Hippocratic PostgreSQL

Jalaja Padma[1], Yasin N. Silva[1], Muhammad U. Arshad[2], Walid G. Aref[1]
Department of Computer Science, Department of Electrical and Computer Engineering
Purdue University, West Lafayette, Indiana

Motivation

- **Privacy preservation:**
  - An important component of information systems that deal with personal data
- Laws recognize the right of data owners to control
- Personal data is handled in compliance with its associated privacy definition
  - Hippocratic database system (HDB):
    - Has privacy as a core principle
    - Allows automated, fine-grained data disclosure at the database level
- HDBs are an answer to data owners' privacy requirements
  - With whom their data is shared with
  - Purposes for which data can be shared

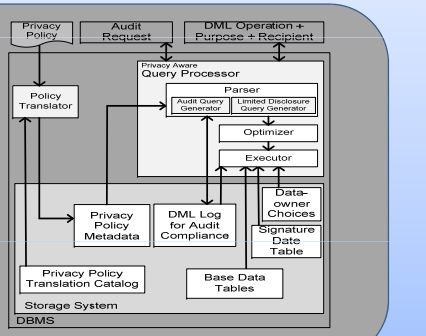
- Previous work has discussed the main guidelines and proposed an initial architecture
  - We implement HDB components as an integral part of an open-source DBMS
  - Study the problems faced to realize HDBs
- The requirements of privacy are addressed at the database level
  - → Developers can build more easily information systems that protect the privacy and ownership of data

Main Components

- Open-source privacy-aware DBMS
- Framework to test-and-integrate new privacy related components

- The project includes the implementation of components to support:
  - Limited Disclosure
  - Retention Time
  - Management of Multiple Policies and Policy Versions
  - Support for K-anonymity and Generalization Hierarchies

Architecture of Hippocratic PostgreSQL



EXAMPLE SCENARIO

PRIVACY MAPPINGS

Purpose	Recipient	P3P Type	Tabname	Colname
Treatment	Doctors	#patient.pno	options_patient	pno_option

Table 1: Stores the mappings between P3P element and database table.

POLICIES

Policy ID	Primary Table	ID_Column_Name
Medical PatientPolicy1.0.0.1	Patient	pno

Table 2: Stores the policies (could be active/inactive)

PATIENT_CHOICES

pno	pno_option	name_option	birthsex_option	SA Gen_Level	phone_option
P1	t	t	t	1	t
P2	t	f	t	1	t
P3	t	t	f	0	f
P4	t	f	f	0	f

Table 3: Stores the opt-in/out choices of each data owner. (One/more tables specific to each application

PATIENT

pno	name	birth	sex	phone	disease
P1	N1	1965	M	765 111 1111	Gastritis
P2	N2	1966	M	765 222 2222	Bronchitis
P3	N3	1967	F	765 333 3333	StomachUlcer
P4	N4	1966	F	765 444 4444	Indigestion

Table 4: Primary table that stores the data owner's data.

(a) The above tables are created and populated. Domain Generalization Hierarchy for the Sensitive Attribute 'Disease' is stored in another table.
(b) The P3P policy gets translated into the metadata tables POLICIES, PATIENT_CHOICES and a few more not given above.
(c) Query: SELECT pno, phone, disease from patient;

pno	phone	Disease
P1	765 111 1111	Gastritis
P2	765 222 2222	Bronchitis
P3		Stomach Ulcer
P4		Indigestion

Intermediate Result 1: Opt-in/out choices specified in the policy are ensured through limited disclosure component.

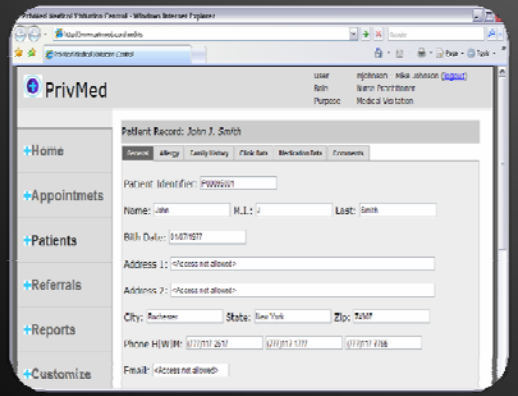
pno	phone	Disease
P1	*765 xxx xxxx*	Gastritis
P2	*765 xxx xxxx*	Bronchitis
P3		Stomach Ulcer
P4		Indigestion

Intermediate Result 2: After 2-anonymization of Intermediate Result 1.

pno	phone	Disease
P1	*765 xxx xxx*	*Stomach Infection*
P2	*765 xxx xxx*	*Respiratory Infection*
P3		Stomach Ulcer
P4		Indigestion

Final Result: After Sensitive Attribute Generalization of Intermediate Result 2.

APPLICATION



Limited Disclosure and Retention Time

- Hippocratic PostgreSQL's 'Privacy-Aware Query Processor'
- Privacy Policy Metadata tables and the data owner preferences

DML-Operation> PURPOSE <Purpose> RECIPIENT <Recipient>

- Example: Obtain information about patients and their diseases that will be shared with a research lab

SELECT P.name, P.birth, P.sex, P.disease FROM PATIENT P
PURPOSE research RECIPIENT lab

- The result is restricted to include only the columns that the combination of purpose and recipient is allowed to access according to the policy specification

Multiple Policies and Policy Versions

- Policy Translator

The Hippocratic PostgreSQL command to perform the translation:

TRANSLATE POLICY <policy-path> [FROM <language>]
[POLICY_ID <policy-id> POLICY_VERSION <policy-version>]

- Translation Process
- Privacy Policy Translation Catalog
- Privacy Policy Metadata