Published in E. Simion and R. Géraud-Stewart, Eds., *Innovative Security Solutions for Information Technology and Communications (SecITC 2019)*, vol 12001 of *Lecture Notes in Computer Science*, pp. xviii–xx, 2020.

Privacy... Please!
(Extended abstract)

Fari Assaderaghi[1] and Marc Joye[2]

[1] NXP Semiconductors, San Jose, USA
fari.assaderaghi@nxp.com
[2] OneSpan, Brussels, Belgium
marc.joye@onespan.com

Abstract. The Internet-of-Things does not only refer to a wide variety of inter-connected devices but also to the data generated by these devices. This large amount of data is an opportunity but is also a threat: for example, information collected about the physical health or behavior of the consumer can be very detailed and poses a real privacy risk. This paper discusses privacy-preserving approaches which might play a differentiating role in the success and deployment of IoT solutions.

With the growing Internet-of-Things and its billions of connected devices, one of the main challenges the industry is facing is how to make sense of the enormous amount of data generated by the IoT devices. This is where machine learning techniques come into play. The basic premise of learning from data is to uncover a process from a set of observations. In that sense, machine learning is different from traditional statistics. Although applying traditional statistical methods is very efficient at extracting information from a huge amount of information it needs a built-in model. Machine learning, on the other hand, can dynamically adapt to a certain task given the data and the desired goal. Hence, it learns the important impact factors of the model from the data itself. Machine learning enables the development of a multitude of new applications: regression, classification, recommender systems, clustering, personal assistants, monitoring systems, and more [1,6].

The EU General Data Protection Regulation (GDPR) [9] that took effect in all EU countries in May 2018, aims at giving users control over their data. Companies need to comply to a set of rules, including the requirements of (i) obtaining the clear consent of users for processing their personal data; (ii) offering means to users for accessing, rectifying and erasing their personal data. Likewise, in the US, California has passed the California Consumer Privacy Act (CCPA) [8] that will take effect in January 2020. It grants users the right to know what personal information a business has collected and with whom it is shared. It also provides more control by granting users the right to opt-out to have their personal data sold or made available to third parties.

The combination of increasing public awareness of privacy threats and the ongoing implementation of compliance rules are creating momentum in the development of privacy technologies. This is the right time for IoT companies to properly address privacy issues in the design of their products and solutions. Two different approaches are available: differential privacy and data encryption.

Differential privacy As famously exemplified by the Netflix competition [10], it is well known that anonymizing a dataset is insufficient to conceal the users' identity. Differential privacy [3] is a technique that guarantees that the distribution of the system's output is insensitive to any individual's record, preventing the inference of any single user's data from the output. But this comes at a price. Differential privacy works by incorporating noise to the data. More noise injected in the data implies better privacy guarantees but also less precision in the system's output. Differential privacy is therefore essentially a trade-off between privacy and accuracy.

Working over encrypted data Data encryption is an alternative way to enable privacy. However, one limitation and fundamental property of traditional encryption schemes is that data first needs to be decrypted prior to being processed. The privacy control therefore lies in the hands of the recipient of the encrypted data. A fundamentally different approach is to rely on (fully) homomorphic encryption [5]. This allows the recipient to directly operate over encrypted data.

Other useful cryptographic tools to work on encrypted data include functional encryption [2], garbled circuits [7] and secure multi-party computation techniques [4].

We note that most known practical implementations for machine learning over encrypted data require two non-colluding entities (this is known as the two-server model). It is also important to stress to that, although significant progresses have been made, working over encrypted data remains a topic of intense development in the research community. Known techniques in general involve heavy computing resources and do not offer a one-solution-fits-all breakthrough solution. Only certain use-cases can be shown to be practical. The current situation can be compared to the 1980's, when at the start of the era of public-key cryptography the algorithms were also too slow for general purposes. New advances made public-key cryptography one of the foundational building blocks in modern computer security and the same is expected for these privacy-preserving techniques.

References

1. Abu-Mostafa, Y.S., Magdon-Ismail, M., Lin, H.T.: Learning From Data: A Short Course. AMLbook.com (2012), http://amlbook.com
2. Boneh, D., Sahai, A., Waters, B.: Functional encryption: Definitions and challenges. In: Ishai, Y. (ed.) Theory of Cryptography (TCC 2011). Lecture Notes in Computer Science, vol. 6597, pp. 253–273. Springer (2011). https://doi.org/10.1007/978-3-642-19571-6_16

3. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Halevi, S., Rabin, T. (eds.) Theory of Cryptography (TCC 2006). Lecture Notes in Computer Science, vol. 3876, pp. 265–284. Springer (2006). https://doi.org/10.1007/11681878_14
4. Evans, D., Kolesnikov, V., Rosulek, M.: A Pragmatic Introduction to Secure Multi-Party Computation. Now Publishers (2019). https://doi.org/10.1561/3300000019
5. Gentry, C.: Fully homomorphic encryption using ideal lattices. In: Mitzenmacher, M. (ed.) 41st Annual ACM Symposium on Theory of Computing (STOC). pp. 169–178. ACM (2009). https://doi.org/10.1145/1536414.1536440
6. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics, Springer, 2nd edn. (2009). https://doi.org/10.1007/978-0-387-84858-7
7. Yao, A.C.C.: How to generate and exchange secrets. In: 27th Annual Symposium on Foundations of Computer Science (FOCS). pp. 162–167. IEEE (1986). https://doi.org/10.1109/SFCS.1986.25
8. The California consumer privacy act of 2018. `https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375`
9. The EU general data protection regulation (GDPR). `https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN`
10. The Netflix prize, `https://www.netflixprize.com`