Technology & Society



Whose Name is it Anyway?

Decentralized Identity Systems on the Web

Daniel J.Weitzner • MIT Computer Science and Artificial Intelligence Laboratory

new form of personal identity is emerging on the Web. Decentralized identification protocols are a departure from traditional distributed authentication approaches developed for the Internet. From a technical perspective, they're quite similar to distributed systems based on public-key infrastructures or federated identity systems, such as that proposed by the Liberty Alliance or Microsoft's Passport. What distinguishes the new decentralized approach is its use of URIs as the underlying identifier. The Web – using URIs as the basis for a global hypertext system of documents - set off an unprecedented explosion of human communication and knowledge sharing. So, too, could these new decentralized identity systems potentially augment the Web to let us reliably and scalably communicate with each other about our identities in a more trustworthy manner.

Being Known on the Web as your URI

Why should we even consider using the somewhat clumsy-looking URI (mine is relatively short at http://www.w3.org/People/Weitzner.html) for a personal identifier? Consider what the Web achieved by labeling every document with an identifier that's at once globally unique and individually "mint-able" without needing permission from any central authority. Among other benefits, the URI on every document lets search engines mine links and point users to the documents that are most likely to match their queries. Given the number of documents on the Web and how many different queries are possible, it's pretty remarkable that any search engine can return results that are even faintly reliable. When it comes to identity, we're desperately in need of better reliability that can scale to Web size. But can we expect URIbased identity architectures to help with problems such as email spam, blog-comment spam, and other Web applications that require knowing that people are who they say they are?

Whether this new breed of identity systems works or not will depend in large part on whether we design and deploy them in a way that heeds the Web's basic architectural principles. Starting with URIs is an important first step but won't be sufficient. If we expect hundreds of millions of people to represent their identity in this decentralized style, the software tools must be very simple to use (as early HTML was) and based on a suite of technical standards that meet basic openness requirements, such as royalty-free patent licenses that allow full open source implementation. Most important, we must avoid centralized bottlenecks, such as new registration services. The beauty of using URIs as personal identifiers is that everyone can create identities as they see fit, without needing specific permission beyond simply having a DNS name (or using the services of someone who already has one). A truly decentralized, Web-like design will avoid creating new centralized registration requirements.

Just as the Web brought profound changes to the social dynamics of information sharing resulting in real pressure on existing legal rules such as copyright – we can expect that decentralized identity systems, by lowering the boundaries to creating, expressing, and analyzing identity information, will raise fundamental privacy questions. Although decentralized identity systems give individuals more control over how they express their own identities, they also open the field for some to attach their own views to other people's identities, or to mine links among them. We probably can't design away these problems, but we can be sure that decentralized identity systems keep basic privacy requirements in mind and provide individuals and society tools to protect privacy principles in the face of personal information's inevitably increasing transparency. (I introduce this discussion on Web-style identity systems in an earlier column, "In Search of Manageable Identity Systems."¹

Identity Approaches, Old and New

On the Internet, we generally authenticate ourselves with username-password pairs, cryptographic keys, or even hardware tokens such as smart cards. Although these accounts might in some sense belong to users, such systems require them to rely on someone else to control and manage the identifiers and associated authentication mechanisms. When we sign into a Web site or create a new email address, we might designate the name under which we're known, but the thirdparty service controls and effectively owns that name. Each system manages its own namespace of identities. Despite many efforts at interoperability (or federation) between different name spaces, they've never successfully scaled up to anything even close to Web size. Third-party identification systems certainly have great value. Some originally offline ones have been adapted quite effectively for online use for example, we can use our credit cards (identifiers created and controlled by Visa, Mastercard, American Express, and others), bank account identities, or even drivers licenses, passports, and national identity cards online as well as off.

Yet when it comes to representing our identity on the Web, we long for a unified approach to authentication that can simplify the tangle of different usernames and passwords that we try to maintain in our heads or on actual or virtual sticky notes. Efforts to solve this so-called "single-sign-on" problem have yet to achieve any noticeable deployment.

The absence of a usable, widely deployed identity system for the Web has real social costs, beyond the sim-



Figure 1.The OpenID authentication interface on a blog. This Web form allows users to authenticate themselves to the blog either using an OpenID identifier or a username–password pair.

ple inconvenience of having to remember several username-password pairs (or the security risk when you just use one pair for all transactions on all sites). Our inability to distinguish between individuals on the Web whom we trust and those we know nothing about or have reason to distrust is now hampering the growth and utility of blogs, Web-scale social networking, wikis, and various other forms of interpersonal interaction. These new information-exchange modes, often lumped under the technical rubric Web 2.0, are all designed to let users easily generate and publish their own content, and combine information they create with that of others. The social benefits of these extremely open, globally interconnected online fora are clear, but their realization is limited when content that we have reason to trust, from people whom we trust, is intermixed with untrustworthy information from unknown or untrustworthy users. In the end, the lack of trustworthy identity infrastructure on the Web almost effectively eliminates the possibility of personal accountability in an environment in which it's sorely needed. What we need is a way to manage identity and share trust

information across the Web based on a common set of identifiers.

Into this breach steps OpenID (http://openid.net/), an emerging set of specifications from a very creative group of Web developers. The technical details of its operation are available on the Web site. From a functional standpoint, the OpenID protocol lets users securely identify themselves to any Web site using a URI that the user proves he or she can control. So, when I sign into the blog shown in Figure 1, I use my homepage URI as my OpenID login. The blog site, called the relying party, checks my URI and finds a pointer to my identity provider. Through a series of http methods, the relying party asks my identity provider to get a passphrase from me (I only need a single passphrase for all the sites I visit with that OpenID identifier) to prove that I created the URI I've given. If I provide the correct passphrase, then the relying party knows that my identity provider believes I created and control the Web page my URI identifies. This provides a persistent, traceable identifier that the relying party can associate with me. If the site determines that the person (or bot) associated with the URI is a spammer, or isn't trustworthy for some other reason, then access might be blocked.

OpenID Enhances Social Networks

OpenID alone won't close the trust gap on the Web, but it will provide an important architectural foundation on which we can build a more trustworthy, reliable view of people's online identities. Consider blog-comment spam. Blogs promote an interactive exchange of ideas among people on the Web, but, sadly, abusive advertisers and other spammers have found that they can use the blogs' comment facility as a free platform for advertisements - never mind that bloggers don't usually intend to operate free advertising bulletin boards. Some comment spammers also use comments to place links to their own Web sites in an effort to artificially increase their pageranks on search engines. Requiring blog commenters to sign in to a blog before commenting is neither effective nor scalable. Malicious comment bots go through the motions of creating a username and password with a blog server and appear to be real people. Expecting bloggers to manually screen every subscription request would considerably reduce the amount of useful commenting, defeating the blog's original purpose.

Rather, a better way to address blog spam and other Web trust challenges, as suggested by blog software hackers and Semantic Web developers, would combine OpenID's lightidentity weight features with Semantic Web techniques to let users describe their social relationships in machine-readable format that can be easily reasoned over in order to explore individual trust relationships. For my own blog, I'd like to open up comments to anyone I know and trust, plus those trusted by the people I trust. Using OpenIDs to identify those who want to comment on my

blog, I can assess whether those comments should be posted based on whether the person with a given OpenID appears on my list of trusted individuals or on those individuals' trusted lists. Using the Semantic Web data format called friend-of-a-friend (FOAF), I can publish the IDs of the people I know and follow those links to their FOAF files. Then, anyone can use that social-network data to make inferences about people's trustworthiness. Figure 2 shows a variety of trust rules that I can implement using FOAF with a Web identifier system such as OpenID. The blue circles show people who I know directly. I could set up my blog to allow posts only from those people I know (second degree relationships), with the result that those inside the blue dotted circle would be allowed to post.

The second degree rule might be too restrictive, so I could modify the rule to allow anyone who I know plus those who are known to those I know to post. With that rule, all of those inside the purple dashed circle can post. Finally, if that's still too limiting, I can add another rule allowing anyone who is a member of an organization I trust to post as well.

The critical role that OpenID identifiers play here is to provide a readily visible, easily referenced name for each person on the Web, while allowing for the fact that we might all be associated with any number of identities. The OpenID becomes a hook on which to hang an arbitrary number of assertions from an unlimited number of people. From a user standpoint, I know that if I show up at a site that lets me authenticate myself with an OpenID URI, I can do so using whatever identity I choose, while still relying on the fact that the site can connect me to numerous assertions that will, hopefully, allow others to recognize me as trustworthy. From an implementation and administrative standpoint, OpenID gives service

providers an identifier for everyone who visits that site without requiring the site or the user to either create a new identity token or negotiate technical and policy agreements to use tokens that some other identity system maintains. The major distributed identity systems such as Liberty and Passport could certainly handle this job, but would incur considerable coordination costs to get hundreds of thousands of blogs to all deploy the relatively complex set of Web services protocols required for such federated identity networks. Because OpenID is based on http and URI standards already deployed through the Web, it would be much easier to achieve broad deployment.

User-Created vs. Decentralized or Distributed ID Systems

The socially significant difference between decentralized identity systems such as OpenID and all the other systems we've mentioned (including offline and online ones) becomes truly apparent when we examine allocating responsibilities and capabilities with respect to identifiers. Traditional distributed identity systems – such as credit cards and most current implementations of online federated identity systems – tend to

- exercise a fair amount of control over how users create identities;
- invest a lot of centralized effort in securing and vouching for the identity token;
- work hard to assure the quality of the information asserted about the person whose identity they're managing; and
- place considerable restrictions on who can use those identities.

These are large systems that affect numerous people, but the number of users and types of uses to which the identity and associated attributes are put are all highly constrained.

By contrast, decentralized systems such as OpenID give complete control over creating the identifier to the user (who just mints a URI). The identity provider invests a comparatively small amount of resources in vouching for the association between the identifier and the user presenting it at any given site. Notably, the identity provider is virtually uninvolved with the nature or quality of assertions others make about the identity holder. My identity provider helps any relying party be sure that I'm the person actually presenting my ID in a given place, but any assertions made about that identity are generally entirely beyond the provider's competence and control.

This strict separation of concerns between identity verification and vouching for assertions about any given identity is an architectural quality that seems essential if such systems are going to scale to Web size. Yet, we can also imagine that this would render the data in the system useless in terms of quality. With no one institution or individual responsible for assuring the integrity and reliability of assertions made about any individual, we might wonder how and why anyone would decide to trust the information in such an unruly, unregulated environment. Yet, this lack of centralized control should be familiar - it's the way the Web works. Many doubted that the Web would work at all, but clearly, its success can teach us something.

Heed Lessons from the Web

If we want to develop decentralized, Web-scale identity systems, we can learn some simple social and technical lessons from Web architecture.

First, keep it simple: HTML and http were widely (and correctly) regarded as among the least sophisticated hypertext technologies available when the Web was designed. That Web



Figure 2. Friend-of-a-Friend (FOAF) network. Relationships described in FOAF files can be used to determine trust relationships for Web applications such as blogs.

standards were easy to implement in browsers and servers was key to the technology's rapid profusion, and the ease of HTML authoring was essential to the rapid appearance of what are now billions of Web pages.

Next, stick to nonproprietary standards with royalty-free access to all necessary patent rights - the technology must not only be easy to implement, it must be free of patent and other intellectual property barriers. Otherwise, potential OpenID adopters will get sidetracked in licensing negotiations or end up out of the market all together. When it became clear in the late 1990s that Web standards risked becoming bogged down in patent traps, the W3C declared that it would only standardize technology that everyone can implement and use on a royalty-free basis. The OpenID community should be sure to follow this example.

Finally, avoid centralized registries. The Web's hallmark is that anyone can create a Web page, without payment, permission, or registration with any centralized entity. (Of course the Internet depends on the Domain Name System, which ICANN manages, but this single registry should be enough.) Some have proposed adding a new kind of URI, called an XRI (www.oasisopen.org/committees/tc_home.php?wg _abbrev=xri), to the OpenID standards. XRIs require registration with a central authority, and the technology comes with a great deal of patent licensing uncertainty. There's no reason to have to register XRIs with a new ICANNlike entity when we already have enough addressing services on the Internet to support other kinds of URIs. Such bottlenecks should be avoided at all costs.

Privacy Risks

A final hurdle that OpenID-style identity systems will have to face is the challenge to individual privacy rights. By definition, OpenIDs are public declarations of at least some personally identifiable information. When linked with social-network tools, such as FOAF, the resulting information architecture will reveal quite a lot about us as individuals, who we associate with, and what we do online. In fact, this system's very purpose is to reveal more about users to facilitate trust.

We must eventually face the fact that links on the Web, especially if we add personal information to them, will reveal a lot. Protecting privacy in the face of this increased transparency is a challenge that requires new strategies. Kim Cameron, who has coalesced much of the Web community's thinking on identity systems, has suggested in his "laws of identity" that, "technical identity systems must only reveal information identifying a user with the user's consent" (see www.identityblog.com/?page id=352/#lawsofiden t opic3). Minimization is a sound privacy principle, but it might not do the job here. When we use OpenID systems, we'll reveal a lot about ourselves. but that doesn't necessarily mean we want it available for all conceivable purposes. I've written elsewhere that privacy protection on complex information systems such as the Web must occur through increased reliance on information usage rules and mechanisms to assure accountability to such rules.² We won't be able to protect privacy just by limiting information access or disclosure simply because we have reasons to disclose so much information today. So, it's not clear how granting or withholding consent to release the identifier will contribute to privacy protection when what is revealing is all the information associated with the identifier scattered around the Web, not the identifier itself.What's more, our tools for inferring sensitive, personal facts by analyzing our behavior in an increasingly transparent world are becoming dramatically more powerful.

n the first decade of the Web's development, most efforts to provide better identity and trust management proceeded on the assumption that the right application of Internet security techniques (PKIs, PGP Web of Trust, and so on) would eventually lead it to be more secure and trustworthy. Much progress has occurred in enterprise security for the Internet during those 10 years. Yet individual Web users and those who participate in new forms of social interaction on the Web have ended up with very poor security and increasing concerns about how to assess the trustworthiness of both people and services online. As the growing interest in approaches such as OpenID shows, we have a chance to make real progress toward a more trustworthy Web experience if we can just take seriously the architectural and social lessons that helped the Web itself to grow as richly as it has. Identity systems based on decentralized architectures, URIs, and easy-to-implement, royalty-free standards have a real chance of bringing personal identity to the Web as a first class object. We have privacy challenges to face to be sure, but will be well-rewarded with a more trust-based, humane environment in which to share information and create new services. G

Acknowledgments

I thank Phil Hallam-Baker, Tim Berners-Lee, Dan Connolly, Jim Hendler, Kyle Young, Mary-Ellen Zurko and many other colleagues for the opportunity to talk through many of the ideas that appear here. My acknowledgment of their wise council doesn't mean that they agree with any or all of what is said here, of course. I've also learned a lot from the efforts of those in the OpenID/FOAF/Semantic Web community who've been leading the way in exploring new identity architectures.

References

 D. Weitzner, "In Search of Manageable Identity Systems," *IEEE Internet Computing*, vol. 10, no. 6, 2006, pp. 84–86.

- D. Weitzner et al., Transparent Accountable Data Mining: New Strategies for Privacy Protection, tech. report MIT-CSAIL-TR-2006-007, MIT CSAIL, Jan. 2006; http://dig. csail.mit.edu/2006/01/tami-privacy-strategies -aaai.pdf.
- Daniel J. Weitzner is principal research scientist at the MIT Computer Science and Artificial Intelligence Lab, co-director of the MIT Decentralized Information Group, and the Technology & Society Policy Director of the W3C. You can find his blog at http://people. w3.org/~djweitzner/blog/. Contact him via www.w3.org/People/Weitzner.html.