Readings in Computer Science

Comparison of two scientific papers

Niyaz Iralin

Table of contents

1 Introduction

Can a Smartphone be specifically identified according to the behaviour or actions of the user? Smartphones are nowadays widely used especially due to their numerous functionalities. These devices act like small computers as they have many functions such as internet connection, emailing, chats and social media among others. A lot of traffic is therefore generated from these devices and adversaries may use the traffic to attack the smartphones. Traffic generated by smartphones may be used to interfere with user privacy. Information collected from the traffic can be used to create specific fingerprints that will identify a particular device. This is seen from the two articles reviewed in this paper. Both papers investigate how traffic from a smartphone can be used by attackers to violate user privacy. Although different methods are used in the studies, the results indicate that user privacy can be infringed.

The paper starts with a summary of both articles to be compared and looks at the problem question that was being solved. Methods used to come up with a solution for the problem are also discussed in the summary. Lastly the summaries also show the results from each study and its contribution towards the common problem. Contextualization of the papers is done by looking at other research developments conducted using related literature. Finally, the paper looks at how similar the articles are and whether they are related as well as their results.

2 Summary

2.1 Article 1: Who do you sync you are? Smartphone Fingerprinting via Application Behaviour

In their research Stober, Frank, Schmitt and Martinovic (2013) study how fingerprinting in smartphones can be done by the installed application's behavior. They make use of side-channel information generated from network traffic such as timing and data volume leaked from periodic traffic patterns. The extracted data from popular applications is used to evaluate whether a smartphone can be identified through the background information extracted. The study was motivated by the huge number of people who use the smartphone with the number continually increasing. Research questions that the study intended to answer were:

1) The discriminatory nature of Smartphone traffic features;

2) Whether configurations of installed Apps be used to tell apart different Smartphones;

3) The time it would take to identify a Smartphone.

The study assumes that the eavesdropper location is within the 3G/UMTS transmission range. Another assumption is that an attacker has a way of capturing the encrypted traffic being transmitted and is able to decrypt it, and use it to acquire Smartphone fingerprints.

Data used in the research was acquired from existing datasets of traffic recorded from five different users. The 3G traffic had been captured for about eight hours in the background.

Another dataset consisted 8 hour record from 20 devices that had different combinations of the applications installed. Devices used were running on the Android operating system.

Conclusion from the study indicates that background communication from the installed applications generates a distinct pattern that can be used by an eavesdropper to correctly identify a Smartphone. An attacker would only need to use approximately 15 minutes of users' traffic to obtain 90 percents of accuracy. Therefore, user privacy is affected since an attacker is able to identify a particular Smartphone.

2.2 Article 2: Can't You Hear Me Knocking: Identification of User Actions on Android Apps via Traffic Analysis

In their research Conti, Mancini, Spolaor and Verde (2015) study how user actions on Android applications can be identified by analyzing traffic. The researchers investigate the extent to which it is possible to recognize specific user actions using mobile applications through eavesdropping on their network traffic even when it is encrypted. Motivation for the study is due to concerns that smartphones can be used as tracking devices. A framework to deduce what actions were carried out on installed apps is proposed. The framework uses traffic generated by a smartphone. Three commonly used apps were used to test the framework, that is Facebook, Twitter and Gmail. A Samsung Galaxy smartphone that runs on Android 4.1.2 operating system was used together with Wi-Fi for wireless connectivity and a server to route traffic. To capture the traffic a Wireshark software was used with files being stored in csv format. For each of the three applications used ten accounts were created and were divided into active and passive categories. A script was used to capture the traffic and record when each action was executed. A collection of 220 action sequences for individual app was used. With each sequence containing 50 different types of actions.

Results from the study show that it is possible to determine specific user actions executed on the installed apps even with encrypted traffic. With such a possibility an attacker can accumulate data from many users and use it against a competitor. Such a framework may be used to personally identify anonymous user actions by governments. The research indicates that user privacy can easily be undermined.

3 Comparison

Conti et al study whether a user can be identified by the actions performed on their Android apps when traffic is analyzed. This is related to Stober et al study of using application behavior to fingerprint a Smartphone. An increase in the usage of smartphones has resulted in a need for research on the traffic they generated. According to Stober et al the number of devices purchased in the second quarter of the year 2012 was 419 million. The high usage is attributed to the capability of mobile phones to use the internet, improved performance, low prices, and battery life. A measure of traffic from different Smartphones in 24 hours showed that 30 percents of the traffic is interactive while 70 percents is from background activities that is generated by installed apps. Background activities result to traffic patterns that can be distinguished.

From the two articles encryption is mentioned as a way to protect user data from eavesdropping. However, Conti et al highlight that even when users adapt good practice to enhance privacy it does not stop malicious attackers from tracing users. This is especially due to the wireless nature of Smartphones that creates many options that adversaries can use. Their work shows that encrypted traffic can be analyzed to profile a user action on their device. Similarly, Stober et al argue that despite encryption at the data link-layer, wireless eavesdropping can be used to analyze side-channel information from traffic that is encrypted. Analysis of the traffic would lead to the identification of a particular user Smartphone.

In their model Stober et al assume that an attacker's location is within UMTS transmission range. Wireless communications, broadcast its signals, allowing an eavesdropper to access the signals. The model also assumes that an attacker can demodulate as well as demultiplex communication on the physical channel so as to measure information in the side-channel. For fingerprinting a burst notion is introduced in the model framework, with the bursts extracted in regard to the timing information. Classifiers are used to represent a phone's fingerprint. An experiment is carried out to investigate whether the model identification approach is feasible.

Conti et al on the other hand use a model that pre-processes network. Apps on mobile devices rely on SSL/TLS for secure communication with their peers where the protocols are put atop the TCP/IP suite. Encryption of data received by the TCP layer is done and the proposed model uses each network flow as a set of time series. In the framework domain filtering is done using the WHOIS protocol. Packet filtering is then done for packets that do not have information that can help in characterizing the flow. The generated time series length is limited by a timeout technique as well as a packet interval. A managed learning approach is then used to classify user actions from the collected datasets.

Both papers, use Galaxy Nexus Smartphones running on the Android operating system to perform their experiments. Conti et al use a Wireshark software to capture network packets. Stober et al aimed to find from their experiment the time it would take for an attacker to collect a given amount of traffic. Traffic captured is from commonly used applications such as Facebook, Gmail, and Twitter. Conti et all specifically used datasets from three official apps, that is Facebook v3.8, Gmail v4.7.2, as well as Twitter v4.1.10. On the other hand, Sober et al used datasets from 14 top free apps found in Google play. The selected apps are those that have background transmissions that is not affected by user actions.

Stober et al aimed to find out if a smartphone can be identified by traffic generated in the background. Conti et al in their case aimed to find out whether an analysis of encrypted data can be done to find out the particular actions that a user performs on their Smartphones. From the research conducted on both papers it shows that encrypted data can easily be captured by an adversary. The two papers present a common problem that involves user privacy while using a Smartphone.

4 Contextualisation

4.1 Security issues

Evolution in technology has resulted in an increased demand in the use of smartphones. This is mostly due to the opportunities offered by the smart devices that have features such as powerful processors, big storage space, multiple radio and network interfaces, and sensors (Hogben and Dekker, 2010). Smartphones have become a target for attackers and are now ridden with risks such as:

1) Leakage of data from a phone without memory protection

2) Data disclosure done unintentionally as some users do not set privacy settings on the installed applications. Some users are not even aware of privacy settings,

3) Phishing where an attacker uses fake applications to collect user credentials such as passwords.

4) Spyware can be installed on a smartphone and an attacker is then able to collect personal data without the knowledge of a user.

5) Network spoofing, which is done by an attacker creating a fake network access point for a user to connect to. Such a connection will be used for further attacks.

6) Surveillance is possible because smartphones are built with multiple sensors that include GPS, camera, and microphone.

Hogben and Dekker (2010) note that some users are not aware of the functionalities of applications on the smartphones. Some of the functionalities include the ability of an app to collect and publish personal data. Social media applications are able to transmit location data that will enable an attacker to trace a smartphone user.

An increase in the risks associated with using Smartphones necessitated the commissioning of the Goode Intelligence (2013) for an expert report. The purpose was to investigate the security of operating systems used on Smartphones. Areas that were investigated include security threats facing Smartphones, security vulnerabilities, how to deal with security issues, how new technologies affect the security of the phones, mobile app store security, and application security analysis.

The Goode report acknowledges that Smartphones have played a great role in how digital information is created and consumed. Smartphones contain operating systems that can be upgraded and have the capability to run programs known as mobile apps. In essence, they work like computers. Such developments have resulted to smartphones becoming a target for illegal and fraudulent actions. Operating systems used by the phones are also vulnerable to malicious exploitation. Like computer operating systems those used by phones should also be frequently updated and patched. To protect users, there should be legislation and regulation in the usage of the phones as well as user awareness and control on technology.

The Goode (2013) report gives a detailed description of the Android operating system. Android a product of Google is an open source software and is available for different manufactured devices. However, managing the OS updates and patches is a challenge for Google. Nexus devices update their OS by using a FOTA process wirelessly or by a manual download. Samsung devices using Android manage their OS through a computer program known as Samsung Kies. Mobile apps are delivered to smartphone through App stores. Apps can be categorized as official and unofficial with security for unofficial apps being considered poor. This is because major Smartphone OS vendors run and manage official apps and are sure to protect app integrity. It is possible to install an app without using the official platform normally associated with Android. This kind of downloading is referred to as side loading and produces a security risk to Android Smartphones' users.

Developers have a chance to develop apps for the Android OS with most apps available for free download. This creates a possible risk whereby the apps can Trojanized by attackers for some gain. Smartphones based on Android are gaining popularity due to its advanced capabilities and to secure an Android smartphone Trend Micro (2011) suggest five steps that include:

1) Making use of built-in security features by configuring security settings and location.

2) Disabling the Wi-Fi option for auto-connect helps to avoid free flow of a wireless router or an access point.

3) Users should block unofficial apps and download apps from the Android market as they can be more trusted.

4) Before allowing permissions a user should understand them first as some apps could be used to create backdoors. Such a backdoor will be used to collect personal data or perform other unauthorized functions.

5) Installation of a good security app for the smartphone for added security.

4.2 Fingerprinting

Failure to distinguish app traffic from other types of HTTP data exchange raises a security concern as well as inefficient network management. Miskovic, Lee, Liao and Baldi (2015) developed an AppPrint system that analyses traffic and learn app fingerprints for their identification. The proposed system claims to use a wider coverage of traffic by the use of two features, the tokens and traffic flow groups. Strings or parameters on HTTP headers are the tokens that can be used to identify an app as well as flow grouping. The study uses two algorithms, MAP for discovering and learning fingerprints for new apps as well as SCORE to identify apps in observed traffic. Evaluation of the system used two types of datasets, individual apps were run to create lab traffic and anonymous traffic from Android users. AppPrint is able to identify large instances of apps and achieves a precision of 93.7 percents.

Dai, Tongaonkar, Wang, Nucci and Song (2013) developed a technique to extract fingerprints to detect and identify apps. The technique automatically identifies Android apps from network traffic. Mobile traffic fingerprinting is crucial since protocol identification is not able to distinguish the smartphone traffic. A NetworkProfiler developed by the researchers consists of DroidDriver that collects network traces and a fingerprint extractor to extract fingerprints from the collected network traces. According to the results of the study it is possible to identify apps with a high accuracy.

Smartphone fingerprints can be created by a range of sensors found on the devices as observed by Bojinov, Boneh, Michalevsky and Nakibly (2014). A mobile fingerprint would help identify a device as well as identify authorized users connecting to a server. For the study, the researchers analyze speakerphone-microphone frequency response and calibration errors from the device specific accelerometer. Identification of devices can be used for malicious as well as well intended purposes. A malicious attacker can track a user by fingerprinting devices that contact a website. Users who leave their devices connected to the internet are more likely to be fingerprinted without their knowledge. Imperfections on sensors made during manufacturing or the assembly process result in a variation of biases such as timing, linear bias, and tolerance. The ability of a device to transmit through a speaker and receive through a microphone makes it possible to create unique fingerprints. Results from the study show that it could be dangerous when an untrusted web code is running on a mobile browser.

User smartphones can be identified through the configurations that are set for a specific device as reported by Kurtz, Gascon, Becker, Rieck and Freiling (2016). The researchers use Apple iOS to show that a device can be fingerprinted using features obtained from third party apps. Most free apps include adverts or a tracking library that collects user personal information and behavior. Collected information may be used by advertising networks to target specific users based on their behavior. Users may not notice the background activities that collect their information which can be considered an infringement on user privacy. Matching collected data with a specific user requires the unique identification of a device. Devices that allow access to ID identifying devices makes it easy to correctly fingerprint a device.

Kurtz et al. (2016) study how fingerprinting can be done based on configurations on a device set by a user. Focus on personalized configurations assumes that a user willingly installed an app, but the app has other operations that a user is not aware about. 13,000 data records were collected from 8,000 devices in a period of 140 days. By using identified fingerprint features, the researchers created their own app which was made available for download in the App store. About 57 percents of the collected data was from returning devices. The research work aims to show a technique based on software that correctly fingerprints a mobile device by using information from a service provider. It also answers the question concerning multi-class classification issue for fingerprints originating from one device and show how the problem can be solved by threshold classifier. Lastly the study investigates how accurate and robust the proposed approach can evaluate about 13,000 records in 8,000 different devices.

4.3 Eavesdropping

A wireless network is highly prone to eavesdropping as the network is not secure enough (Shriraghavan, Sundaragopalan, Yang and Jun, 2003). Lack of proper security open the network to any unauthorized user who can easily access sensitive data on transit. An attacker only needs to be within the access point range. Eavesdropping is categorized into passive and active types. A passive attacker monitors the network for transmitting message content to learn about the network activities. An active attacker on the other hand, modifies the data transmitting on the communication channel. By injecting the packets in a specific pattern the time necessary to establish the message content is reduced. The modifications may involve changing the destination IP address of an encrypted message to the attacker host.

The widespread use of mobile devices has also increased the risk of sensitive information being exposed. In their study Maggi, Gasparini and Boracchi (2011) demonstrate how touchscreen phones can be attacked. The researchers design a mechanism to show how keystrokes can be detected from a touchscreen. The technique filters and validates the detected keys that are later reconstructed to create an accurate keystroke sequence. Their model requires that an attacker needs to point a camera from a smartphone towards the targeted device. However, the model would be effective if the target keyboard must display a feedback and the attacker should be able to know the model of the target phone. Such a model shows how sensitive data can be captured through eavesdropping.

Ahmed, et al. (2009) refers to mobile eavesdropping as signal interception which is used to acquire data being transmitted using wireless connections. Hardware devices that intercept signals may be used as well as software techniques. Making eavesdropping a major threat to mobile device users. Wireless signals are intercepted and an attacker decrypts the signals in case they are encrypted. The signal can also be tapped from transmission medium, from switches, or servers. Tapping the information can be made by mobile operators, security enforcement agencies, as well as individuals or organizations that have the necessary equipment and skills.

The security concern of wireless communication eavesdropping is also highlighted by Halevi and Saxena (2013). Wireless communications are considered easy to eavesdrop due to the use of RFID, Bluetooth, microphones, and WiFi. In their research they look at bootstrapping two wireless devices in a secure environment, how eavesdropping can be resisted and attacks from a man-in-the-middle. Their study also investigates acoustic eavesdropping attacks whereby the acoustic signals result from vibration or button clicking. Results from the study indicate that security levels in a paring operation are weak and opens up devices for attacks.

4.4 Network Traffic

Traffic identification and its measurement are important in creating network policy as well as in network business intelligence (Sandvine, 2015). Different traffic types include traffic from a website, application, provider, service, and protocol. The traffic can then be categorized in terms of gaming, storage, file-sharing, communications, administration, marketplaces, administration, social networking, tunneling, web browsing and real-time entertainment. Identification of traffic may be faced with some challenges, including stateful protocols, related flows and sessions, routing asymmetry, tunnels and encapsulation as well as devices and tethering. There are many techniques available that can be used to identify traffic as well as extract information that can be used to measure traffic.

Encryption of the transmitted payload is not enough security as shown in the study conducted by Park and Kim (2015). In their work they show that encrypted traffic is prone to analysis to classify user activities. The proposed model uses a learning machine technique on an instant messaging service to classify user activities. Results from their experiment showed that activities generate a unique packet sequence that can be used to accurately infer user activities. Such results show that privacy of data, such as chats, contacts, photographs and other personal information can be accessed from captured traffic.

Analysis of traffic to collect information that can be useful can be done while looking for digital evidence as seen in a study by Walnycky, Baggili, Marrington, Moore and Breitinger (2015). 20 popular Android apps for instant messaging were used in the study with most of the tests able to reconstruct entire messages. Such a study shows that the applications on the smartphones have poor security, and may be used for malicious purpose. Network traffic to and from the test devices was used to perform the experiments. From the network analysis, it was possible to capture text messages, multimedia content, URLs, and chat logs. Results from the study indicate that most of the apps used on smartphones have vulnerabilities that make them easy to infringe user privacy.

5 Conclusion

Smartphone usage is on the rise due to technological advancement. However, this has resulted in new forms of security risks for such devices. Smartphones are built to perform tasks that can be done using personal computers. This means that a lot of sensitive data can be stored on the devices. The works compared in this paper indicate that different methods can be used to identify a particular smartphone as well as the actions of a user. Fingerprinting of a device poses a safety threat to user information. Although methods such as encryption are used the studies show that such security measures can be bypassed. An attacker will be able to analyze traffic generated by the phones and use it to fingerprint a specific device. Using mobile apps that produce background information that a user is not aware of makes it easier for an attacker to collect useful data. Fingerprinting Smarphone devices may be useful to security agencies in deanonymizing criminal activities, but may also be used for ulterior motives.

The two articles compared are related as they show how a user privacy can be infringed. While different methods are used the overall result can be attributed smartphone security. Although users may adapt safety precautions when using their devices, there is no guarantee that their information is safe. Therefore, it is necessary for measures to be developed to protect user information from attackers or other malicious actions.

6 References

1) Ahmed, M., Penney, J., Ikki, S., Salami, A., Bath, T., Allah, M., et al., 2009, *Threats to Mobile Phone Users' Privacy*, Canada.

2) Bojinov, H., Boneh, D., Michalevsky, Y., and Nakibly, G., 2014, *Mobile Device Identification via Sensor Fingerprinting.*

3) Conti, M., Mancini, L., Spolaor, R., and Verde, N., 2015, *Can't You Hear Me Knocking: Identification of User Actions on Android Apps via Traffic Analysis.* Texas.

4) Dai, S., Tongaonkar, A., Wang, X., Nucci, A., and Song, D., 2013, *NetworkProfiler: Towards Automatic Fingerprinting of Android Apps.* IEEE.

5) Goode., 2013, *Study into the implications of Smartphone operating system security.* London: Goode Intelligence.

6) Halevi, T., and Saxena, N., 2013, Acoustic Eavesdropping Attacks on Constrained Wireless Device Pairing – Final, *IEEE Transactions on Information Forensics and Security,*

pp.563-577.

7) Hogben, G., and Dekker, M., 2010, *Smartphones: Information security risks, opportunities and recommendations for users*, European Union: ENISA.

8) Kurtz, A., Gascon, H., Becker, T., Rieck, K., and Freiling, F., 2016, Fingerprinting Mobile Devices Using Personalized Configurations, *Proceedings on Privacy Enhancing Technologies*, pp. 4-19.

9) Maggi, F., Gasparini, S., and Boracchi, G., 2011, A Fast Eavesdropping Attack Against Touchscreens. *Information Assurance and Security (IAS), 2011 7th International Conference*, pp. 320 – 325, Melaka: IEEE.

10) Miskovic, S., Lee, G., Liao, Y., and Baldi, M., 2015, *AppPrint: Automatic Fingerprinting of Mobile Applications in Network Traffic*, Springer International Publishing, Switzerland.

11) Park, K., and Kim, H., 2015, Encryption is Not Enough: Inferring User Activities on KakaoTalk with Traffic Analysis, *16th International Workshop*, WISA 2015, pp. 254-265, Korea.

12) Sandvine., 2015, *Identifying and Measuring Internet Traffic: Techniques and Considerations.*, [online], Available at:¡https://www.sandvine.com/downloads/general/whitepapers/identifying-and-measuring-internet-traffic.pdf¿ [Accessed 30 April 2016].

13) Shriraghavan, S., Sundaragopalan, S., Yang, F., and Jun, J., 2003,*Introduction to Information Security.*

14) Stober, T., Frank, M., Schmitt, J., and Martinovic, I., 2013, *Who do you sync you are? Smartphone Fingerprinting via Application Behaviour*, Budapest.

15) Trend., 2011, *5 Simple Steps to Secure your Android-Based Smartphones.* [online] Available at : ¡ http://www.trendmicro.com/media/misc/secure-your-android-based-smartphones-en.pdf¿ [Accessed 29 April 2016].