Federated Identity for CyberInfrastructure: Workshop Report
http://workshops.cilogon.org/2010
November 2010

Contents

Executive Summary

The Federated Identity for CyberInfrastructure workshop was held November 4th and 5th at the 2010 Internet2 Fall Member Meeting in Atlanta, Georgia. It brought together campus identity providers, cyberinfrastructure (CI) providers, and CI users to discuss requirements, challenges, and approaches for using federated identity to access CI (such as TeraGrid). The 69 workshop attendees represented 53 unique institutions, including 30 universities, 11 R&E network providers, and other R&E labs and institutes. The workshop was organized by Jim Basney, Randy Butler, Ken Klingenstein, Scott Koranda, and Von Welch. It was hosted by Internet2 and sponsored by the NSF-funded CILogon project. The workshop web site (http://workshops.cilogon.org/2010) includes links to all presentation materials.

We highlight the following workshop conclusions and outcomes:

- Federated identity is a technology of value to CI projects, but they need assistance to realize its full potential.
- We cannot expect scientists to come to technologists asking for federated identity solutions. The technologists need to bring the solutions to the scientists.
- Federated identity has value even if identities are managed internal to the specific science project or virtual organization (VO), for single sign-on across the members of the VO and to set a foundation for federation with external identity providers in the future.
- CI projects are using different technologies for identity federation (SAML/Shibboleth, OpenID, PKI, and RADIUS) and developing bridges between them.
- Applications of interest to use with federated identity include wikis, secure shell (SSH), email lists, and data access and analysis.
- Level of assurance requirements vary across CI providers, from the requirement to gather simple usage statistics to the requirement for strong authentication to protect

high value resources from compromise.

- Many CI projects share a common need for a unique (persistent, non-reassigned) user identifier, and more consistency across SAML federations in meeting this need would be helpful.

Further discussion of workshop conclusions and outcomes follows.

Definitions

The workshop operated under the following definitions for *federated identity* and *cyberinfrastructure*:

> "Identity management refers to the **policies, processes, and technologies that establish user identities and enforce rules about access to digital resources**… Federated identity management permits extending this approach above the enterprise level, creating a trusted authority for digital identities **across multiple organizations**. In a federated system, participating institutions **share identity attributes based on agreed-upon standards**, facilitating authentication from other members of the federation and granting appropriate access to online resources. This approach streamlines access to digital assets while protecting restricted resources."
>
> - EDUCAUSE: 7 Things You Should Know About Federated Identity Management (September 2009)
>
> "The comprehensive infrastructure needed to capitalize on dramatic advances in information technology has been termed cyberinfrastructure (CI). **Cyberinfrastructure integrates hardware for computing, data and networks, digitally-enabled sensors, observatories and experimental facilities, and an interoperable suite of software and middleware services and tools.** Investments in interdisciplinary teams and cyberinfrastructure professionals with expertise in algorithm development, system operations, and applications development are also essential to exploit the full power of cyberinfrastructure **to create, disseminate, and preserve scientific data, information and knowledge.**"
>
> - NSF's Cyberinfrastructure Vision for 21st Century Discovery (March 2007)

An illustrative example of federated identity in practice is the InCommon Federation, a community of over 180 higher education institutions plus laboratories, research centers, agencies, and sponsored partners, serving over 5 million end users, that enables "shared management of access to online resources in support of education and research in the United States." An illustrative example of cyberinfrastructure in practice is TeraGrid, "an open scientific discovery infrastructure combining leadership class resources at eleven partner sites to create an integrated, persistent computational resource."

Science Perspective

The workshop included presentations from representatives of science projects: Scott Koranda of the LIGO project (Physics), Rachana Ananthakrishnan of the Earth System Grid (ESG) project (Climate Science), Tom Barton of Project Bamboo (Arts and Humanities), and Edwin Skidmore of the iPlant Collaborative (Plant Science). As one would expect, the workshop attracted representatives of science projects with an interest in federated identity. Scott Koranda raised the question of how best to engage additional science projects. While identity management is a challenge for many science projects, it seems few would link the phrase "federated identity" with this challenge and few can spare the effort to engage in a workshop or conference devoted to federated identity. The federated identity champions need to go to the science communities, rather than (or in addition to) trying to bring the different science communities together on the topic of federated identity.

A common thread running through the science project presentations was an interest in using identity providers external to the project, but an acknowledgment that currently the projects are each internally managing user identities, because of the additional effort that is required to work with external identity providers and the difficulty of finding existing external identity providers that meet project requirements (such as coverage of the user community, reliability, and strength of authentication). Three of the projects are currently using federated identity technologies internally (SAML/Shibboleth in LIGO and iPlant; OpenID in ESG) to provide single sign-on and potentially enable federation with external identity providers in the future. Other projects may consider following this example of using federated identity internally then expanding to use of external identity providers.

CI Provider Perspective

The workshop also included presentations from representatives of CI providers: Jim Basney of the NSF TeraGrid and CILogon projects, Patricia Kovatch of the National Institute for Computational Science (NICS, a TeraGrid resource provider), Debbie Bucci of the NIH Center for Information Technology, Mike Helm of the DOE Science Identity Federation (SIF), and Milan Sova of the TERENA Certificate Service (TCS) project. These efforts are incorporating federated identity into CI in different ways, with opportunities for interoperability across CI providers, agencies, and nations.

One thread running through the CI provider presentations was support for (bridging between) multiple federated identity technologies, specifically: SAML/Shibboleth, OpenID, PKI, and RADIUS. NIH iTrust and NSF CILogon support both SAML and OpenID authentication. NICS supports federation of two-factor authentication tokens via RADIUS and PKI. TCS, TeraGrid, and CILogon each support issuing certificates based on SAML authentication, while SIF has an experimental identity provider that issues SAML assertions based on certificate-based authentication.

Applications

The success of federated identity for CI depends critically on supporting the CI applications that researchers care about. Workshop participants identified wikis as popular federated applications (for example: wiki.ligo.org and ctsawiki.org). Supporting the many different wiki software platforms is a major challenge (for example, LIGO has Shibboleth-enabled Foswiki, Moin, and aLOG wiki instances).

Secure Shell (SSH) was identified as another top application. Rhys Smith presented Project Moonshot which is working to enable federated authentication to non-web application such as SSH. Benn Oshrin presented planned work in the COmanage project to support federation-aware provisioning of SSH accounts. Jim Basney described the use of federated authentication to certificate-enabled SSH in TeraGrid and CILogon (see go.teragrid.org).

Also, Scott Koranda described successful use of federated identity management for email lists in LIGO, where group membership changes are reflected automatically in email list memberships, and Rachana Ananthakrishnan described federated identity for data access and analysis in ESG (based on OpenID).

It appears that there is significant duplication of effort in "domesticating" (i.e., federation-enabling) applications across CI projects. The COmanage project is developing a registry of domesticated applications which can help share results across projects. This is a step toward the idea discussed at the workshop of an "application marketplace" (i.e., an "app store") where domesticated applications can be published, found, and easily installed and used.

Scope of Federated Identity

Federated identity by definition enables use of digital identities *across multiple organizations*. In workshop presentations and discussions, we heard about federated identity implemented by a specific science project across multiple institutions (e.g., LIGO, ESG), federated identity across government agencies (e.g., NIH iTrust), federated identity across campuses in a national federation (e.g., InCommon and TeraGrid/CILogon), and international use of federated identity across national federations (TERENA Certificate Service, eduGAIN). It is important to not lose sight of the benefits of federated identity in these different scopes. A CI project may today benefit from federated identity across a small number of institutions participating in the project, without addressing the complexities of national-scale or international federations, knowing that the choice of standards-based federated identity technologies will enable the project to expand to other identity providers and federations in the future.

Levels of Assurance

The workshop included a session focused on levels of assurance. Levels of assurance (LOA) specify the strength of security credentials, protocols, and procedures. NIST SP 800-63 defines Levels 1-4, starting at Level 1 which provides a basic strength of authentication with no identity proofing requirements, and proceeding up to Level 4 which requires strong in-person identity proofing and strong hardware cryptographic security tokens. Workshop participants discussed the variety of levels of assurance supported by identity providers and required by CI projects. Rachana Ananthakrishnan presented ESG's requirements for ease of use and minimal identity vetting (roughly analogous to 800-63 Level 1), primarily needed to facilitate calculation of usage statistics for reporting to funding agencies and data providers. Tom Barton presented the CIC InCommon Silver project, which is an effort to implement the InCommon Silver Identity Assurance Profile (roughly analogous to 800-63 Level 2) across the Big Ten universities plus the University of Chicago, the University of Washington, and Virginia Tech, with a target date of Fall 2011, to enable federated access to applications (such as payroll, benefits, and grant administration) requiring a higher LOA. Patricia Kovatch presented the use of two-factor authentication at NICS (corresponding to 800-63 Level 3), to protect the Cray supercomputer from compromise by attackers, thereby providing high availability to the researchers using that system. Other high performance computing systems such as Blue Waters at NCSA and supercomputers at the DOE Leadership Computing Facilities also require (or will require) two-factor authentication for access for similar reasons, which motivates federation of the two-factor authentication mechanisms, so researchers that use multiple high performance computing systems such as these can use a single authentication token consistently across the systems. There are multiple technologies available for federating these tokens, including RADIUS, PKI, and SAML, and likely CI providers will (continue to) support (bridging between) multiple mechanisms for greater flexibility for their user community.

The workshop presenters in the LOA session illustrated that LOA requirements vary across cyberinfrastructure. Many CI projects have requirements similar to ESG, for using authentication to gather statistics about resource usage and to personalize the user experience. However, some CI resources, such as the NICS supercomputer, require higher LOA credentials for access. If an identity provider (e.g., a university) can support higher LOA credentials for these applications, for example through the InCommon Identity Assurance program, then this avoids the need for CI users to obtain higher LOA credentials elsewhere.

User Identifiers

While the workshop did not have a session specifically devoted to it, the topic of user identifiers received significant attention during panel discussions. Many CI providers expressed a common requirement for persistent, non-reassigned user identifiers that can be used to accurately count CI users and identify users in groups and access control lists. The properties of *persistence* and *non-reassignment* ensure that a returning user has the same identifier over time, i.e., a user's identifier persists over time and is not re-assigned to someone else (which could give some other person the user's access rights).

In the eduPerson schema, the eduPersonTargetedID (ePTID) is a persistent, non-reassigned identifier, but it is not consistently available across today's identity providers. Therefore, service providers must make do with eduPersonPrincipalName (ePPN), with an understanding of the different re-assignment policies and practices for that attribute across the different identity providers. Many InCommon identity providers today never re-assign ePPN or have a well-defined hiatus period (of a year or more) during which an ePPN must go unused before it may be re-assigned. Surveying the international workshop participants, we learned that Sweden has implemented ePPN without re-assignment, Finland has added an additional "ePPN timestamp" attribute which indicates when the ePPN value was assigned, and Japan requires ePTID. Additionally, in the UK Federation, most participants assert that they have a minimum two year hiatus period for ePPN reassignment.

It was suggested that the Internet2 MACE-Directories Working Group (MACE-dir) could provide greater clarity and/or standardization regarding this identifier re-assignment issue.

Conclusions and Next Steps

In conclusion, the Federated Identity for CyberInfrastructure workshop enabled a wide-ranging discussion of a variety of requirements, challenges, and approaches in this area. Together with the CyberInfrastructure and Focus on Federations tracks at the Internet2 Member Meeting, we explored the intersection of these two areas of strong interest. Attendees generally agreed that follow-on meetings on this topic would be worthwhile. One specific suggestion was for a follow-on meeting co-located with the Internet2 Spring Member Meeting, held each year in Washington, DC, to enable more participation from government agencies to address the topic of cross-agency and international federation activities. A second suggestion was for a Campus Architecture and Middleware Planning (CAMP) meeting focused on Virtual Organizations (i.e., a VO CAMP or VAMP), tentatively scheduled for Summer 2011.

Acknowledgments

We very much appreciate the active participation of the workshop presenters/panelists and other workshop attendees. The workshop would not have been a success without them.

We gratefully acknowledge support from Internet2 for hosting the workshop and from the National Science Foundation under grant number 0850557. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Internet2 or the National Science Foundation.