Data Security and Privacy on Intelligent Environments

Leandro Marin
Area of Applied Mathemetics (DITEC)
Faculty of Computer Science
University of Murcia
leandro@um.es

Abstract

In this paper we present some of the problems related with the data protection and privacy on intelligent environments and some of the solutions that have been considered.

1 Introduction

Information has always been an object of desire for malicious attacker. This is not something new. We can see multiple examples during the history, but nowadays, intelligent environments are making the problem much more dangerous.

We can divide the information management in three different problems: storage, manipulation and transport.

In the past, it could be enough to have the information hidden in a safe place and to reduce the number of people allowed to manipulate the information to zero or almost zero. The communication process was based on cryptographic protocols applied by the people allowed to manipulate the information under very restrictive circumstances.

On a first etage, computers replaced humans in these tasks. This was not a big problem if the computers used are safe. In order to make a computer safe, the first idea is to reduce to the minimum the interactions of this computer with other computers or humans. If the computer is manipulated only by trusted users and we use strong cryptographic algorithms to transfer the information between these safe computers, the security level could be consider high.

The reduction of the number of trusted users and strong restrictions on the software and hardware used for the system can increase in the security level, but the price paid is the reduction of the usability and nowadays, the technology is moving in the opposite direction.

Nowadays, the computers have been replaced by much more complex intelligent environments, in which it is almost impossible to have a control over the software running in our system or the devices connected to our network. It is not realistic to consider isolated environments or even ones in which only trusted users can manipulate the data. The great challenge is to balance security and usability.

2 Multiple faces, only one problem

We are going to consider some different cases that, at the end, can be considered only one single problem:

2.1 Protection of Public Databases

Consider a database with registers linked to persons and some kind of medical information. It is clear that the names and the ID-numbers are identifiers that should be protected, but other information like postal codes or illness could be interesting to obtain legitimate statistical information, for example, an illness related with pollution levels in certain areas. The problem is that this information can reveal the actual identity of the patient.

2.2 Multiparty Computation

Consider the case of a group of two or more users that want to make some kind of computation based on common data, but without revealing to the others their own information. The typical example is the Yao's Millionaire Problem, in which two millionaire

that want to know who is the richest without revealing to the other the amount of money that they have.

2.3 White Box Cryptography

Consider a program that should be able to make a computation using some kind of secret information. The result can be given, but the secret information should not be revealed. During the computation, the attacker has access to all memory positions. A typical example is the encryption of some kind of information without revealing the keys.

3 Protection techniques

There are several techniques that can be used to protect the information:

3.1 Encryption

This is probably the first idea that we can consider to protect the information. This is a good solution when the computation can be made in safe environments, but it cannot be applied to all cases. For example, the information in databases that should be used by learning algorithms or partially modified by legitimate users. Multiparty computation and white box environments are also a problem for standard encryptions.

3.2 Anonymized Data

It is necessary to make the information useless to the attacker, specially in the case of active attackers that can introduce records or information linking data, in order to make visible any transformation made in the data. When the data is given in categories, it is possible to use a bijection as a method of anonymization.

3.3 Statistical Protection

One of the ideas to protect information is to introduce noise in order to make records indistinguishable. Many of the definitions of privacy are given in terms of probability, so randomness seems to be a good choice, but this randomness should be introduced carefully in order to avoid interferences with legitimate learning algorithms.

3.4 Algebraic Protection

Another fruitful method is to analyze the kind of legitimate operations that should be performed on the data to generate algebraic transformations compatible with the operations. For example, linear transformations are quite useful when the data should be added and multiplied by constants.

3.5 Multiple Implementations

In some cases, it is really impossible to generate safe implementations. For example in the case of Internet of Things. We can have a huge amount of computational devices with very limited resources and in that case, we can be sure that the security of some of our devices will be broken. If the information retrieved from one of these devices can be used to break the following ones, the damage can grow exponentially. So, it is better to have multiple implementations that make the information of one device, not usable to break other ones.

4 Conclusions

It is non longer reasonable to consider that information can be kept safe in isolated places. Multiple devices collect information all around us, even inside our own computers, cell phones or intelligent devices. Preserving security and privacy in these environments is a serious task and it is connecting different areas of research.

4.0.1 Acknowledgements

This research is partially financed by the project TIN2017-86885-R

References

[1] Alfredo Cuzzocrea. Big Data Provenance: State-Of-The-Art Analysis and Emerging Research Challenges. In Workshop Proceedings of the EDBT/ICDT 2016 Joint Conference (March 15, 2016, Bordeaux, France)

[2] Sjouke Mauw, Yunior Ramrez-Cruz, Rolando Trujillo-Rasua Anonymising social graphs in the presence of active attackers. Transactions on Data Privacy 11 (2018) 169198

[3] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. In Joe Kilian, editor, *Advances in Cryptology - CRYPTO 2001, 21st Annual International Cryptology Conference, Santa Barbara, California, USA, August 19-23, 2001, Proceedings*, volume 2139 of *Lecture Notes in Computer Science*, pages 1–18. Springer, 2001.

[4] Leandro Marin. White Box Implementations Using Non-Commutative Cryptography. Sensors. 19. 1122 (2019).