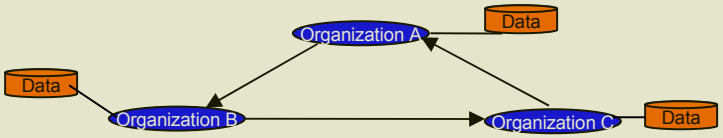
Privacy Preserving Distributed Data Mining: A Game Theoretic Approach

Kamalika Das*, Hillol Kargupta**
*University of Maryland, Baltimore County
**University of Maryland, Baltimore County & Agnik, LLC

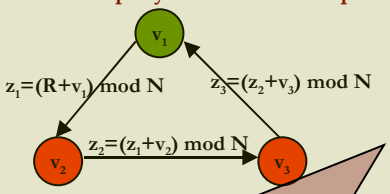


Compare, match, and analyze data from different organizations without disclosing the private data to any other party

Multi-Party PPDM as Games

- Computation Strategies: Perform or not perform local computation
- Communication Strategies: Send/Receive messages to other nodes in the network or not
- Privacy Compromise due to Collusion: Whether or not to be part of a colluding group to reveal others' private data

Illustration: 3-party Secure Sum Computation



- Each party has an array of n numbers
- Compute n sums without divulging individual numbers
- Scenario: Sequence of secure sum computations

We can arrange the sites in the following order:

Site worried about privacy

$\underbrace{v1 v2 \ldots v_{s-k-1}}_{\text{honest sites}} v_i \underbrace{v_{i+1} \ldots v_{i+k}}_{\text{colluding sites}}$

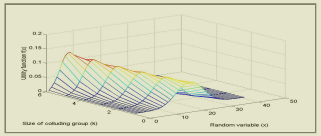
We have

$\underbrace{\sum_{j=1}^{s-k-1} v_j}_{\text{denoted by X}} + v_i = \underbrace{v - \sum_{j=i+1}^{i+k} v_j}_{\text{denoted by C}}$,

k is the number of colluders

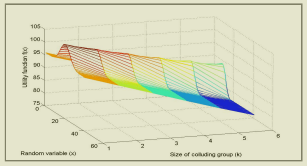
where v is the total sum of the s values.

$u_i(\{\sigma_i, \sigma_{-i}\}) = w_{im} c_{im}(Mi) + w_{ir} c_{ir}(Ri) + w_{is} c_{is}(Si) + w_{ig} c_{ig}(Gi)$,



Overall utility for classical secure sum computation. The optimal strategy takes a value of k>1.

$u_i(\{\sigma_i, \sigma_{-i}\}) = u_i(\{\sigma_i, \sigma_{-i}\}) - *\alpha_k'$, where $\alpha > 0$



Overall utility for secure sum computation with punishment strategy. The optimal strategy takes a value of k=1.

Personalized Privacy in Distributed Environment

- Privacy: a social concept
- Amount of resources vary across users
- Distributed multi-objective optimization gives parameter values for privacy model
- Mechanism design to incorporate penalty in protocol

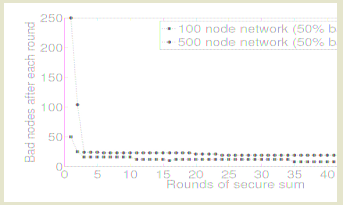
Penalty for Desired Equilibrium

- Centralized Control
  - Global Synchronization
  - Trusted Third Party
  - Auditing Device
- Distributed Control
  - Distributed Decision
  - Keep nodes in the system

Secure Sum with Penalty Algorithm

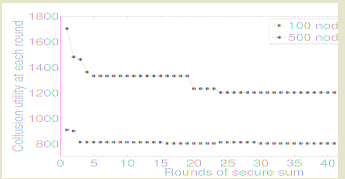
- Network has n nodes: nodes are *good (n-k)* or *bad (k)*. *Bad* nodes form one colluding group
- *Good* nodes solve local objective function based on estimated threat, desired privacy and cost constraints to decide on amount of penalty (k').
- To penalize *bad* nodes, *good* nodes split their data into αk' parts.
- *Bad* nodes turn *good* at end of sum computation if cost is too high.

WORKS FOR REPEATED GAMES



Rate of decrease of bad nodes

Collusion Utility vs. Total Cost



Applications

- Distributed privacy preserving ranking: Application in P2P web advertising
- Distributed privacy preserving feature selection: Application in P2P decision tree induction