Data Mining of Social Media Specific Strings for Rapid Forensic Investigation

A. Vinu ¹ and A. Kumaravel²*

¹Department of Information Technology, Jerusalem College of Engineering, Chennai - 600100, Tamil Nadu, India
²Department of Information Technology, Bharath University, Chennai - 600073, Tamil Nadu, India;
kumaravel.cse@bharathuniv.ac.in

Abstract

Instant Messengers become an important means of communication. Millions of people, regardless of age, nationality, gender and computer skills, spend a lot of time using them every day. Thus, Social networks already occupied the place of traditional messaging systems of the past. More and more communications are migrating from public chat rooms and private messengers into online Social Networking Sites (SNSs). As cybercrimes mushroom in recent years, more and more digital crime investigations have strong relations to these SNSs. Communications extracted from SNSs can be extremely valuable and useful to all kinds of investigators including forensic investigators. We now apply the spotlight on distinct strings specific to each SNSs and volatile memory analysis and display of any case detail as a key to a successful digital investigation through rapid & improved retrieval performance.

Keywords: Digital Evidence, Distinct String Analysis, Forensic Investigators, Instant Messengers, Volatile Memory Analysis

1. Introduction

Social Network Sites (SNSs) such as Facebook, Twitter, MySpace etc., have attracted millions of users. It serves as a platform for interacting and sharing data with other people. As cybercrimes mushroom in recent years, more and more digital crime investigations have strong relations to these SNSs. Many criminal investigations will include computers at some point in the case. Email and Internet activities need to be analyzed most of the times to find evidence about the motives of murder and other criminal activities. Therefore our project is useful to Digital Forensic Investigations (DFIs) when the PC of a suspect has been seized for collecting evidence. A Digital Foensic Investigation consists of three steps: Acquisition or Imaging, Analysis and Reporting or Documentation of case files.

Data mining (also referred as data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining is applicable in many different fields. An important application area for data mining techniques is the field of criminal forensics. Examples include detecting deceptive criminal identities, identifying groups of criminals who are engaged in various illegal activities and many more. Data mining techniques typically aim to produce insight from large volumes of data.

Computers require that a certain amount of computer memory called “Random Access Memory” (RAM) which is used by the operating system and its applications when the computer is in operation. The data in RAM is called “volatile data”. This Volatile data can contain information of interest (running processes, Passwords in clear text, unencrypted data, Instant Messages (IMs), date and time, user details) to the investigator. Thus distinct strings specific to each SNS plays a very vital role in resolving the criminal case rapidly.

* Author for correspondence
2. Related Work

2.1 Reference 1: Social Network Sites: Definition, History and Scholarship

SNS such as Facebook, Twitter, MySpace etc., have attracted millions of users. Variety of SNS was addressed from multiple theoretical and methodological angles. This paper highlights the significance of SNS in the lives of users and as a topic of research. Participation on social network sites leaves online traces which offer unprecedented opportunities for researchers. Methodologically, SNS researcher’s ability to make causal claims is limited by a lack of experimental studies.

Each SNS providers has its own access control mechanism which restricts the unauthorized access of User Generated Content (UGC). Although, it is ultimately exploited by criminals for their own intent. In such circumstances, live data acquisition plays an imperative role in reconstructing the digital evidences of the crime scene.

2.2 Reference 2: Volatools: Integrating Volatile Memory Forensics into the Digital Investigation Process

Volatools is a toolkit for volatile memory analysis intended to assist with the survey phase of a digital investigation. Volatools are used to address some of the limitations associated with popular tools and techniques for live response. Information obtained from volatools toolkit are date and time, running processes, open ports, etc.

Recent research works support the live data acquisition technique by saying, the RAM chips may maintain some of their memory states even when the power is removed. When they are connected to main electricity or have a battery, they remain in stand-by state. Generally RAM loses its contents when power is turned off, but not necessarily all of its contents. This distinction is of paramount importance as RAM contents are getting overridden.

2.3 Reference 3: Content Sharing based on Personal Information in Virtually Secured Space

It provides a method for sharing User Generated Contents (UGC) securely based on the personal information of users. Data being shared can be encrypted with secret key or uploaded directly. The virtual secure space allows UGC creator to deliver contents to users who have similar personal information. Thus personnel information leakage can be prevented. There is a slight chance to break the virtually secure space using brute force attack.

2.4 Reference 4: Digital forensics: Defining a Research Agenda

This paper gives a detailed study of Network forensics, Data Collection and Analysis and Live Acquisitions. The goal is to educate and train academic researchers regarding digital forensics and the tools and practices developed for digital analysis. The research work serves as a foundation for further advances in Digital Forensics area.

3. System Architecture

System architecture is the conceptual model that defines the structure, behavior and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structure of the system which comprises system components, the externally visible properties of those components, the relationships (e.g. the behavior) between them, and provides a plan from which products can be procured and systems developed, that will work together to implement the overall system.

Digital Forensic (DF) investigators respond to a homicide on a street corner. The victim was shot once in the chest, and there was no witness. The victim was identified and a search was conducted at his residence in an attempt to determine a suspect and motive. A computer was found at his residence. The computer was already turned on and was running Windows XP.
Figure 2. Live acquisition of physical memory in Helix

Step 1
The DF investigators utilized Helix2008R16 to make an image of the RAM as in Figure 2. The imaged file crime.dd was generated.

Step 2
Access Data FTK7,9 was utilized to import crime.dd. To carry out the digital evidence recovery in a more efficient and effective manner, the search keyword www was utilized to find the type of SNS used. Keyword www resulted in 4577 hits in 432 files but only three SNS were found on tracking through all 4577 hits namely “www.facebook.com”, “www.myspace.com” and “www.friendster.com”.

Step 3
Now, DF obtained the distinct strings specific to Friendster to carry out the analysis of the crime scene7. First he examined the occurrences of the search keyword email? email= and he obtained the digital trails of the sender id around offset 11a064a of the imaged file crime.dd as shown in Figure 3(a).

Step 4
Similarly, DF found another precious digital trail using the search keyword password= at offset 11b069a of the imaged file, crime.dd as Figure 3(b) indicates.

Step 5
And finally, DF found the digital evidence9 to prove the crime using keyword post%5Bbody%5D around offset 11c08e of the imaged file crime.dd as Figure 3(c) indicates.

Step 6
Then DF applied the search keyword www.myspace.com which reveals the profile name of the user in other side around offset 11c421a of the imaged file, crime.dd as Figure 3(d) indicates.
Data Mining of Social Media Specific Strings for Rapid Forensic Investigation

Step 7
Moreover, DF found another digital evidence using the keyword message list and cracked the reply message around offset 0079ee4 of the imaged file, crime.dd as Figure 3(e) indicates.

Step 8
Additionally, DF used search keyword www.facebook.com which resulted as 577 hits in 38 files as shown in

Step 9
On tracking through the Hits obtained with www.facebook.com, DF disclosed another important digital trail around offset 06deb48, ramya raji, who was also a part of the crime as Figure 3(g) indicates.

Step 10
At last DF found the interaction date, with the keyword Date: around offset 06c66a3 of the imaged file, crime.dd as Figure 3(h) indicates.

Finally, DF combined all the digital trails obtained and found that the sender n.i.v.a.e.d.i.t.a and receiver r.a.m.y.a... r.a.j.i were involved in crime activity on Saturday, 09 February, 2013.

5. Conclusion
The use of computer and digital devices in the act of crime is continuously growing day by day. So this gives challenges to forensic how to collect information from the system after an incident. In this research, we focus on identifying most of the legitimate SNS footprints from RAM for several Common activities such as comments, events and chats using the open source forensic tools (Helix & FTK). These footprints include user profile ID, the message contents, sender, receiver and corresponding time stamps which were detected using the keywords specific to each SNS. Thus the crime scene was successfully reconstructed. However, further investigation may be required to verify whether the genuine account owner is involved in the case.

6. References
7. Shipley TG, CFE, CFCE, Reeve HR. Collecting Evidence from a Running Computer: A Technical and Legal Primer for the Justice Community.
10. Available from: http://www.accessdata.com