Forensic Science International 255 (2015) 64–71

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

A cloud-based forensics tracking scheme for online social network clients

Feng-Yu Lin [a,*], Chien-Cheng Huang [b], Pei-Ying Chang [c,d]

[a] Department of Criminal Investigation, Central Police University, No. 56, Shujen Road, Takang Village, Kueishan District, Taoyuan City 33304, Taiwan, ROC
[b] Department of Information Management, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan, ROC
[c] Department of Forensic Science, University of New Haven, 300 Boston Post Rd, West Haven, CT 06516, USA
[d] Institute of Biophotonics, National Yang-Ming University, No. 155, Sec. 2, Linong Street, Taipei 11221, Taiwan, ROC

ARTICLE INFO

Article history:
Available online 28 August 2015

Keywords:
Online social network
IP location
Network forensics
Law Enforcement Agency

ABSTRACT

In recent years, with significant changes in the communication modes, most users are diverted to cloud-based applications, especially online social networks (OSNs), which applications are mostly hosted on the outside and available to criminals, enabling them to impede criminal investigations and intelligence gathering. In the virtual world, how the Law Enforcement Agency (LEA) identifies the "actual" identity of criminal suspects, and their geolocation in social networks, is a major challenge to current digital investigation. In view of this, this paper proposes a scheme, based on the concepts of IP location and network forensics, which aims to develop forensics tracking on OSNs. According to our empirical analysis, the proposed mechanism can instantly trace the "physical location" of a targeted service resource identifier (SRI), when the target client is using online social network applications (Facebook, Twitter, etc.), and can analyze the probable target client "identity" associatively. To the best of our knowledge, this is the first individualized location method and architecture developed and evaluated in OSNs.

1. Introduction

In recent years, with significant changes in communication modes, most users are diverted to cloud-based applications, especially online social networks (OSNs), such as Facebook and Twitter. The worldwide number of unique users of Facebook social network was about 1.23 billion at the end of 2013. Individuals can access social network messaging communications and remain within Facebook or other social networks, and through chat windows, can write on someone else's wall or their own wall, and as social network data center/servers may be in another country, unlike call detail records, metadata are not available and contents are encrypted. Despite being primarily used to communicate and socialize with friends, the diverse and anonymous nature of social networking websites makes them highly vulnerable to cyber-crimes. Phishers, fraudsters, child predators, and other cyber criminals, can register at these services with fake identities, hiding their malicious intentions behind innocent appearing profiles [1,2]. The large number of criminal acts that can be performed through social networks raises the importance of identifying the "actual" identity and location of cyber-criminal suspects in the digital virtual world.

However, the aforesaid topic, to the best of our knowledge, has not been thoroughly studied. Related research on how to gather digital evidence of OSNs mainly uses traditional digital forensics for acquisition of smartphones, and is focused on acquisition techniques and general forensic analysis. They believe that potential evidence can be held on user devices and recovered with the right tools and examination methods. The data that could be extracted from the internal memory of these devices include call logs, SMS, MMS, emails, webpage bookmarks, photos, videos, and calendar notes [3]. Recent scientific research has focuses on individual types of smartphones, and investigation methods that could be used to acquire and analyze data, through either physical or logical methods [4–8]. The above mentioned data sources were organized in a taxonomy of evidence types: identity evidence, location evidence, time evidence, context evidence, motivation evidence, and means evidence, which are derived from a set of questions i.e.: who, where, when, what, why, and how [9]. However, the application of this method is limited to searching the mobile phones of target clients. Unfortunately, this assumption is

* Corresponding author. Tel.: +886 3 3282321; fax: +886 3 3284118.
E-mail addresses: fengyulin@ntu.edu.tw (F.-Y. Lin), chienchenghuang@ntu.edu.tw (C.-C. Huang), peiychang@gmail.com (P.-Y. Chang).

http://dx.doi.org/10.1016/j.forsciint.2015.08.011
0379-0738/

inappropriate for many real-world applications. The related research indicated that traditional approaches to forensics on cloud computing and social network forensics are insufficient from organizational and technical perspectives, while traditional digital forensics is based on the analysis of file systems and captured network traffic [10–14].

To individually analyze a targeted service resource identifier (SRI) and locate its location in online social network applications, there are some challenges that still require solutions, as follows: (1) multiple cyber-identities and applications; (2) the majority of social network services switch rapidly over to hypertext transfer protocol secure (HTTPS) versions, and there is a growing range of sophisticated, encrypted communication channels to exploit; (3) relevant information is obtained only from the payload of the application layer, is spread over multiple packets/sessions, and must be highly correlated, in order to obtain the metadata related to SRI location measurement (i.e. application layer location measurement) and individualized analysis; (4) IP location issue. This is a difficult problem, even putting mobility aside, as the decentralized management of the Internet means that there is no authoritative database of host locations. The databases that do exist are derived by combining a mix of sources (including domain name system (DNS) type mnemonic records, who the site is registered to, and DNS hostname parsing rules), which are all manually maintained, and thus, subject to inconsistencies and outdated information [15].

The "CloudTracker" mechanism, as proposed in this study, is based on the concepts of IP Location and Network Forensics, which develops forensics tracking aims to instantly trace the "physical location" of targeted SRI, when the target client is using online social network applications, and associatively analyze probable "identity". The remainder of this paper is organized as follows. Section 2 reviews related works; Section 3 outlines the proposed scheme system architecture and main elements; Section 4 introduces the proposed CloudTracker mechanism; Section 5 presents the empirical evaluation results for the proposed CloudTracker mechanism, and discusses its strengths and weaknesses; Section 6 offers conclusions and suggestions for future research.

2. Related works

2.1. Mobile device forensics

Smartphones constantly interweave into everyday life, as they accompany individuals in different contexts [9]. It is believed that smartphones include a combination of heterogeneous data sources, which can prove essential when combating crime. Forensic examination of smartphones is challenging, as they are always active and are constantly updating data, which can cause faster loss of evidentiary data. Second, the operating systems (OS) of smartphones are generally closed sources, with the notable exception of Linux-based smartphones, which makes creating custom tools to retrieve evidence a difficult task for forensic examiners. In addition, smartphone vendors tend to release OS updates very often, making it hard for forensic examiners to keep up with the examination methods and tools required to forensically examine each release. The variety of proprietary hardware of smartphones is another issue faced by forensic examiners [16]. From the initial works of the mobile device forensics field, later research provided foundational concepts on forensic analyses of new generations of smartphones (e.g. BlackBerry and iPhone), to recent scientific research focused on individual types of smartphones and investigating methods, there has been complete review in Mutawa et al.'s work [1]. They further focus on conducting forensic analyses on three widely used social networking applications of smartphones: Facebook, Twitter, and MySpace, which were aimed at determining whether activities conducted through these applications were stored on the device's internal memory. Mylonas et al. [9] proposed a proactive smartphone investigation scheme that focuses on ad hoc acquisition of smartphone evidence. They also consider the legal implications of the proposed scheme. Lee and Hong [17] introduced a service concept called "forensic cloud" to develop new paradigms in digital forensics. Furthermore, in order to show the feasibility of the concept, the paper suggested a technology framework for forensic analysis, as based on the mobile cloud, in order to describe the current status of its development. However, previous approaches suffered from the problem of assuming the availability of the target client's smartphone. The forensic tracking mechanism, as discussed in his paper, is underpinned by Network Forensics rather than by Mobile Device Forensics.

2.2. Man-in-the-middle attack

There are two general approaches to meet HTTPS protocol challenges. First, in some cases, we can capture traffic and use the server's private key to recover the session keys and decrypt the contents (depending on the method of key exchange), which requires that we have access to server's private key either before or after the traffic capture. Second, we can intercept the transport layer security/secure sockets layer (TLS/SSL) session using a man-in-the-middle (MITM) proxy. As most well-known applications (services) are foreign venders, the practitioners are unwilling to provide private keys without jurisdiction; therefore, Law Enforcement Agency (LEA) only adopts approach number two to solve the problem of HTTPS encrypted communication protocol.

Web-based applications rely on the HTTPS protocol to guarantee privacy and security. Users trust this protocol to prevent unauthorized viewing of their personal and confidential information over the web. The MITM attack exploits the fact that the HTTPS server sends a certificate with its public key to the web browser. If this certificate is not trustworthy, the entire communication path is vulnerable. Such an attack replaces the original certificate authenticating the HTTPS server with a modified certificate. The attack is successful if the user neglects to double-check the certificate when the browser sends a warning notification [18]. Research activities already have various works to deal with HTTPS. Burkholder [19] analyzed the SSL handshake defect and verified the possibility of attack to SSL. Callegati et al. [20] described conducting SSL attacks by webmitm. Marlingspike [20] declared that, in practical applications, the redirection from hypertext transfer protocol (HTTP) to HTTPS connections would be a security risk. Soghoian and Stamm [18] have shown that it is possible to attack web-based connections secured via HTTPS by exploiting some properties of common local area networks (LANs), as well as the typical behaviors of inexperienced users. Two kinds of drawbacks in the SSL handshake, which respectively conduct fake certificate and conversion from HTTPS to HTTP data to attack is analyzed in [21]. Both are effective to defeat HTTPS communication. Alternative compelled certificate creation attacks are introduced in [18], in which government agencies may compel a certificate authority to issue false SSL certificates in order to covertly intercept and hijack individuals' secure web-based communications, as all web browsers will trust without warning. However, previous research has been limited to the LAN environment. The MITM proxy, as proposed by this study, is designed with reference to the Suga's model [21], to proposes an Inline Redirection model, which attempts to be implemented in large-scale third generation (3G) universal mobile telecommunications system (UMTS) networks.