Privacy-Preserving Set Operations

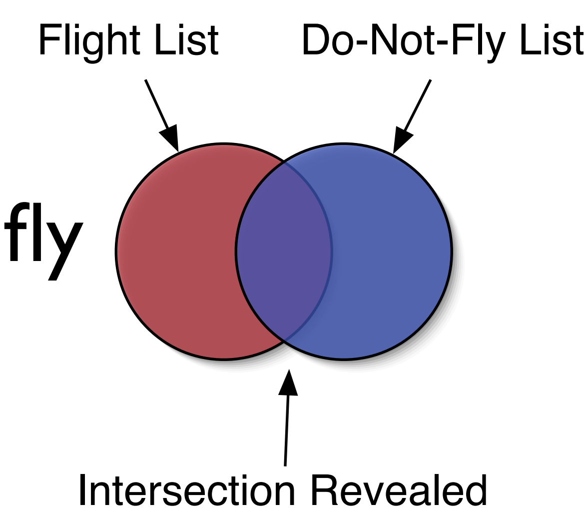
Lea Kissner
leak@cs.cmu.edu

Dawn Song
dawnsong@cmu.edu

Carnegie Mellon University

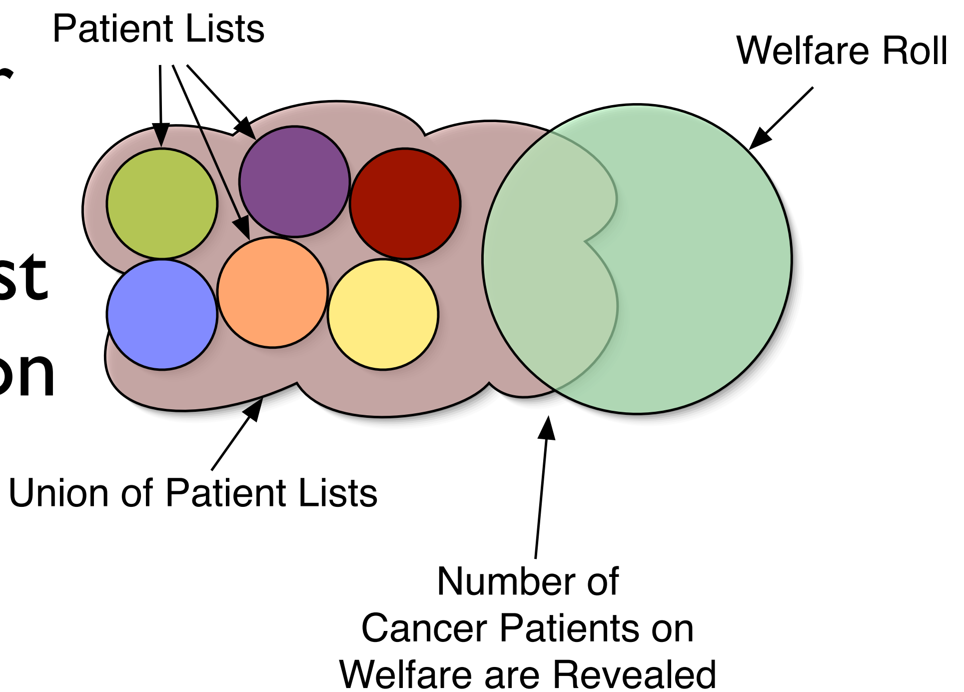
Motivation (1)

- 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



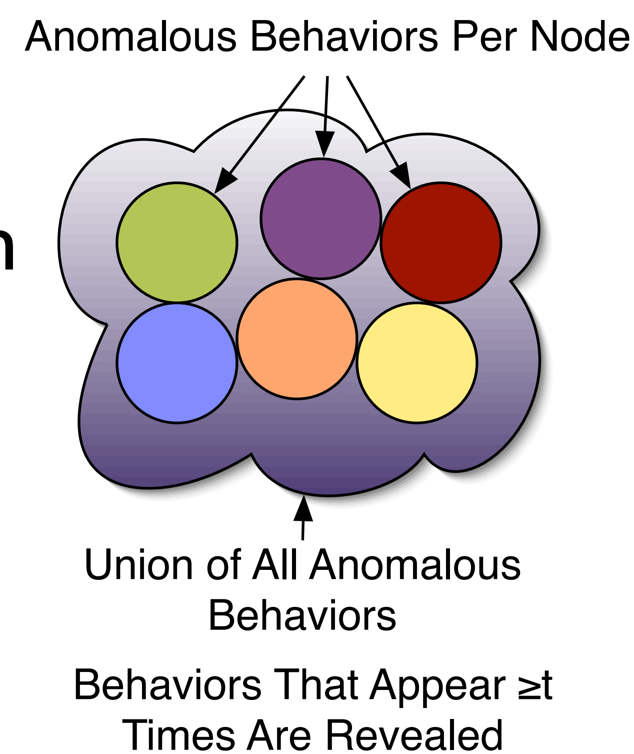
Motivation (2)

- 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



Outline

- Introduction
  - Motivation
  - *Contributions*
  - *Related work*
- Techniques for privacy-preserving operations
- General computation with multisets
- Conclusion

Contributions

- 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

Related Work

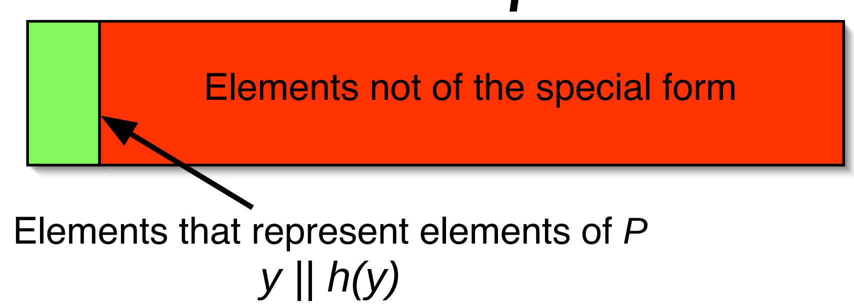
- Two-party intersection (and related problems): [AES03] [FNP04]
- Disjointness of sets: [KM05]
- Single-element intersection: [FNW96] [NP99] [BST01] [L03]
- For most of the problems we address, the most efficient previous work is general MPC [Y82] [BGW88]

Outline

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

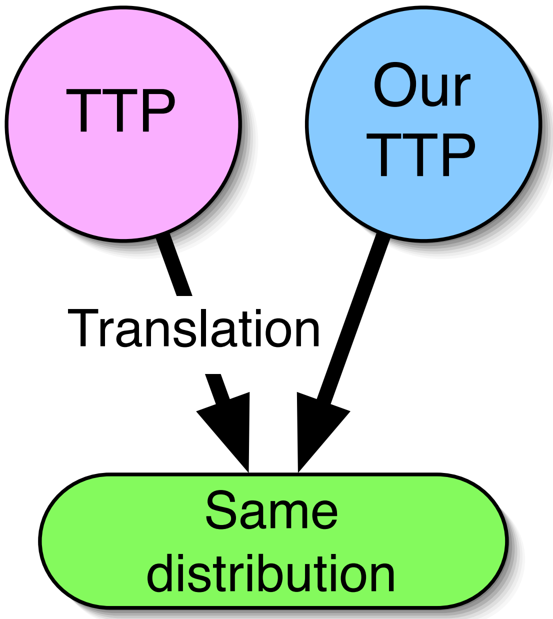
Sets as Polynomials

- To represent multiset S as a polynomial over ring R, compute $\prod_{a \in S} (x - a)$
- The elements of the set represented by the polynomial *f* is 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 $\boldsymbol{gcd(v,w)=1}$,

$Deg(v)=Deg(w)$,

$y \geq Deg(v)$,

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

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

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*,*s* be uniformly distributed polynomials from $R^{\mathrm{Deg}(f)}[x]$, F be a public random polynomial Deg(F)=d

- We calculate $Rd_d(S)$ as $\mathbf{f^{(d)}*F*r + f*s}$
  - According to standard lemma, desired result is obtained by calculating intersection of *f*, $f^{(d)}$
  - If $f(a) = 0, f^{(d)}(a) = 0 \Leftrightarrow (x-a)^{d+1} \mid f$

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 $f^{(d)}*F*r+f*s$.*
  - By our earlier lemma, $f^{(d)}*F*r+f*s=$ gcd$(f^{(d)},f)*u$ where u is uniformly distributed
  - Note that by standard lemma, gcd$(f^{(d)},f)$ is the polynomial representation of $Rd_d(S)$

Outline

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

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

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

Without TTP (2)

- Formal derivative
  -
  -
  -
- Multiplication

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

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

Outline

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

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

Conclusion (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)

Conclusion (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

Thank you

http://www.cs.cmu.edu/~leak/

Multiset Intersection

- Let each player i ($1 \leq i \leq n$) hold an input multiset S_i
- Each player calculates the polynomial f_i representing S_i and broadcasts $E(f_i)$
- For each i, each player j ($1 \leq j \leq n$) chooses uniformly distributed polynomial $r_{i,j}$, and broadcasts $E(f_i * r_{i,j})$
- All players calculate and decrypt
- $$E\left(\sum_{i=1}^{n} f_i * \left(\sum_{j=1}^{n} r_{i,j}\right)\right) = E(p)$$
- Players determine the intersection multiset: if $(x-a)^b \mid p$ then a appears b times in the result

Multiset Intersection

- Let each player i ($1 \leq i \leq n$) hold an input multiset S_i
- Each player calculates the polynomial f_i representing S_i and broadcasts $E(f_i)$
- For each i, each player j ($1 \leq j \leq n$) chooses uniformly distributed polynomial $r_{i,j}$, and broadcasts $E(f_i * r_{i,j})$
- All players calculate and decrypt
- $$E\left(\sum_{i=1}^{n} f_i * \left(\sum_{j=1}^{n} r_{i,j}\right)\right) = E(p)$$
- Players determine the intersection multiset: if $(x-a)^b \mid p$ then a appears b times in the result

Multiset Intersection

- Let each player i ($1 \leq i \leq n$) hold an input multiset S_i
- Each player calculates the polynomial f_i representing S_i and broadcasts $E(f_i)$
- For each i, each player j ($1 \leq j \leq n$) chooses uniformly distributed polynomial $r_{i,j}$, and broadcasts $\boxed{E(f_i * r_{i,j})}$
- All players calculate and decrypt
- $$E\left(\sum_{i=1}^{n} f_i * \left(\sum_{j=1}^{n} r_{i,j}\right)\right) = E(p)$$
- Players determine the intersection multiset: if $(x-a)^b \mid p$ then a appears b times in the result

Multiset Intersection

- Let each player i ($1 \leq i \leq n$) hold an input multiset S_i
- Each player calculates the polynomial f_i representing S_i and broadcasts $E(f_i)$
- For each i, each player j ($1 \leq j \leq n$) chooses uniformly distributed polynomial $r_{i,j}$, and broadcasts $E(f_i * r_{i,j})$
- All players calculate and decrypt
- $$\boxed{E\left(\sum_{i=1}^{n} f_i * \left(\sum_{j=1}^{n} r_{i,j}\right)\right) = E(p)}$$
- Players determine the intersection multiset: if $(x-a)^b \mid p$ then a appears b times in the result

Multiset Intersection

- Let each player i ($1 \leq i \leq n$) hold an input multiset S_i
- Each player calculates the polynomial f_i representing S_i and broadcasts $E(f_i)$
- For each i, each player j ($1 \leq j \leq n$) chooses uniformly distributed polynomial $r_{i,j}$, and broadcasts $E(f_i * r_{i,j})$
- All players calculate and decrypt
- $$E\left(\sum_{i=1}^{n} f_i * \left(\sum_{j=1}^{n} r_{i,j}\right)\right) = E(p)$$
- Players determine the intersection multiset: if $\boxed{(x-a)^b \mid p}$ then *a* appears b times in the result