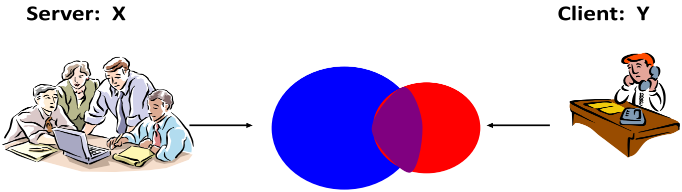
Secure and Efficient Private Set Intersection Cardinality using Bloom Filter

Sumit Kumar Debnath and Ratna Dutta

Department of Mathematics
Indian Institute of Technology Kharagpur
Kharagpur-721302, India

Outline

Private Set Intersection (PSI) Protocol



- At the end of the protocol, either one of them gets the intersection, yielding-one-way PSI, or both of them get the intersection yielding-mutual PSI (mPSI)

Private Set Intersection Cardinality(PSI-CA)

This is a variant of PSI, where the participants wish to learn the cardinality of the intersection rather than the content.

Private Set Intersection (PSI) Protocol

The applications of PSI and PSI-CA protocols are as follows:

- Two real estate companies would like to identify customers (e.g., home owners) who are double-dealing, i.e., have signed exclusive contracts with both companies to assist them in selling their properties.
- Two different health organizations want to know the number of common villagers who are suffering from a particular disease in a village. None of the organizations will reveal their list of suspects but they may learn the number of common suspects by running an PSI-CA.

Cryptographic Building Blocks

- Bloom Filter of [1]
- Homomorphic Encryption of [2]

[1]: B. H. Bloom, Communications of the ACM 1970.
[2]: S. Goldwasser and S. Micali, Journal of computer and system sciences, 1984

Bloom Filter (BF)

Bloom filter (BF) is a data structure that represents a set $X = \{x_1, ..., x_v\}$ of v elements by an array of m bits and uses k independent hash functions $H = \{h_0, h_1, ..., h_{k-1}\}$ with $h_i : \{0,1\}^* \rightarrow \{0, 1, ..., m-1\}$ for $i = 0, 1, ..., k-1$. Bloom filter of X is denoted by BF_X

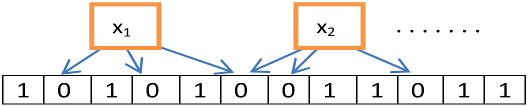
Bloom Filter (BF)

Choose $m = 12$ and $k = 3$.

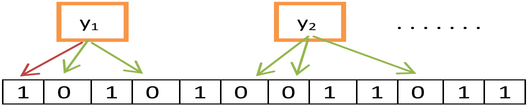
Initialization:

1	1	1	1	1	1	1	1	1	1	1	1

Add step: Suppose $(h_0(x_1) = 5, h_1(x_1) = 1, h_2(x_1) = 3)$, $(h_0(x_2) = 9, h_1(x_2) = 6, h_2(x_2) = 5)$......



Check step: Suppose $(h_0(y_1) = 0, h_1(y_1) = 3, h_2(y_1) = 1)$, $(h_0(y_2) = 9, h_1(y_2) = 6, h_2(y_2) = 5)$......



Goldwasser-Micali (GM) Encryption

This is a homomorphic encryption under the X-OR operation and consists the algorithms (**KGen**, **Enc**, **Dec**):

• $(pk = (n, u), sk = (P, Q)) \leftarrow \mathbf{KGen}(1^\kappa)$, where $n = PQ$ is an RSA modulus, $L(\frac{u}{P}) = -1$ and $L(\frac{u}{Q}) = -1$ but $J(\frac{u}{n}) = 1$.

• $c \leftarrow \mathbf{Enc}(m \in \{0, 1\}, pk)$, where

$$c = \mathrm{Enc}_{pk}(x) = \begin{cases} r^2 \bmod n \text{ if } m = 0 \\ ur^2 \bmod n \text{ if } m = 1 \end{cases}$$

• $m \leftarrow \mathbf{Dec}(c, sk = (P, Q))$, where $L(\frac{c}{P}) = 1$ implies the decryptor outputs the message m as 0 else, the decryptor outputs the message m as 1.

Quadratic Residuosity (QR) Assumption

Let X be the subgroup of $\mathbb{Z}_n^*$ of elements having Jacobi symbol equal to 1. The QR assumption states that, given an *RSA* modulus n (without its factorization), it is computationally infeasible to distinguish a random element u of $X \subseteq \mathbb{Z}_n^*$ from an element of the subgroup $\{x^2 | x \in \mathbb{Z}_n^*\}$ of quadratic residues modulo n for every PPT algorithm $\mathcal{A}$.

PSI-CA Protocol

C's private input		S's private input
$Y = \{c_1, c_2, ..., c_w\} \subseteq \{0,1\}^*$		$X = \{s_1, s_2, ..., s_v\} \subseteq \{0,1\}^*$
$(pk_C, sk_C) \leftarrow \mathbf{KGen}$		
constructs BF_Y		
For $j = 0, 1, ..., m-1$,		
computes $b_j = \mathsf{Enc}_{pk_C}(BF_Y[j])$		
$\overline{Y} = \{\mathsf{Enc}_{pk_C}(BF_Y[j])\}_{j=0}^{m-1}$		For $i = 1, 2, ..., v$,
	$\xrightarrow{\overline{Y}, pk_C}$	(i) for $j = 0, 1, ..., k-1$,
		(a) computes $h_j(s_i) \in \{0, 1, ..., m-1\}$,
		(b) extracts $b_{h_j(s_i)}$ from $\overline{Y}$,
		(ii) sets $E(\bar{s}_i) = \{b_{h_j(s_i)} \cdot r_{i,j}^2\}_{j=0}^{k-1}$,
Sets $card = 0$		where $r_{i,0}, ..., r_{i,k-1} \leftarrow \mathbb{Z}_n$.
For $i = 1, 2, ..., v$,		$\overline{X} = \{E(\bar{s}_i)\}_{i=1}^{v}$
(i) for $j = 0, 1, ..., k-1$,	$\xleftarrow{\overline{X}}$	
$\mathsf{Dec}_{sk_C}(\mathsf{Enc}_{pk_C}(\bar{s}_{i,j})) = \bar{s}_{i,j}$,		
(ii) if $\bar{s}_i$ is all-zero string		
then $card = card + 1$.		
Outputs $card$ as $\lvert X \cap Y \rvert$		

PSI-CA Protocol contd...

Correctness: $E(\bar{s}_i) = \{b_{h_0(s_i)} \cdot r_{i,0}^2 \mod n, ..., b_{h_{k-1}(s_i)} \cdot r_{i,k-1}^2 \mod n\}$
$= \{\mathrm{Enc}_{pk_C}(\mathrm{BF}_Y[h_0(s_i)]) \cdot \mathrm{Enc}_{pk_C}(0), ..., \mathrm{Enc}_{pk_C}(\mathrm{BF}_Y[h_{k-1}(s_i)]) \cdot \mathrm{Enc}_{pk_C}(0)\}$
$= \{\mathrm{Enc}_{pk_C}(\mathrm{BF}_Y[h_0(s_i)] \oplus 0), ..., \mathrm{Enc}_{pk_C}(\mathrm{BF}_Y[h_{k-1}(s_i)] \oplus 0)\}$
$= \{\mathrm{Enc}_{pk_C}(\mathrm{BF}_Y[h_0(s_i)]), ..., \mathrm{Enc}_{pk_C}(\mathrm{BF}_Y[h_{k-1}(s_i)])\}$
Therefore $\bar{s}_i = \{\mathrm{BF}_Y[h_0(s_i)], ..., \mathrm{BF}_Y[h_{k-1}(s_i)]\} \in \{0,1\}^k$ for all $i = 1, 2, ..., v$.
Now it can be easily shown that $\bar{s}_i \in \{0,1\}^k$ is a all-zero string if and only if $s_i \in X \cap Y$.

APSI-CA Protocol

Off-line Phase:

C	:	$(pk_C, sk_C) \leftarrow$ **KGen**
$C \longrightarrow$ CA	:	$Y = \{c_1, c_2, ..., c_w\}, pk_C$
CA	:	Generates $(pk_{DSig}, sk_{DSig}) \leftarrow \mathbf{KGen}.DSig$, constructs BF_Y, sets $b_i = \mathrm{Enc}_{pk_C}(\mathrm{BF}_Y[i])$ for each $i = 0, 1, ..., m-1$ and computes $\Omega = \{Sig(b_0), ..., Sig(b_{m-1})\}$ using sk_{DSig}
CA $\longrightarrow C$	:	$\overline{Y} = \{b_0, ..., b_{m-1}\}, \Omega, pk_{DSig}$
CA $\longrightarrow S$	:	pk_{DSig}

APSI-CA Protocol

Online Phase:

C's private input
$Y = \{c_1, c_2, \ldots, c_w\} \subseteq \{0,1\}^*$
$\overline{Y} = \{b_0, \ldots, b_{m-1}\}$,
$\Omega = \{Sig(b_0), \ldots, Sig(b_{m-1})\}$

$\xrightarrow[\Omega]{\overline{Y}, pk_C}$

S's private input
$X = \{s_1, s_2, \ldots, s_v\} \subseteq \{0,1\}^*$

Verifies Ω

If verification fails, then aborts
Otherwise, for $i = 1, 2, \ldots, v$,
(i) for $j = 0, 1, \ldots, k-1$,
(a) computes $h_j(s_i) \in \{0, 1, \ldots, m-1\}$,
(b) extracts $b_{h_j(s_i)}$ from $\overline{Y}$,
(ii) sets $E(\bar{s}_i) = \{b_{h_j(s_i)} \cdot r_{i,j}^2\}_{j=0}^{k-1}$,
where $r_{i,0}, \ldots, r_{i,k-1} \hookleftarrow \mathbb{Z}_n$.
$\overline{X} = \{E(\bar{s}_i)\}_{i=1}^{v}$

$\xleftarrow{\overline{X}}$

Sets $card = 0$
For $i = 1, 2, \ldots, v$,
(i) for $j = 0, 1, \ldots, k-1$,
$\mathsf{Dec}_{sk_C}(\mathsf{Enc}_{pk_C}(\bar{s}_{i,j})) = \bar{s}_{i,j}$,
(ii) if $\bar{s}_i$ is all-zero string
then $card = card + 1$.
Outputs $card$ as $|X \cap Y|$

PSI Protocol

C's private input

$Y = \{c_1, c_2, ..., c_w\} \subseteq \{0,1\}^*$

$(pk_C, sk_C) \leftarrow$ **KGen**

constructs BF_Y

For $j = 0, 1, ..., m-1$,

computes $\mathsf{Enc}_{pk_C}(BF_Y[j])$

$\overline{Y} = \{\mathsf{Enc}_{pk_C}(BF_Y[j])\}_{j=0}^{m-1}$

$\xrightarrow{\overline{Y}, pk_C}$

S's private input

$X = \{s_1, s_2, ..., s_v\} \subseteq \{0,1\}^*$

For $i = 1, 2, ..., v$,

(i) for $j = 0, 1, ..., k-1$,

(a) computes $h_j(s_i) \in \{0, 1, ..., m-1\}$,

(b) extracts $b_{h_j(s_i)}$ from $\overline{Y}$,

(ii) generates $\mathsf{Enc}_{pk_C}(s_{i,0}), ..., \mathsf{Enc}_{pk_C}(s_{i,k-1})$,

where $s_{i,j}$ is j-th bit of $\phi(s_i) \in \{0,1\}^k$,

(iii) sets $E(\bar{s}_i) = \{b_{h_j(s_i)} \cdot \mathsf{Enc}_{pk_C}(s_{i,j})\}_{j=0}^{k-1}$

$\overline{X} = \{E(\bar{s}_i)\}_{i=1}^{v}$

$\xleftarrow{\overline{X}}$

computes $\widetilde{Y} = \{\phi(c_i)\}_{i=1}^{w}$

for $i = 1, 2, ..., v$,

(i) for $j = 0, 1, ..., k-1$,

$\mathsf{Dec}_{sk_C}(\mathsf{Enc}_{pk_C}(\bar{s}_{i,j})) = \bar{s}_{i,j}$

sets $\widehat{X} = \{\bar{s}_1, ..., \bar{s}_v\}$

outputs $\{c_i \in Y | \phi(c_i) \in \widehat{X}\}$ as $X \cap Y$

APSI Protocol

Online Phase:

C's private input
$Y = \{c_1, c_2, ..., c_w\} \subseteq \{0,1\}^*$

S's private input
$X = \{s_1, s_2, ..., s_v\} \subseteq \{0,1\}^k$

$\xrightarrow[\Omega]{\overline{Y}, pk_C}$

Verifies $Sig(\bar{h}(b_0), ..., \bar{h}(b_{m-1}))$.
If verification fails, then aborts.
Otherwise, for $i = 1, 2, ..., v$,
(i) for $j = 0, 1, ..., k-1$,
(a) computes $h_j(s_i) \in \{0, 1, ..., m-1\}$,
(b) extracts $b_{h_j(s_i)}$ from $\overline{Y}$,
(ii) generates $\mathsf{Enc}_{pk_C}(s_{i,0}), ..., \mathsf{Enc}_{pk_C}(s_{i,k-1})$,
where $s_{i,j}$ is j-th bit of $s_i \in \{0,1\}^k$,
(iii) sets $E(\bar{s}_i) = \{b_{h_j(s_i)} \cdot \mathsf{Enc}_{pk_C}(s_{i,j})\}_{j=0}^{k-1}$,
$\overline{X} = \{E(\bar{s}_i)\}_{i=1}^{v}$

$\xleftarrow{\overline{X}}$

for $i = 1, 2, ..., v$,
(i) for $j = 0, 1, ..., k-1$,
$\mathsf{Dec}_{sk_C}(\mathsf{Enc}_{pk_C}(\bar{s}_{i,j})) = \bar{s}_{i,j}$,
sets $\widehat{X} = \{\bar{s}_1, ..., \bar{s}_v\}$
outputs $\{c_i \in Y | c_i \in \widehat{X}\}$
as $X \cap Y$

Security

The security definition is based on a comparison between the ideal model and real model.

Security Requirements

- **Privacy:** Each party should learn whatever prescribed in the protocol, not more than that.
- **Correctness:** At the end of interaction, each party should receive correct output.

Theorems

Theorem

If the quadratic residuosity assumption holds, then PSI-CA protocol is a secure computation protocol for functionality $\mathcal{F}_{card} : (Y, X) \longrightarrow (|X \cap Y|, \perp)$ in the standard model against semi-honest server and semi-honest client except with negligible probability $\frac{1}{2^k}$, where $Y = \{c_1, c_2, ..., c_w\} \subseteq \{0,1\}^*$ and $X = \{s_1, s_2, ..., s_v\} \subseteq \{0,1\}^*$ with $w \leq v$.

- APSI-CA is secure in the standard model against semi-honest server and malicious client except with negligible probability ϵ under QR assumption.
- PSI is secure in the standard model against semi-honest server and semi-honest client except with negligible probability ϵ under QR assumption.
- APSI is secure in the standard model against malicious server and malicious client except with negligible probability ϵ under QR assumption.

Efficiency

Table: Comparison of PSI-CA and APSI-CA protocols

PSI-CA Protocol	Security model	Adv. model	Security assumption	Comm.	Comp.	Based on	Size hiding
Sch. 1 of [1]	ROM	SH	DDH and GOMDH	$O(w+v)$	$O(w+v)$		no
Sch. 2 of [1]	ROM	MS, SHC	GOMDH	$O(w+v)$	$O(w+v)$		no
Our	Std	SH	QR	$O(w+v)$	$O(w+v)$	BF	yes
APSI-CA Protocol	**Security model**	**Adv. model**	**Security assumption**	**Comm.**	**Comp.**	**Based on**	**Size hiding**
[2]	Std	Mal	Strong RSA	$O(wv)$	$O(wv)$	OPE	no
[1]	ROM	SH	GOMDH	$O(w+v)$	$O(w+v)$		no
Our	Std	MC, SHS	QR	$O(w+v)$	$O(w+v)$	BF	yes

[1] E. De Cristofaro, P. Gasti, and G. Tsudik, In Cryptology and Network Security 2012.

[2] J. Camenisch and G. M. Zaverucha, In Financial Cryptography and Data Security 2009.

Conclusion

- We have presented efficient constructions for PSI-CA, APSI-CA, PSI and APSI protocols with linear complexities based on Bloom filter and homomorphic GM encryption.
- In our protocols, client's input set size need not be revealed to the server.
- Proposed PSI-CA and APSI-CA are the *first* cardinality set intersection protocols secure in *standard model* with *linear complexity* and preserving client's input set size *independency*.


ThaNK
YOU

For any query mail at sd.iitkgp@gmail.com