Brave New World: Privacy Risks for Mobile Users

Paarijaat Aditya[†] Bobby Bhattacharjee[‡] Peter Druschel[†]
Viktor Erdélyi[†] Matthew Lentz[‡]
[†]Max Planck Institute for Software Systems (MPI-SWS) [‡]University of Maryland-College Park

I. Introduction

Sophisticated mobile computing, sensing and recording devices are commonplace. Smart phones have achieved significant penetration and novel devices like Google Glass are imminent. These devices can serve most functions of a conventional notebook computer, but also have a range of additional capabilities, including image/audio/video recording, GPS location, compass, accelerometer, near-range radio (NFC and Bluetooth), and soon health and fitness monitors.

Moreover, these devices are carried by their users virtually around the clock, blurring the distinction between the online and offline world and enabling transformative new applications and services. For instance, mobile apps can provide location and activity-sensitive services and information, in the case of Google Glass overlaid right onto a user's field of view. They can record what the user does, sees and hears for future reference; and they can keep track of a user's encounters with nearby users' devices to enable communication related to a shared experience or event.

However, these applications and services also introduce a range of new threats to users' privacy. While a user carries it, a mobile device can capture a complete record of the user's location, online and offline activities, and social encounters, potentially including an audio-visual record. While such a record is very useful to a user for their own reference and to enable new applications, it is also highly sensitive and inherently private. Unlike information users post on Facebook or Twitter, most users would likely not want to share such a comprehensive record with anyone.

In this paper, we catalog privacy threats introduced by these devices and applications. Our survey of threats underlines how privacy threats from mobile devices are fundamentally different and inherently more dangerous than in prior systems. For each specific risk vector, we describe technical challenges that, if solved, can mitigate its effects. We note that technical innovations merely provide a starting point: an end-to-end privacy-preserving infrastructure will require changes in how basic services are deployed, how laws are written and interpreted, and most importantly, a broad societal conversation about the value of user privacy. Then, we briefly sketch our own work on *secure encounters*, which provides a powerful primitive for secure communication among mobile devices. We conclude with a description of the current state of a user and their rights (and lack thereof) in this "brave new world"[1] of smart mobile devices with ubiquitous connectivity.

II. Risks

In this section, we catalog primary threats to users' privacy that arise from data captured by portable devices. Our taxonomy covers risks that are seemingly benign (service providers logging users' location to improve coverage) to legal (user's data stored on devices later subpoenaed by a court) to malicious (stolen device or data). As we list the risks, a theme that will recur is how little (or no) control users have over what personal data is collected, how this data is collected, where it is stored, how it is shared, and when it is deleted.

Whenever any organization collects, aggregates and stores users' personal information, three broad types of threats emerge:

Data exploit The organization may willingly exploit users' data to provide customized services, perform targeted advertisement, or simply sell the data (or derived knowledge) to third parties. Often, this use and its privacy implications are not transparent to users, and may not respect their wishes.

Data loss Software bugs, misconfigurations, theft, operator mistakes or rogue employees may cause the unintended loss, leakage or corruption of users' data.

Governmental (over)reach The organization can be compelled by government agencies (security services, law enforcement, courts) to reveal users' personal data. Recent events have shown that even democratic elected governments show little self-restraint in

[1] with apologies to Aldous Huxley.

sacrificing citizens' rights to privacy in the face of real or perceived security threats.

Not surprisingly, these risks increase with both the detail of personal information collected and the coverage the collecting organizations have over the general population. In the remainder of this section, we categorize risks by the primary vector for leakage.

II.A. Cellular service provider

In most countries, users choose cellular service providers relatively infrequently, and are often bound to annual or longer contracts that are expensive to void. In return, they receive initially subsidized mobile devices. Decisions made by service providers for mobile devices have enormous, perhaps disproportionate, influence on user's privacy. Just as importantly, users have essentially no say about service provider policies, and in many cases, do not know how their data is being stored or shared. Even when information comes to light, financial disincentives make it difficult for users to switch providers. Even without such disincentives, no providers' policies are transparent, and users literally do not have a provably privacy-preserving choice.

Cellular service providers can keep a log of all conversations (calls, TCP connections) and indeed packets generated by or incident upon a device. While some of this data is required for billing, privacy conscious providers could delete such data after the billing cycle ends, store it anonymized, or use other cryptographic techniques to preserve user privacy. Similarly, service providers log which cell towers a cellular radio equipped device communicates with. This log provides a detailed location track for individual devices, and coupled with the connection and packet data, provides a thorough glimpse into users' whereabouts and habits [14].

While the problem is not new (cellular providers have been able to track subscriber movement and call activity all along), the problem has been aggravated by the emergence of cellular data service, the ubiquity of increasingly capable mobile devices, and their large penetration in most parts of the world.

Research Challenge The key research challenge is to enable public infrastructure, such as a cellular network, that collects and stores only the minimum information required for correct operation. For example, when a user connects to the cell data network, it suffices to verify that the user has valid credentials for data transfer (i.e., has a valid contract), but not *who* the user is. If transaction history must be recorded, it could be stored such that it can be decrypted only with the user's consent. A similar design is feasible even for phone calls, whereby calls are routed to ephemeral numbers distributed by the user to their contacts. While the underlying cryptographic blocks may already exist [18], before such a drastic change can be deployed, basic research is required in demonstrating the scalability and economic viability of a "minimum-information" infrastructure. A more likely short-term measure is stricter government regulation on data that is collected and stored by service providers.

II.B. Device system software and OS

Along with the cellular provider, the device's operating system and system software provider have the most unfettered access to users' data. There are relatively few OS providers that account for the vast majority of all mobile device OSs: Google and the device manufacturer on Android devices, Apple on iOS devices, Microsoft on Windows devices, and Blackberry on Blackberry devices. System software is provided jointly by the device manufacturer and the service provider.

Data collection and retention policies within the OS and systems software are opaque to the user (unless the device is "rooted' and the user chooses to install a third-party, unsupported OS). As with data collected by the service provider, users are not aware of which data is collected, or how it is used. There are well publicized incidents, for instance, of the iPhone collecting and logging user's location data into hidden files, which were then uploaded back to Apple. The location data was initially stored unencrypted on the iPhone. According to Apple, the collected data was used to refine the WiFi access point database for augmenting GPS tracking [2].

This example highlights a fundamental problem: devices collect sensitive data (location), which is uploaded to unknown parties, and is used for unknown purposes. While Apple's published reasons for collecting user's data is plausible, such surreptitious collection still puts a user at risk because the data was stored on the phone (and could provide sensitive information to third parties who compromised the device.) The list of viruses and Trojans running on desktop and laptop computers surreptitiously collecting data is long; however, the difference in cases of mobile devices is the range of sensitive data available on a single device (location, activity, search history, friends) over all time (as opposed to when the user is interacting with a desktop or laptop).

Mobile devices ship with a standard software suite, including mail applications, calendars and schedulers, web browsers, navigation software (for GPS-enabled devices), and software for managing contacts. Often a single entity (the device manufacturer) provides this entire suite of applications, and is able to collect and correlate data across the entire application suite. Manufacturers and software provider end-user agreements do not explicitly identify what data is collected (or not), and how it is stored or shared.

Research Challenges Protecting user privacy without trust in the device OS platform is particularly challenging. Open-source platforms have an advantage here, as they can be inspected and certified by independent parties. Semantic attestation [20] of the OS based on a TPM might provide a solution for untrusted OS platforms. Another challenge is how to protect device integrity from untrusted third-party apps. Without it, attackers can compromise the user's system and collect information at the source, bypassing upstream security and privacy mechanisms. Potentially promising approaches include sandboxing of untrusted third-party apps, and information flow control to prevent the leakage of private information across the network.

II.C. Third-party Applications

A common paradigm in mobile applications is that apps connect to a Cloud-based backend service. In the simplest case, the backend stores a copy of the user's profile and preferences for convenience, durability, and availability across the user's different devices. In other cases, the backend provides a database queried by the app (e.g., a map service), or processes users' live audio/video (e.g., Siri speech recognition, augmented reality apps).

Depending on the nature of the application, highly sensitive data may end up being processed and stored by the app provider's site. In many cases, the information collected is more comprehensive than the information accessible to cellular providers. Unlike the telecom industry, however, which is subject to government regulation in most countries, app providers have fewer legal restrictions regarding the collection, handling, retention, and user of customer's data. In extreme cases, this highly sensitive data could be leaked by the provider due to a change in policy [3] or due to a internal misconfiguration [10].

The Friday mobile app [5], for instance, collects a device's time/location trace annotated with all of a user's online actions (phone calls, emails, chats, posts, note taking), and stores this data at a Cloud site, so a user can browse and annotate the trace conveniently from any of her devices. With platforms like AllJoyn [1], which enable device-to-device communication capabilities via Bluetooth and WiFi Direct, such traces will soon include a user's encounters with other devices.

In addition, numerous mobile apps are known to collect and upload to the Cloud more information than is required to perform the service they offer [13]. When an app is installed, mobile OSs requires users to grant permission for the app to access certain sensors and devices, e.g., camera, microphone or location information. However, users often lack the technical expertise to decide whether an app has a legitimate need to access an information source, or grasp the potential risks. Once installed, an app has unlimited access to the sources it requested. In addition, many apps do not follow the rule of least privilege when requesting permissions [19].

Research Challenges The primary research challenge is providing mechanisms that allow users to set reasonable and safe policies that protect private information collected by mobile apps. In particular, we need to distinguish different use cases and threat models. Data intended for exclusive use by the user can be encrypted prior to storage or upload to the Cloud provider. However, data that is to be processed by the app provider must (currently) be revealed. Techniques such as homomorphic encryption, oblivious RAM, and trusted hardware features like Intel SGX [22] or ARM TrustZone [16] may provide a solution that allows the provider to operate on encrypted information, but in general, efficient solutions are some ways off.

II.D. Tracking and recording by other devices

Additional privacy risks arise from the tracking and recording capabilities of mobile devices. Devices can track and be tracked by other devices, and can use their audio/video/image recording capabilities to capture nearby users.

Tracking Mobile devices have active radios and these radios can be used to track a user. Current devices do not cloak their protocol-specific information, and applications that try to continuously create local groups by actively connecting to peer applications allow the devices to be tracked by third parties. An attacker can simply record the

MAC address of a device at different places and track a user. All applications that create Bluetooth groups are susceptible [1, 8] as are applications that use NFC. The Bluetooth working groups have recognized this problem and the Bluetooth 4.0 standard includes randomized addressing modes that thwart such tracking at the Bluetooth link later.

Even if a user does not run a Bluetooth application, they may still be vulnerable to a MAC address tracking attack if they connect to WiFi access points; in this case the attacker can spoof well known provider access points (say `att-wifi` [9]) with high signal strength, or simply capture a packet dump of ongoing WiFi communication.

Moreover, these capabilities can easily be exploited to track devices using WiFi base stations. The larger the network of base stations that share tracking data, the larger the region over which devices and their owners can be tracked.

Research Challenges Randomized addresses, such as in Bluetooth 4.0, make tracking using packet captures much more difficult, but are not yet widely supported, and must be combined with measures to avoid leaking linkable information at higher protocol layers. Moreover, analog fingerprinting, whereby a device's radio is identified by variations during manufacturing, is a viable technique [17], and can be used to track devices, even those that do not leave a trackable signature (such as a MAC address) at the link layer or above.

To prevent device tracking via radio, existing technology like address rotation in Bluetooth 4 must be combined with protocols that enable communication with encountered devices, yet divulge no linkable information. A key research challenge is to provide an end-to-end, usable, MAC and upper-layer protocol stack that does not leak device-specific information.

Recording An additional risk arises from the fact that mobile devices are increasingly used to take images and audiovisual recordings, which may capture bystanders without their consent. The problem is aggravated by new devices like Google Glass, which can capture a continuous audiovisual record of everything the wearer sees and hears. Existing battery technology limits recordings to 30 minutes [6], but this limitation will be lifted with time.

The resulting trace of tracked devices and recorded audiovisual data is uploaded to an application provider, exposing the data to all of the risks described above.

Mitigation To mitigate privacy risks related to audiovisual recordings by other users' devices, techniques are needed to allow a user to automatically announce their privacy preferences (if and under what conditions they agree to be recorded and for what purpose). As a condition of agreeing to be recorded, a user's device might require an addressable identifier from a recording device, which can be used later to hold the recorder accountable for respecting the recorded user's preferences.

II.E. Compromised, stolen or seized devices

Mobile devices get lost, are stolen, and can also be seized by various authorities. Once an attacker has physical access to a device, it is usually relatively easy to obtain complete access to all data and software on the device. Along with access to user's personal data (location, passwords, online accounts), the attacker may also be able to access networks and services that were otherwise inaccessible. Obviously, such an attack, whereby a stolen or compromised device enables access to sensitive services, applies to laptops just as well; however, it is easier to misplace (or steal?) a small phone or a pair of glasses than a laptop.

Research Challenge Ideally, a device containing personal information would simply be invulnerable to software and hardware analysis without the owner's consent. Smart phone theft is so prevalent that governments have asked manufacturers to specifically address this issue [11]. Many carriers will not activate phones reported stolen. However, data on the devices are still vulnerable to simply being copied and exploited. Techniques such as full-disk encryption are relatively common, and yet surprisingly easy to subvert [21]. A key research challenge is to design OS and lower techniques that (1) provide strong user authentication (to identify valid users), and (2) render the device entirely useless when in a non-authenticated mode, without adversely affecting performance or usability.

III. Secure Encounters

Next, we briefly sketch our own work to address some of the challenges outlined in this paper. SDDR [23, 15] provides a protocol for device discovery and recognition via short range radio. SDDR establishes a *secure encounter* with every device in radio range.

An encounter establishes a secret key that is shared among a pair of devices. By default, no linkable information is revealed during an encounter: devices that meet repeatedly are unable to recognize each other. However, SDDR allows users to explicitly pair their devices, thus enabling them to identify each other in subsequent encounters, while remaining unlinkable to other devices. Finally, SDDR allows users to revoke such linkability selectively and unilaterally. SDDR is power-efficient and can run continuously, forming encounters with all devices in Bluetooth range.

Secure encounters prevent tracking of mobile devices, while enabling recognition among consenting users. Peers can communicate securely, during and after an encounter, via untrusted networks or Cloud services, and even if they haven't exchanged any linkable information. An attacker who obtains the database of encounters on Alice's device learns no useful information about Alice's encounters, unless he also has access to devices of users that Alice has previously met.

Research Challenges

We believe secure encounters provide a powerful primitive for privacy-preserving communication among personal devices, but many challenges remain. For e.g., not all encounters among devices in radio range are meaningful, and users must be able to identify those encounters relevant to a social situation. Contextual information associated with encounters can provide useful indicators for identifying relevant encounters. These include (but are not limited to) information such as location, time, duration, physical proximity, audio and visual inputs, user annotations, selective linkability, etc. EnCore [15] provides a communication platform that integrates secure encounters with the associated contextual information and enables users to create and communicate within groups of socially meaningful encounters called *events*. However, identifying relevant encounters in a larger and denser environment with many unlinkable devices is still an open challenge. Conversely, we would also like to enable secure communication and selective linkability among devices that participate in an event but are not within radio range. Additionally, a user's database of encounters must be carefully protected from attackers with the ability to combine different users' databases.

IV. The Transparent Citizen

Deployment of mobile devices and sensors coupled with pervasive data collection and analysis is threatening to bring the age of the "transparent citizen". It is increasingly difficult to function in modern Western societies without interacting with a mobile device or the Internet; however, as we have catalogued, users have little control over their personal data: how it is stored, how it accessed, and how it is used. Devices capture location traces in conjunction with search histories, social interactions both online and device-to-device. Taken together, this data divulges every salient aspect of one's life, including location, eating, shopping, sexual habits, ailments, likes, dislikes, and perhaps information they themselves don't know to be important yet.

The utility and capabilities of mobile devices cannot be overstated. They have revolutionized how information is accessed, and enabled functionality that was otherwise simply impossible. However, in deploying these services and applications, providers and developers have not had much incentive to restrain their appetite for users' personal information. Indeed, incentives point towards gathering, storing and monetizing as much as possible, e.g. to personalize services and advertisement.

Unfortunately, in the process of creating this always connected and data-driven society, technological developments are threatening to overrun citizens' constitutional rights. In many countries, citizens enjoy a constitutional right to withhold information that would incriminate them. Whether this right covers information that was digitally recorded by the user's mobile devices is uncertain. Without this protection, however, the right against self-incrimination is diminished in proportion to the extent of information recorded by such devices. Also, many countries grant special protection to the privacy of their citizens' home, requiring a specific and reasonable suspicion for a search warrant to be issued. How this right extends to a citizen's mobile devices and the information recorded by them is unclear.

Until now, users have had little recourse: neither laws nor deployed technological solutions help strengthen, or even maintain, user privacy. While some technically feasible solutions do exist, the lack of financial incentive has held back the deployment of privacy-preserving alternatives. Data sharing policies were often changed without user input [7, 4], or were not made public at all [2, 14]. Positive changes towards protecting user data were motivated by poor public relations [12], and even these changes never provide provable guarantees to the user. In light of recent revelations about pervasive data gathering and analysis, data privacy has perhaps finally entered the general consciousness. More than ever, users now explicitly want to understand what data is being collected and analyzed. Going forward, we hope a better understanding and broader education about the leakage vectors and risks associated with mobile devices, the data being collected, and how it can be used, will lead to a broader conversation, and ultimately reform, in both the technical and legal sectors.

References

[1] AllJoyn: The Fast Track to the Internet of Everything. https://www.alljoyn.org/.

[2] Apple Q/A on Location Data. http://www.apple.com/pr/library/2011/04/27Apple-Q-A-on-Location-Data.html.

[3] Dear Fitbit Users, Kudos On the 30 Minutes of "Vigorous Sexual Activity" Last Night. http://gizmodo.com/5817784/dear-fitbit-users-kudos-on-the-30-minutes-of-vigorous-sexual-activity-last-night.

[4] Facebook Changes Privacy Settings for Millions of Users - Facial Recognition is Enabled. http://nakedsecurity.sophos.com/2011/06/07/facebook-privacy-settings-facial-recognition-enabled/.

[5] Friday. http://www.fridayed.com/.

[6] Google Glass has a 30-Minute Battery Life While Shooting Video. http://www.digitaltrends.com/mobile/google-glass-30-minute-videobattery/.

[7] Google Privacy Policies Get Major Revamp. http://www.huffingtonpost.com/2012/01/24/google-privacy-policies_n_1229470.html.

[8] Haggle. http://code.google.com/p/haggle/.

[9] Precautions to Take When Using Public WiFi Service. www.att.com/esupport/article.jsp?sid=KB112237.

[10] Thousands of Amazon on S3 Data Stores Left Unsecured Due to Misconfiguration. http://www.engadget.com/2013/03/27/thousands-of-amazon-s3-data-stores-left-unsecured/.

[11] US Prosecutors Want 'Kill Switch' to Stop Smartphone Theft. http://spectrum.ieee.org/tech-talk/consumer-electronics/gadgets/us-prosecutors-want-kill-switch-stop-smartphone-theft.

[12] User Backlash Forces Facebook Privacy Tweaks, Again. http://www.fastcompany.com/1484919/user-backlash-forces-facebook-privacy-tweaks-again.

[13] What They Know - Mobile. http://blogs.wsj.com/wtk-mobile/.

[14] ACLU. Cell Phone Company Data Rentention Chart. http://www.aclu.org/cell-phone-location-tracking-request-response-cell-phone-company-data-retention-chart.

[15] P. Aditya, *et al.* EnCore: Private, Context-based Communication for Mobile Social Apps. In *MobiSys*, 2014.

[16] ARM. ARM Security Technology. http://infocenter.arm.com/help/topic/com.arm.doc.prd29-genc-009492c/PRD29-GENC-009492C_trustzone_security_whitepaper.pdf, 2009.

[17] V. Brik, S. Banerjee, M. Gruteser, and S. Oh. Wireless Device Identification with Radiometric Signatures. In *MOBICOM*, 2008.

[18] J. Camenisch and A. Lysyanskaya. Signature Schemes and Anonymous Credentials from Bilinear Maps. In *Advances in Cryptology - CRYPTO 2004*, 2004.

[19] A. P. Felt, E. Chin, S. Hanna, D. Song, and D. Wagner. Android Permissions Demystified. In *CCS*, 2011.

[20] V. Haldar, D. Chandra, and M. Franz. Semantic remote attestation: A virtual machine directed approach to trusted computing. In *Proceedings of the 3rd USENIX Virtual Machine Research And Technology Symposium*, 2004.

[21] J. A. Halderman, *et al.* Lest we Remember: Cold-boot Attacks on Encryption Keys. *Communications of the ACM*, 52(5):91–98, 2009.

[22] Intel Corp. Software Guard Extension Programming Reference. http://software.intel.com/sites/default/files/329298-001.pdf, 2012.

[23] M. Lentz, *et al.* SDDR: Light-Weight, Secure Mobile Encounters. In *USENIX Security*, 2014.