

1

# HIDING YOUR SIZE

---

In Private Set Intersection and Related Protocols

Gene Tsudik, UC Irvine

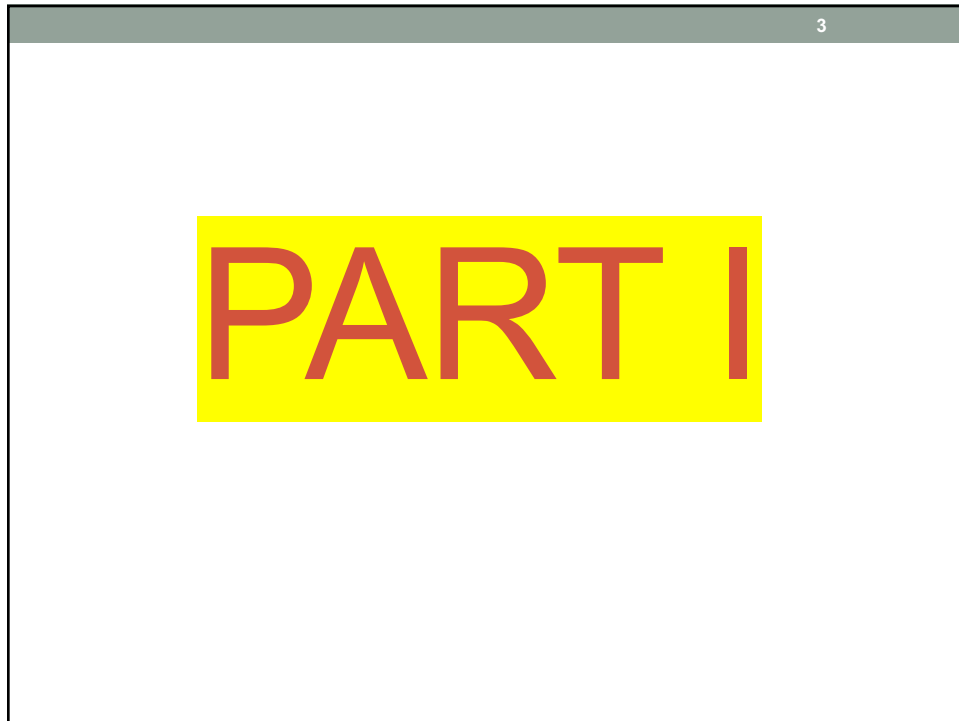
1

2



## Outline

- Private Set Intersection (PSI)
- Why does size matter?
- Size-Hiding via other techniques?
- Is Size-Hiding even possible?
- SHI-PSI: Size-Hiding PSI
  - Security of SHI-PSI
  - Extensions and costs
- Upper-Bounded Size-Hiding PSI
- Lower-Bounded Size-Hiding PSI
- Lower- and Upper-Bounded Size-Hiding PSI
- So what?

2



3



**(If Size Matters)  
SIZE-HIDING PRIVATE  
SET INTERSECTION**

Giuseppe Ateniese<sup>1</sup>, Emiliano De Cristofaro<sup>2</sup>, and **Gene Tsudik**<sup>2</sup>

<sup>1</sup> Johns Hopkins University  
<sup>2</sup> University of California, Irvine

14<sup>th</sup> IACR International Conference on Practice and Theory of Public Key Cryptography (PKC 2011)

4

5

## Outline

- Private Set Intersection (PSI)
- Why does size matter?
- Size-Hiding via other techniques?
- Is Size-Hiding even possible?
  
- SHI-PSI: Size-Hiding PSI
- Security of SHI-PSI
- Extensions
  
- The cost of Size-Hiding
- Conclusion

5

6

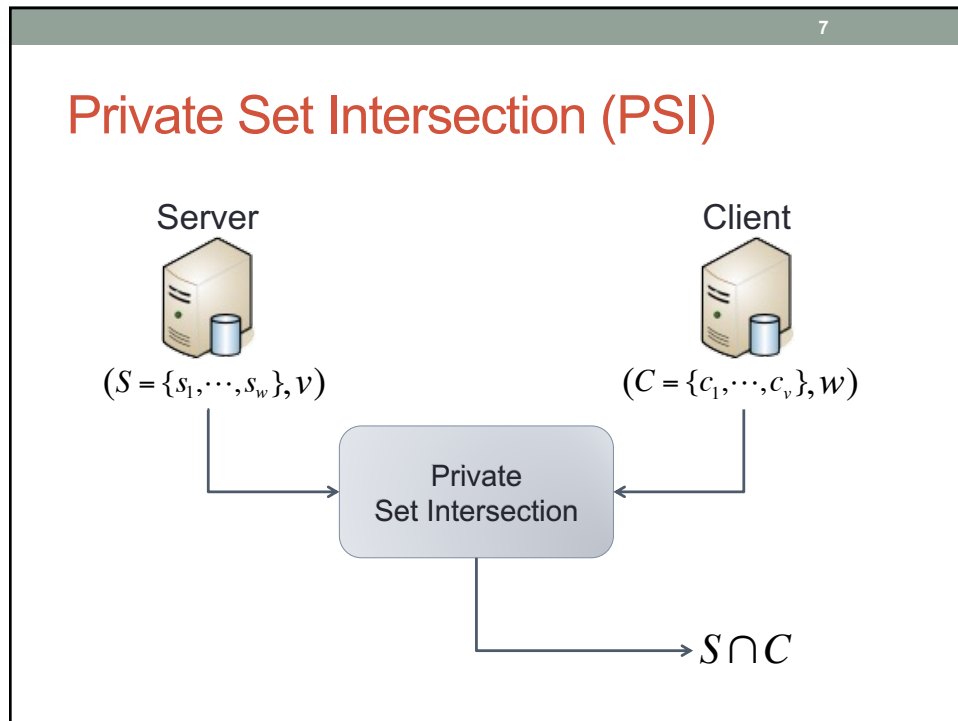
## Privacy

- Privacy and society
  - Basic individual right & desire
  - Relevant to entities, e.g., corporations & governments
  - Increasing awareness
- Privacy and technology
  - >> Information disclosed (mostly on the Internet)
  - >> Handling and transfer of sensitive information
  - << Privacy and accountability
- Yet, sensitive information must be shared at times
- **How to “share” only what must be shared and nothing (or as little as possible) else?**

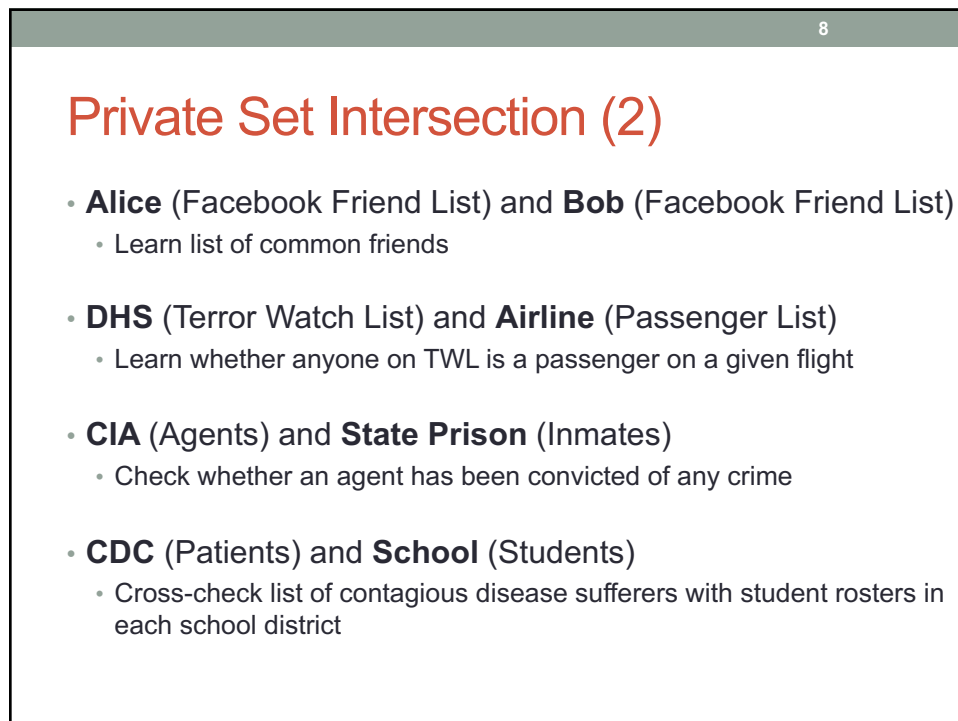


(Image from geekologie.com)

6



7



8

9

## Why does size matter?

- **DHS** can't disclose the size of the **TWL**
  - TWL is dynamic: revealing size leaks sensitive information
- **CIA** can't divulge # of agents
  - Prevented by law
- **CDC** can't reveal the number of infected school-kids
  - Disclosure can cause panic
  - Health insurance rates would go up
- **Fluctuations** in set size might be even more sensitive
- Also, in one-way PSI, (ideally) server **workload** should be independent of client input size

9

10

## Size-Hiding PSI with current tools?

- **Private Set Intersection (PSI)** has been extensively studied
  - Semi-Honest Adversaries: [FNP04], [KS05], [DT10]
  - Malicious Adversaries: [JL09], [DMRS09], [HN10], [DJT10], [JL10]
  - Authorized Inputs (APSI): [DT10], [DKT10]
  - All of them expose input set sizes
- **What if we run PSI with random padding?**
  - Client pads its input up to some fixed size
  - Upper bound would be known
  - If client set is dynamic, fixed upper bound must reflect maximum possible set size
    - Wasted computation and communication

10

## Is Size-Hiding Possible?

- **Secure Two-Party Computation [GMW87]**
  - Input sizes are mutually known
- **Zero-Knowledge Sets (ZKS) [MRK03]**
  - Server publishes a commitment to its database
  - Client asks server to show whether a specific item is in database
  - Neither commitment nor proof reveal database size
  - But, client input is *not* private; ZKS does not offer 2-party functionality
- **One exception:**

[IP07] Y. Ishai and A. Paskin, "Evaluating branching programs on encrypted data", *TCC* 2007.

12

## SHI-PSI: Building Blocks

- **RSA accumulator:**  $g^{\prod_i x_i} \bmod N$ 
  - [Baric-Pfitzmann'97]
- **Unpredictable function:**  $f_{p,q}(x, y) = x^{(1/y) \bmod \phi(N)} \bmod N$ 
  - Unpredictable if p & q are unknown
  - Under RSA assumption on safe moduli
  - Cannot invert in the exponent (if p & q are unknown)

13

## SHI-PSI: Intuition

- **Server** generates its RSA modulus:  $N=pq$
- **Client:** (can't factor  $N$ )
  - Computes a global witness ( $\mathbf{X}$ ) for its input set as an RSA accumulator based on all of its (hashed) items:  $x_i = H(c_i)$

$$X = g^{\prod_i x_i} \bmod N$$

This unconditionally hides client items (and their number)

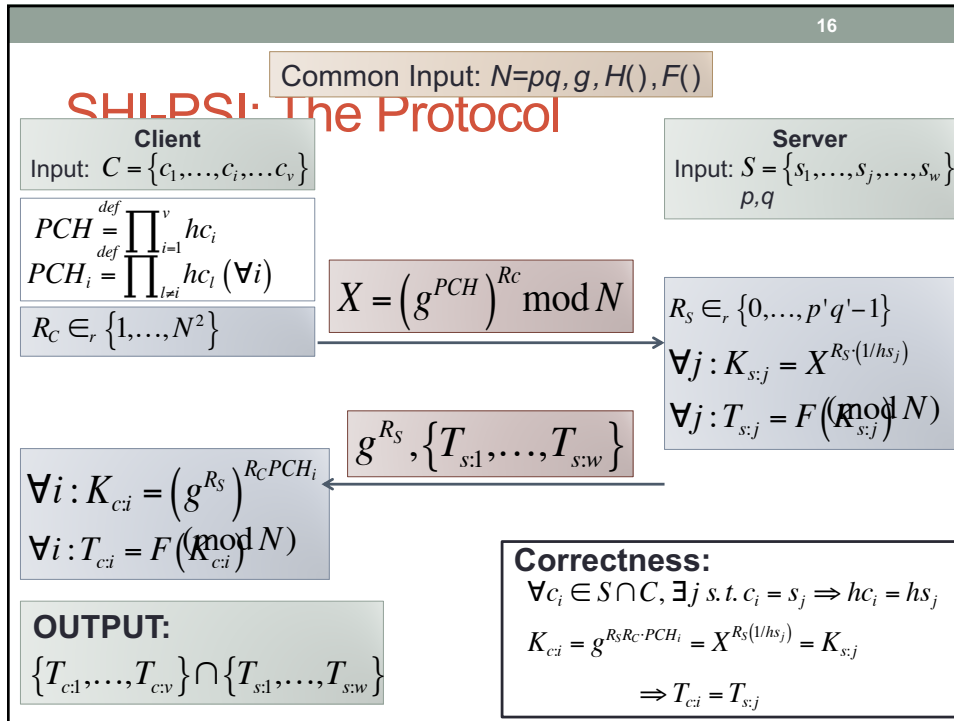
- **Server:** (receives  $\mathbf{X}$ , knows  $p, q$ )
  - Computes:  $f_{p,q}(X, s_j) = X^{1/H(s_j)}$
  - Applies one-way function (cryptographic hash)
  - Hash of unpredictable function is a PRF (in ROM)

14

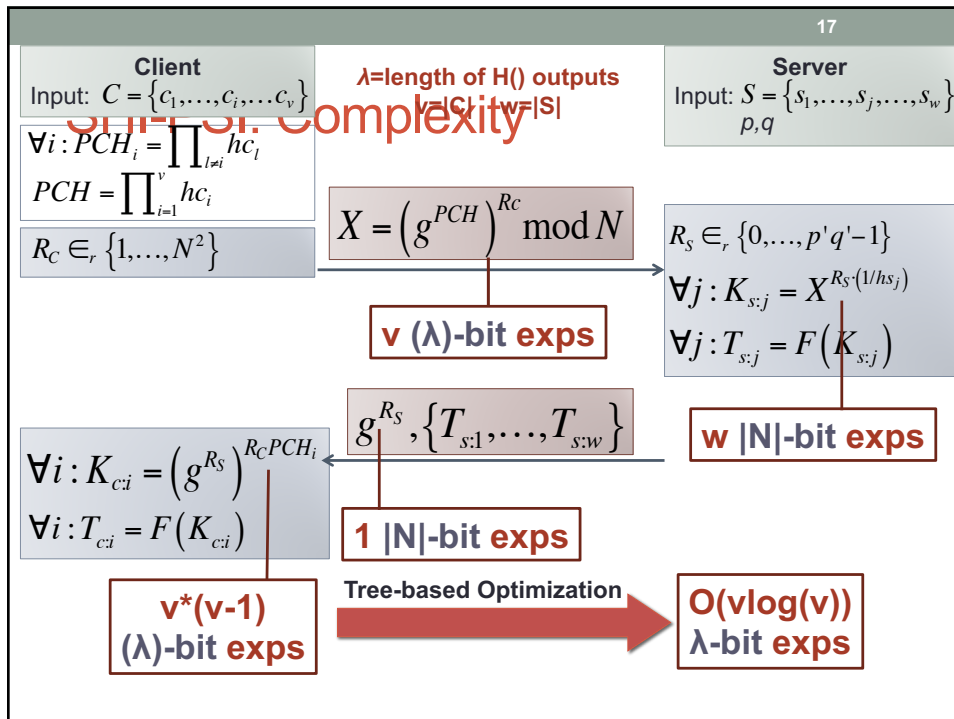
## SHI-PSI: Notation

Symbol	Meaning
$\lambda, \lambda_1, \lambda_2$	Security Parameters
$p, q$	Safe primes
$N=pq$	Safe RSA modulus
$g$	Generator of $QR_N$
$H()$	Random Oracle $H: \{0, 1\}^* \rightarrow \{0, 1\}^{\lambda_1}$
$F()$	Random Oracle $F: \{0, 1\}^* \rightarrow \{0, 1\}^{\lambda_2}$
$C, S$	Client and Server sets
$v, w$	Sizes of $C$ and $S$
$i \in [1, v]$	Index of elements of $C$
$j \in [1, w]$	Index of elements of $S$
$c_i, s_j$	Generic elements of $C$ and $S$
$hc_i, hs_j$	$H(c_i), H(s_j)$
$\pi$	Random permutation

15



16



17



18

## SHI-PSI: Complexity

- **Communication**
  - $O(1)$  client  $\rightarrow$  server
  - $O(w)$  server  $\rightarrow$  client
- **Client Computation**
  - $O(v \cdot \log(v))$  modular exps
- **Server Computation**
  - $O(w)$  modular exps

$$v = |C|$$

$$w = |S|$$

18

19

## PSI Security Requirements (informal)

- **Correctness**
  - Client outputs intersection (if any)
- **Server Privacy**
  - Client learns nothing about server elements not in intersection
- **Client Privacy**
  - Server learns nothing about client input (except size?)
- **(opt) Server Unlinkability**
  - Client cannot tell if any two (or more) protocol instances are related, i.e., run on same server input
- **(opt) Client Unlinkability**
  - Server cannot tell if any two (or more) protocol instances are related, i.e., run on same client input

19

20

## SHI-PSI: Security

- **Assumptions**
  - Random Oracle Model (ROM)
  - Honest-but-Curious (HbC) adversaries
  - RSA assumption on safe moduli
- **Client Privacy: Indistinguishability**
  - For every PPT  $S^*$  that plays the role of the server, for every input set  $S$ , and for any client input set  $(C^{(0)}, C^{(1)})$ , two views of  $S^*$  corresponding to client's inputs:  $C^{(0)}$  and  $C^{(1)}$  are computationally indistinguishable. (**Not even if  $|C^{(0)}| \neq |C^{(1)}|$** ).
- **Server Privacy: Comparison to Ideal Model**
  - Let  $View_{Client}(C, S)$ , be a random variable representing Client's view during execution of SHI-PSI with inputs  $(C, S)$ . There exists a PPT algorithm  $C^*$  s.t.:
 
$$\{C^*(C, S \cap C)\}_{(C, S)} \equiv \{View_{Client}(C, S)\}_{(C, S)}$$

20

24

## The Cost of Hiding Size: PSIs vs SHI-PSI

	Tools	Model	Adv	Server Op	Client Op	Bandwidth
[FNP04]	Oblivious Poly Eval	Standard/ROM	HbC/Malicious	$O(w \log \log(v))$ 160-bit mod 1024 exps	$O(w+v)$ 160-bit mod 1024 exps	$O(w+v)$
[KS05]	Oblivious Poly Eval	Standard	HbC Malicious*	$O(w \cdot v)$ $m$ -bit mod 2048 exps	$O(w+v)$ $m$ -bit mod 2048 exps	$O(w+v)$
[JL09]	OPRF q-DDH	Standard CRS	Malicious	$O(w)$ $m$ -bit mod 2048 exps	$O(v)$ $m$ -bit mod 2048 exps	$O(w+v)$
[HN10]	DDH	Standard	Malicious	$O(w \log \log(v))$ 160-bit mod 1024-bit exps	$O(w+v)$ 160-bit mod 1024 exps	$O(w+v)$
[JL10]	OneMore-DH	ROM	Malicious	$O(w+v)$ 160-bit mod 1024 exps	$O(v)$ 160-bit mod 1024 exps	$O(w+v)$
[DT10]	OneMore-RSA	ROM	HbC	$O(w+v)$ 1024-bit mod 1024 exps	$O(v)$ mod <b>mults</b>	$O(w+v)$
[DKT10]	DDH	ROM	Malicious	$O(w+v)$ 160-bit mod 1024 exps	$O(v)$ 160-bit mod 1024 exps	$O(w+v)$
SHIPSI	RSA	ROM	<b>HbC</b>	$O(w)$ 1024-bit mod 1024 exps	$O(v \log(v))$ 1024-bit mod 1024 exps	$O(w)$

**$v = |C| \quad w = |S|$**

24

25

## Summary:

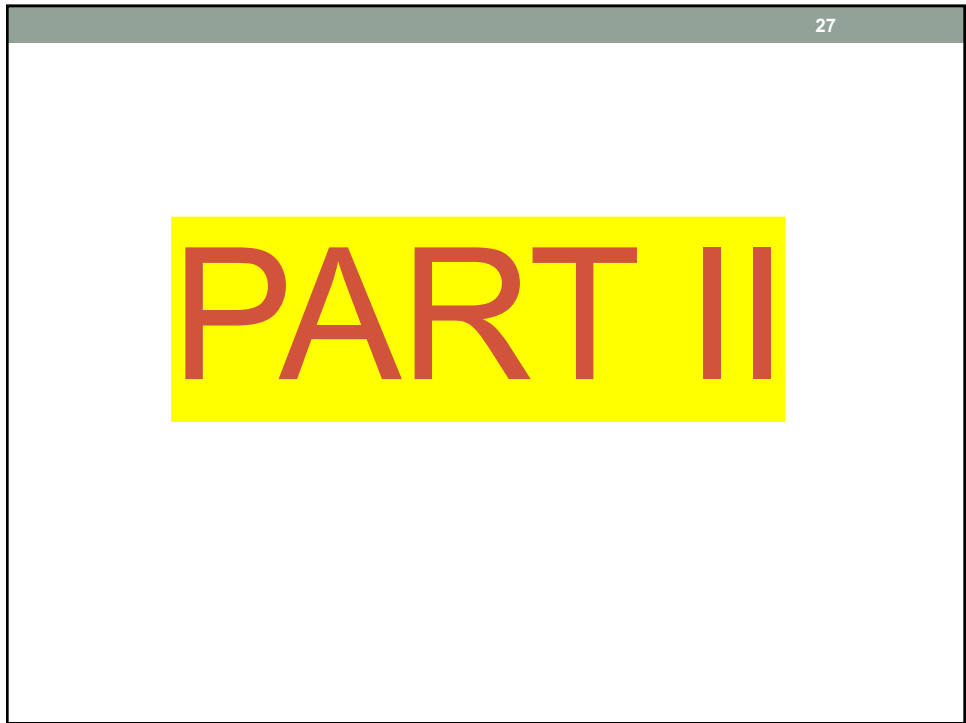
- New concept of **Size-Hiding** Private Set Intersection
- A **secure** and **efficient** SHI-PSI construct
  - First of its kind
  - Despite 2PC defs, input sizes do not have to be revealed
- **Current/Future Work:**
  - Eliminate ROM
  - SHI-PSI in presence of malicious participants
  - Authorization of client input
  - Extend to multiple clients (multi-party SHI-PSI)

25

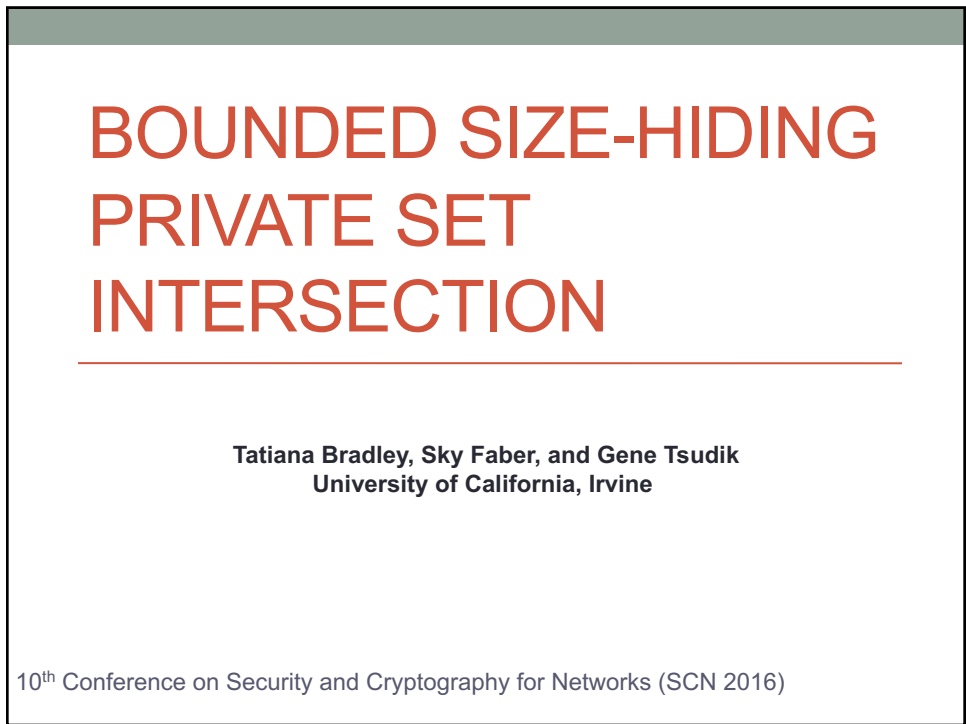
26

## Questions so far?

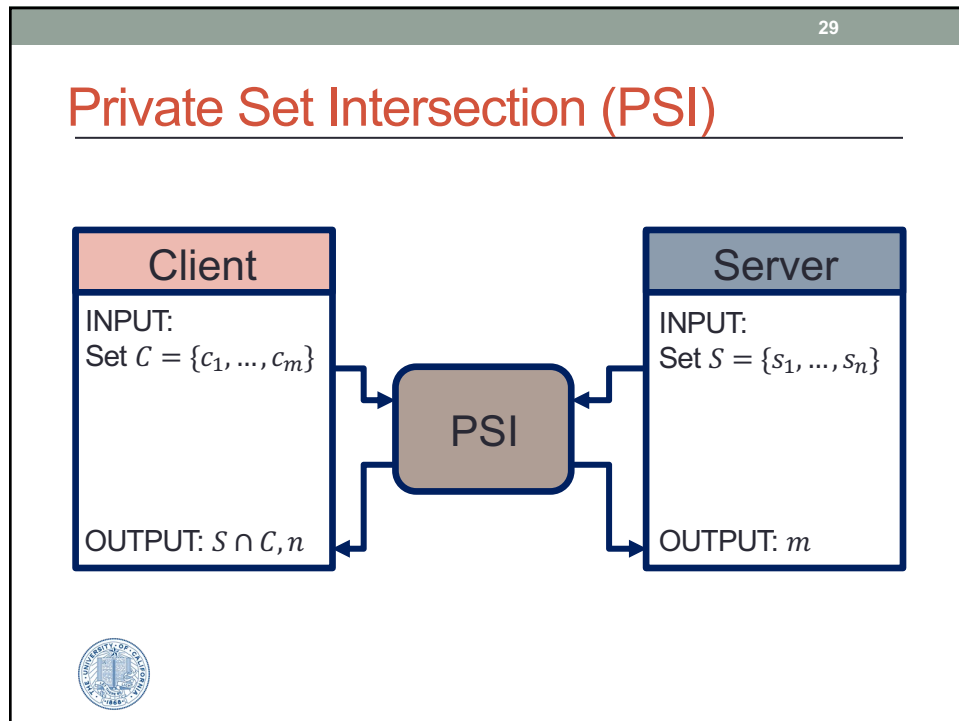
26



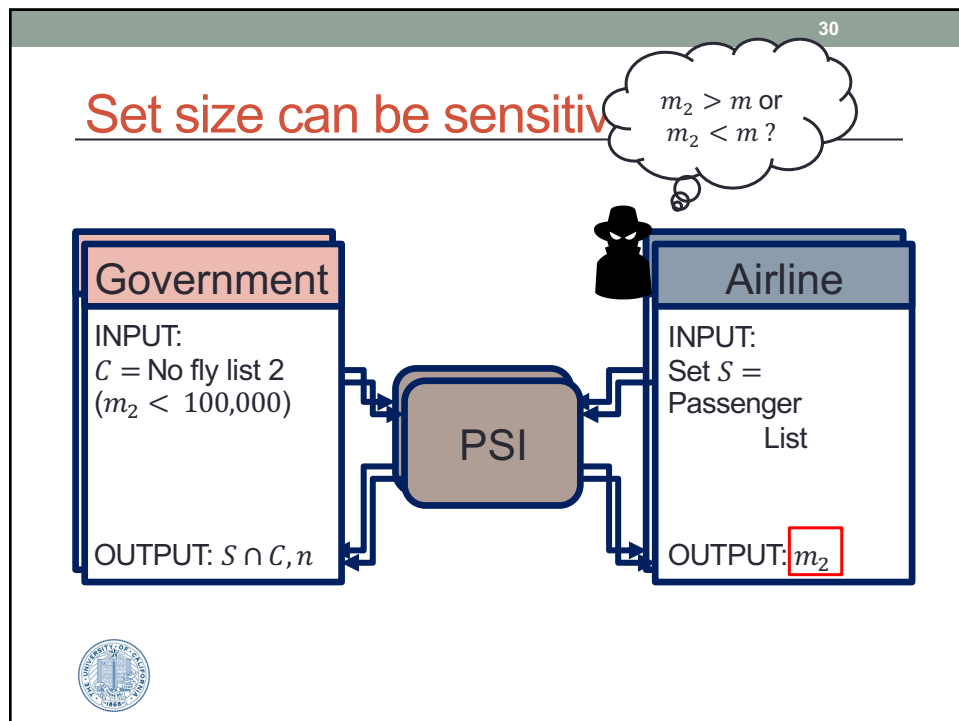
27



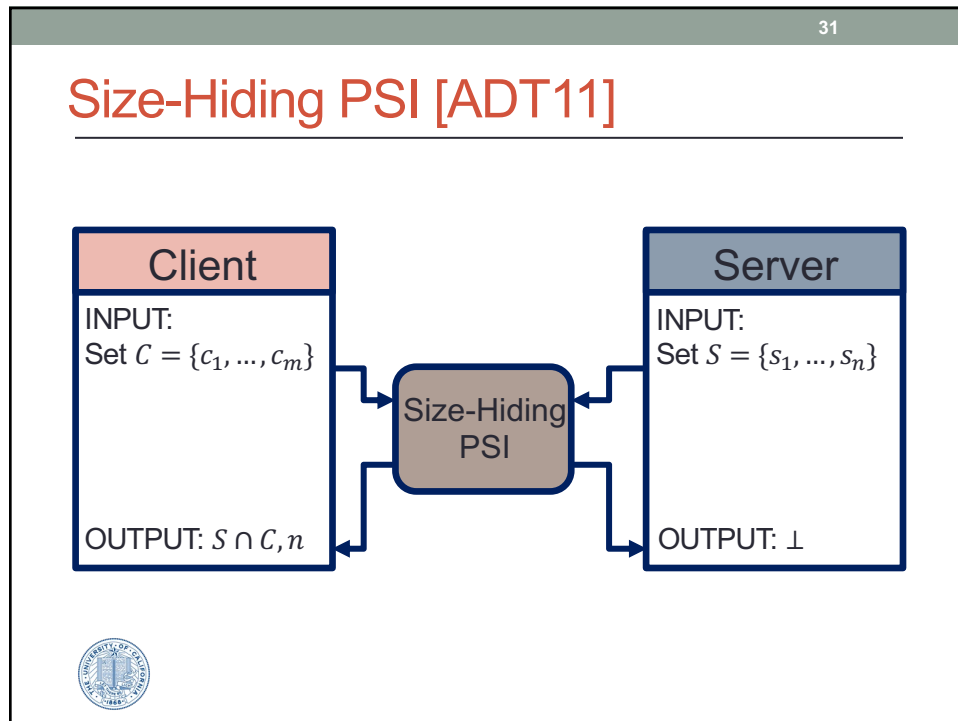
28



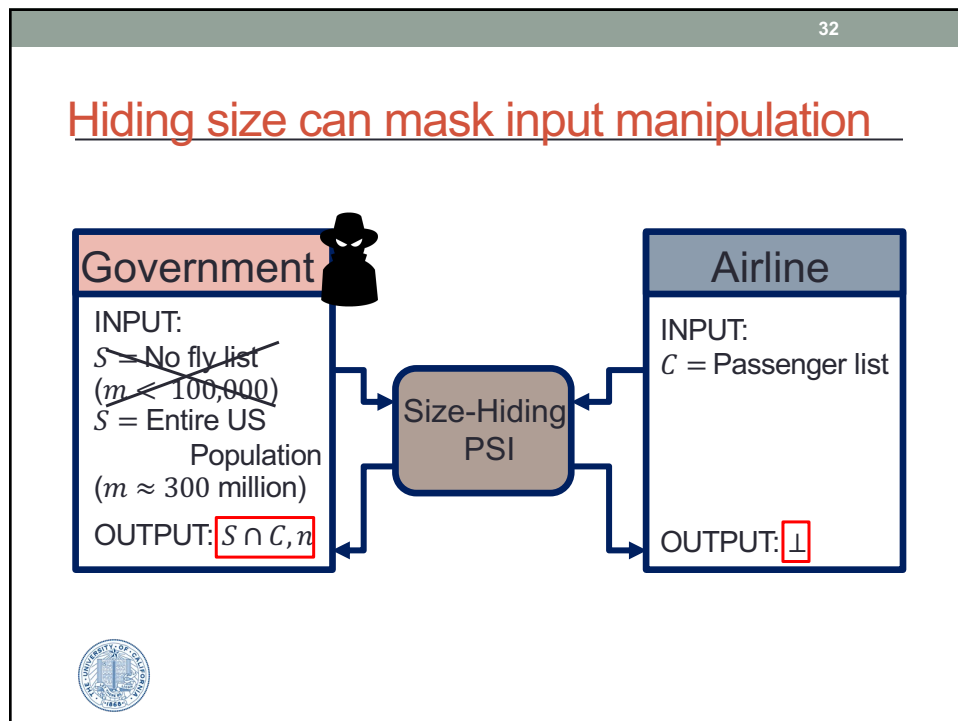
29



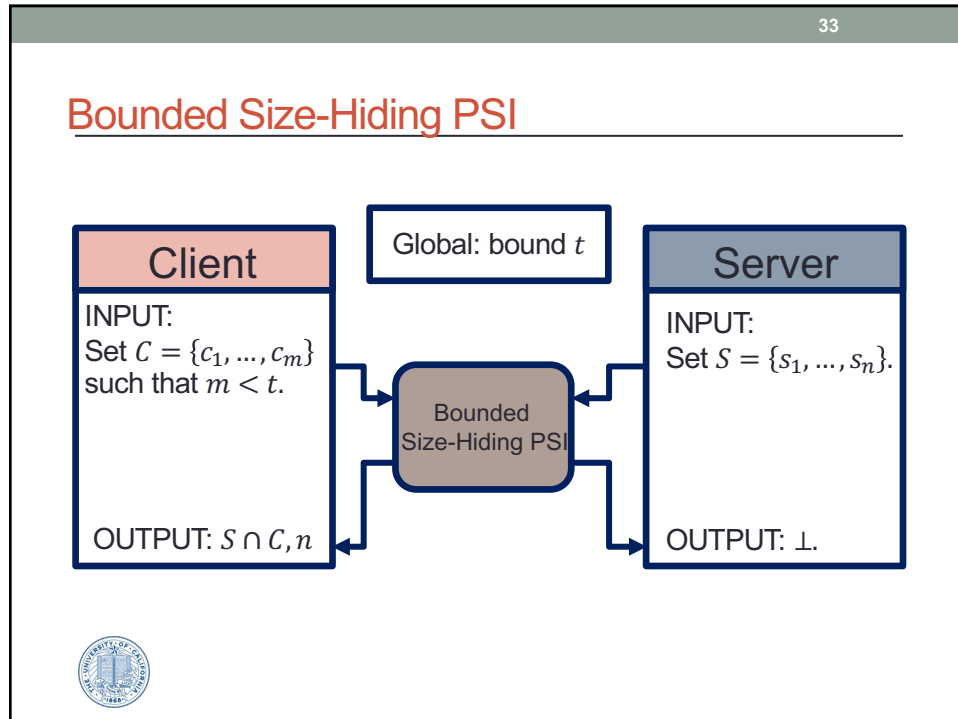
30



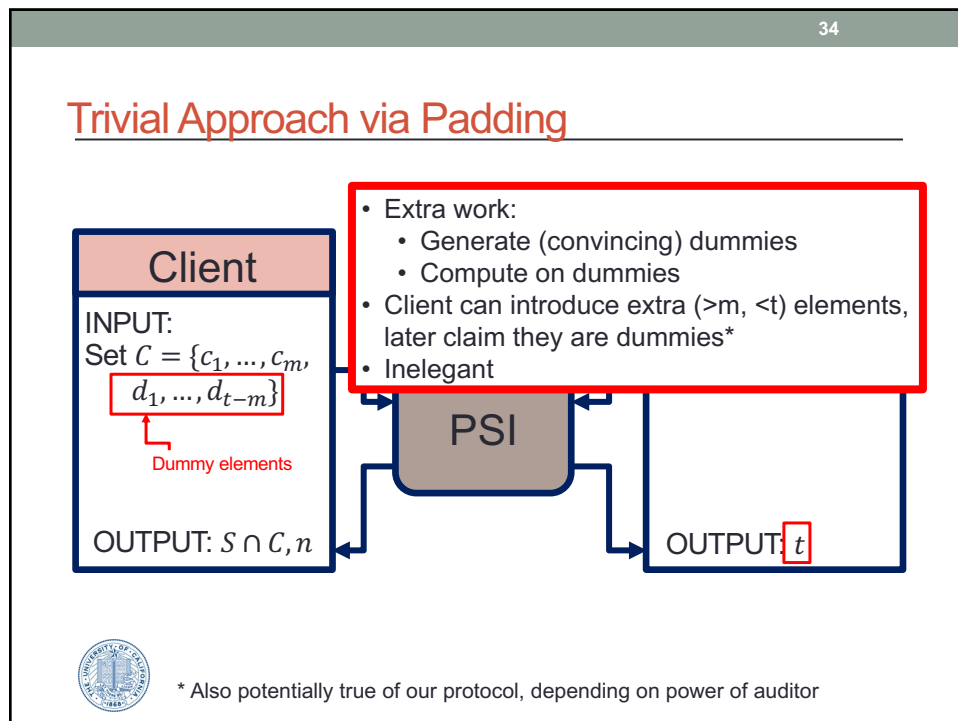
31



32



33



34

35

## Related Work and Contribution

- PSI, e.g., [DT10, DCW13, FNP04, DMR12]
- **Size-Hiding PSI** [ADT11, DVP15]
- Size hiding in 2PC/MPC
  - [COV15, DFT13, GOS+14, IP07, LNO13, MRK03]
  - Generic techniques; padding; specific non-PSI protocols
- **Contribution of this work:**
  - Notion of Bounded Size-Hiding PSI (bSH-PSI)
  - First provably secure+efficient bSH-PSI **not** based on padding

35

36

## Model and Requirements

- Random Oracle Model
- Relaxed HbC model = stronger adversary
  - Client can try to cheat on set size
- Correctness
- Boundedness
  - Client can't learn anything extra info by inputting  $> t$  elements
- Client Privacy
  - Server learns nothing about client set
- Server Privacy
  - Client learns nothing about server set beyond intersection and size

36



37

## Hardness Assumptions

Given  $[g, g^z, g^{(z^2)}, \dots, g^{(z^q)}]$  ← All mod  $p$

Generator of  $Z_{p'}^* \leq Z_p^*$

- **One generator** q-Strong Diffie-Hellman problem [BB04]
  - Find  $[c, g^{\frac{1}{z+c}}]$  where  $c \in Z_{p'}^*$
  - Polynomial-generalized: find  $[c, g^{\frac{P_n(z)}{z+c}}]$  where  $c \in Z_{p'}^*, n \leq q, (z+c) \nmid P_n(z)$  ↪ Equivalent
- **Exponent** q-Strong Diffie-Hellman problem [TS10]
  - Find  $g^{(z^{q+1})}$  ↪ Equivalent
  - Polynomial-generalized: Find  $g^{P_n(z)}$  where  $n > q$  ↪ Equivalent

37

38

## Bounded Size-Hiding PSI

**Client**

INPUT:  
Set  $C = \{c_1, \dots, c_m\}$   
such that  $m < t$ .

OUTPUT:  $S \cap C, n$

Global: bound  $t$

**Server**

INPUT:  
Set  $S = \{s_1, \dots, s_n\}$ .

OUTPUT:  $\perp$ .

Bounded Size Hiding PSI

38

39

## Protocol Overview

### Client

- Embed all set elements into accumulator  $X$  (based on [N04]) using public key
- Send  $X$  to server
- Compute a tag (derived from witness) for each set element using public key

### Server

- Compute a tag (based on  $X$  and secret  $z$ ) for each set element
- Send tags (with order permuted) to client

### Client

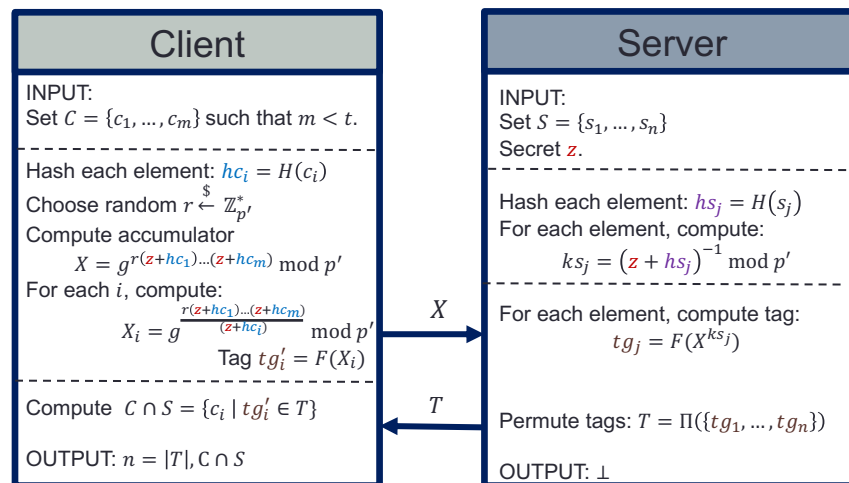
- Compute normal intersection of client and server tags

39

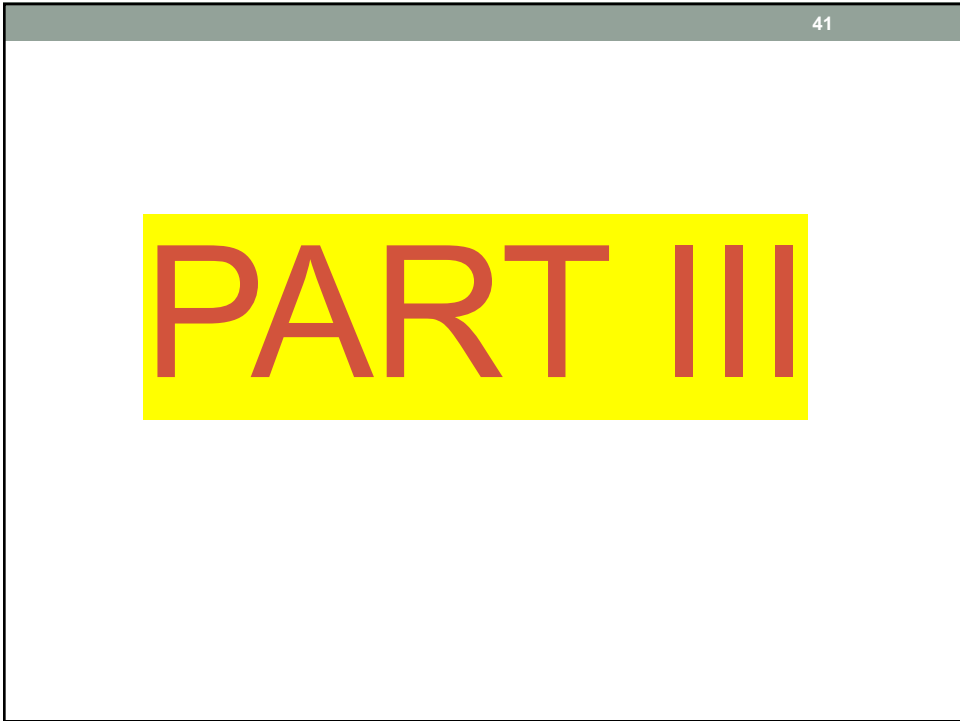
40

## Protocol


**Public:** Bound  $t$ , Primes  $p, p'$ , Random oracles  $H(\cdot), F(\cdot)$ ,  
Key  $[g, g^z, g^{z^2}, \dots, g^{z^t}]$



40



41



**ELEMENT DISTINCTNESS AND  
BOUNDED INPUT  
SIZE IN PSI AND VARIANTS**

---

Xavier Carpent, University of Nottingham  
Seoyeon Hwang, University of California, Irvine  
Gene Tsudik, University of California, Irvine

22nd International Conference on Applied Cryptography and Network Security (ACNS 2024)

42

43

# SWITCH

43

44

## So what?

- Size-Hiding is interesting and sometimes important
- Unbounded size-hiding is cool
  - 13+ years later there have been no other concrete SH-PSI constructs
  - RSA accumulator SH-PSI seems to be a weird singleton
- Upper-Bounded Size-Hiding is more practical/realistic
- Lower-Bounded Size-Hiding has some applications
  
- Raises a major issue: What is the role of input validity in PSI and related protocols?

44