Research on Forensics of Social Network Relationship Based on Big Data

Tian Junjing1*, Bi Yan1 and Ma Jinqiang1

1China People’s Police University, Lang fang, Hebei, 065000, China

* a88241@163.com

Abstract. With the development of technologies such as big data, cloud computing, and mobile Internet, many cybercrimes have changed from the previous crime mode of single-person and single-machine to the crime mode of multi-person and multi-machine. Due to the popularity of mobile phones and various mobile terminal applications, many criminals often communicate with other members and commit crimes by mobile phones. Therefore, it is very important for the forensics of the electronic data in the instant communication tools. Much valuable information such as personal common contacts, chat content can be found in the mobile phone, which can help the investigators take the evidence of social network relationship and find many key clues. This paper expounds the forensics mode of social network relationship and the forensics process of mobile phones, and puts forward the forensics method of social network relationship based on Wechat platform, analyses the instance data set, obtains the social network diagram, and intuitively and clearly shows the relationship and intimacy between multiple members. The forensics of Social network relationship can help the investigation organizations narrow the scope of investigation and improve the efficiency of handling cases.

1 Introduction

With the improvement of justice in China and the enhancement of people’s cyber security awareness, the "Criminal Procedure Law of the People's Republic of China", "Civil Procedure Law of the People's Republic of China" and "Administrative Procedure Law of the People's Republic of China" have all listed electronic data as the type of evidence, and clarify the legal evidence position of "electronic data". Therefore, electronic data forensics, as a key technology for case investigation and judicial appraisal, has been highly valued by China's administrative law enforcement departments and the enterprises, and has become a weapon for combating criminal activities such as cybercrime and financial fraud [1].

With the development of the big data, cloud computing, the objects of forensics, forensics tools, forensics methods, forensics organizations and other aspects have all changed. The forensics objects have shifted from a single computer to a variety of media such as the cloud servers and multiple mobile terminals. The types of forensics objects have grown by leaps and bounds in recent years. In terms of forensics tools, forensic personnel used forensics software such as Encase, FTK in the early days, and nowadays they mostly use online forensics analysis software (for example, F-Response) and correlation analysis software. The hardware devices for forensics have developed from write-protection devices, mirror devices, forensics towers, mobile terminal forensics toolboxes and other equipment. The forensics tools must be compatible with multiple operating systems, with
forensic resources and computing platforms as the core to set up a cross-platform, distributed forensics model. In terms of forensic methods, it has developed from single computer forensics to a distributed comprehensive forensic system, and the investigators can use data mining, artificial intelligence and other technologies to achieve data correlation analysis, collision comparison, and intelligently display electronic data evidence in multiple dimensions. In terms of the forensics organizations, social-oriented judicial appraisal agencies, security companies, and scientific research institutes have also carried out training and research on electronic data forensics in addition to the law enforcement departments. Evidence collection departments have expanded from the law enforcement departments to multiple relevant departments.

The development of big data technology and products provides a more comprehensive and complete data source for the forensics analysis of electronic data, but it also brings many problems and challenges to the electronic data forensics. First of all, the volume of forensic data has increased dramatically, that requires the support of large-capacity servers. The forensic data may be stored in a variety of distributed and heterogeneous systems, and some data may be located in remote clouds, the forensics personnel should accomplish a lot of integration work. Due to the large scale of the data, the analysis and inspection of these data also need to spend a lot of time. Secondly, many big data platforms currently use the NoSQL databases such as HBase and MongoDB. The data storage methods and retrieval methods are different from traditional databases. The corresponding technologies about data preservation and data recovery should be updated synchronously [2-3]. Finally, the widespread use of encryption technology also increases the difficulty of obtaining data. The criminals often delete the criminal data, or overwrite the data, remove the traces of crime, and use anti-forensics techniques such as information hiding, making the forensics of these data extremely difficult. Therefore, the forensics personnel should use algorithms of artificial intelligence and data mining to establish forensic models, improve search efficiency, and improve the storage, recovery, and analysis capabilities of evidence data which is an important developing direction of forensics in the era of big data.

2. The forensics of Social network relationship

2.1 Social network analysis

The forensics of social network relationships can be carried out by means of social network analysis. Social network analysis (SNA) is a set of norms and methods for analyzing the relationship structure and the attributes of social networks. The position of each member in the network structure and the relationship with other members can be determined through social network analysis.

The social network can be represented by the graph \( G = (V, E) \). \( V = \{v_1, v_2, \ldots, v_m\} \) represents the set of all nodes in the social network graph. \( E = \{e_1, e_2, \ldots, e_n\} \) represents the set of all edges between the nodes in the social network graph. If \( e_{ij} = e_{ji} \), the network graph \( G \) is an undirected graph; if \( e_{ij} \neq e_{ji} \), the network graph \( G \) is a directed graph. For the weighted graph \( G \), it can be represented by \( G = \{V, E, W\} \), \( W \) represents the weight of the edge \( E \), usually indicates the closeness between the two nodes in the social network graph. The social networks can be represented by weighted and directed graphs, as shown in Figure 1.

![Figure 1. Social networks graph](image-url)
2.2 The forensics model of multi-dimensional social network relationship

Each person plays various roles in the social network relationship and establishes contact with various types of members. Therefore, the personal information can be collected from multiple dimensions to determine its social network relationship in the stage of social network relationship forensics. The social network information related to the individuals includes the family information, the education information, the residence information, and the information about social identity and network roles. The forensics model of multi-dimensional social network relationship is shown in Figure 2.

![Figure 2. The forensics model of multi-dimensional social network relationship](image)

3. The forensics of social network relationship based on WeChat platform

According to the forty-fourth China Internet Development Statistics Report released by the China Internet Network Information Center, the number of mobile Internet users reached 847 million, accounting for 99.1%, and the usage of mobile Internet continues to rise [4]. The WeChat platform is a feature-rich instant communication platform that has emerged in recent years, and has surpassed QQ and it gradually becomes the main tool for people's instant communication. Therefore, the forensics of social network relationship based on WeChat platform is very valuable.

3.1 Forensics process

The National Institute of Standards and Technology (NIST) issued the "Mobile Device Forensics Guide" in 2014, which divides the forensics process of mobile terminal device into four stages: preservation of evidence, extraction of evidence, analysis of evidence, and generation of visual reports [5]. Based on these four stages, the mobile phone forensics process can be derived as shown in Figure 3.

![Figure 3. Mobile forensics process](image)
(1) Getting the material evidence of mobile phone. The investigators should collect the material evidence of the mobile phone related to the case, and get the SIM card and SD card of the mobile phone at the same time. The signal of the mobile phone should be shielded, the mobile phone needs to be protected against water and dust and static electricity.

(2) Forensics preparation. The investigators must ensure that the phone has sufficient power and that the phone is in airplane mode or disconnected. And then, they should check the data status of this mobile phone and list the data sources for mobile phones to prevent data loss or data overwriting. If the data of the mobile phone is deleted, the data needs to be restored.

(3) Evidence extraction. The forensics personnel should obtain the Root permission of the mobile phone in order to extract the data, and then store the evidence data and submit to the data analysis interface to get valuable data clues.

(4) Evidence analysis. The investigators can perform data analysis or data mining for the extracted electronic data to determine the internal connections between the data, and then use some algorithms to establish the rules or models based on the data.

(5) Evidence visualization. The forensics personnel should use visualization tools to show the rules or the models established in the previous step. They can use charts, network diagrams and other tools to visualize the rules of data and the relationships contained in the data.

(6) Evidence Submission. The forensics personnel should submit the key electronic data of the case to the judicial organization, and elaborate various operations of each forensics step, form the evidence chain that can prove the facts of the case in combination with other evidences, and submit the supporting documents and evaluation reports.

3.2 Electronic data forensics on WeChat platform

By analyzing the tables and the fields in the database, the tables related to personal social network relationships are the message table, the rcontact table, and the chatroom table. The key fields should be extracted from these tables for the next step to establish a network diagram and analyze the society Network relationship. The key tables and fields are shown in Table 1 [6].

| Table 1. Key tables and fields of social network diagram |
|--------------------------------------------------------|
| table name | fields name | explanation |
| message | talker | The other side of the communication |
| | isSend | Is the message sent or received? 0 received ;1 sent |
| | createTime | Generation time of this message |
| | content | Chat content |
| rcontact | username | User name |
| | conRemark | The note name of the friend |
| | nickname | Nickname |
| chatroom | chatroomname | Group chat name |
| | roomowner | Group owner name |
| | memberlist | Group member list |
| | modifytime | Establishment time of the chat group |

3.3 Analysis and visualization of social network relationship

In the analysis of social network, since each one may send or receive messages in the WeChat platform, the formed social network is a directed network. The network can be represented by a directed weighted graph \( G = (V, E, W) \), and \( V = \{v_1, v_2, ..., v_m\} \) represents the set of all accounts in WeChat that communicate with other accounts. \( E = \{e_1, e_2, ..., e_n\} \) represents the set of all edges between two accounts. \( e_i = \{<v_i, v_j | v_i \text{ sends a message to } v_j>\} \). When the two accounts communicate with each other, the compactness between them can be expressed by \( W (v_i, v_j) \). \( W (v_i, v_j) \) can be measured by the total number of communication between them, which can be expressed by Equation (1). \( N_k (v_i, v_j) \)
indicates whether there are some communication messages between \(v_i\) and \(v_j\). \(N\) is the total number of communication records in the message table.

\[
W(v_i, v_j) = \sum_{k=1}^{N} N_k (v_i, v_j)
\]  

(1)

In the social network diagram, the length \(L(v_i, v_j)\) of the edge between the two accounts and the weight \(W(v_i, v_j)\) are in inverse proportion. \(L(v_i, v_j)\) can be expressed by Equation (2). \(m\) is the distance parameter in the Equation (2), which is used to optimize the results of the social network graph in the screen.

\[
L(v_i, v_j) = \frac{m}{W(v_i, v_j)}
\]  

(2)

4. Case analysis

Firstly, the investigators should get the mobile phone, and then obtain the Root permissions to get the database files and the configuration files in the WeChat platform. The information about the talker field and isSend field of the message table, the username field, conRemark field and nickname field of the rcontact table, and the chatroomname field of the chatroom table can be extracted from the database. The investigators can merge the fields with the same meaning, and calculate \(W(v_i, v_j)\) and \(L(v_i, v_j)\). The instance data set is shown in Table 2.

| talker               | isSend | conRemark | nickname | \(W(v_i, v_j)\) | \(L(v_i, v_j)\) |
|----------------------|--------|-----------|----------|----------------|----------------|
| wxid_i5mg78wz6tx21   | 1      | Li Xue    | snow     | 13875          | 0.8            |
| wxid_zqgj1mu540di22  | 1      | Feng Ge   | Wind     | 22200          | 0.5            |
| ...                  | ...    | ...       | ...      | ...            | ...            |
| wxid_m3k70des6trx34  | 1      | Shi Tou   | jetian   | 500            | 22.2           |
| wxid_ugvr500uz6nq21  | 0      | Liu Yun   | Baiyun   | 8538           | 1.3            |

According to the data in the instance data set, the social network diagram can be obtained as shown in Figure 4.

![Figure 4. The social network diagram](image)

5. Conclusion

At present, the focus of social network forensics should be the acquisition of real-time data, the analysis of massive data content, the display of complex interactive data and the credibility of...
evidence [7]. The mining and analysis of user relationship in social network is an important part of social network forensics. By using the mining and analysis of social network relationship, it can provide valuable clues for network crime investigation. The forensics of social network relationship can help the investigation organizations to narrow the scope of investigation through the analysis of social network, find potential evidence and reduce the difficulty of evidence collection. The Wechat chat content can also be analyzed in the future research work, the popular words and sensitive words can be filtered, the user-defined words can be queried. The investigators can master the communication data more accurately, and find the potential dangerous molecules quickly.

Acknowledgments
This paper is supported by the project of” Research on Key Technologies of Electronic Data Forensics in the Context of Big Data” (Project ID: QN2019332)

References
[1] Jin Bo 2016. Overview of the development of electronic data forensics and identification. Chinese Journal of Forensic Sciences.01 62-74
[2] Zhang Qiqian 2017. overview of big data forensics technology. Journal of Information Security Research.3 795-802
[3] Li Tao 2017. Data mining technology in network security(Beijing: tsinghua university press) p 225-228
[4] China Internet Network Information Center 2019.the forty-fourth China Internet Development Statistics Report[EB/OL]
[5] Chen Ruijun 2019 Research on Android Mobile Phone Data Forensics Application(Gansu university of political science and law) pp 18-19
[6] Wang Haoyuan 2017 Research of Mobile Phone Forensics Technology Based on Android Platform(Information Engineering University) pp 36-39
[7] Wu Xindong 2014. Study on Social Network Forensics. Journal of Software.25 2877-2892