Elisa Bertino
Purdue University

Privacy in the Era of 5G, IoT, Big Data, and Machine Learning

We have today a number of technologies that, when combined, can support unprecedented applications and significantly enhance existing applications. These technologies include 5G cellular networks, big data, machine learning, and the Internet of Things (IoT). The combination of those technologies allows us to: a) increase our capacity for pervasive, fine-grained, and continuous data gathering and for the effective and efficient processing of these data (even with real-time guarantees); b) generate knowledge from data and continuously evolve this knowledge, thus supporting recommendation systems and decision processes—also accompanied by suitable explanations; and c) make devices, control systems, and cyberphysical systems intelligent and autonomous.

The increased adoption of wearable devices and continuous data streaming from these devices allows a party to collect fine-grained geo-temporal data about individuals.

However, many such technologies collect and/or use data, which often contain privacy-sensitive data. Collected data, even if anonymized by removing identifiers such as names or Social Security numbers, when linked with other data may lead to reidentifying the individuals to which specific data items are related. Also, as organizations, such as governmental agencies, often need to collaborate, they exchange datasets, resulting in these datasets being available to many different parties. Privacy breaches also occur at many different layers (for example, networks, hosts, and applications) and components in our interconnected systems. An example of a privacy attack in the context of a cellular network is the ToRPEDO side-channel attack,[1] which exploits the paging protocol to track users.

On the other hand, security techniques implemented by applications, especially the mobile ones, often have vulnerabilities, which undermine privacy. Notable examples are vulnerabilities in authentication protocols, such as in conventional login-password-based authentication and in SMS-based one-time passwords, or the use of covert channels and side channels to bypass the permission systems of the underlying operating systems.[2] It is important to emphasize that security and privacy are two different requirements. However, security is a prerequisite for privacy. The use of machine learning techniques further threatens privacy because of attacks such as the inversion ones by which a party can infer the sensitive contents of the data samples used for training. Finally, the increased adoption of wearable devices and continuous data streaming from these devices allows a party to collect fine-grained geo-temporal data about individuals.

Given the many ways by which data privacy can be breached, one may wonder whether the battle for privacy is lost or whether something can be done. In this respect, it is important to notice that research and industry have proposed many privacy-preserving techniques over the last 20 years, ranging from cryptographic techniques, such as oblivious data structures, which hide data access patterns, and homomorphic encryption to

continued on p. 91

Digital Object Identifier 10.1109/MSEC.2022.3221171
Date of current version: 19 January 2023

 Copublished by the IEEE Computer and Reliability Societies

data anonymization techniques, which transform data to make it more difficult to link specific data records to specific individuals or perturb the data. The problem of location privacy has also been the focus of extensive research both in the past and recently. Research efforts have also been devoted to investigating privacy-preserving techniques for data in the cloud, on smartphones, and in social networks. Finally, trusted environments have been developed that represent an important building block for privacy-preserving techniques.

However, despite the availability of many privacy-preserving techniques, we are still far from satisfactory solutions to privacy. The first reason is that privacy depends very much on user personal preferences, contexts, and culture. Therefore, we need privacy-preserving techniques that can be personalized. The other reason is that several privacy regulations have been defined over the years, such as the Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR). Technical solutions need to comply with these regulations—and in some cases, coming up with approaches is challenging, and also, these approaches need to be complemented by proper organizational processes.

Finally, it is important to notice that different privacy techniques must be used depending on the tasks to be executed on the data, such as record linkage, data analytic, and operational tasks, and on the specific transactions a user is executing, such as browsing the web, getting recommendations on movies, and buying products online. In the latter context, techniques proposed for privacy-preserving digital and for privacy-preserving e-commerce transactions are critical. Those techniques combined with network anonymizers and application-level privacy techniques are critical building blocks for holistic online privacy.

How can we make it possible for people to make decisions about "personal privacy versus collective safety"?

Given that we have all those privacy-preserving technologies, what more do we need for privacy? What we need are holistic privacy-preserving environments able to combine privacy-preserving approaches with adaptation to different user contexts and tasks to achieve "privacy protection in depth." Users consider privacy important, but they often feel that privacy is complex to manage.

However, there is also the key question of "personal privacy versus collective safety," as ultimately, the choice of making available (some of) our personal data, and thus renouncing some privacy, to benefit society is a personal choice. Users must be able to make informed decisions. Therefore, two challenging questions need to be addressed: 1) How can we make it possible for people to make decisions about "personal privacy versus collective safety"? 2) How can we make it possible to reconcile those two seemingly opposing goals? We believe that data transparency and policy-based use of data are two key elements relevant to answering these questions. ■

References

1. S. R. Hussain, "Privacy attacks to the 4G and 5G cellular paging protocols using side channel information," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2019, pp. 1–15, doi: 10.14722/ndss.2019.23442.
2. J. Reardon et al., "50 ways to leak your data: An exploration of apps' circumvention of the android permissions system," in *Proc. 28th USENIX Secur. Symp.*, 2019, pp. 1–10.

Elisa Bertino is a professor with Purdue University, West Lafayette, IN 47907 USA. Contact her at: bertino@purdue.edu.