Sharing confidential information while preserving privacy

Luca Melis

March 20, 2014

1 Motivations

In today's world, data is generally shared in a digital format and gathered by third parties. Thus the leak of private information is becoming a critical concern involving forensic and financial privacy disputes.

The need for privacy-preserving sharing of confidential information happens in many realistic situations. A typical scenario may involve two entities, one of which is either motivated, or obliged, to share only the information required from the other entity. It is clear that in this case there is a strong contradiction between the need of information sharing and the need of privacy.

Consider the following examples:

- **Law authorities** Two law enforcement agencies having different lists of terrorist suspects are motivated to make a comparison between their respective databases. Of course, they cannot unveil their entire databases without breaking national privacy laws. They only need to uncover a list of potential terrorists of mutual interest.

- **Genomics** Genome Sequencing is a process that determines the complete DNA sequence of an organism's genome at a single time. On the one hand the availability of DNA sequences enables efficient testing and specialized treatments based on genomics features, but on the other hand the possible disclosure of personal information raises serious privacy concerns.

- **Face recognition** Automatic recognition of human faces techniques are employed in a wide range of applications ranging from surveillance systems to social networks and thus, raising several privacy concerns. In such a scenario, a system for privacy-preserving face recognition may involve two parties: a *client* searching for a human face in the database of a *server*, and the *server* itself.

- **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** Copyright concerns prevent two parties, willing to compute the similarity of their multimedia files, to share the actual content of these files. Privacy-preserving techniques may represent a valid solution to overcome this problem.

The aforementioned examples motivate the need for techniques for sharing confidential information and at the same time preserving privacy. They also present two interesting questions:

1. how to allow privacy-preserving sharing of information, allowing parties to learn only the strictly required information,

2. how to do build efficient techniques for privacy-preserving sharing of information.

2 Privacy-preserving cryptographic protocols

Following the privacy concerns introduced by modern technologies, a lot of research activities has recently focused on the field of *Privacy-Enhancing Technologies* (PETs). Modern cryptography has contributed to PETs, producing several cryptographic protocols for privacy-preserving.

Retrieving messages from a server or querying a database records are examples of situations in which entities may seek confidential data from other entities.

A number of cryptographic protocols can be defined as the secure implementation of a public functionality f.

In the famous protocol 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. The parties achieve this without the help of a trusted third party.

An important primitive related to MPC is Oblivious Transfer (OT) [3]. OT allows a client to transfer one of potentially many pieces of information to a receiver, but the client itself remains oblivious as to what message (if any) has been transferred.

Related techniques are *Private Information Retrieval* (PIR) [2] and *Private Set Intersection* (PSI) [4].

In particular, PSI involves two parties, a server and a client, each with a private input set. At the end of a PSI protocol, the client will only learn the set intersection, but no information will be available to the server besides the client set size.

In PSI, an interesting privacy property is that of hiding the size of the set held by one party from the other one [5].

In [6, 7], the authors designed PSI protocols with linear computational complexities.

My research objective will be focusing on the secure computation of specific functionalities, using specialized protocols based on PSI techniques rather than generic solutions such as those derived from MPC. Two valid motivations are:

1. it is sometimes difficult to use generic solutions for implementing all the functionalities needed for sharing information,

2. ad-hoc protocols represent, in many cases, a more efficient way to do that.

2.1 Open Problems

In our highly digitally and networked world, privacy-preserving protocols are becoming increasingly important. However, there are some problems [8], either theoretical or practical, that are still open and that represent the main directions for my future researches.

2.1.1 Multi-party PSI

The usual PSI protocols allow the participation of only two entities. An interesting extension of PSI protocols is that of allowing a group of $n > 2$ parties to share the intersection of their respective sets. The main challenge for future is that of designing an efficient multi-party PSI with linear complexity without the need for a trusted authority.

My approach to reduce the complexity of a Multi-party PSI will be to find an efficient way to represent an input set. Choosing a good set representation may reduce the computation and communication overheads of the protocol [14].

2.1.2 Size-Hiding PSI Secure in Malicious Model

One important factor on the security of cryptographic protocols is the adversarial model which can be 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, *malicious adversaries* can deviate from the protocol without restrictions.

In [5], the PSI with the *size-hiding* property, is secure only against semi-honest adversaries. Future research is needed in order to design a size-hiding PSI secure against malicious adversaries.

A solution might be to ask the (malicious) sender to commit his private set and then run a *zero-knowledge* (or *quasi-knowledge* [13]) proof of knowledge of the committed set. In this case, the witness provided in the proof of knowledge must not reveal the size of the set.

2.1.3 Applications in genomic information testing with privacy

Paternity tests and personalized medicine testing are examples of genomic testing applications in which privacy issues are becoming increasingly relevant.

In [9], the authors show how to enable privacy in these applications, but the practical implementation of ad-hoc cryptographic frameworks for genomic privacy protection is still an open direction for future research.

In particular, my future researches will be focused in realizing techniques for certified forensic identification. In this scenario, the subject of investigation must be able to prove the authenticity of its input.

3 Personal background

My background is mainly that of a Computer Engineer. I'm able to write software 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 in the field of cryptography and information technology security.

In particular, I took 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, where I deal with their correctness and security issues.

Finally, I would definitely like to apply for a departmental grant so as to have 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] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval, Journal of ACM, 45(6):965981, 1998.

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

[4] E. De Cristofaro, P. Gasti, and G. Tsudik. Fast and Private Computation of Cardinality of Set Intersection and Union, Cryptology ePrint Archive, Report 2011/141, 2011.

[5] G. Ateniese, E. De Cristofaro, and G. Tsudik. (If) Size Matters: Size-Hiding Private Set Intersection, Cryptology ePrint Archive, Report 2010/220, 2010.

[6] E. De Cristofaro, J. Kim, and G. Tsudik. Linear-Complexity Private Set Intersection Protocols Secure in Malicious Model, Cryptology ePrint Archive, Report 2010/469, 2010.

[7] C. Dong, L. Chen, and Z. Wen. When Private Set Intersection Meets Big Data: An Efficient and Scalable Protocol, Cryptology ePrint Archive, Report 2013/515, 2013.

[8] E. De Cristofaro. Sharing Sensitive Information with Privacy, PhD dissertation, 2011

[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] M. Deng, T. Bianchi, A. Piva, and B. Preneel. An Efficient Buyer-seller Watermarking Protocol Based on Composite Signal Representation, Proceedings of the 11th ACM Workshop on Multimedia and Security, 2009.

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

[12] L. Melis, C. Orlandi, and I. Damgård. On The Learning Parity with Noise Problem, Master Thesis dissertation, Available online: `http://www.cs.au.dk/~orlandi/masterprojects/LucaMelis.pdf`, 2013.

[13] M. Chase, I. Visconti. Secure Database Commitments and Universal Arguments of Quasi Knowledge, Cryptology ePrint Archive, 2012.

[14] J. Hee Cheon, S. Jarecki, and J. Hong Seo. Multi-Party Privacy-Preserving Set Intersection with Quasi-Linear Complexity, Cryptology ePrint Archive, Report 2010/512, 2010.