W3C Workshop on Usability and Transparency of Web Authentication

Observations on Distributed Authentication and Authorization in Web Applications

Eve Maler
eve.maler@sun.com
Sun Microsystems, Inc.
15–16 March 2006

Abstract: A number of modern-day technologies, such as the LID, OpenID, and YADIS community efforts; the SXIP protocols; Microsoft's Infocard technology; and SAML, Liberty, and Shibboleth tackle various aspects of distributing user identity information (identifiers, credentials, attributes, and entitlements) across Internet domain boundaries. To varying degrees, they attempt to tackle two complementary goals: allowing web applications and services to use the identity information for personalization and access control, and allowing users to shape their digital identity and control the sharing of personal information with those same web applications and services. At the same time, many of these solutions proceed from different assumptions or embody special usage scenarios (for example, making different choices around the relatively priorities of security, ease of integration, privacy, and so on). This paper examines some of the trends and tensions in the space.

Agreeing on common concepts is an urgent priority

In the Internet security field, authoritative and stable sources exist for terminology [SecGloss]. And there have been more recent advances in defining precise terminology for the identity field as it relates to security concepts, of which [AnonTerm] is an excellent example. However, a number of terms that are in wide use in this growing field have little agreement about their meaning. This problem is likely due both to the vigorous intellectual debate around identity that is currently taking place, and to the marketing-speak that growing fields tend to attract as a matter of course.

Terminology questions matter to the people whose identity information is being shared and manipulated precisely because these people are, at various times, exposed to these processes. Users have little chance of becoming better stewards of information about themselves if they get mixed signals from different systems. The terms also matter to people who are in the position of making software acquisition and architecture decisions – it's easier to compare features when the terms line up. But most importantly in this relatively early phase of intellectual inquiry around identity, the designers of the features themselves need to agree on terms simply to facilitate the conversation.

One source of considerable confusion is “identity provider,” a fairly common term (one formal definition can be found in [SAMLGloss]) – with multiple concepts attached to it and multiple other terms in use for some of those concepts. The SAML [SAML] and Liberty [Liberty] worlds and systems derived from them abbreviate it IdP, and reserve it for the system entity that does credential-checking and providing evidence of this to requesters. Microsoft abbreviates it IP, and includes the supplying of attributes among its jobs. SXIP instead talks about a “homesite.” SAML would say the attribute-providing job is done by an “attribute authority,” and Liberty would talk about “web service providers” of identity information to consumer service endpoints. All these differences are not trivial; they manifest themselves in differences in levels of security, trust, liability, and services.

Other terminological issues abound. What does “user-centric identity” mean? What about it is

dependent on an “identity selector”? What is its relationship with privacy and anonymity?

What does it mean to be an “identity metasystem”? Is it meta because it involves a new level of abstraction? Does it require the same infrastructure to be put in place across participating systems?

Have we agreed yet on what “(an) identity,” “identifier,” and “identity federation” mean? Some technologists use “identifier” colloquially to stand for the entire set of identity information managed under that identifier; is this workable or too confusing? Is it possible for these difference to remain a purely technical consideration, or will users somehow have to learn the distinctions?

What's the deal with “assertions,” “assertion statements,” and “claims”? Is it possible to settle on common terms and meanings?

And...what about “personae”? [LoC]

Identity information exchange needs to be solved for the general case

[OpenID] says, “Logging in to a dozen websites a day is lame.” The use case most frequently addressed by identity systems of all kinds is distributed authentication, often called cross-domain single sign-on (SSO); for this discussion it's emblematic of other use cases that require identity information transfer (although it should be noted that SSO does exercises only a small portion of the potential value in identity information exchange).

The intra-domain SSO problem was solved a long time ago with cookie-based technologies. Now that the reach of modern identity systems extends outside single domains, other barriers to the free flow of authentication information come to the fore – for example, differences in the ability to protect the privacy of a user's information, in the ability to negotiate security assurances between domains, and in flexibility around handling “legacy” identity information.

OpenID is, in fact, an example of a lightweight identity system that has chosen to trade off a great deal of security and privacy assurance for additional simplicity, which makes it attractive for some current personal-identity scenarios – but would be unacceptable in others, and in most enterprise or government scenarios. Like several other lightweight identity systems, it has made an architectural bet on URL-based identifiers and on protocols that work exclusively with them. By contrast, many of the “heavyweight” identity scenarios involve anonymity-preserving federation of existing identities whose identifiers – for example, employee numbers and email addresses – and attribute names are predetermined. These scenarios are often required to address many more security risks directly (in addition to spelling out any dependencies on security provided at lower protocol layers) because of greater liability considerations. SAML is an example of a system that was developed in response to these extensive requirements (see [SAMLSecPriv] for more information about SAML's security and privacy considerations), for example with its ability to capture a wide range of authentication types in its “authentication context” descriptors [SAMLAuthnCxt].

For these reasons, the sphere of personal identity is in danger of becoming a silo of its own unless these systems and the other more-complex ones can be made to integrate with each other somehow, allowing “lightweight” acts of authentication to move upward into higher-stakes uses where this makes sense – and also allowing the complex systems to learn from the very simplicity of the others.

SAML may offer a useful integration point because of its extensibility options. For example:

- SAML's various protocols have bindings to underlying transport and messaging protocols, and these bindings contribute significantly to building in security. XML Signature, relied on heavily

in SAML, is considered by some to be problematic and "heavyweight." New bindings could be written to provide for message and assertion signing (or other means of achieving data origin authentication) in less onerous ways. (Johannes Ernst's "Really simple Signatures" proposal [XML-RSig] is provoking interesting discussion about a way forward.)

- For flexibility, SAML uses URIs to label the format of an identifier. It predefines URI labels for email addresses, X.509 subject names, Windows domain qualified names, Kerberos principal names, and others, but it does not yet define a label representing a "URI-based identifier." Such a label could easily be defined, and used immediately.

Users' identity habits are far more promiscuous than we might wish

Users consistently show a tendency to favor convenience and potential gain over complexity and potential loss; this is demonstrated every time they distractedly click through dialog boxes full of warnings, but happily spend five minutes filling in personally identifiable data on a web form just to earn a chance to win a token prize. The requirement to create yet another account on every new website (this author has *hundreds* of logins) has only exacerbated the tendency, encouraged poor maintenance hygiene for already-weak authentication methods (primarily passwords), and even trained many users that they have to supply a raft of information to get anything of value from the web. By contrast, providers of services on the web typically have more incentive to authenticate and authorize other providers with which they interact.

This situation suggests some lessons:

- Ensure that creating a new user account is a signal event rather than a daily one. Average users should ideally be able to count all their identity providers on their fingers, and explain why they went with these providers (just as they can probably tell you their reasons for choosing their bank or even their dry cleaner). This is a very aggressive goal, of course, since it would require massive uptake of SSO and attribute exchange technologies, along with the attendant legal and contractual relationships.
- In concert with the above, strive for consistent application of human-computer interaction (HCI) principles in the various identity management interfaces presented to users and maintainers. Paul Madsen recently noted [Madsen] six examples of a user-facing identity "conversation" pattern, which use flows and technologies that differ to greater or lesser degrees, and most of these allow for presentation variation. The variations go deeper than surface considerations, and are subject to a host of the weaknesses discussed in, for example, [SOUPS]. Add to this the variety of challenging environments for users and maintainers, from mobile to rapid response, we have the case for subjecting identity designs to HCI criteria.
- Where possible, avoid relying on synchronous interactions that interrupt the flow of a user getting something positive from a web experience (for example, buying something or playing a game), because many will click "I Agree" or "Use the default identity" or whatever the nearest button is, if it gets in the way of the benefit they are eager to experience. Initial policy configuration, reasonable defaults, and the new notion of identity rights agreements [IDRights] may help here, as well as exposing audit logs that let users backtrack and "debug" interactions gone wrong.
- Try to reserve such interactions for garnering user consent when they are exposed to significant loss (according to their own policy definition) in machine-to-machine identity interactions that

would normally be silent from their perspective. An example of a system for this is the Interaction Service [LibInteract] within Liberty's identity web services framework.

References

[AnonTerm] *Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management - A Consolidated Proposal for Terminology* (http://dud.inf.tu-dresden.de/Anon_Terminology.shtml)

[IDRights] *Identity Rights Agreements and Provider Reputation* (http://www.w3.org/2005/Security/usability-ws/papers/26-idcommons/)

[Liberty] Liberty Alliance Project specifications (http://projectliberty.org/resources/specifications.php)

[LibGloss] *Liberty Technical Glossary* (http://projectliberty.org/specs/draft-liberty-glossary-v2.0-03.pdf)

[LibInteract] *Liberty ID-WSF Interaction Service Specification* (http://projectliberty.org/resources/specifications.php#box2)

[LID] Light-Weight Identity Wiki (http://lid.netmesh.org/wiki/Main_Page)

[LoC] After *Love of Chair* from The Electric Company (http://en.wikipedia.org/wiki/Love_Of_Chair)

[Madsen] *A generalized identity conversation* (http://connectid.blogspot.com/2006/02/generalized-identity-conversation.html)

[OpenID] From OpenID's “Why?” section (http://openid.net/)

[SAML] SAML V2.0 information available on the OASIS Security Services TC website (http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security#samlv20)

[SAMLAuthnCxt] *Authentication Context for the OASIS Security Assertion Markup Language (SAML) V2.0* (http://docs.oasis-open.org/security/saml/v2.0/saml-authn-context-2.0-os.pdf)

[SAMLGloss] *Glossary for the OASIS Security Assertion Markup Language (SAML) V2.0* (http://docs.oasis-open.org/security/saml/v2.0/saml-glossary-2.0-os.pdf)

[SAMLSecPriv] *Security and Privacy Considerations for the OASIS Security Assertion Markup Language (SAML) V2.0* (http://docs.oasis-open.org/security/saml/v2.0/saml-sec-consider-2.0-os.pdf)

[SecGloss] *Internet Security Glosssary* (http://www.ietf.org/rfc/rfc2828.txt)

[SOUPS] *Symposium on Usable Privacy and Security Conference Report* (http://cups.cs.cmu.edu/soups/2005/SOUPS_2005_Conference_Report.html)

[XML-RSig] *So what about really simple XML Signatures?* (http://netmesh.info/jernst/Technical/really-simple-xml-signatures.html)

Acknowledgments

The author would like to thank Hubert Le Van Gong, Paul Madsen, Nick Ragouzis, and Bill Smith for their comments on earlier versions of this paper.