Database Privacy Research @Stanford --- An Overview

Krishnaram Kenthapadi
kngk@cs.stanford.edu

Hector Garcia-Molina, Rajeev Motwani
G. Aggarwal, M. Bawa, C. Dwork, P. Ganesan, E-J. Goh, N. Mishra, S. Nabar, U. Srivastava, D. Thomas, Y. Xu

Private Information Management

- Individual centric privacy
- Search over access-controlled data
- Aggregates on vertically-partitioned databases
- Approximations for k-anonymity

Private Information Management

- Individual centric privacy
- Search over access-controlled data
- Aggregates on vertically-partitioned databases
- Approximations for k-anonymity
- Secure indexes
- Secure quantile computation
- …

Individual Centric Privacy

(P4P: *Paranoid* Platform for Privacy Preferences)

[ABG+04]

Managing Personal Information

- Status
  - P3P: organization declares privacy policies
  - Hippocratic DB: organization's datastore implements policies
- Critique
  - Individual must trust each organization
  - Ex of misuse: *Acxiom*, *JetBlue*, *Northwest*,…

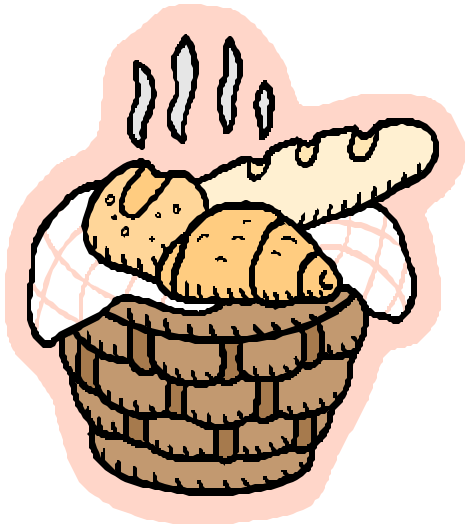
Managing Personal Information

- Thesis
  - Enable an individual to retain "control" over his/her information, *even after it has been released to an organization*
- Plan
  - Design models and mechanisms for release, acquisition, use and update of personal information (*the P4P framework*)

Example: Managing Credit Card

George

Credit Card Number



CafeDay

Control: [a] Permission, [b] No copies, [c] No Integration, …

Information Types

- Ownership
  - Individual, Organization
- Function
  - Identifier, Service Handle, Input to Predicate, Copy
- Control
  - Complete Privacy, Limited Use, No Predicate Input, No Integration, Accountable, Sharable

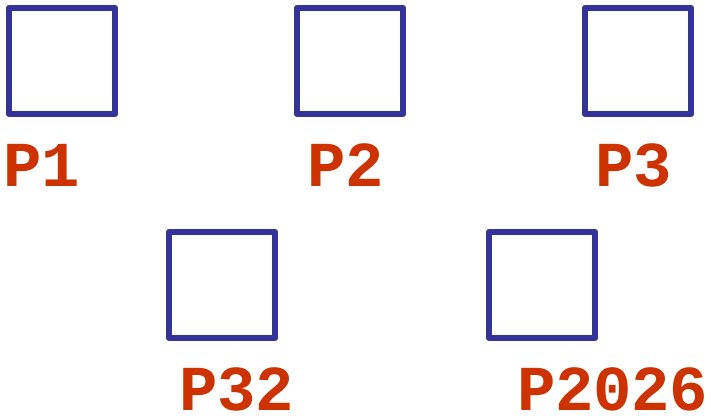
Goal: Mechanisms for each information type to enforce desired properties

Search over access-controlled data

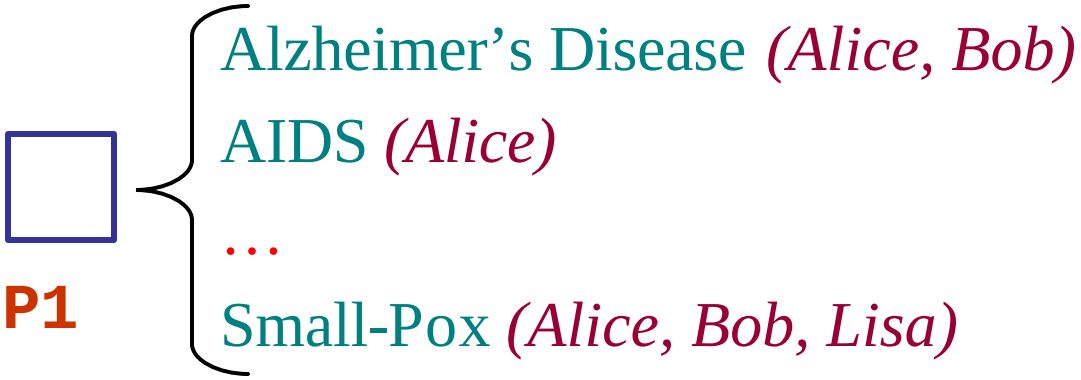
(PPI: Privacy-Preserving Indexing)

[BBA03]

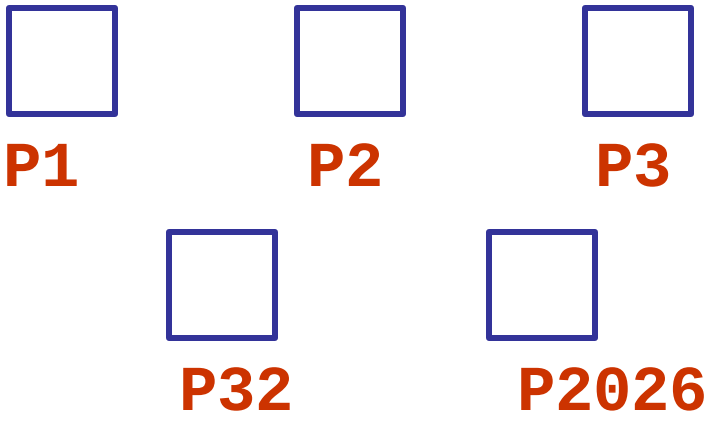
Provider



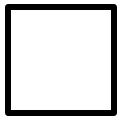
- Shares documents
- Enforces *access policy*



Searcher



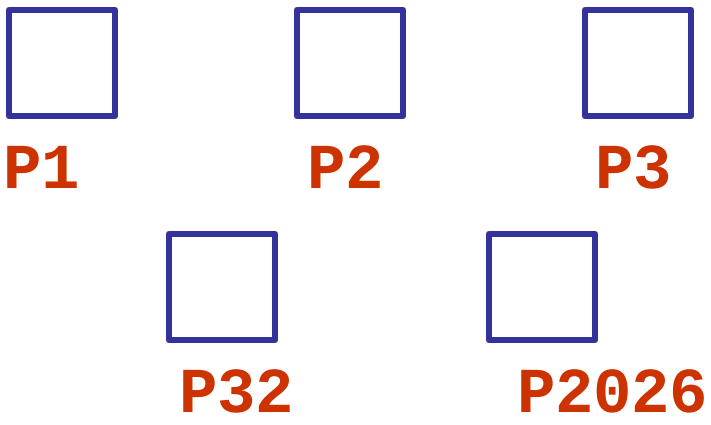
- Has an identity
- Wants documents
  - That match a keyword query Q; and
  - With appropriate access-rights



Q = “AIDS”

Alice

Search Engine



- Engine not trusted by providers:
  - Providers do not want to send documents to search engine
  - Providers do not want to reveal access-lists to search engine

How do we enable search?

Aggregates on vertically-partitioned databases [AST04]

Vertically-Partitioned Databases

Name	*State*	*Sex*
John	NJ	M
Alice	NJ	F
Mary	CA	F
Tom	CA	M

Census Bureau

Name	*Salary*	*Age*
John	120K	35
Alice	80K	22
Mary	100K	26
Tom	200K	72

Dept. of HRD

Q: Select State, Avg(Salary)
Where Census.Name = HRD.Name
From Census, HRD
Groupby State

A:

State	*Salary*
NJ	100K
CA	150K

Vertically-Partitioned Databases

- Privacy concerns
  - Databases cannot be released as-is
  - Databases can be released after data has been perturbed

Goal: Return high precision aggregate answers

Approximations for k-anonymity [AFK+04]

k-anonymity

SSN	*Name*	*Age*	*Sex*	*Zip*	*Symptom*
614	Joe	23	M	94305	Flu
615	Alice	32	F	94301	Flu
629	Jen	18	F	95102	Cold
710	Kate	22	F	95103	Rashes
840	Eve	20	F	95103	Cold

k-anonymity: suppress keys

Age	*Sex*	*Zip*	*Symptom*
23	M	94305	Flu
32	F	94301	Flu
18	F	95102	Cold
22	F	95103	Rashes
20	F	95103	Cold

k-anonymity: generalize attributes

k = 2

Age	*Sex*	*Zip*	*Symptom*
[20-35]	*	9430*	Flu
[20-35]	*	9430*	Flu
[15-25]	F	9510*	*
[15-25]	F	9510*	*
[15-25]	F	9510*	*

k-anonymity – work in progress

- [MW04]
  - NP-hardness
  - O(k log k) - approximation algorithm
- O(k) - approximation algorithm

References

http://theory.stanford.edu/~rajeev/privacy.html

- [ABG+04]: *Stanford Database Privacy Group*. Enabling privacy for the paranoids.
- [BBA03]: *Bawa*, *Bayardo Jr.*, *Agrawal*. Privacy-preserving indexing of documents on the network.
- [AST04]: *Agrawal*, *Srikant*, *Thomas*. Privacy preserving OLAP.
- [AFK+04]: *Aggarwal*, *Feder*, *Kenthapadi*, *Motwani*, *Panigrahy*, *Thomas*, *Zhu*. k-anonymity: Hardness and approximation results.
- [AMP04]: *Aggarwal*, *Mishra*, *Pinkas*. Privacy-preserving computation of the k^{th}-ranked element.
- [Goh03]: *Goh*. Secure indexes.