Design of Privacy Preserving Secure Set Intersection Protocols

Thesis submitted to the
Indian Institute of Technology Kharagpur
for the award of the degree
of

Doctor of Philosophy

by

Sumit Kumar Debnath

Under the guidance of

Dr. Ratna Dutta

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY
KHARAGPUR

November **2017**

ABSTRACT

Nowadays, modern society has become dependent on the necessity of electronic information. There are numerous real life scenarios where electronic information is often shared among unreliable entities preserving the security and privacy issues. As a consequence, there is a strong need of privacy preserving cryptographic techniques to enable secret sharing of sensitive information. Among these, *Private Set Intersection* (PSI) and *Private Set Intersection Cardinality* (PSI-CA) have received considerable attention to the recent research community due to their importance and wide applications. PSI and PSI-CA are appealing when two entities seek to determine the intersection and cardinality of the intersection respectively of their associated private data sets without disclosing any additional information (beyond the set sizes) to each other. This thesis presents a variety of solutions in the context of secure set intersection.

There are several existing works on Mutual PSI (mPSI) and Mutual PSI-CA (mPSI-CA), where both the entities receive the output in contrast to one-way PSI and one-way PSI-CA which allow only one of the entities to learn the output. The major issues in designing mPSI and mPSI-CA are to attain complexity *linear* to the set sizes in terms of both communication and computation along with the security in the presence of malicious adversaries. In this thesis, we came up with solutions to these problems and provide two constructions for mPSI and two constructions for mPSI-CA achieving the above features. Each of these constructions preserve fairness ensuring that either both the entities or none of them receive the output on completion of the protocol. Fairness is achieved by using an off-line semi-trusted third party, called arbiter, who cannot get access to the private information of the entities while follows the protocol honestly. One of our proposed mPSI and mPSI-CA do not use any random oracles, thereby are secure in standard model. The other two constructions yield security in the random oracle model (ROM). The underlying group in all these constructions is of prime order except one mPSI that uses composite order group.

Besides, we design size-hiding PSI and PSI-CA employing Bloom filter. The term size-hiding implies that the size of the set held by the client never gets disclosed to the server. These PSI and PSI-CA are further extended to authorized PSI (APSI) and authorized PSI-CA (APSI-CA), where the client's set is made authorized by a trusted third party at the beginning of the protocol. Authorization of the client's set is required to prevent the client from arbitrary input manipulation. Similar to PSI and PSI-CA, our APSI and APSI-CA preserve size-hiding property. All these constructions achieve linear complexity. Our PSI and PSI-CA are proven to be secure in standard model against semi-honest adversaries. On the

other hand, APSI and APSI-CA achieve their security in fully malicious model and in the presence of malicious client respectively without using random oracles. In particular, our PSI-CA and APSI-CA are the *first* to achieve *size-hiding* property.

There is no construction of PSI-CA so far that attains *linear* complexity and security against *malicious client* in *standard* model. We address this issue by designing two PSI-CA protocols with the aforementioned properties. One of these PSI-CA uses Bloom filter and is significantly more efficient than the other one in terms of communication. These PSI-CA protocols are then converted to PSI constructions retaining linear computation and communication overheads, achieving security against malicious server as well as malicious client. The PSI protocol employing Bloom filter is proven to be secure in the ROM in contrast to the other PSI which is in standard model.

Finally, we focus on constructing PSI scheme with linear computation cost while attaining constant communication overhead. We design the first such construction by combining somewhere statistically binding (SSB) hash function with indistinguishability obfuscation (iO) and Bloom filter. The scheme achieves its security in standard model against semi-honest adversaries. Due to the use of Bloom filter, it works fast even for big data subject to the availability of efficient iO construction.

Keywords: private set intersection; fairness; Bloom filter; SSB hash; iO; zero-knowledge proofs.