A Scalable Multiparty Private Set Intersection

Atsuko Miyaji\textsuperscript{1,2,3(✉)} and Shohei Nishida\textsuperscript{1}

\textsuperscript{1} Japan Advanced Institute of Science and Technology, Nomi, Japan \{miyaji,shohei-n\}@jaist.ac.jp
\textsuperscript{2} Japan Science and Technology Agency (JST) CREST, Tokyo, Japan
\textsuperscript{3} Graduate School of Engineering, Osaka University, Suita, Japan

Abstract. Both scalability and flexibility become crucial for privacy preserving protocols in the age of Big Data. Private Set Intersection (PSI) is one of important privacy preserving protocols. Usually, PSI is executed by 2-parties, a client and a server, where both a client and a server compute jointly the intersection of their private sets and at the end only the client learns the intersection and the server learns nothing. From the scalable point of view, however, the number of parties are not limited to two. In this paper, we propose a scalable and flexible multiparty PSI (MPSI) for the first time: the data size of each party is independent to each other and the computational complexity is independent to the number of parties. We also propose \(d\)-and-over MPSI for the first time.

1 Introduction

Both scalability and flexibility become crucial for privacy preserving protocols in the age of Big Data. Private Set Intersection (PSI) is one of important privacy preserving protocols. PSI is executed by 2-parties, a client and a server, where both compute jointly the intersection of their private sets and, at the end, only the client learns the intersection and the server learns nothing. From the scalable point of view, however, the number of parties are not limited to two. This is why a multiparty PSI (MPSI) \cite{8,14} becomes important. However, both are far from scalability: the computational complexity depends on the number of parties, and the data size of each party is equal to each other in \cite{14} and \cite{8} computes only the approximate number of intersection.

In this paper, we propose a scalable and flexible MPSI: the data size of each party is independent to each other and the computational complexity is independent to the number of parties. Furthermore we also propose a new notion of \(d\)-and-over multiparty PSI (\(d\)-and-over MPSI) for \(d \leq n\). A \(d\)-and-over MPSI means to compute securely \(\bigcap_{j=1}^{\geq d} S_j = \bigcup_{n \geq \ell \geq d} (S_{j_1} \cap \cdots \cap S_{j_\ell})\), where \(S_i\) is a set of \(P_i\). Let us think the following scenario: There are \(n\) shops \(P_i\) in a shopping mall whose customers’ list is \(S_i\). Shops think to promote number of customers each other and plan to have a promotion campaign. In the promotion campaign, a

\begin{center}
This study is partly supported by Grant-in-Aid for Scientific Research (C) (15K00183) and (15K00189).
\end{center}

© Springer International Publishing Switzerland 2015
M. Qiu et al. (Eds.): NSS 2015, LNCS 9408, pp. 376–385, 2015.
DOI: 10.1007/978-3-319-25645-0_26
shop $P_i$ wants to know customers who joins an intersection of 3-and-over shops including $P_i$ without learning any information about customers that are not in the intersection. Such a scalable MPSI has, however, not proposed yet as far as authors know.

This paper is organized as follows. Section 2 summarises security assumption and building blocks used in our proposal. Section 3 explains the previous results. Then, after investigating set operations required in the case of $n$ parties in Section 4, we propose concrete schemes of MPSI and $d$-and-over MPSI in Section 5. Comparison with the previous MPSI [14] is shown in Section 6.

2 Preliminary

This section summarises security assumption and building blocks used in our proposal.

2.1 Security Assumption

We describe two standard adversary models [10]: semi-honest adversaries and malicious adversaries. In semi-honest adversaries model, all players act according to their prescribed actions in the protocol. If a protocol is secure in a semi-honest model, then no player gains information about other player’s private input sets, other than what can be deduced from the result of the protocol. On the other hand, in malicious adversaries model, an adversary player can behave arbitrarily. In particular, we cannot hope to prevent a malicious player from refusing to participate in the protocol, substituting an input with an arbitrary value, and aborting the protocol prematurely.

The security assumptions used in our protocol are defined as follows.

**Definition 1 (DDH Assumption).** Let $\mathbb{F}_p$ be a finite field, $g \in \mathbb{F}_p$ with prime order $q$ and size of $q$ is $\ell$. The DDH (Decisional Diffie-Hellman) problem is hard in $G$ if, for any efficient algorithm $A$, there exists $\epsilon > 0$ and the following probability is satisfied: $|\Pr[x, y \leftarrow \{0, 1\}^\ell : A(g, g^x, g^y, g^{xy}) = 1] - \Pr[x, y, z \leftarrow \{0, 1\}^\ell : A(g, g^x, g^y, g^z) = 1]| < \epsilon$.

2.2 Bloom Filter

A Bloom filter [2], denoted by $\text{BF}$, is a space-efficient probabilistic data structure, that is used to test whether an element $x$ is included in a set $S$. False positive matches are possible, but false negatives are not, thus a Bloom filter has a 100% recall rate. Elements can be added to the set, but not removed. A Bloom filter is an array of $m$ bits that can represent a set $S$ with at most $w$ elements. A Bloom filter uses a set of $k$ independent uniform hash functions $\mathcal{H} = \{H_0, ..., H_{k-1}\}$, where $H_i : \{0, 1\}^* \rightarrow \{0, 1, \cdots , m - 1\}(0 \leq \forall i \leq k - 1)$. Here after, we denote a Bloom filter parametrised $(m, k)$ by $\text{BF}_{m,k}(S)$ that encodes a set $S$. Let us explain how $\text{BF}$ is constructed, which is given by $\text{const.BF}$ (see Algorithm 1):