Locational wireless and social media-based surveillance

Maxim Chernyshev

*Edith Cowan University*

Follow this and additional works at: [https://ro.ecu.edu.au/ADF](https://ro.ecu.edu.au/ADF)

Part of the Computer Engineering Commons, and the Computer Sciences Commons

DOI: [10.4225/75/57b3dd30fb879](http://dx.doi.org/10.4225/75/57b3dd30fb879)

12th Australian Digital Forensics Conference. Held on the 1-3 December, 2014 at Edith Cowan University, Joondalup Campus, Perth, Western Australia.

This Conference Proceeding is posted at Research Online.

Abstract

The number of smartphones and tablets as well as the volume of traffic generated by these devices has been growing constantly over the past decade and this growth is predicted to continue at an increasing rate over the next five years. Numerous native features built into contemporary smart devices enable highly accurate digital fingerprinting techniques. Furthermore, software developers have been taking advantage of locational capabilities of these devices by building applications and social media services that enable convenient sharing of information tied to geographical locations. Mass online sharing resulted in a large volume of locational and personal data being publicly available for extraction. A number of researchers have used this opportunity to design and build tools for a variety of uses – both respectable and nefarious. Furthermore, due to the peculiarities of the IEEE 802.11 specification, wireless-enabled smart devices disclose a number of attributes, which can be observed via passive monitoring. These attributes coupled with the information that can be extracted using social media APIs present an opportunity for research into locational surveillance, device fingerprinting and device user identification techniques. This paper presents an in-progress research study and details the findings to date.

Keywords
Surveillance, digital fingerprinting, smartphone, GPS, geotagging, Wi-Fi probe requests, Bluetooth, traffic inspection, eavesdropping, social media, personal information

INTRODUCTION

Modern electronic technologies enable various types of surveillance activities such as phone wiretapping, video surveillance, cellular tower-based movement tracking, and implementation of databases for storage, aggregation and processing of various types of digital surveillance information (Marwick, 2012). These databases may be utilised by law enforcement organisations, commercial entities and individuals with nefarious objectives to collect and analyse large amounts of data, sometimes as many as 2 terabytes of textual content per day (Alberton, 2012; Hannay & Baatard, 2011). The more personal information is shared publicly on the Internet, the easier it becomes to exploit this information (Timm & Perez, 2010).

A recent Internet traffic report suggests a 70% volume growth in 2012 (Cisco, 2013). The same report also indicates that the volume of mobile traffic in that year was around twelve times the volume of the entire Internet traffic in 2000. In addition, Cisco (2013) also predict that mobile traffic originating from smartphones and tablets will yield a compound annual growth rate of 81% and 113% respectively between 2012 and 2017, with the worldwide number of smartphone and tablet subscribers growing constantly at the same time (Meeker & Wu, 2013). It is arguable that smartphones and tablets are being used as a primary means of Internet access by a significant portion of their users (Brunty, Helenek, & Miller, 2012).

When it comes to identification and tracking in the digital world, numerous research efforts have focused on being able to link digital traces to device users (Shavers, 2013; Takeda, 2012; Wilkinson, 2012). Omnipresent mobile devices with wireless network connectivity leak data on their own (Humphreys, 2011; Kramer & Haines, 2010; Takeda, 2012). Unintentionally or by design, these devices expose information that can be used to track their users (Lincoln, 2013; Storm, 2013). This information, combined with the data publicly available through social media application programming interfaces (APIs), could be utilised to facilitate multichannel digital surveillance techniques. This paper presents an in-progress study on acquisition and aggregation of various types of electronic surveillance information. The primary objective of the study is to determine whether public locational data from selected social media channels can be combined with digital fingerprints and wireless signals through the development of a consistent collection technique that delivers a number of digital locational surveillance artefacts.
Contemporary smartphones and tablets come equipped with a variety of features, which commonly include but are not limited to:

- Speakers and microphone
- Wireless Local Area Network (WLAN) interface
- Bluetooth
- Locational capability based on a combination of Global Positioning System (GPS), Global System for Mobile Communications (GSM) and WLAN
- Web browser
- Ability to install and run third party mobile applications (apps)

In 2012, the predicted number of GPS-enabled handsets – 560 million – was expected to be more than three times than in 2007 (Crawford & Goggin, 2009, p. 101). It is anticipated that around 30% of apps available via the current major distribution platforms – the App Store and Google Play (previously known as the Android Market) by Apple and Google respectively – potentially rely on the ability to access user location (Apple, 2013; Google, 2013). While there has been a slight decrease in the number of apps that require locational capabilities, the number of apps that may access this data is still significant (Lookout, 2013).

**Smart Device Fingerprinting**

The features provided by modern smartphones and tablets facilitate a wide range of software and hardware fingerprinting techniques. In essence, device fingerprinting is used to derive a unique identifier via means of monitoring and processing selected externally observable characteristics (Desmond, Yuan, Pheng, & Lee, 2008). These characteristics range from low-level attributes, such as internal clock skews, to higher-level ones, such as viewport resolution and installed system fonts (Eckersley, 2010; Kohno, Broido, & Claffy, 2005). Ultimately, the aim of digital fingerprinting is to provide the basis for subsequent tracking with the potential to connect the fingerprint to a real identity (Mowery, Bogenreif, Yilek, & Shacham, 2011).

![Figure 1. Conceptual framework for device fingerprinting and user tracking](image)

Figure 1 depicts the selected features of smartphones and tablets that enable various digital fingerprinting methods. A number of fingerprinting approaches based on the characteristics of the WLAN interface have been found to be highly reliable (Desmond et al., 2008; Kohno et al., 2005). Also, in a paper titled “Blueprinting”, Herfurt and Mulliner (2004) suggested a fingerprinting approach based on hashing of attributes disclosed through the Service Description Protocol (SDP) profiles. Simply searching for Bluetooth devices in the immediate vicinity can uniquely identify a certain percentage of nearby devices that have been configured as detectable (Takeda, 2012). Specialised tools such as Ubertooth One, deployed across a number of locations can facilitate tracking of people carrying Bluetooth devices, even if they are not discoverable but are actively transmitting (Ossmann, 2012).

Hardware sensors built into smartphones can also be used for fingerprinting purposes. Due to the peculiarities of the sensor manufacturing process that leads to minor performance differences, data from accelerometers based on micro-electro-mechanical systems (MEMS) can be used to create unique digital fingerprints with 96%
A novel fingerprinting approach based on smartphone speakers and microphones also suggests that manufacturing imperfections can enable audio-based fingerprinting with 90% accuracy (Das, Borisov, & Caesar, 2014).

Fully featured browsers with JavaScript support pre-installed on smart devices can also be used to create fingerprints based on a variety of disclosed characteristics, such as the user agent string, screen resolution and installed plugins and fonts (Eckersley, 2010). For instance, the user agent string that is sent with every Hypertext Transfer Protocol (HTTP) request contains an array of useful characteristics that can be used to identify the specifics of the client platform down to the minor version of the underlying operating system. However, while the user agent string alone is fairly effective, it can be easily spoofed. Therefore, successful browser fingerprinting techniques often rely on a combination of both browser and system-dependent attributes (Boda, Földes, Gulyás, & Imre, 2012). It has also been suggested that analysing the JavaScript browser engine conformance using readily available complex test suites can be used to identify a particular browser type and version with negligible overhead (Mulazzani et al., 2013). Furthermore, the “canvas” element defined in the HyperText Markup Language (HTML) 5 specification enables the creation of reliable fingerprints via the means of pixel extraction from rendered fonts and images (Mowery & Shacham, 2012).

Locational IEEE 802.11 Wireless Surveillance

Multiple security researchers have taken advantage of the 802.11 specification by focusing on implementing tools that enable locational wireless surveillance (Goodin, 2012; O’Connor, 2013; Seiwert, 2012; Wilkinson, 2012; Wuergler, 2012). Specifically, projects such as Snoopy and CreepyDOL exploit the fact that wireless-enabled devices periodically scan for previously connected Wi-Fi networks. These scans are referred to as the “probe requests” that consist of a number of mandatory data attributes, which are transmitted in plan text (Gast, 2005).

Two of the transmitted data attributes are of particular interest – the Service Set Identifier (SSID) and the Source Address (SA), otherwise known as the media access control (MAC) address. The SSID value can be used to determine what network the device is probing for and, thus, was previously connected to. While it can be spoofed, the MAC address generally uniquely identifies the device. Relying on probe requests is effective because users might often leave Wi-Fi enabled on their devices for convenience of access (Wilkinson, 2012). This is not surprising considering that the primary Australian telecommunications retailer has announced plans to roll out one of the world’s largest networks with more than 8,000 additional wireless hotspots (Turner, 2014).

In what is described as a trial of new technology for collection of aggregated footfall data, a dozen probe request sniffing bins were placed in central London and collected over 4 million probe requests and over 500,000 unique MAC addresses in a week (Hamill, 2013). The creators of Snoopy disclose having collected almost 80,000 unique MAC addresses in less than a day at a crowded public location. A similar people movement tracking technology has been utilised by a major Australian retailer (Battersby, 2013). Other researchers have used probe requests to uncover offline social networks and track visitors at mass events (Barbera, Epasto, Mei, Perta, & Stefa, 2013; Bonné, Barzan, Quax, & Lamotte, 2013; Cunche, Kaafar, & Boreli, 2012).
Adding the geographic location as an additional data attribute is made possible by incorporating the GPS capability into the collecting sensors (Wilkinson, 2012). Furthermore, it is possible to infer the geographic location of the networks being probed for based on the SSID using a freely available online service (“WiGLE: Wireless Geographic Logging Engine,” 2013). It should be noted that the future ability to track smart devices manufactured by Apple via the means of passive probe request sniffing may be limited with the introduction of MAC address randomisation during network scanning, which is expected to be built into the upcoming iOS 8 (Hutchinson, 2014; Stites & Skinner, 2014). At the time of writing, the presence and exact behaviour of the proposed privacy-enabling change has not been confirmed based on the available iOS 8 beta distribution (Cox, 2014). Nevertheless, the potential change in network probing behaviour of future iOS-based devices further highlights the significance of multi-protocol surveillance and tracking techniques.

Social Media-Based Locational Tracking

Major social networking services have been taking advantage of locational capabilities of smart devices by allowing and encouraging users to ‘geotag’ the uploaded content or simply share their location (Hannay & Baatard, 2011). In essence, ‘geotagging’ can be described as labelling of online content with geographic coordinates of the originating location (Crawford & Goggin, 2009; Joshi, Gallagher, Yu, & Luo, 2012).

Having access to geotagged data means that not only is it possible to identify the individual data source and the time of creation, but also the place of origin (Hannay, 2009). Data collected over a period of time can be analysed for patterns and exceptions, which could provide insights into more than what the original author had intended to share (Espinhara & Albuquerque, 2013; Hannay & Baatard, 2011; Oculus, 2013; Omand, Bartlett, & Miller, 2012; Trottier, 2012). Geotagged public data presents an additional layer of intelligence through the prospect of developing comprehensive locational profiles on individuals (Valli & Hannay, 2010). In fact, device manufacturers themselves are known for building locational tracking capabilities into their products. For instance, the Motorola Droid X2 phone has been discovered to contain a potential location-reporting component codenamed “Little Sister” (Lincoln, 2013). Furthermore, location-aware devices based on iOS 4 have been known to passively maintain a local database of user movements in an unencrypted format (Cheng 2011). Finally, having access to the complete set of individual locational data may not even be necessary, as some researchers suggest that as little as four unique spatiotemporal points are needed to identify 95% of individual mobility patterns (de Montjoye, Hidalgo, Verleysen, & Blondel, 2013).

Mass sharing of personal information publicly also presents new opportunities to commit conventional crimes through social engineering, spear phishing, identity theft, defamation, cyber-bullying and other activities that can lead to damaging and fatal acts (Brunty et al., 2012; Espinhara & Albuquerque, 2013; Hill, 2012; Jacobson & Idziorek, 2012; Timm & Perez, 2010).

SIGNIFICANCE OF RESEARCH

Turnbull and Slay (2008) proposed the use of wireless network signals as a source of digital evidence and suggested the need to develop specialised technical collection processes. Furthermore, Turnbull, Osborne, and Simon (2009) examined the legal and technical implications of using wireless signals for evidential purposes and suggested that collection mechanisms would need to operate with minimal interaction, provide accuracy and repeatability and function in accordance with all applicable legal requirements.

Device fingerprints coupled with spatiotemporal information can be used to suggest the answers to the “who (which device), where, and when” questions. Reliability classification of digital locational evidence has been proposed by Hannay (2013), which further highlights this potential. In a similar vein, Minnaard (2014) suggest that residual probe request traces left in volatile memory of access points may be of potential forensic value, especially in remote locations.

This study focuses on the low-level aspects of collecting and processing of the relevant data attributes and ways to develop a consistent collection technique that meets the aforementioned requirements. Furthermore, it is expected that additional attributes and protocols will be incorporated into the study or subsequent works.

RESEARCH METHOD

Research Questions

The research consists of two parts and will aim to seek answers to the following research questions:
1. Can a consistent collection technique be developed for acquisition of geotagged social media data and wireless signals?

2. Can the content and characteristics of freely available geotagged data be mapped to a standard set of attributes?

Research Design

This is an exploratory empirical study that employs a multi-method approach based on systems development and experimentation. A functional prototype will be required to conduct the experiments. To shorten the initial prototype development phase, an existing tool – Snoopy – will be examined and its architecture extracted as the basis for subsequent prototyping. Snoopy has already been used in production across a variety of disperse geographical locations, and has further been enhanced as a proprietary tool that can support a number of protocols and physical deployment models (Goodin, 2014). The source code of the original Snoopy proof of concept, as well as its successor – Snoopy-ng – is publicly available. The latter uses an improved modular framework and provides a number of additional data collection plugins. Both versions are functional and are considered an appropriate basis for architecture extraction and reuse.

Once available, the prototype will be used to conduct laboratory experiments involving multiple devices and participants. The experimentation is expected to feed into the iterative process of architecture validation, systems refactoring and conceptual framework alignment as part of answering the research questions.

PROTOTYPE DEVELOPMENT

Logical Architecture

At the time of writing, the initial conceptual, logical and execution architecture specifications have been developed. The execution architecture has been created on the basis of Snoopy through source code analysis and isolated experimentation. The resulting multilayer logical architecture is presented Figure 2.

![Logical Architecture Diagram](image)

Figure 2. Logical architecture

The capture layer is responsible for facilitating the capture of the required data attributes. It has been broken down into two components – device sensors and public data extractors. The former is responsible for the
acquisition of surveillance-enabling data attributes transmitted by smart devices over various number of wireless communication protocols, such as 802.11. The latter is responsible for the extraction of locational public data and metadata using the available APIs.

The storage and processing layer is made up of the central Database Management System (DBMS) back-end and a number of processing agents that turn the raw captured data into usable surveillance artefacts. The layer is broken down into three components to represent the data transformation process facilitated by the processing agents. It is also anticipated that raw data may be captured and sent over to the data store in a variety of different formats. Using the raw captured data as input, the agents apply the relevant processing logic to produce the required surveillance outputs, which are stored in the central DBMS in a structured format. The data extractor components from the capture layer can also be considered a special subclass of processing agents, which have been separated into the capture layer due to their designated purpose. Extraction of public data does not occur en-masse and is generally triggered when relevant identifiers are discovered in the raw data sets.

Finally, the presentation layer includes two components that are responsible for delivering the derived surveillance artefacts to their consumers. The user interface component is responsible for web-based data presentation to real users. The integration interface is responsible for API endpoint provisioning to facilitate integration with other systems, so that this solution can be used as a building block of a larger structure.

Custom Sensor Prototype

At the time of writing, a custom sensor prototype based on a Raspberry Pi and a number of third-party add-ons is in the initial stages of the implementation. Specifically, in this context, the study aims to develop a specification for a low-cost portable, extensible and easily constructible sensor that can support the incorporation of additional components in the future.

CONCLUSION & ONGOING RESEARCH

The omnipresent smartphones and tablets come equipped with a wide range of features that facilitate a variety of reliable digital fingerprinting and loctional tracking techniques. In addition, the proliferation of social media services resulted in vast amounts of personal geotagged information publicly available online. These sources of locational and other data can be used as a basis for the development of digital surveillance techniques and potential collection of digital evidence.

At this stage of the study, the sensor prototype development phase has been initiated and will be followed by experimentation to answer the stated research questions. While the original scope of the proposed study is limited to 802.11 wireless signals and publicly available social data and metadata from selected providers, the derived conceptual framework for fingerprinting of smart devices suggests the need to consider other features and protocols. This need is expected to result in a subsequent scope expansion or will be addressed through additional future research.

REFERENCES


22


