Web Information Management With Access Control

Serge Abiteboul[1], **Alban Galland[1],** Neoklis Polyzotis[2]

[1]INRIA Saclay and LSV ENS-Cachan
[2]University of California Santa Cruz

Organization

- Introduction
- Representing Web information as logical sentences
- Representing Web data management tasks as logical rules
- Link with the *current* Web
- Conclusion
-

Introduction

Motivating Example

- Alice : get me the pictures of my friends where I am with Bob?
- What needs to be done:
  - Find the friends of Alice (The iPhone of Alice may have that information)
  - For each friend, say Sue, find where Sue keeps her pictures (She may keep her pictures on Picasa)
  - Find the means to access Sue's pictures (Alice may obtain the private URL from a common friend)
  - Find the photos with Bob and Alice (e.g. by querying the meta-data)

Key Issue: Heterogeneity

- Alice : get me the pictures of my friends where I am with Bob?
- Heterogeneity of **hosting**: Some keep their pictures on trusted servers such as Picasa, some put in on untrusted DHT, some have them on their smartphones…
- Heterogeneity of **access-control**: Some are public, some use login-password, some use private URL, some use cryptography…
- Heterogeneity of **data description**: they may use different models of meta-data (taxonomies, ontologies…)
-

Context: Web Data Management

- **Scale**: lots of users, servers, large volume of data…
- **Heterogeneity:**
  - **Distribution** heterogeneity: Cloud (social networks), P2P (DHT, gossiping)…
  - **Security** heterogeneity: login, https, crypto, hidden URL…
  - **Terminology** heterogeneity: annotation, semantic Web, ontologies…
  - The heterogeneity keeps increasing with new systems and new applications arriving
- **Incomplete information:** inconsistencies, belief, trust…
-
- **Consequence 1: difficulty to perform data integration/management**
- **Consequence 2: impossibility to keep control over one's own data**

Thesis: Web Data = Distributed Knowledge

- Work plan
  1. Represent Web information as logical sentences
  2. Represent Web data management as logical rules
  3. Develop a system to validate these ideas
- **Global idea:** use a declarative framework to manage access control, distribution and data integration
- Motivation for the approach
  - Facilitate the design/implementation/evolution of complex systems
  - Facilitate the control/surveillance/analyze of complex systems
  - Facilitate optimization of query evaluation

Representing Web Information As Logical Sentences

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

centre de recherche
SACLAY – ÎLE-DE-FRANCE

The Information Belongs To Someone

- Each information belongs to a **principal**
  - A principal has an identity (URI) which can be **authenticated**
  - Two kinds of principal: peer and virtual principal
- **A peer**: alice-laptop, alice-iPhone, picasa, facebook, dht-peer-124, …
  - Storage and processing capabilities
  - A peer typically has a URL and can be sent query/update requests
- **A virtual principal**: users (alice, bob), groups (alice-friends, roc14), …
  - A virtual principal relies on peers for storage and processing

The Kind Of Information We Are Talking About

- **Data**: pictures, movies, music, emails, ebooks, reports
  - **Document:** picture34@alice-iPhone = {fileName:picture34.jpg, date:09/12/2009,…}
  - **Collection:** pictures@alice += picture34@alice-iPhone
- **Localization**: bookmarks, knowledge such as “Alice puts her pictures in Picasa’’
  - where@alice(picture37) += picasa/alice
- **Access**: login/password, access rights on servers
  - **Access right**: owner@picasa/alice += alice
  - **Access secret** : ownSecret@picasa/alice = {login:alice, password:HG-FT23)
- And more (services, ontologies, beliefs, …)
  -

WebdamExchange Statement: Authenticated Knowledge

- Statement: **alice-laptop** states picture37@alice = {….} **requester bob at 12:30, 10/08/2009**
  - The performer (alice-laptop) did the update
  - The requester (bob) requested the update
  - The time is the time of the performer, there is **no global clock**
- The content is possibly encrypted**:**
  - alice-laptop states picture37@alice = {….} **protected for reader@alice** requester bob at 12:30, 10/08/2009
- Annotated with a proof that the update was valid
  - E.g. an RSA signature of the statement using the update secret key of the owner (Alice)

WebdamExchange External Knowledge: Communication Between Peers

- External knowledge: **alice-laptop says (**alice-laptop states …**) to sue-iphone**
  - The performer (alice-laptop) sent the message to the receiver (sue-iphone)
  - External knowledge is **authenticated** by the performer.
- We maintain **a trusted trace of the provenance** when knowledge is forwarded:
  - **sue-iphone says (**alice-laptop says (alice-laptop states …) to sue-iphone**) to bob-iphone**

Representing Web Data Management As Logical Rules

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

centre de recherche
SACLAY – ÎLE-DE-FRANCE

WebdamExchange On Top Of Webdamlog

- WebdamExchange statements can be easily translated into datalog facts.
- Datalog enables reasoning on localization, exchange, and update information
  - Example: use logical reasoning among peers to locate the pictures of Alice's friends in which she appears with Bob
- This motivates **Webdamlog**, a rule-based language for web data management, presented in PODS'11.

Webdamlog Example

Alice: get me the pictures of my friends where I am with Bob

- result@alice-iphone($photo) :- friends@alice-iphone($X),
  findPhotos@alice-iphone($X, $R, $P),
  $R@$P($X, $Photo, $Meta),
  contains@$P($Meta, “Alice”) ,
  contains@$P($Meta, “Bob”)
  - findPhotos@alice-iphone($X, photos, picasa) :- member@picasa($X)
  - friends@alice-iphone(Sue)
  - member@picasa(Sue)

Link With The *Current* Web

Authentication

- The model is based on an abstract notion of secret (and possibly hint)
- Different means of enforcing access control
  - RSA cryptography (or other asymmetric authentication)
  - Login/password (or more refined schemes based on symmetric authentication)
  - URL-based
- The main peers of the system use RSA Cryptography and communicate with the current Web using wrappers
-

Distribution

- The model is based on an abstract notion of localization and does not prescribe any particular architecture for distribution
- Different means of distribution depending of the rules
  - Centralized server
  - Gossiping
  - DHT
- One may combine different authentication and distribution policies to support existing Web applications.
  - Example of ICDE 2011 demo: WebdamExchange peer communicating with Facebook, a blog and a trusted pastry DHT

Conclusion

Summary Of WebdamExchange

- All the information forms a trusted knowledge base
- Each peer manages some portion of the knowledge base using a rule-based language
- The current Web is mapped to the system using special policies and wrappers
-
- The system has been demonstrated at ICDE 2011
- The WebdamLog language will be presented at PODS 2011

Thanks!

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SACLAY – ÎLE-DE-FRANCE