Privacy Assistants

Dave Raggett <dsr@w3.org>

Helping users to manage the information they disclose to websites

Disclaimer: Ideas describing work in progress

Why?

- Websites collect all kinds of personal data, potentially leading to
  - Inappropriate collection of personal data
  - Inappropriate use of personal data
  - Aggregation of personal data across sites
- This may be subject to data protection laws
  - Varies by jurisdiction, but the Web is world-wide
- You need help in asserting your rights!
  - How to determine what personal data a given website holds on you?
  - How to correct errors in the personal data they hold?
  - How to determine what their privacy policies are?

Credential-Based Access Control

- Credential as an attestation by a trusted party as to properties of the bearer
  - *X says that I am over 21 and a UK resident*
- Cryptographic credentials
  - Can provide proof of properties without directly revealing your identity
    - Can even reveal selected subset of properties in a credential
  - Credentials increase privacy by reducing the kinds of personal data that need to be collected
  - Facilitate use of anonymous or partial identities
- Users control what credentials they release

Privacy Assistant

- Firefox add-on that tracks what personal data you've released
  - Which sites have I given my email address to?
  - What personal data have I released to example.com?
- Support for PrimeLife project ideas
  - Control over privacy preferences and credentials
  - What credentials does this site want from me?
  - What purposes will my personal data be used for, and for how long will it be retained?
  - Viewing notifications from data controllers

Privacy Policies

- Provided by website in a machine interpretable XML format
  - Plus pointer to Lawyer-readable plain text equivalent
  - Neither are intelligible to ordinary people
- How to present the policy to the end-user?
  - Without becoming a nuisance!
    - Only bother user when user is expected to do something
    - Otherwise allow user to view privacy policy via
      - Clicking on icon on browser status bar
      - Or selecting a menu item on menu bar
  - Automatic generation of plain language descriptions
  - Plus easy to understand icons

Independent Outlook

- Websites have a business model to protect
  - They will slant things to suit their own interests
- A privacy assistant...
  - Is independent of the websites you deal with
  - Will use plain language for describing policies
  - Use consistent wording across different websites
  - Could consult 3rd party for independent advice
    - Trusted authorities
    - Wisdom of crowds via reputation system

Natural Language Generation

- Manually create corpus of plain language texts and the XML policies they should be generated from
  - XML policy as representation of semantics
    - And/Or tree for what needs to be disclosed
      - Which properties in a credential will be revealed to website
    - Details of purposes and retention periods
    - What kinds of notifications are available
  - Use examples to “train” realizer
  - This is a limited domain, which makes it easier
- Generation is a multi-stage process
  - First stage is text planning
    - Use templates for generating candidate phrases
      - Use of pronouns, connectors, conjunctions...
  - Second stage is sentence generation
    - Instantiate words with appropriate morphology

3rd Party Privacy Assistants

- Keeping your privacy tracking data in one computer is risky and restrictive
  - What if you drop it, or it breaks or is stolen?
  - What happens when you want to upgrade the OS?
  - What if you want to use one desktop at home, another at work, an iPhone on the train, and an Internet Cafe when on vacation?
- 3rd Party Privacy Assistant can solve all of those
  - As well has helping with trust relationships
    - Which websites have trustworthy privacy practices?
    - Avoid weaknesses of OpenID/email addresses as global ids
    - Support 24x7 authorizations for web 'bots acting on your behalf
  - Work with any browser without needing an add-on

How to obtain Policies?

- One idea is to use HTTP Link header
  - Corresponds to HTML <link> element
    - Describe relation between requested resource and some other resource
      - `Link: <http://www.example.com/Policy>; rel="PrivacyPolicy"`
  - Proposed in HTTP 1.1, but removed in RFC2616
    - Now back as draft-nottingham-http-link-header
- Could be indexed by search engines
  - Search results could indicate which links are privacy enabled

Credentials and HTTP

- HTTP defines headers for access control
- Server sends 401 Unauthorized response
  - Plus info in WWW-Authenticate header
- Client resends request with requested info
  - Authorization header
- Further work needed for defining how to use these headers with credentials
  - Could embed XML format …

Avoiding wasted round trips

- Web page has many subsidiary resources
  - Style sheets, scripts, images, …
  - Avoiding 401 Unauthorized on each resource
- Get policy to state which resources it applies to
  - Some kind of wild-card patterns
  - Some resources are from offsite
    - Load balancing with Akamai
- Learning from P3P

Policy Negotiation

- PrimeLife assumes server proposes policy and client accepts or declines
  - Issue: how does client bind personal data to policy?
- More flexible approach allows policies to be sent in both HTTP requests and responses
  - Client could send policy which broadens server's
    - Expanded set of purposes, or longer retention period
  - Or client could send a more restrictive proposal
  - If server doesn't like client's proposal it could return 412 Precondition Failed response along with server's (revised) proposal
  - Not clear if this flexibility is really justified

Reality or Illusion?

- Do websites and end users really want all this?
  - End users focus on immediate benefits and downplay privacy risks
  - Lack of interest in setting personal preferences
- Perhaps we should instead focus on providing a strong legal framework to discourage abuses
- Users would still need a means to track their personal data and make corrections to errors
  - Privacy assistant is still valuable
    - Privacy preferences could be set by 3rd party