Mobile Security Behavior Observatory: Long-term Monitoring of Mobile User Behavior

Akira Yamada
KDDI Research, Inc.

Shoma Tanaka
KDDI Research, Inc.

Yukiko Sawaya
KDDI Research, Inc.

Ayumu Kubota
KDDI Research, Inc.

So Matsuda
KAYAC Inc.

Reo Matsumura
karakuri products Inc.

Shun Umemoto
SecureBrain Corporation

Jun Nakajima
SecureBrain Corporation

Kyle Crichton
Carnegie Mellon University

Jin-Dong Dong
Carnegie Mellon University

Nicolas Christin
Carnegie Mellon University

Abstract

Rather than due to a lack of security enforcement mechanisms, security issues often result from a combination of attacker deception and user inability to follow proper security precautions. In the case of smartphone security, the centralized app store model, through which the majority of software is delivered to users, provides substantial protections by filtering out (most) malware. As such, the security of a smartphone device significantly relies on user decision-making; where they download their applications from, which links they choose to follow, which websites they visit. To observe these decisions, we develop a mobile security behavior observatory (MSBO) platform for smartphone devices predicated on previous work studying the behavior of home computer users [4]. Through the platform, participants consent to smartphone use data collection, and also elect to receive questionnaires on their device. This paper describes our architecture, reports initial engagement results, and discusses how we plan on using the MSBO for future user studies.

1 Introduction and Related Work

Over the past decade, the penetration of mobile devices in the global market and the use of these devices to connect to the internet has grown to the extent that mobile internet use has exceeded that of personal computers [1, 10]. With these trends projected to increase over the next five years, it is vital for researchers to understand how users interact with their mobile devices and how their behavior differs from that on desktop computers. While observational studies of user behavior through instrumented client machines have been conducted on personal computers [4,7], such an approach has not been applied to smartphone devices.

Our proposed architecture fills this gap by providing a light-weight client-side application that can be installed on a user's smartphone and collect data without interfering with the normal use of the device. Our work builds on previous observational data collection efforts conducted at Carnegie Mellon University. The project, called the Security Behavior Observatory (SBO), implemented a similar client-server architecture for collecting data on user interactions with their home computer [4]. Based on lessons learned from the SBO study, we design our system to collect data only on metrics that were found to capture a depth of insight into user behavior, specifically, data on the user's web browsing, different application usage, and the installation of new applications. This limits the amount of data collected, while providing a similar level of insight into user behavior as the desktop SBO version. This design enables the application to be deployed on mobile devices with fewer computing resources than personal computers, and to remain invisible to the user while conducting background data collection.

With the data collected through our Mobile Security Behavior Observatory (MSBO), we will be able to investigate, at scale, how users interact with their mobile devices, and, for instance, to build on previous studies that predict whether a user will be exposed to malicious web content [2, 3, 5, 8, 9]. With a larger, more diverse participant pool, we will be able to further identify behavioral differences across cultures and devices. Beyond investigating security issues, our work will hopefully provide a blueprint for empirical smartphone user research across a broad spectrum of disciplines.

2 Data Collection

The MSBO is a client-server architecture where participants install a "sensor" app on their smartphones. The app captures multiple user metrics to assess security, and uploads the captured data to a central server. We target Android OS

USENIX Symposium on Usable Privacy and Security (SOUPS) 2020.
August 9–11, 2020, Virtual Conference.

Data	Description
Web browsing behavior	Navigation URLs, anchor tag text (Chrome browser, and Chrome component apps)
App usage	App usage history and activity history
Installed apps	Installed app list: name, hash values, certificates
Network type	Network type (cell, wi-fi), public IP address
SMS	URLs embedded in SMS messages and fuzzy hashes of these messages
Device info.	Device name, OS version, and patch level
Other	Third-party market app installation.

Table 1: Collected data.

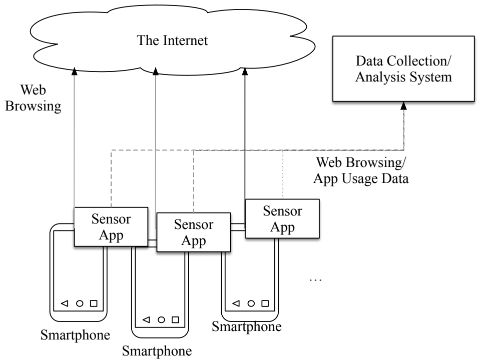

Figure 1: Data flow between sensor app and collection server.

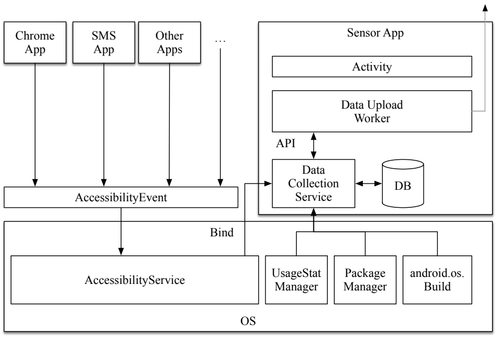

Figure 2: App Configuration.

versions 6, 7, 9, and 10; Android is not only the most popular smartphone operating system, but also provide flexible monitoring functionalities for app developers. Tab. 1 lists the data the MSBO collects. Users reportedly install malicious apps through web downloads, in-app installs, and SMS links. Thus, we capture both web browsing behavior, and app usage and installation history. We capture URLs in SMS messages, along with fuzzy hashes of the corresponding messages. We are also interested in app installation activity from third-party stores. Finally, we collect device and network information, including IP addresses, to understand whether user environment plays a role in malware distribution.

Besides the data collection process, participants can self-report malicious web sites or message uploading screenshots via the app interface. The app also disseminates optional weekly technology survey URLs, and shares survey results graphically.

3 Data Flow

Fig. 1 presents the data flow between our sensor apps and the data collection server. Participants install the sensor app from Google Play, and then the app monitor other apps as a background app. It has two monitoring functionalities: realtime display inspection and periodic data collection. The sensor app starts capturing display text when users run specific apps such as Google Chrome, SMS, and Settings, and stores the data in the database. The app also periodically collects app usage history and install apps. The client app uploads the data stored in the database when the internet connection is available. The entire collection and update processes are entirely transparent to the user and do not affect their usual users' activities.

Fig 2 shows the sensor app configuration. Our sensor app extensively relies on Android's Accessibility Service, which is designed to provide alternative navigation feedback to applications installed on Android devices. For example, the Accessibility Service can be used to convert text to speech, or to warn of malicious web sites in addition to other tools (e.g., Google Safe Browsing). Most apps (e.g., Chrome, SMS, ...) fire AccessibilityEvents to communicate UI changes to the Accessibility Service.

The MSBO app binds its own Data Collection Service to the Accessibility Service. That way, as long as the user grants Accessibility Service permission to the MSBO app, the Data Collection Service can capture whatever text is displayed in the app the user is running; e.g., the URL in the navigation bar, any anchor text in the browser, or any URL in an SMS.

The second major component of the MSBO is a DataUpload Worker. This worker, under Android's WorkerManager, uploads collected data as a background service. These uploads are scheduled, deferrable, asynchronous tasks, and are resilient to app crashes or device restarts.

4 User Study Methodology

We do not incent user participation with financial rewards. Instead, we offer popular anime character icons to participants; users can directly interact with these characters through the phone UI. As users continue participating, their characters can display additional emotions, and feature new color schemes and other decorative add-ons as rewards.

We recruit users through the websites of eight Japanese organizations which participate in this research project. We then publish the MSBO app to the Google Play store (Japan only). The recruiting process is entirely anonymous and passive, and we do not communicate with participants at this point. We use an (IARC) content-rating badge on the store page to dissuade minors from participating.

When participants download the app and start it, it displays this research overview, the data collection agreement, and the consent form. At that point, participants can self report their age, so that we can further screen out minors. Participants also have the option of leaving this study after reading the conditions. We only start the data collection process after participants understand and consent to the terms.

5 Ethics and Participant Privacy

Data are collected in Japan, by Japanese companies. In lieu of an academic IRB, we rely on an ethics board approval, which includes external, independent privacy experts; US researchers in the team do not collect data and are covered by an IRB-approved data-sharing protocol.

Any (non-minor) Google Play Store (Japan) user can join the experiment. Participants can also withdraw at any time by simply uninstalling the app. Additionally, the app provides users with a one-click option to request deletion of the collected data. Since deletion requires interaction with our back-end server, it cannot be immediate. Instead, users need to provide an email address so that they can be notified when their data has been purged.

We deploy a simple heuristic PII filter inside the app, which attempts to identify email addresses, phone number, credit card number, SNS account names, and passwords, and automatically purges these strings from the collected data. To preserve participant privacy, we only collect a hash of the SMS sender information in SMS that contain a URL. Additionally, we only capture the fuzzy hash [6] of the SMS, rather than its textual content.

6 User Engagement

We started the app distribution via Google Play Store, on March 16, 2020. We had multiple advertisements on that day. As of May 14, 2020, 2,031 participants had installed the app. During the first week (March 16–22, 2020), we recorded 1,502 installations, but approximately 25% dropped out by not opening the app, not consenting to the agreement, refusing to grant the requested permissions, or not configuring the app settings. After about a month, about 20% of participants have continued to use the app and became stable active users. As a result, we currently have about 500 daily active users.

Acknowledgement

The research results have been achieved by WarpDrive: Web-based Attack Response with Practical and Deployable Research InitiatiVE, the Commissioned Research of National Institute of Information and Communications Technology (NICT), JAPAN. K. Crichton was partially supported by a gift from Bosch, and J.-D. Dong was partially supported by a Carnegie Institute of Technology Dean's Fellowship.

References

[1] Monica Anderson. Mobile technology and home broadband 2019, Dec 2019. https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/.

[2] Leyla Bilge, Yufei Han, and Matteo Dell'Amico. Riskteller: Predicting the risk of cyber incidents. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, pages 1299–1311, New York, NY, USA, 2017. ACM.

[3] Davide Canali, Leyla Bilge, and Davide Balzarotti. On the effectiveness of risk prediction based on users browsing behavior. In *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security*, ASIA CCS '14, pages 171–182, New York, NY, USA, 2014. ACM.

[4] Alain Forget, Saranga Komanduri, Alessandro Acquisti, Nicolas Christin, Lorrie Cranor, and Rahul Telang. Security Behavior Observatory: Infrastructure for Long-term Monitoring of Client Machines (CMU-CyLab-14-009). Jul 2014.

[5] Chanhyun Kang, Noseong Park, B. Aditya Prakash, Edoardo Serra, and V. S. Subrahmanian. Ensemble models for data-driven prediction of malware infections. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, WSDM '16, page 583–592, New York, NY, USA, 2016. Association for Computing Machinery.

[6] Jesse Kornblum and 1 Tsukasa Oi. Ssdeep - Fuzzy hashing program, Apr 2018. https://ssdeep-project.github.io/ssdeep/index.html.

[7] Fanny Lalonde Levesque, Jude Nsiempba, Jose Fernandez, Sonia Chiasson, and Anil Somayaji. A clinical study of risk factors related to malware infections. pages 97–108, Nov 2013.

[8] F. L. Lévesque, J. M. Fernandez, and A. Somayaji. Risk prediction of malware victimization based on user behavior. In *2014 9th International Conference on Ma-*

licious and Unwanted Software: The Americas (MALWARE), pages 128–134, Oct 2014.

[9] Mahmood Sharif, Jumpei Urakawa, Nicolas Christin, Ayumu Kubota, and Akira Yamada. Predicting impending exposure to malicious content from user behavior. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, pages 1487–1501, New York, NY, USA, 2018. ACM.

[10] StatCounter. Mobile and tablet internet usage exceeds desktop for first time worldwide, Nov 2016. https://gs.statcounter.com/press/mobile-and-tablet-internet-usage-exceeds-desktop-for-