2014

# Security with Privacy - Opportunities and Challenges

Elisa Bertino
*Purdue University*, bertino@cs.purdue.edu



Bertino, Elisa, "Security with Privacy - Opportunities and Challenges" (2014). *Cyber Center Publications.* Paper 626.
http://dx.doi.org/10.1109/COMPSAC.2014.98

# Security with Privacy - Opportunities and Challenges

## Panel Position Paper

Elisa Bertino
Cyber Center, CERIAS and CS Department
PurdueUniversity
West Lafayette, Indiana (USA)
bertino@cs.purdue.edu

***Abstract*— this paper summarizes opportunities and challenges concerning how we can achieve security while still ensuring privacy. It identifies research directions and includes a number of questions that have been debated by the panel.**

***Keywords—cyber security; information privacy; privacy-preserving data analytics***

## I. MOTIVATIONS

Technological advances and novel applications, such as sensors, cyber-physical systems, smart mobile devices, cloud systems, data analytics, and social networks, are making possible to capture, and to quickly process and analyze huge amounts of data from which to extract information critical for security-related tasks. In the area of cyber security, such tasks include user authentication, access control, anomaly detection, user monitoring, and protection from insider threat [1]. By analyzing and integrating data collected on the Internet and Web one can identify connections and relationships among individuals that may in turn help with homeland protection. By collecting and mining data concerning user travels and disease outbreaks one can predict disease spreading across geographical areas. And those are just a few examples; there are certainly many other domains where data technologies can play a major role in enhancing security.

The use of data for security tasks is however raising major privacy concerns. Collected data, even if anonymized by removing identifiers such as names or social security numbers, when linked with other data may lead to re-identify the individuals to which specific data items are related to. Also, as organizations, such as governmental agencies, often need to collaborate on security tasks, data sets are exchanged across different organizations, resulting in these data sets being available to many different parties. Apart from the use of data for analytics, security tasks such as authentication and access control may require detailed information about users. An example is multi-factor authentication that may require, in addition to a password or a certificate, user biometrics. Recently proposed continuous authentication techniques extend user authentication to include information such as user keystroke dynamics to constantly verify the user identity. Another example is location-based access control [2] that requires users to provide to the access control system information about their current location. As a result, detailed user mobility information may be collected over time by the access control system. This information if misused or stolen can lead to privacy breaches.

It would then seem that in order to achieve privacy we must give up privacy. However this may not be necessarily the case. Recent advances in cryptography are making possible to work on encrypted data – for example for performing analytics on encrypted data [3]. However much more needs to be done as the specific data privacy techniques to use heavily depend on the specific use of data and the security tasks at hand. Also current techniques are not still able to meet the efficiency requirement for use with big data sets.

The goal of this panel is to debate whether security and privacy can be reconciled and if so to identify methods and techniques to make this reconciliation possible.

In what follows, we first discuss a few examples of approaches that help with reconciling security with privacy. We then summarize questions addressed by the panel.

## II. EXAMPLES OF PRIVACY ENHANCING TECHINQUES

Many privacy enhancing techniques have been proposed over the last fifteen years, ranging from cryptographic techniques such as oblivious data structures [4] that hide data access patterns to data anonymization techniques that transform the data to make more difficult to link specific data records to specific individuals [5]; and we refer the reader for further references to specialized conferences, such as the Privacy-Enhancing Symposium (PET)[1] series, and journals, such as Transactions on Data Privacy[2].

In what follows, we focus on a few examples which focus on efficiently reconciling security with privacy.

- Privacy-preserving data matching: Record matching is typical performed across different data sources with the aim of identifying common information shared among these sources. An example is matching a list of passengers on a flight with a list of suspicious individuals. However matching records from different data sources is often in contrast with privacy requirements concerning the data owned by the sources. Cryptographic approaches, like

[1] https://petsymposium.org/2014/
[2] http://www.tdp.cat/

0730-3157/14 $31.00 © 2014 IEEE
DOI 10.1109/COMPSAC.2014.98

secure set intersection protocols, may alleviate such concerns. However, these techniques do not scale for large data sets. Recent approaches based on data transformation and mapping into vector spaces [6], and combination of secure multiparty computation (SMC) and differential privacy [7] have addressed scalability. However, work needs to be done concerning the development of privacy-preserving techniques suitable for complex matching techniques, based for example on semantic matching. Security models and definitions also need to be developed supporting security analysis and proofs for solutions combining different security techniques, such as SMC and differential privacy.

- Privacy-preserving collaborative data mining: Conventional data mining is typically performed on big centralized data warehouses collecting all the data of interest. However, centrally collecting all data poses several privacy and confidentiality concerns when data belongs to different organizations. An approach to address such concerns is based on distributed collaborative approaches by which the organizations retain their own data sets and cooperate to learn the global data mining results without revealing the data in their own individual data sets. Fundamental work in this area includes: (i) techniques allowing two parties to build a decision tree without learning anything about each other data sets except for what can be learned by the final decision tree [8]; (ii) specialized collaborative privacy-preserving techniques for association rules, clustering, k-nearest neighbor classification [9]. These techniques are however still very inefficient. Novel approaches based on cloud computing and new cryptographic primitives should be investigated.
- Privacy-preserving biometrics authentication: Conventional approaches to biometrics authentication require recording biometrics templates of enrolled users and then using these templates for matching with the templates provided by users at authentication time. Templates of user biometrics represent sensitive information that needs to be strongly protected. In distributed environments in which users have to interact with many different service providers the protection of biometric templates becomes even more complex. A recent approach addresses such issue by using a combination of perceptual hashing techniques, classification techniques, and zero-knowledge proof of knowledge (ZKPK) protocols [10]. Under such approach, the biometric template of a user is processed to extract from it a string of bits which is then further processed by classification and some other transformation. The resulting bit string is then used, together with a random number, to generate a cryptographic commitment. This commitment represents an identification token that does not reveal anything about the original input biometrics. The commitment is then used in the ZKPK protocol to authenticate the user. This approach has been engineered for secure use on mobile phones. Much work remains however to be done in order to reduce the false rejection rates. Also different approaches to authentication need to be investigated based on recent homomorphic encryption techniques.

## III. Panel Questions

The panel debated several aspects related to security with privacy focusing not only on research perspectives but also on regulatory and organizational perspectives. Questions asked to the panelists include:

- Are there additional domains for which security with privacy is critical?
- Which research advances are needed to make it possible to reconcile security with privacy?
- Which are policy issues related to the use of data for security, in addition the well-known privacy policies?
- How can academia, industry and governments engage in projects and initiatives focusing on the use of big data for security and its privacy implications?
- Are there national and international initiatives that we should engage with?

## Acknowledgments

The work reported here has been partially supported by the Purdue Cyber Center (Discovery Park), by NSF under awards CNS-1111512 and CNS-1016722, and by NIST under the project "Advancing Commercial Participation in the NSTIC Ecosystem".

## References

[1] E. Bertino, Data Protection from Insider Threats. Morgan&Claypool, 2012.

[2] M. Damiani, E. Bertino, B. Catania, P. Perlasca, "GEO-RBAC: A Spatially Aware RBAC", ACM Transactions on Information and System Security 10(1), 2007.

[3] D. Liu, E. Bertino, X. Yi, "Privacy of Outsourced K-Means Clustering", Proceedings of the 9th ACM Symposium on Information, Computer and Communication Security, Kyoto (Japan), June 4-6, 2014.

[4] H. X. Wang, K. Nayak, C. Liu, E. Shi, E. Stefanov, Y. Huang, "Oblivious Data Structures", IACR Cryptology ePrint Archive 2014: 185.

[5] J.-W. Byun, A. Kamra, E. Bertino, N. Li, "Efficiently k-Anonymization Using Clustering Techniques", Proceedings of 12th International Conference on Database Systems for Advanced Applications, DASFAA 2007, Bangkok, Thailand, April 9-12, 2007. LNCS, Springer.

[6] M. Scannapieco, I. Figotin, E. Bertino, A. Elmagarmid, "Privacy Preserving Schema and Data Matching", Proceedings of 2007 ACM SIGMOD International Conference on Management of Data.

[7] M. Kuzu et al. "Efficient Privacy-aware Record Integration", Proceedings of Joint 2013 EDBT/ICDT Conferences, EDBT'13, Genoa, Italy, March 18-22, 2013, ACM.

[8] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining", in Advances in Cryptology, Springer-Verlag, Aug. 20-24 2000.

[9] J. Vaidya, Y. Zhu, C. Clifton, "Privacy Preserving Data Mining", Advances in Information Security 19, Springer 2006, pp.1-121.

[10] H. Gunasinghe, E. Bertino, "Privacy Preserving Biometrics-Based and User Centric Authentication Protocol for Mobile Devices, submitted for publication, 2014.