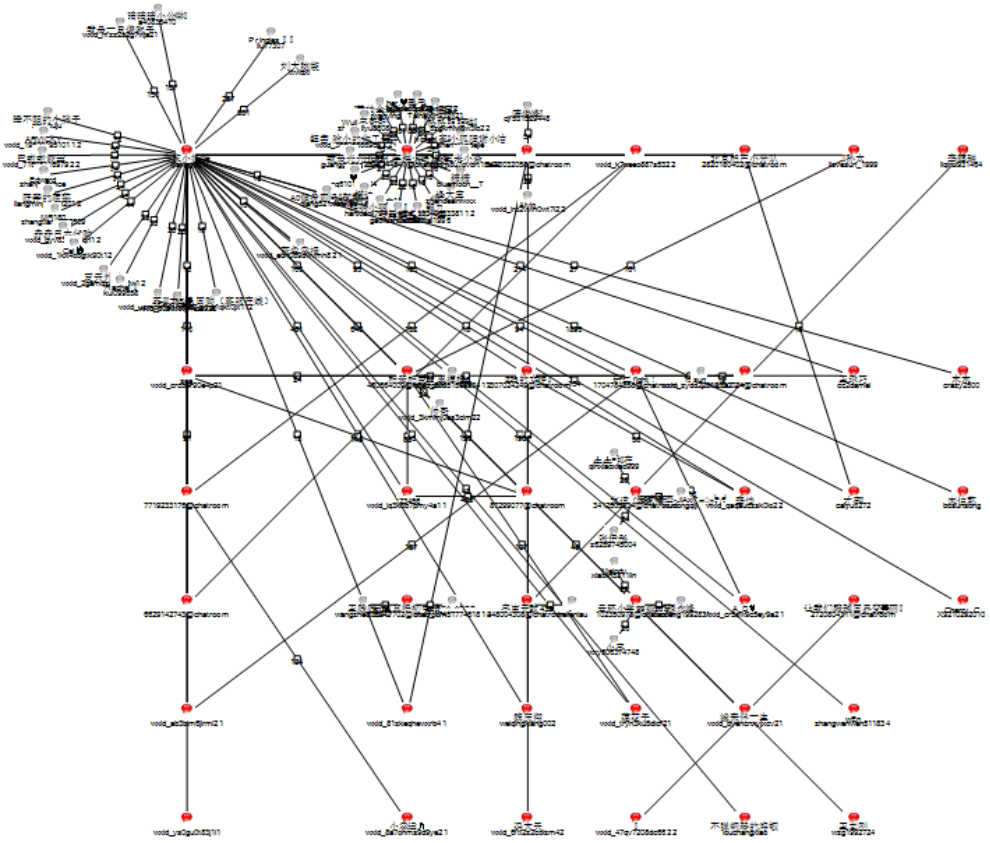

Figure 6 Result of WeChat Records

In the figure 4 and 5, the central node is the mobile phone user, the circumference node is the user contact person, and the number of times is the communication number. In the Figure 6, the central node is the mobile phone user, the circumference node is the micro contact person, and the number of times for the communication. The remaining circles are micro-credit groups, cross-line that micro-credit group members and the user is a friend relationship. Intuitive image of the user's social network.

3.3. Comparative and Analysis

Figure 7 is based on the vector clock data correlation model forensic analysis of visualization results. As can be seen from the figure, the cause of event E1-2 is E1-1, the result is E1-3, E2-2. The cause of event E2-1 is E11 for E2-2. Through the time vector based visualization method can be found directly from the visual results of the reasons and results of the time. Figure 8 is visualization result of social networks. In the figure, the red node is the center user, and the blue node of the circumference is the user who communicates with it. Through the visualization of the size of the weight in the side can determine the degree of intimacy of the two, and then narrow the scope of the next step to determine the phone forensics.

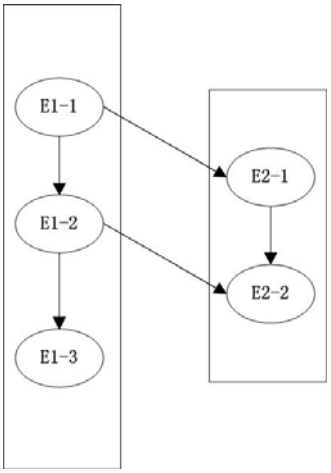

Figure 7 Visualization Result of Vector Clock

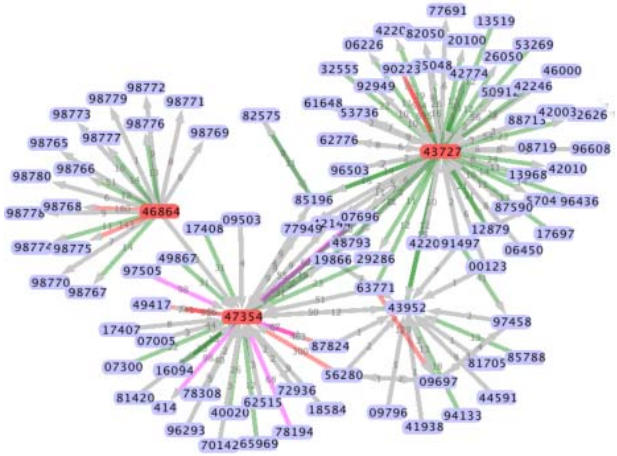

Figure 8 Visualization Result of Social Network

Compared with figure 7 and figure 8, we can find that visualization method shows more intuitive image based on the mobile information network, for large data volume of the evidence can be more intuitive display, we can quickly find the results from the evidence and the case-related information, and find close relationship and abnormal contact. It can greatly improve the efficiency of forensics, so that forensic staff could more easily access to the case of effective information.

4. Conclusion and Future work

The paper constructs the mobile information social network model based on the extracted data, and obtains the contact information of the mobile phone users' two-dimensional space, which is easy for forensic personnel to obtain valuable and potentially accurate evidence from a large number of disorganized data. For example, SMS, call records, instant messaging and so on. Geographic location information is also an important part of mobile phone evidence, the next study needs to associate the geographic location information, time information, and event information in the phone for a more comprehensive analysis.

Acknowledgments

This work is supported by National Natural Science Foundation of China (No. 61472097, 61173008, 61402124, 61303244), Strategic Pilot Technology Chinese Academy of Sciences (No. XDA06010703), Young Scholar Foundation of Institute (No. 1104005704).

References

[1] Sun C, Cai W. Capturing Causality by Compressed Vector Clock in Real-time Group Editors. Parallel and Distributed Processing Symposium. Proceeding International, IPDPS. 59-66. IEEE Computer Society. U.S.A 2002.

[2] http://en.wikipedia.org/wiki/Vector_clocks.

[3] Tannenbaum A, van Steen M. Distributed Systems. Principles and Paradigms. Pearson Prentice Hall. U.S.A. 2007.

[4] Graphviz - Graph Visualization Software, http://www.graphviz.org.

[5] T.M.J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. Software: Practice and experience, 21(11):1129 – 1164, 1991.

[6] S G Kobourev, K Wampler. Non- Euclidean spring embedders [J]. IEEE Transactions on Visualization and Computer Graphics , 2005 , 11(6) : 757 - 767.

[7] G D Battista , et al.Graph Drawing Algorithms for the Visualization of Graphs［M］. New York: Prentice Hall Press，1999.

[8] M. Ghoniem, J. D. Fekete, and P. Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In Symposium on Information Visualization. IEEE, 2004.

[9] Tomar V, Asnani H, Karandikar A, et al. Social network analysis of the short message service[C]//Communications (NCC), 2010 National Conference on. IEEE, 2010:1-5.

[10] Shiravi H, Shiravi A, Ghorbani A A. A survey of visualization systems for network security [J]. Visualization and Computer Graphics, IEEE Transactions on, 2012, 18(8): 1313-1329

[11] Crnovrsanin T, Muelder C W, Faris R, et al. Visualization techniques for categorical analysis of social networks with multiple edge sets [J]. Social Networks, 2014, 37: 56-64.

[12] von Landesberger T, Kuijper A, Schreck T, Kohlhammer J, van Wijk JJ, Fekete JD, Fellner DW (2011) Visual analysis of large graphs: state-of-the-art and future research challenges. In: graphics forum.

[13] Wang X, Zhou X, Lan W, et al. An improved graph drawing algorithm for email networks[C]//Asian Control Conference, 2009. ASCC 2009. 7th. IEEE, 2009: 1667-1672.

[14] Catanese S, Ferrara E, Fiumara G. Forensic analysis of phone call networks [J]. Social Network Analysis and Mining, 2013, 3(1): 15-33.

[15] Chen H, Zeng D, Atabakhsh H, et al. COPLINK: Managing Law Enforcement Data and Knowledge[J]. Communications of the Acm, 2003, 46(1):pάgs. 28-34.

[16] Jonker D, Wright W. Visualization and Comprehension[J]. 2010:285-309.

[17] Zhao Haoliang, Design and Implementation of Mobile Software Forensics Software Framework [D]. Dalian Maritime University, 2012.

[18] Choi J, Lee S. A study of user relationships in smartphone forensics [J]. Multimedia Tools & Applications, 2016, 75(22):1-13.

[19] Ferrara E, De Meo P, Catanese S, Fiumara G (2014) Detecting criminal organizations in mobile phone networks. Digital Investigation 41(13):5733–5750

[20] Anglano C, Canonico M, Guazzone M. Forensic analysis of the ChatSecure instant messaging application on android smartphones [J]. Digital Investigation, 2016, 19:44-59.