Device Analyzer: a privacy-aware platform to support research on the Android ecosystem

Daniel T. Wagner
Computer Laboratory
University of Cambridge
dtw30@cam.ac.uk

Daniel R. Thomas
Computer Laboratory
University of Cambridge
drt24@cam.ac.uk

Alastair R. Beresford
Computer Laboratory
University of Cambridge
arb33@cam.ac.uk

Andrew Rice
Computer Laboratory
University of Cambridge
acr31@cam.ac.uk

ABSTRACT

Device Analyzer is an Android app available from the Google Play store. It is designed to collect a large range of data from the handset and, with agreement from our contributors, share it with researchers around the world. Researchers can access the data collected, and can also use the platform to support their own user studies. In this paper we provide an overview of the privacy-enhancing techniques used in Device Analzyer, including transparency, consent, purpose, access, withdrawal, and accountability. We also demonstrate the utility of our platform by assessing the security of the Android ecosystem to privilege escalation attacks and determine that 88% of Android devices are, on average, vulnerable to one or more of these type of attacks.

1. INTRODUCTION

Device Analyzer [1] is an Android app available from the Google Play store. It is designed to collect a large range of data from the handset and, with agreement from our contributors, share it with researchers around the world. Our app collects data on: apps installed, when they are run and the resources they consume; data transferred and the air interfaces used; phone calls placed and received as well as text messages sent and received; battery levels, charging cycles and power usage; Bluetooth devices and WiFi access points seen and connected to; handset location; as well as other system parameters [2].

Our contributors install the app from the Google Play store and in return we provide them with summary statistics in the app, and a copy of their data from the project web page. We have received contributions from over 23,000 devices around the world since May 2011 and our archive now contains over 100 billion records.

WiSec'15 , June 22-26, 2015, New York, NY, USA
ACM 978-1-4503-3623-9/15/06.
http://dx.doi.org/10.1145/2766498.2774992

2. PRIVACY-AWARE DATA SHARING

We encourage third-party researchers to use Device Analyzer in two ways: (i) researchers can sign a legal agreement with the University of Cambridge to access data from contributors who consent to data sharing; and (ii) contributors can enter a participation code to take part in a study run by external researchers. The difference between the two is the amount of data available: in the former case sensitive data items are not made available; in the latter case they are. The Device Analyzer website [3] contains instructions on how to access the data. External researchers should go through their own ethics review and in all cases contributors install our app from the Google Play store.

Our approach is based on both the European legal framework, and the seven principles of Privacy by Design [4]. Our approach is compatible with an earlier set of recommendations made to researchers in ubiquitous computing [5]. An overview of our approach follows.

Transparency, consent, and purpose: Our data collection is both transparent and explicitly given by the contributor. Device Analyzer is distributed as a standalone app on the Google Play store; it is never bundled with other software or pre-installed on devices. We require consent inside the app to activate data collection and remind the contributor of on-going data collection via a monthly notification. These notifications also raise awareness of the app if it is installed by someone else without the owner's consent (e.g. a stalker installing the app on a victim's phone). Data collection can be suspended at any time inside the app.

Security: Communication is secured with TLS and data stored on local servers in the Computer Laboratory.

Access and withdrawal: Contributors can view summary data on their smartphone. Space constraints prohibit accessing the full archive on the device itself, but contributors can download their data from the project website. Contributors can delete their data and withdraw from the study at any time from within the app. We do not provide external researchers with access to any data until it is at least three months old. This is a "try before you buy" approach with the added advantage that older data is less invasive in many cases. We do see active use of withdrawal and opt-out features. Between May 2011 and December 2013 approximately 40% of our 26,800 installations were never activated, 4% of contributors withdrew from further collection and 2.5% re-

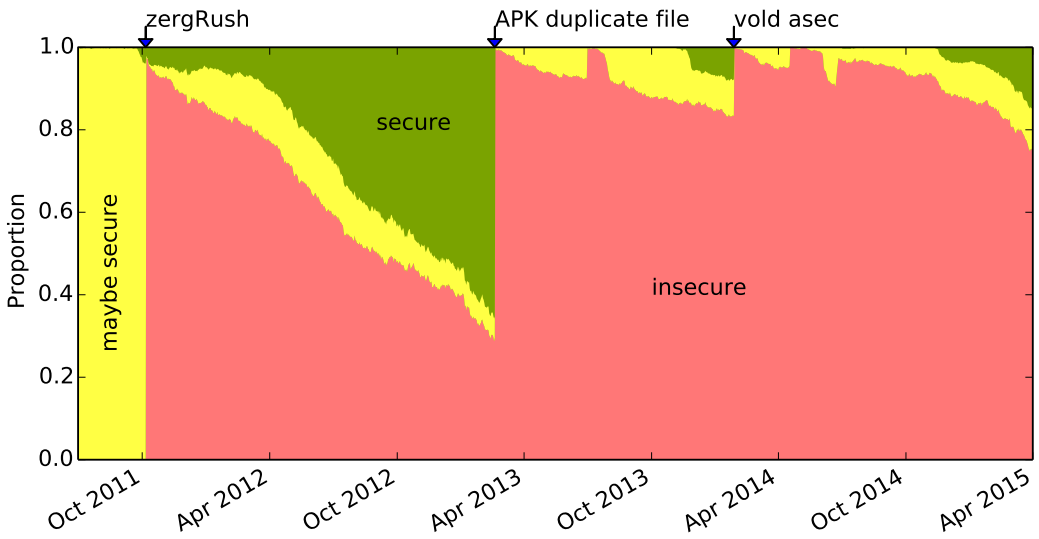

Figure 1: Proportion of devices running insecure, maybe secure and secure versions of Android against time. Large increases in insecure devices occur when new vulnerabilities are discovered, with those which have the biggest impact (zergRush, APK duplicate file, vold asec) annotated.

quested all of their historic data be deleted.

Accountability: We provide a quick feedback feature inside the app to collect anonymous feedback and a working email address in the app and on the project website.

Proactive privacy: Device Analyzer collects metadata rather than direct personal identifiers. To do so, we apply a salted hash to direct personal identifiers on the device itself before they leave the handset. For example, we do not collect an audio recording of a phone call; we record time, duration and a salted hash of the phone number called. This means that we do not know the number called, but we can tell when a contributor calls the same number twice. This also means we cannot trivially construct a social network between our contributors.

Privacy by default: Contributors must explicitly opt-in to reveal some data items. For example, by default, GSM cell tower identifiers and Wifi AP names are preprocessed by a salted hash function on the device. By default, location data is only available to researchers in Cambridge.

3. EXAMPLE: ANDROID SECURITY

We have used Device Analyzer to answer many research questions. Here is one: *Can my phone prevent a malicious app from gaining root on my device?* Device Analyzer collects OS version strings and build numbers, and we have compared these with 11 publicly disclosed privilege-escalation attacks drawn from our database on AndroidVulnerabilities.org. We have looked at data over 3 years and 20,100 devices and found that, on average, 88% of devices were exposed to known privilege-escalation attacks.

Figure 1 provides a more detailed view. A device is marked as *insecure* if both: (1) it is running a version of Android which is vulnerable to at least one privilege escalation attack; and (2) the device has not received an update which might fix the vulnerability; a device is marked *maybe secure* if it has received an update which might fix the vulnerability; and a device is *secure* if it is running a version of Android which is not vulnerable to any publicly disclosed vulnerabilities on the date in question.

4. CONCLUSION

In this paper we have introduced Device Analyzer and described how external researchers can use the platform to answer research questions and conduct new user studies. We have provided an overview of the privacy-enhancing techniques we have integrated into Device Analyzer in order to protect the privacy of our contributors and provide an example usage of the data, showing that 88% of Android devices are, on average, vulnerable to one or more privilege escalation attacks.

5. ACKNOWLEDGEMENTS

This work was supported by the University of Cambridge Computer Laboratory Premium Studentship scheme, a Google focussed research award and the EPSRC grants EP/P505445/1 and EP/L504920/1.

References

[1] D. T. Wagner, A. Rice, and A. R. Beresford, "Device Analyzer: Understanding smartphone usage.," in *Mobile and Ubiquitous Systems: Computing, Networking and Services (Mobiquitous 2013)*, Springer, 2014, pp. 195–208, ISBN: 978-3-319-11568-9.

[2] `https://deviceanalyzer.cl.cam.ac.uk/keyValuePairs.htm`.

[3] `https://deviceanalyzer.cl.cam.ac.uk/`.

[4] C. Information Commissioner of Ontario, *Privacy by design*, See `https://privacybydesign.ca/`.

[5] M. Langheinrich, "Privacy by design—principles of privacy-aware ubiquitous systems," in *Proceedings of the Third International Conference on Ubiquitous Computing (UbiComp)*, Atlanta, Georgia, USA: Springer-Verlag, 2001, pp. 273–291, ISBN: 3-540-42614-0.