Towards A Privacy-Aware, Trusted Web

Position paper for the on Privacy for Advanced Web APIs. [1]

Authors: Henry Story, Michael Hausenblas, Nathan Rixham, Bruno Harbulot

INTRODUCTION

A fundamental requirement for enabling privacy on the Web is that publishers need to be able to control who has access to their information resources. Once such a decision is made, it should follow that in actual fact only those agents have access. This applies not just to documents but recursively to identity information too. Therefore, we will argue that for this to work at web scale, it requires global distributed identity, authentication, and access control.

This paper introduces Web IDs as global identifiers and FOAF+SSL as an authentication protocol that seamlessly integrates into the current architecture of the world wide web with a low adoption cost, whilst enabling privacy, access control, information accountability, distributed social networks whilst putting each agent in full control of their own data.

A TYPICAL DAY IN THE PRIVACY-AWARE WEB

Sue, a 40 year-old with two children, hosts her personal and family profile on her ADSL box. There, she also hosts her family and party photos, blogs, home videos and work content, as she is self employed. Sue is in full control of the content on her server.

Sue's family is spread throughout the world, she does not know how most of them host their profiles. Some use large service providers, others host their profiles directly off their cell phone, yet others use technology she has never heard of. All she - or more precisely her Server Agent - knows, are the Web IDs of her close family.

Her direct family get access to the family pictures, videos and blogs in addition to public content, but not to her friend list or work related content.

The Server Agent, itself identified as a robot under Sue's responsibility, crawls the profiles of her family from time to time in order to keep up to date with new family members, and allow the extended family access to the less personal parts of her family profile.

Today, Sue's Agent leaves her a message informing her of the birth of a new family member from her brother's wife's sister Teresa in Brazil. She drags Teresa's icon on her mail box, and a web mail client opens up. (Web mail feels like normal mail but uses a FOAF+SSL enhanced Atom Publishing protocol). When she has finished editing the congratulations mail, her mail client just POSTs the content to her local personal collection, creating a new entry that is only visible to Sue and her close family. Her server then pings Teresa's server agent informing it of the new entry. The entry is protected by HTTPS, strong cryptography across the international frontiers, and no intermediary stalker can peep in on their conversation; especially important as kidnapping children has grown in some quarters of the world.

This morning Sue is pinged by a new person named Mr Smith who is related to a medium sized company she has not worked for before. Her server agent followed the linked data network and proves that Smith does indeed work for that company, and also shows a summary of the publicly available information about both the company and Mr Smith's job description. The contact seems genuine enough so Sue adds Smith to her business contacts, and her Agent fetches the web mail (atom entry)

from their site.

The request is for a job to put together a design for the company's new logo, and if she accepts will get paid €1000 immediately, to come to the company headquarters. The web mail contains a link to a contract published by a Notary (verified by the agent looking up the Notary ontology that points to the specific country Notary lookup). She connects securely to that resource authenticating herself thereby, reads the contract, and presses the accept button.

The company approves of the payment by connecting to her bank (information available from her Web ID) with the Company's Web ID referencing the contract. A few minutes later the money is in her bank account, and her agent is working with the rail and bus network on her travel arrangements.

PRINCIPLES OF WEB PRIVACY

For privacy on the web to be meaningful a number of criteria need to be fulfilled:

FREEDOM OF EXPRESSION

An agent should be able to control statements he makes (including statements about himself) and decide who can view them. Without this, privacy would always be compromised by being viewable or controllable by a third party. This does not mean that one is not to be held accountable for what one says.

IDENTITY HAS TO BE CONTROLLABLE BY THE IDENTIFIED ENTITY

an architecture where identity MUST be controlled by a third party would give that party control in what an agent can say and who that agent is speaking to. This does not mean that all identities should be controllable by the user. An employee working for a company could have his work identity controlled by the company for whom he speaks. Similarly a state can create identities for it's citizens that constitute it. But states or companies do not have ontological priority over individuals or robots, which should also be able to have their own identity and control it. From this it follows that an agent should be able to have multiple identities. This is good because one cannot always speak with the same voice to every one, neither are we in control of our life enough to be able to easily make the correct identity decisions early on.

IDENTIFIER HAS TO BE UNIVERSAL AND LINKABLE

Just as a phone number, email address, or web page can be identified globally so an Agent has to be identified globally without ambiguity. It should be easy to coin such IDs with the certainty that there will not be name clashes. Or else one may accidentally give the wrong person access. URIs are thus the appropriate type of identifier for a Web ID.

This is a web architectural principle for being able to create a social web, where each agent can control their identity. The URI should be dereferenceable, so that one can meaningfully link to it, and find more information about the agent at that URI. This does not mean that one can necessarily know a lot about someone from such a handle - one may indeed be able to glean nothing from it. An email address, for example, tells us very little about the user, but it is enough to communicate consistently over time.

COMMUNICATION SECURITY

Not only should resources be access controlled, but the communication channels should be secure and only visible to the identified parties.

ADDITIONAL USER INTERFACE REQUIREMENTS:

IT SHOULD BE EASY TO USE

The WWW was enabled by graphical user interfaces allowing users to click on a link and reach another web page, without ever having to remember the URLs for each page. So selecting an identity should be built into the client and be as easy to select as a point and click.

THE USER SHOULD KNOW WHAT IDENTITY HE IS USING AT ANY TIME

When interacting with another agent. When connecting to a web site (web server agent), the user agent should display to the user information about his revealed identity.

These are fundamental principles of privacy consistent with web architecture. To this one should add a requirement of being unencumbered by patents, allowing the maximum number of people to join.

A COMPLIANT PROTOCOL: FOAF+SSL

The FOAF+SSL protocol [2] is consistent with all of the aforementioned principles of web privacy.

FOAF+SSL has a very low adoption cost, indeed it already works in most current browsers, a number of mobile devices, and can easily be supported by any client that supports HTTPS. On the server side support has already been implemented for Apache, and the protocol has been implemented in a number of the most common programming languages.

Community support is ever growing and new projects to implement FOAF+SSL are appearing frequently.

FOAF+SSL IS BUILT ON A NUMBER OF OPEN AND WIDELY DEPLOYED WEB STANDARDS

- URL's and URIs: Each Agent/Person is identified by one ore more Web IDs [3], which is a globally unique dereferenceable URL as used in the Semantic Web.
- SSL v3 (Secure Socket Layer) and its successor TLS (Transport Level Security) [4] guarantee security of communication as well as Identification, but with a twist making this well used technology widely deployable.
    - X.509 certificates which tie an Agent's Identity to a Public Key and a Web ID
    - Public Key Cryptography
- HTTP and HTTPS as the communication protocol.
- RDF (Resource Description Framework) [5] providing the semantics for the descriptions
- Linked Data pattern of deployment of documents, allowing agents to interlink into a global web of trust.
- A number of ontologies such as:
    - required: the cert ontology [6], and public key ontologies (only RSA [7] published at the moment)
    - an Access Control Ontology [8]
    - FOAF [9] for basic description of people, relations between them
    - Any other semantic web ontology can be made to work seamlessly with it
- Client certificate selection is widely deployed in browsers [10]

All of the technologies FOAF+SSL utilizes are both widely deployed and believed to be without patent encumbrances.

FOAF+SSL uses TLS as-is, and adds a Web ID to an X.509 certificate, made possible by the subject

alternative name extension. It then removes the requirement for Certificate Authorities by a change of perspective on what client authentication really means.

The FOAF+SSL protocol was presented at the 2008 Workshop on the future of Social Networking (see FOAF & SSL: creating a global decentralised authentication protocols).

FOAF+SSL SATISFIES THE ABOVE REQUIREMENTS IN THE FOLLOWING WAY:

- Freedom of expression: being based on the web, it inherits the freedom of expression that comes with it. Anyone can buy a domain name, set up an HTTP server, and publish any document they wish. FOAF+SSL enhances the web of documents, by making it easier for people to specify who can view their documents, in a dynamic decentralised way: one can for example specify that only friends of friends can view or edit a resource, where the decisions of who one's friends friends are, clearly cannot be within one's control. HTTPS guarantees the security of the conversation.
- Control of identity: A Web ID is related to a profile document published on the web. Therefore, publishers can also choose which of their properties are seen by whom. It is also possible for someone else to host a profile, if the user trusts that entity, removing the potentially identifying link between the domain owner and the identified agent.
    - A Web ID is a URI, and so is globally addressable
    - Being a specific form of URI - a URL - it is globally linkable
    - An agent can create multiple profiles, each hosting a Web Id. These may but need not be linked, using relations such as owl:sameAs.
    - Web IDs are very easy to create and to use. [11]
- By a simple and necessary user interface improvement to browsers it would be possible to know which identity they were using when identified at a web site. [12]

NEXT STEPS WITHIN W3C

We understand that there are several interdependent areas we need to address, in order to realise the overall vision. During the workshop, and in conjunction with other proposals, we would like to discuss some of the next steps that the W3C could lead, in this context. In particular, we suggest that the following topics be addressed in an upcoming Incubator Group or Working Group:

- Carry out a state-of-the-art review concerning privacy and the Web, taking into account experiences made within W3C, such as P3P, from the TAG, the Web Apps WG, etc.
- Gather requirements and use cases with input from the W3C Social XG, the Geolocation WG, the Device API and Policy WG, as well as the Web App WG concerning privacy and accountability.
- Compile a good-practice catalog for privacy-aware Networking behaviour (such as robots.txt or an RDF equivalent for data sharing/interlinking).
- research the relation of Web IDs to other identification technologies such as OpenId, SAML ids, XRIs, ...
- Standardise the FOAF+SSL protocol and associated vocabularies.
- Develop additional relations/protocols to allow adding remote agents to groups of users with certain rights, such as described in the Sketch of a restful photo printing Service.