IJCSN International Journal of Computer Science and Network, Volume 4, Issue 2, April 2015
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 0.417

A Review Paper on Forensic Retrieval of Deleted Information from SIM Card in a Mobile Phone

[1] Nupur Atrey, [2] Pratiksha Patre

[1] Computer Technology, R.T.M.N.U, R.G.C.E.R, Nagpur,

[2] Computer Technology, R.T.M.N.U, R.G.C.E.R, Nagpur

Abstract - It is a forensic software framework for extraction and decoding of data stored in electronic devices. It is designed for data acquisition from SIM card and Mobile phones through Data cables, Infrared, Bluetooth and SIM Card Reader. It retrieves Phonebook, SMS (read, deleted, sent,), Timestamp, SIM services, Call History and all possible data available in SIM card. It also retrieves Real Time Clock, Audios, Videos, Themes, Wallpapers, Images and all other possible data available in Mobile phone.

Keywords - *SIM Card, Mobile Forensics, IMSI, GSM, ICC, TULP, PCSC, AT-ETSI and Framework.*

1. Introduction

Many times there may be a chances of data loses due to some accident happens, and therefore to regain the loss data is necessary for user. After understanding the problems of all users we are designing a tool to recover data and messages. Sim card and mobile forensic tool is a forensic software framework to assist forensic investigators with their examinations of electronic devices. It is not a "push one button" tool automating the complete forensic analysis process. It is assumed that trained examiners who know how to investigate a device, but they are in need of some assistance to speed up the forensic investigation process and to minimize the human errors.

SIM and mobile forensics data recovery tool offers prominent flexible solution for regaining lost or deleted text messages and contact numbers from mobile phone & sim cards. In Mobile phone & sim card the utility of undeletion recovers accidentally deleted text messages, contact numbers and the other important data stored in our mobile phone SIM card. The utility of sim card restoration supports easy recovery of sim card data in the cases like accidental deletion or intentional deletion of data, virus corrupted sim cards, software/ hardware failure and many more. The salvager program of Mobile sim card even displays ICC Identification number of mobile, the IMSI number, name of the service provider etc. the manager utility of Sim card data supports all GSM sim cards subscribed to any national or international sim card service provider.

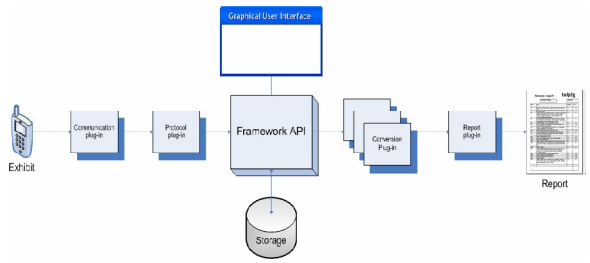

Fig 1: Framework Architecture

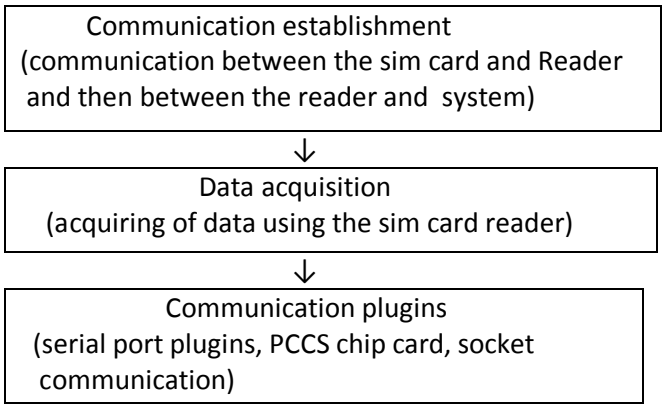

Fig2: Flow Chart

In the literature, to the best of our knowledge, there are some works done on data recovery. Cards4Labs is forensic software originating from 1998 for investigating chip cards, and mainly used for investigating SIM cards from mobile GSM phones. TULP were born in 1996 at the Computer Research section of the Dutch Forensic Science Laboratory presently known as Digital Technology (DT) section of the Netherlands Forensic Institute (NFI). At that

time almost all forensic phone examinations were done at the laboratory with manual examination methods. There were three reasons for the development of an automated examination method:

1. The increasing number of forensic phone investigations,
2. The reduction of human errors,
3. The wish to recover all possible phone data (e.g. non-user data, deleted data).

For a few phone types an automated examination method was developed which extracts data via the connector which resides on the downside of most phones. Unfortunately the efforts to find such engineering backdoors stood in no relation to the number of phone models which appeared. By dropping the third reason mentioned above, plans were made for a wizard like program which instructs a novice user how to extract as much as possible from a phone in a forensic sound way. Again these efforts were overruled by the dramatic increase in phone investigations. Based on that situation the following phone strategy was used:

1. TULP was developed as a forensic data retrieval program for mobile equipment based on ETSI standards and other common technology, which can be used by novice users. Because it is based on standards and common technology it works on a large number of phone models but it will only extract a subset of all available data.

2. Depending on the mobile equipment it is sometimes possible, for trained examiners, to extract more data with manual methods.

3. To be sure that all possible data is extracted from a particular mobile phone model, the model has to be examined in depth at a forensic laboratory. This could result in updated instruction for trained examiners or in new tools which can be used in future automated data extraction tools.

The main reasons to develop new software replacing Cards4Labs and TULP:

1. Both tools are based on ASCII text output format. The growing usage of Unicode and the emerging popularity of multi-media data in mobile communication demanded a new storage and output format.

2. Embedded systems specialists want to concentrate on data extraction and data decoding and not on integrating different methods into a user friendly software product. So most tools stay in a "only for laboratory use" stage and can't be used by novice users which is not a desired situation. A framework can release specialists from complex software issues.

3. Embedded specialists at the NFI are busy with complex laboratory examinations and don't have time to implement all requested methods. With this framework more people are able to add their solutions.

2. Data Acquisition

Data acquisition is the automatic collection of data from the device. It starts using a data cable or infrared dongle or Bluetooth dongle or SIM card reader to connect the device to the computer and ends when the report containing the acquired data is generated. The types of data acquisitions for all devices are performed in similar ways and include the same number of main steps with some variations depending on the device. To learn more about the data acquisition processes, explanations of main terms, etc.

Different data is read from different devices during data acquisitions. Of course, their structure and the quantity of data is mainly defined by the characteristics of the device (e.g. only data stored on the device can be acquired) but, at the same time, some data cannot be acquired because there is no known way to access the data at this time.

3. Steps of Data Acquisition

1. Initial Step - In order to start working with software, you should take care of following:

A. Make sure the device is charged in order to prevent its losing power during the acquisition process.
B. Choose the proper Cable, Bluetooth or Infrared for your device. Make sure that the proper drivers for any USB Cable, Bluetooth or Infrared dongle are installed. Connect the device to the computer properly.
C. Insert or remove the SIM card for phones depending on the requirements of the plug-in you are using.
D. Turn the device on or off depending on the requirements of the plug-in you are using.
E. If the acquisition from a device is not being performed for the first time in a case, it is recommended that you should reload the device before starting a new acquisition.

2. Plug-in Selection Step - Start the acquisition. General Steps for Investigation will guide you through the process of the acquisition. Correctly select the following things:

A. Select the proper plug-ins depending on the requirements.

B. Type of connection.
C. Select the proper style sheet depending on the requirements.

3. Data Acquisition Step – This software acquires the information from the device. The process of the acquisition is shown on the progress bar.

4. Final Step - Acquisition completes, report is generated and disconnect the device from the computer.

4. Secure Hash Algorithm (SHA-1)

In cryptography, SHA-1 is a cryptographic hash function designed by the United States National Security Agency and published by the United States NIST as a U.S. Federal Information Processing Standard. SHA is the abbreviation for "secure hash algorithm". There are three types of SHA algorithms each of which are structured differently and are differentiated as SHA-0, SHA-1, and SHA-2. SHA-1 is very similar to SHA-0 type, but it corrects an error in the original SHA hash specification that led to significant weaknesses. The SHA-0 algorithm was not taken into practice by many applications. SHA-2 hash function significantly differs from the SHA-1 hash function. SHA-1 is the most widely used of the existing SHA hash functions, and is brought into practice in several widely used security applications and protocols. In the year 2005, some security flaws were identified in SHA-1, namely that a mathematical weakness may exist, indicating that a stronger hash function would be more desirable. Although no successful attacks have been reported on the SHA-2 variants, they are algorithmically much similar to SHA-1 hash and so efforts are underway to develop new improved alternatives. A new hash standard, SHA-3 hash, is currently under development.

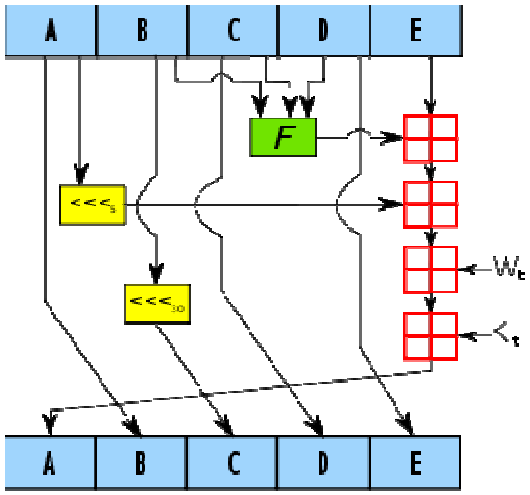

Fig3: An Iteration of Secure Hash Algorithm

SHA-1 hash produces a 160-bit message digest based on some principles similar to those used by Ronald L. Rivest of MIT in the designs of the MD4 and MD5 message digest algorithms, and it has a more conservative design.

The specification of the algorithm was published in 1993 as the Secure Hash Standards, by US government standards agency NIST (National Institute of Standards and Technology). This version is often referred to as SHA-0 hash. It was withdrawn by the NSA just after the publication and was overtaken by the revised version of it, (1995 publication) and commonly referred to as SHA-1 hash. SHA-1 hash differs from SHA-0 only by one single bitwise rotation in the message schedule of its compression function; which was done, according to NSA, to correct a flaw in the original algorithm which resulted in reduced cryptographic security. However, NSA did not provide any kind of further explanation or identify the flaw that was corrected. Weaknesses subsequently have been reported in both SHA-0 and SHA-1 algorithms. SHA-1 appears to provide greater resistance to attacks, thus supporting the NSA's assertion of increased security by the change.

5. Message Digest Algorithm (MD-5)

The MD5 Message-Digest Algorithm is a widely used cryptographic hash function that produces a 128-bits i.e. 16-bytes of the hash value. Specified in RFC 1321, MD5 has been taken into practice in a wide variety of security applications and is also commonly used to check integrity of the data. MD5 was designed by Ron Rivest in 1991 as a replacement or the earlier hash function, MD4. An MD5 is typically represented as a 32-digit hexadecimal number.

However, since it has been shown that MD5 is not collision resistant; MD5 is not suitable for applications like SSL certificates or digital signatures that are dependent on this property. In the year 1996, a flaw was found and since it was not clearly a fatal weakness, cryptographers recommended other algorithms to be used, such as SHA-1—which has been found to be vulnerable. In 2004, more flaws were discovered, making further use of the algorithm questionable—a group of researchers described about creating a pair of files that share the same MD5 checksum. Further upgradaions were made in breaking MD5 in the years 2005, 2006, and 2007. In December 2008, a group of researchers used this technique to fake SSL certificate validity and US-CERT now says that MD5 "should be considered cryptographically broken and suggests not using it further." and most of the U.S. Government applications now require the SHA-2 family of hash functions.

The security of the MD5 hash function is compromised to a large extent. A collision attack exists that able to find collisions within seconds with a 2.6 GHz Pentium 4 processor (complexity of 224.1). There is also a chosen-prefix collision attack that can produce a collision for two arbitrarily chosen different inputs within hours, using off-the-shelf computing hardware (complexity 239). The ability to find collisions has greatly been aided by the use of off-the-shelf GPUs. 16–18 million hashes/second can be computed on NVIDIA GeForce 8400GS graphics processor, 200 million hashes /second can be calculated on NVIDIA GeForce 8800 Ultra.

MD5 processes a variable-length message into a fixed-length output of 128 bits. The input message is broken up into small chunks of 512-bit blocks; the message is padded so that its length is divisible by 512. The padding works as follows: first only a single bit, 1, is added to the end of the message. This is followed by as many zeros as required to bring the length of the message up to 64 bits but less than a multiple of 512. The bits which are remaining are filled up with a 64-bits big Endean integer representing the length of the original message, in bits, modulus 264. The bytes in each 32-bit block are big Endean, but the blocks are arranged in little Endean format.

The MD5 algorithm operates on a 128-bit state, divided into 4, 32-bit words, denoted A, B, C and D. These are initialized to some fixed constants. The main algorithm then operates on each 512-bit message block in turn, each block modifying its state. The processing of a message block consists of 4 similar stages, termed rounds; each of which is composed of 16 similar operations based on a non-linear function F, modulus addition, and left rotation. Figure 1 illustrates one operation in a round. There are four functions F; a different one is used in each round:

6. Communication Plug-In

2.1 Serial Port Communication Plug-In

After selecting the Serial Port plug-in from the Communication-combo box on the Investigation-tab in this software, several options related to which serial port to use as well as how to connect to it and also some testing is available from the configuration screen.

2.2 PCSC Chip Card Communication Plug-In

In the configuration dialog of the PCSC plug-in several options related to chip cards and the chip card readers are shown.

2.3 Socket Communication Plug-In

After selecting the Socket plug-in from the Communication-combo box on the Investigation-tab, several options are available related to which socket to use and how to connect to it and also some testing is available from the configuration screen.

2.4 AT-ETSI Phone Protocol Plug-In

In the configuration dialog of the ETSI-AT plug-in the following options are available:

Command timeout (ms): The timeout interval for checking the response. Default value is 50 milliseconds.

Commands retry attempts: The number of retries for checking the response. Default 1200 retries.
Resends: The number of resending, if the retry attempts are passed. Default 2 resends.

Codec: The codec in which the data is encoded (excluding the SMS codec).

8. Review Table

Table 1 Review

DEVICES USED	USES	TYPE OF DATA AQUIRED
• SIM CARD • DAMAGED SIMCARD • BROKEN SIMCARD	Major part of the project. Used as the source to acquire data.	The type of data acquired from the sim card are the last dialled numbers, phonebook, call records, deleted messages, phone messages.
• THE SYSTEM	The system is feeded with the tool to which works to retrieve data from sim card.	The type of data retrieved is again the deleted messages, phone records, contact

		numbers etc which are internally retrieved in binary form and then converted to readable form using the SHA and MD5 algorithms.
• The SIM CARD READER AND DRVER PLUGINS	Used to build communication between the sim card and the reader then reader and the system.	

9. Conclusion

This paper presents a forensic software framework for acquiring and decoding data stored in electronic devices. The current version of the framework along with the available plug-ins can already be used to assist investigators with their manual investigation tasks.

References

[1] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in Proceedings of ACM CoNEXT, 2009.

[2] Y. Cheng, A. Afanasyev, I. Moiseenko, B. Zhang, L. Wang, and L. Zhang, "Smart forwarding: A case for stateful data plane," Tech. Rep. NDN-0002, May 2012.

[3] L. Zhang et al., "Named data networking (NDN) project 2010 - 2011 progress summary," PARC, http://www.nameddata.net/ndn-ar2011.html, Tech. Rep., November 2011.

[4] D. Rossi, G. Rossini, "Caching performance of content centric networksunder multi-path routing (and more)," Telecom ParisTech, Tech. Rep., 2011.

[5] L. Muscariello. (2011) Content centric networking packet level simulator. Orange Labs. [Online]. Available: http://perso.rd. francetelecom.fr/muscariello/sim.html

[6] A. Carzaniga, M.J. Rutherford, and A.L. Wolf, Ed., A Routing Scheme for Content-Based Networking. IEEE INFOCOM, March 2004. CAIDA. Caida's role in the ndn. [Online]. Available: http://www.caida.org/projects/ndn-fia/

[7] F. Urbani, W. Dabbous, and A. Legout. (2011, November) NS3 DCE CCNx quick start. INRIA. [Online].Available:http://www.sop.inria.fr/members/Frederic.Urbani/ns3dceccnx/index.html

[8] Tony Dearsley, Mobile Phone Forensics – Asking the Right Questions, New Law Journal, July 29, 2005, pp.1164-1165.

[9] The International Telecommunication Charge Card, International Telecommunications Union, Telecommunication Standardization Sector (ITU-T), Recommendation E.118, May 2006.

[11] Svein Willassen, Forensic Analysis of Mobile Phone Internal Memory, IFIP International Conference on Digital Forensics, National Center for Forensic Science, Orlando, Florida, February 13-16, 2005, in Advances in Digital Forensics, Vol. 194,.

Miss: NUPUR ATREY- Bachelor of Engineering, 2015, placed at L&T InfoTech, Pvt ltd.

Miss: PRATIKSHA PATRE- Bachelor of Engineering, 2015.