Efficient Cryptographic Constructions for Privacy-preserving Applications

Dawn Song

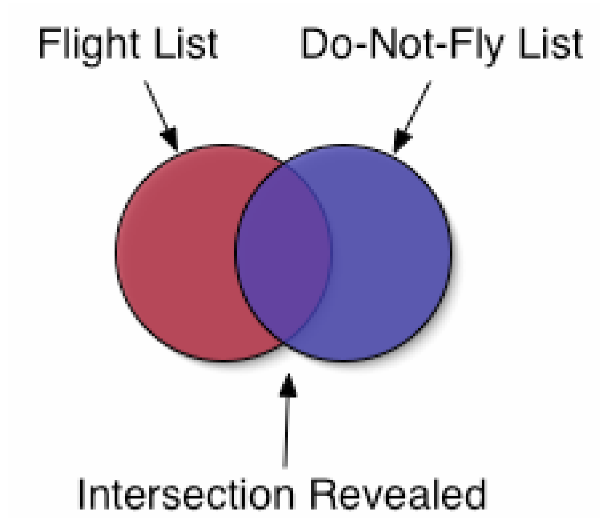
UC Berkeley

Privacy-preserving Computation

- Privacy-preserving set operations
- Computation over encrypted data

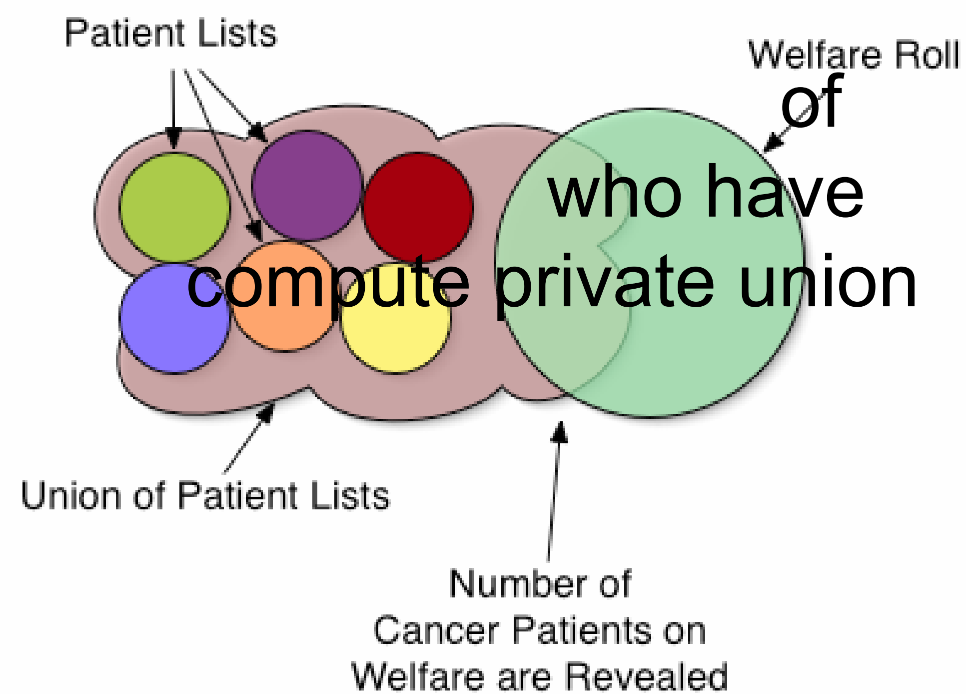
Motivation (1)

- Many bodies of data can be represented as multisets
- The utility of data is greatly increased when shared, but there are often privacy and security concerns
- Do-not-fly list
  - Airlines must determine which passengers cannot fly
  - Government and airlines cannot disclose their lists



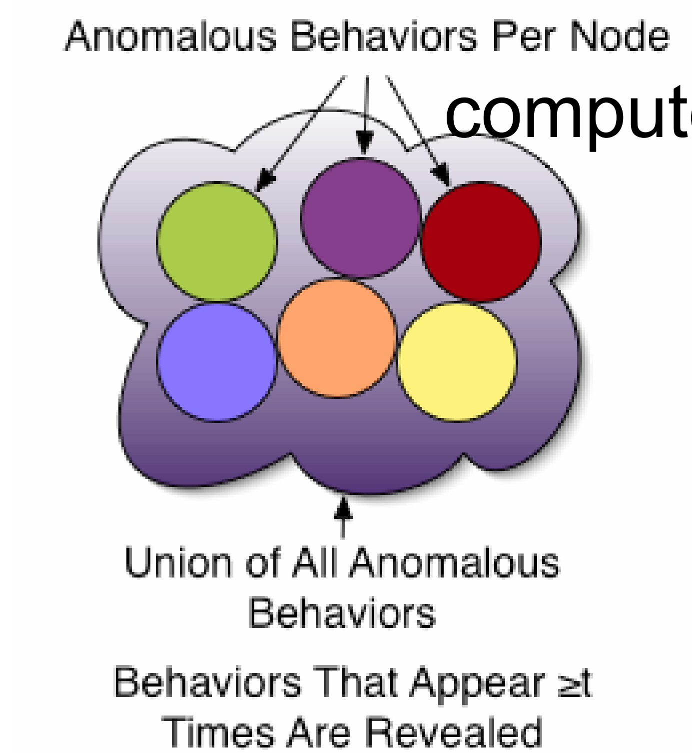
Motivation (2)

- Public welfare survey: how many welfare recipients are being treated for cancer?
  - Cancer patients and welfare rolls are confidential
  - To reveal the number of welfare recipients who have cancer, must compute private union and intersection operations



Motivation (3)

- Distributed network monitoring
  - Nodes in a network identify anomalous behaviors, and filter out uncommon elements
  - The nodes must privately compute element reduction and union operations
  - If an element *a* appears *b* times in S, *a* appears *b-1* times in the reduction of S



Motivation (4)

- Finding friends common in address books
- Finding common interest in social networks
- Finding popular items in social networks

Kissner-Song Construction

- Efficient, composable, privacy-preserving operations on multisets: intersection, union, element reduction
- We use these techniques to give efficient protocols (secure against HBC and malicious adversaries) for practical problems
- Other example applications:
  - General computation on multisets
  - Determining subset relations
  - Evaluating distributed boolean formulas

Outline

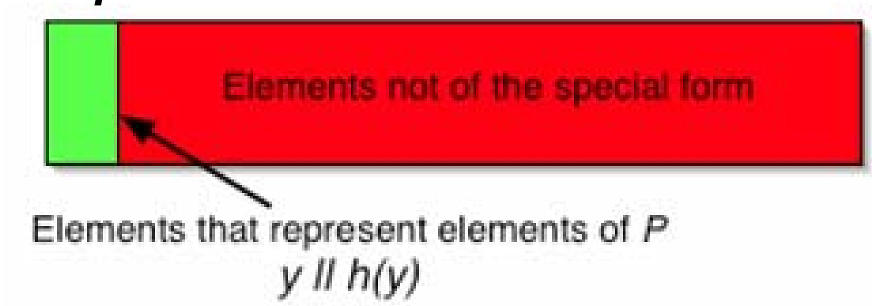
- Techniques for privacy-preserving operations
  - *Polynomial representation*
  - *Indistinguishable TTP security model*
  - *Multiset operations*
  - Multiset operations without a TTP
- General computation with multisets

Sets as Polynomials

- To represent multiset S as a polynomial over ring R, compute $\prod_{a \in S} (x - a)$
- The elements of the set represented by polynomial *f* are the **roots of *f* of a certain form** $y \,||\, h(y)$

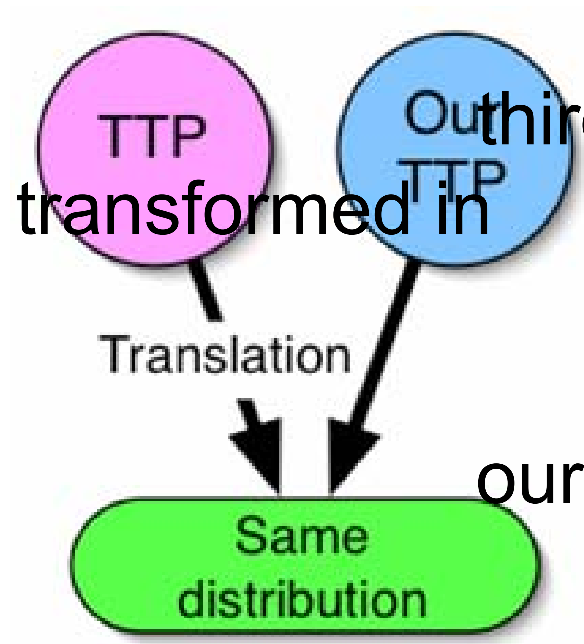
–Random elements are not of this form (with overwhelming probability)

–Let elements of this form *represent elements of P*



Security for Techniques

- We define security (privacy-preservation) for the **techniques** we present as follows:
  - The output of a trusted third party (TTP) can be transformed in probabilistic polynomial time to be distributed identically to a TTP using our techniques



  - This hides all information but the result

Multiset Union

- Let S, T be multisets represented by *f, g*
- We calculate $S \cup T$ as **f*g**
- Theorem: *There exists a PPT translation of the output of a TTP calculating $S \cup T$, such that the translation is distributed identically to f*g.*
- From this theorem we may conclude that our calculation of $S \cup T$ is secure
  - Correct
  - Exposes no additional information

Multiset Intersection

- Let S, T be multisets represented by *f, g*, Deg(*f*)=Deg(*g*)
- Let *r,s* be uniformly distributed polynomials from $R^{Deg(f)}[x]$ (each coefficient chosen u.a.r. from R)
- We calculate S∩T as **f*r+g*s**
  - Polynomial addition preserves shared roots of *f, g*
  - The operation can use ≥2 multisets

Multiset Intersection

Lemma:

If ***gcd(v,w)=1****,*

Deg(v)=Deg(w),

$y \geq Deg(v)$,

$r,s \leftarrow R^{y}[x]$,

then ***v*r+w*s is uniformly distributed over*** $R^{Deg(v)+y}[x]$

Multiset Intersection

- Theorem: *There exists a PPT translation of the output of a TTP calculating $S \cap T$, such that the translation is distributed identically to $f*r+g*s$.*
  - By Lemma,
    $f*r+g*s = gcd(f,g) * (v*r+w*s) = gcd(f,g)*u$, where u is uniformly distributed
  - Note that gcd(f,g) is the polynomial representation of $S \cap T$

Multiset Reduction

- Let S be a multiset represented by *f*,
r_i be uniformly distributed polynomials from $R^{Deg(f)}[x]$, Fi be a public random polynomial $Deg(F_i)=i$ (with a few other properties),
- We calculate $Rd_d(S)$ as $\sum_{0 \leq i \leq d}$ $\mathbf{f^{(i)} * F_i * r_i}$

Multiset Reduction

- Theorem: *There exists a PPT translation of the output of a TTP calculating $Rd_d(S)$, such that the translation is distributed identically to* $\sum_{0 \leq i \leq d} \mathbf{f^{(i)} * F_i * r_i}$

Outline

- Techniques for privacy-preserving operations
  - Polynomial representation
  - Indistinguishable TTP security model
  - Multiset operations
  - *Multiset operations without a TTP*
- General computation with multisets

Without TTP (1)

- Encrypt coefficients of polynomial using a *threshold additively homomorphic* cryptosystem
- We can perform the calculations needed for our techniques with encrypted polynomials (examples use Paillier cryptosystem)

– Addition

$$\begin{aligned} h &= f + g \\ h_i &= f_i + g_i \\ E(h_i) &= E(f_i) * E(g_i) \end{aligned}$$

Without TTP (2)

- Formal derivative

$$\begin{aligned} h &= f' \\ h_i &= (i+1) f_{i+1} \\ E(h_i) &= E(f_i)^{i+1} \end{aligned}$$

- Multiplication

$$\begin{aligned} h &= f * g \\ h_i &= \sum_{j=0}^{k} f_j * g_{i-j} \\ E(h_i) &= \prod_{j=0}^{k} E(f_j)^{g_{i-j}} \end{aligned}$$

Outline

- Techniques for privacy-preserving operations
- General computation with multisets

General Functions

- Using our techniques, efficient protocols can be constructed for any function described by (let s be a privately held set):
  - $\gamma ::= s \mid Rd_d(\gamma) \mid \gamma \cap \gamma \mid s \cup \gamma \mid \gamma \cup s$
- Can less efficiently compute $\gamma ::= \gamma \cup \gamma$
- Additional tricks can be used with our techniques to solve additional problems
- All example protocols deferred to paper

Summary (1)

- Efficient, composable techniques for privacy-preserving multiset intersection, union, and element reduction
- Protocols for $n \geq 2$ players, $c < n$ dishonest
  - Multiset intersection
  - Cardinality of multiset intersection
  - Over-threshold multiset-union
  - Threshold multiset-union (and variants)

Summary (2)

- Protocols secure against malicious players
- Our protocols are fair, if fairness is enforced in threshold decryption
- Efficient computation of many functions over multisets
- General computation over multisets
- Determining subset relations
- Evaluating distributed boolean formulas

Related Work

- Two-party intersection (and related problems): [AES03] [FNP04]
- Disjointness of sets: [KM05]
- Single-element intersection: [FNW96] [NP99] [BST01] [L03]
- For most of the problems we address, the most efficient previous work is general MPC [Y82] [BGW88]

Computation over Encrypted Data

- General computation over encrypted data
- Fully homomorphic encryption by Craig Gentry