Screwbots

by Alessio Malizia and Alan Chamberlain

Editor's Introduction

Have you ever thought of "lying" to your smartphone to protect your privacy? Everyday we face a dilemma about privacy: We take advantage of apps that are able to use our location or data to provide "smart" services at the expense of privacy (we all know our data can be supplied to third parties), or we cling on to our privacy and ignore the benefits of such smart technologies.

It does not have to necessarily be like this, in this article we describe the rise of a new kind of intelligent apps we called Screwbots—programs that can access the personal data we share on the cloud and scramble it ("screw it up"), or intentionally lie by reducing data accuracy, to protect our privacy.

Screwbots

by Alessio Malizia and Alan Chamberlain

Have you ever thought of lying to your smartphone about your location in order to protect your privacy? Do you know that generally this data can be supplied (anonymously) to third parties, both to drive services for the user and to tailor and target advertisements? But what if instead of location-based data, it were data about your personal health being passed on to companies?

The debate about our shrinking right to privacy has been going on ever since the birth of public access to the Internet in the mid-1990s, if not before. However, it is becoming more and more intense with the almost daily introduction of more technologically advanced and flexible smartphones, and the increasing use of cloud-based Web services.

It seems to be a Hobbesian choice: either abandon privacy or forego some of the key benefits of advancing technology. However, there may be another possibility.

Have you ever thought of "lying" to your smartphone to protect your privacy? It is possible, and seemingly increasingly necessary. "Screwbots" are programs that can access the personal data we share on social networks and scramble it ("screw it up"), or even intentionally lie by reducing data accuracy to protect privacy.

Let's envision a diffusion of such programs, for instance, in the form of operating system layers, running on devices (smartphones, ad-hoc, etc.) to filter personal data depending on the application and applying user-defined purposes to protect user's privacy but keeping the benefits of context-aware applications.

But first, just how big is the problem?

The fact is, there are a multitude of times when we are unaware that our data could be sold to third parties. Even more troubling is not knowing if companies have adequate security measures in place to protect, for instance, health information from being passed on to inappropriate third parties from social networks such as Facebook, Twitter, or Foursquare.

Recently, a few examples of applications that can resolve those issues have arisen: apps that are able to "lie" in a smart and intelligent way about our exact location without limiting the use of location-based applications. There are also apps, which are able to generate big virtual traffic

jams on Google maps to screw-up data about, for example, our commuting habits. A hacker was able to cause drivers to divert around certain zones, in essence breaking Google traffic and Waze services and thus clearing the way for him to have a smooth ride. This caused Google and Waze to strengthen their online security protocol preventing such attacks and at the same time preserving the user's privacy. Nevertheless, it was demonstrated that such applications, i.e. Screwbots, are indeed feasible. This breed of application might go some way to help solve the privacy versus benefits of new technologies dilemma framed above.

The Personal Data Sharing Dilemma

Imagine you have just arrived in a foreign city and would like to find directions from the airport to your hotel. You could simply turn on your smartphone and by using the GPS on the device, pinpoint your current location and then get directions to the hotel. Once at the hotel you might like to take a walk and have a coffee, but where is the closest coffee shop? Once again you could use your smartphone and give (share) your location with a navigation app. And the day after? You like jogging and would like to share your newly discovered route in a famous park with friends or maybe review your run performance later. No problem. Your wrist device or smartphone can be used again, but obviously at the cost of sharing personal data about your location, heart rate, your search terms, and so on.

Think about how many people use apps for activities such as running and jogging, in which a plethora of sensors are used to collect data. This data is often recorded and used by specific apps. Such apps generally use smart-watches, wrist devices, or even smartphones to collect data and then use cloud-based Web services. Nowadays, more and more health data is shared in the cloud, for different medical, research, and recreational purposes. With this transparency expected to expand in future, it raises questions about privacy leaking [1]. Think about geneticists who use Amazon cloud services to store petabytes of human genetic data. Health data is clearly at risk of leaking to third party companies [2]. A new breed of devices and smartphone apps allows users to share their data on social networks (and thus there are issues relating to privacy). Even pedometers, which are a mainstream technology nowadays, can share information on the cloud, tracking location at every step. As Eugene Vasserman, a researcher focusing on cyber security and privacy at Kansas State University, said: "They know where I sleep. They know my address" and "I'm aware of the tradeoff I'm making … [but] I don't think people understand what they're giving up by putting this data out there." He continued, "The direct repercussions are not quite clear because the definition of the cloud—excuse the pun—is very nebulous" [2].

A trade-off might exist where apps are allowed to intentionally lie about our personal data (by scrambling, filtering, etc.) prior to sharing it, so the data might still be partially available for a coarse-grained use by corresponding apps or devices without undermining our privacy, i.e. Screwbots.

Where the Apps Dare to Lie

Dewri and Thurimella's "Can a Phone's GPS Lie intelligently" recently introduced default privacy zones as a way of solving the privacy dilemma of sharing personal data [3]. Default privacy zones allow users to define finer-grained privacy controls that let them share just enough locational information to achieve a desired quality of service. The concept is based on the fact that many location-based applications (AroundMe, Loopt, Foursquare, etc.) search for local shops or services, and present the results as a list ordered by the user's proximity to such businesses instead of an exact GPS location. This means the app does not need the user's exact location, just the position within an area, which would suffice to present the user with adequate information in relation to surrounding businesses. For instance, if the user is interested in finding a car shop within a range of say 30 kilometers, this might translate into a list of the 10 closest shops. The area within the closest shops might act as a default privacy zone for that specific query in order for the user to be anywhere within that area and still get related results.

In conclusion, a location-based device could use a certain degree of location inaccuracy to frame a default privacy zone useful enough for an app to retrieve adequate results. The level with which the system could "lie" would vary depending on the application's purpose and user's requirements. In such instances it would be important to make such a system intelligible, that is to say the user would be able to understand the state of the privacy of their data within the system, to ensure the system was using their data in an ethical manner.

Not only could location-based services share some unwanted data, but a plethora of new smartphones equipped with biometric sensors could send personal data with a level of details that might be undesirable for end users. For instance, many smartphones now come with heartbeat rate sensors that users might activate by just touching certain parts of the device itself (normally placed on the back). While this data might be encrypted and used by medical or fitness applications it can also be used by emotion recognition applications as described by Kanjo et al. [4]. For instance, by reducing the heartbeat sampling rate, a screwbot application could make emotion recognition relatively hard but still allow data to be good enough for

statistics on fitness performance. This will indeed free the user from unwontedly sharing their emotional state while at the same time keep track of their fitness progress.

A New Breed of Apps: The Screwbots

Issues and dilemmas have started to emerge relating to user privacy and data sharing. These emerging concerns are most prevalent to research fields such as: location-based services, health informatics, recreational activities and in general to context-aware applications.

Different research areas can possibly contribute to the growth of such applications depending on the service-level support offered to fine tune privacy controls. For example, research in privacy-preserving protocols and fast implementations of cryptographic protocols might help sending filtered ("false") location information without being detected. For example TLS (Transport Layer Security) protocol protects data in transit, but doesn't prevent an app from sending false data through a tunnel created by the app itself. For example, an app can save data packets with information about location, pulse, etc., which could be sent later with modified cookies, platform key, and time stamps.

Algorithmic approaches, such as default privacy zones, can allow users to define finer-grained privacy controls that let them share just enough locational information to achieve a desired quality of service. For instance, studies have appeared in *Nature*'s scientific reports on understanding human mobility as a patterned habitual behaviour. These have suggested it would be possible to identify a user from only four location-based data points. Default privacy zones might help preventing such predictable behaviour thus protecting our privacy.

Database management and data hashing may increase anonymity in medical data by automatically generalizing, substituting, and removing information as appropriate without losing many of the details found within the data. For instance, in recreational areas, for runners wearing specific devices, anonymized data would still be useful to evaluate performance on track or historic data for training purposes. An example might be the notion of a minimal bin size (as defined by The United States Social Security Administration), which reflects the smallest number of individuals matching some specific characteristics. The larger the bin size, the more anonymous the data. The number of people to whom a record may refer increases proportionally to the bin size, thus masking the identity of the actual person. Just like those statistical grouping methods, a privacy zone, for example, might be used to obscure personally identifying location-based data. We can envision a sort of minimal bin size set of location and time-based data that can preserve anonymity while still providing a useful contextual service.

Acknowledgments

This work is partially funded by the project n. 220050/F11 – granted by Research Council of Norway. We would also like to acknowledge the EPSRC awarded grant, Living with Digital Ubiquity EP/M000877/1, the Personal Data and Trust Network (Social & Cultural Innovation Group), and the Tiree Techwave, which gave us valuable time to think about the issues we have raised.

We would like to acknowledge the reviewers for their very valuable suggestions that helped improving this work.

About the Authors

Alessio Malizia is Senior Lecturer in the Department of Computer Science and member of the Human-Centred Design Institute at Brunel University, London, UK.

Alan Chamberlain is Senior Research Fellow in the Mixed Reality Lab, Computer Science, and University of Nottingham, UK.

References

[1] Narayanan, A. and V. Shmatikov. Myths and fallacies of "personally identifiable information." *Communications of the ACM* 53, 6 (2010), 24-26.

[2] Hernandez, D. World's health data patiently awaits inevitable hack. *Wired*. March 25, 2013.

[3] Dewri, R. and R. Thurimella. Can a phone's GPS lie intelligently? *Computer* 46, 2 (2013), 91–93.

[4] Kanjo et al. Emotions in context: Examining pervasive affective sensing, systems, applications and analyses. *Personal and Ubiquitous Computing Journal* 19, 7 (2015).

DOI: 10.1145/3005397

