Sharing confidential information with privacy

Luca Melis

March 10, 2014

1 Motivations

The word *privacy* commonly refers to the ability of a person or a group to make information about themselves private. The availability of sensitive information about an individual often results in having power over that person and thereby generating serious concerns on potential misuse by governments, corporations or even other individuals. Indeed, many countries have laws and constitutions that in some way limit privacy.

In today's world, data is generally shared in a digital format and gathered by third parties. Thus the disclosure of private information is becoming a critical concern raising legal, monetary or even emotional, privacy issues.

The need for privacy-preserving sharing of confidential information happens in many realistic scenarios. A typical scenario may involve two parties: one that needs information from the other that is either motivated, or obliged, to share only the requested information. Consequently, there is a struggle between the need of information sharing and the need of privacy.

Let's consider the following examples:

- **Law enforcement** Two national law enforcement bodies (e.g., USAs FBI and UKs MI5) want to compare their respective databases of terrorist suspects. National privacy laws prevent them from revealing bulk data, however, by treaty, they are allowed to share information on suspects of common interest.

- **Genomics** Genome Sequencing (WGS) is a revolutionary technology that determines the complete genetic blueprint of any organism. Risks of a genome disclosure motivate the need for secure storage and testing of human genomes. In particular, there is a need for genetic testing by physicians and/or laboratories without full access to individuals' genomes. In an idealized future setting, genetic tests should only disclose the required minimum amount of information. At the same time, genomics and biotechnology companies often treat test details as trade secrets, since they represent valuable intellectual property.

- **Face recognition** Automatic recognition of human faces is becoming increasingly popular in civilian and law enforcement applications that require reliable recognition of

humans. A typical application scenario for privacy-preserving face recognition concerns a client who privately searches for a specific face image in the face image database of a server.

- **Internet security** Internet service providers (ISP) usually maintain a *black-list* of potential attackers. Each ISP is motivated to know if others ISPs have some blacklist members in common without revealing the whole list.

- **Multimedia File Similarity** Digital media, e.g., images, audio, video, are increasingly relevant in todays computing ecosystems. Consider two parties that wish to evaluate similarity of their media files, e.g., for plagiarism detection: sensitivity of possibly unreleased material (or copyright issues) may prevent parties from revealing actual content.

The aforementioned examples motivate the need for privacy-preserving sharing of confidential information and pose two main technical challenges:

1. how to enable this type of sharing so that parties learn no information beyond what they are entitled to

2. how to do so efficiently, in real-world practical terms.

2 Privacy-preserving cryptographic protocols

Motivated by the privacy challenges introduced by modern technologies, recently a lot of research activities has been focused in the field of *Privacy-Enhancing Technologies* (PETs). Modern cryptography has contributed to PETs producing several cryptographic protocols for privacy protection. A number of cryptographic protocols can be defined as the secure and privacy-preserving implementation of a public functionality f.

In the famous paradigm referred as *Secure Multi-party Computation* (MPC) [1], two parties, willing to compute a function f over their input, will only learn the output of f and nothing else besides what can be inferred from the output. The parties achieve this without the help of a trusted third party. We expect MPC protocols to run faster and faster in the future, thus allowing to use MPC in more and more scenarios where mutually distrusting parties desire to perform some cooperative computation over their inputs [2].

My research objective will be focusing on the secure computation of specific functionalities, using specialized protocols rather than generic solutions. The motivation is twofold:

1. not all information sharing functionalities can be easily implemented using generic solutions

2. design optimized special-purpose protocols is often more efficient.

In many real applications, parties request sensitive information from other entities, e.g., to retrieve messages, files or database records.

The most prominent technique to address this problem is Oblivious Transfer (OT)[4]. OT permits a client to transfer one of potentially many messages to a server, remaining oblivious about the message transferred (if any).

Similar techniques are *Private Information Retrieval* (PIR) [3] and *Private Set Intersection* (PSI) [5].

In particular, PSI involves two parties, a server and a client, each with a private input set. PSI lets parties run a cryptographic protocol that only disclose to the client the set intersection, and nothing to the server (beyond client set size). In prior works, some variants of PSI have been introduced:

- *Private Set Intersection Cardinality* (PSI-CA) in which the client learns only the cardinality of the set intersection.
- *Authorized Private Set Intersection* (APSI) in which each item in client set must be authorized by some mutually trusted authority in order to ensure that client obtains duly authorized information.
- *Private Set Intersection with Data Transfer* (PSI-DT) in which the client also receives data records associated with each item in the intersection set.

In PSI, an interesting privacy property is that of hiding the size of the set held by one party from the other [6].

In [7, 8], the authors designed PSI protocols with linear computational complexities.

2.1 Open Problems

In recent years, there has been an increased interest in privacy-preserving protocols. Nonetheless there are some problems, both theoretical and practical, that are still open and that represents the main motivation of my PhD work.

2.1.1 Efficient Group Private Set Intersection

The traditional PSI formulation only includes two participants, server and client. However, it is not clear how to efficiently extend such techniques to scenarios where a group of n participants (with $n > 2$) wish to privately compute the intersection of their respective sets (without using a trusted or semi-trusted party). Multi-party PSI protocols with linear complexities still remains a challenging topic for further research.

2.1.2 Size-Hiding Private Set Intersection Secure in Malicious Model

One important factor on the security of cryptographic is the adversarial model which is either semi-honest or malicious. Protocols secure against *semi-honest adversaries* assume that

participants strictly follow the steps of the protocols but they might attempt to infer additional information about other party's input. On the other hand, security against *malicious adversaries* permits arbitrary deviations from the protocols.

In [6], the PSI with the *size-hiding* property, is provably secure under standard assumptions only against semi-honest adversaries. Future research is needed in order to design a size-hiding PSI secure against malicious adversaries.

2.1.3 Privacy in testing genomic information

In [9], the authors show how to enable privacy in genomic applications, e.g., paternity tests, genetic and personalized medicine testing. A further step in that direction is to implement specific cryptographic tools for genomic privacy protection.

3 Personal background

My background is mainly that of a Computer Engineer. I'm able to write programs using the most important programming languages such as $C/C++$, Python, Java and Matlab/Octave.

During my studies, I developed analytical skills and learnt how to deal with complex problems in a more systematic way. Some of the courses I followed turned out to be very relevant for the field of cryptography and information technology security. In particular, I took the courses of cryptography and number theory at the faculty of mathematics and I became familiar with the fundamentals of Cryptography and abstract mathematics. At the faculty of computer engineering I took some courses of applied cryptography such as Network Security and Security of Multimedia Contents.

In the practical realisation of my bachelor's degree thesis, I implemented a prominent cryptographic technique in the field of Multimedia contents security called *Buyer-Seller protocol* [10]. The results of my implementation has been mentioned in [11]. I believe that my thesis has contributed to the improvement of my skills in the practical realization of cryptographic protocols. These techniques also represent the main building blocks of the privacy-preserving protocols that constitute the subject of my PhD work.

During my stay at the University of Aarhus, I worked on my master's degree thesis about the cryptographic problem of *Learning Parity with Noise* (LPN) [12] under the supervision of Ivan Damgård and Claudio Orlandi. This thesis focused on theoretical aspects of cryptography. In this work I propose new cryptographic protocols derived from LPN addressing their correctness and security issues.

Finally, I would like to ask funding through departmental grants for giving me the opportunity to work closely and collaboratively with mentors and other graduate students on the long-term projects mentioned in this research proposal.

References

[1] A. Yao, Protocols for secure computations, In FOCS, pages 160164, 1982.

[2] Orlandi, Claudio, Is multiparty computation any good in practice?, Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, 2011, IEEE.

[3] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, Private information retrieval, Journal of ACM, 45(6):965981, 1998.

[4] M. Rabin, How to exchange secrets by oblivious transfer, TR-81, Harvard Aiken Computation Lab, 1981.

[5] Emiliano De Cristofaro and Paolo Gasti and Gene Tsudik, Fast and Private Computation of Cardinality of Set Intersection and Union, Cryptology ePrint Archive, Report 2011/141, 2011.

[6] Giuseppe Ateniese and Emiliano De Cristofaro and Gene Tsudik, (If) Size Matters: Size-Hiding Private Set Intersection, Cryptology ePrint Archive, Report 2010/220, 2010.

[7] Emiliano De Cristofaro and Jihye Kim and Gene Tsudik, Linear-Complexity Private Set Intersection Protocols Secure in Malicious Model, Cryptology ePrint Archive, Report 2010/469, 2010.

[8] Changyu Dong and Liqun Chen and Zikai Wen, When Private Set Intersection Meets Big Data: An Efficient and Scalable Protocol, Cryptology ePrint Archive, Report 2013/515, 2013.

[9] P. Baldi, R. Baronio, E. De Cristofaro, P. Gasti, and G. Tsudik. Countering GATTACA: Efficient and Secure Testing of Fully-Sequenced Human Genomes. In CCS, 2011.

[10] Deng, Mina and Bianchi, Tiziano and Piva, Alessandro and Preneel, Bart, An Efficient Buyer-seller Watermarking Protocol Based on Composite Signal Representation, Proceedings of the 11th ACM Workshop on Multimedia and Security, 2009.

[11] Rial, A. and Mina Deng and Bianchi, T. and Piva, A. and Preneel, B. A Provably Secure Anonymous Buyer; Seller Watermarking Protocol, Information Forensics and Security, IEEE Transactions on, 2010 .

[12] Krzysztof Pietrzak. Cryptography from learning parity with noise. In SOFSEM 2012: Theory and Practice of Computer Science, 2012.