Simplifying Federated Access to Scholarly Content and Services

Lisa Janicke Hinchliffe (@lisalibrarian)
University of Illinois at Urbana-Champaign

Ralph Youngen (@ryoungen)
American Chemical Society

IAM Online – May 13, 2020
• Publishers license content/services to libraries that are then made available to entitled users.
  • Note: Not all who attempt access to a resource are entitled.
• Authentication/authorization systems support access for entitled users and prevent access for non entitled users.
• Considerations for managing personal data for access – especially with respect to privacy – include legal, contractual, and ethical frameworks.
Terminology

“Discovery”

- **In scholarly communications**: Conducting literature searches to discover disclosed research pertinent to one’s interest
- **In federation operations**: Searching for one’s identity provider to initiate a user authentication workflow

“Metadata”

- **In scholarly communications**: Information about a published research work
- **In federation operations**: Information about Identity Providers and Service Providers
The Scholarly Authentication/Authorization Ecosystem

**Device Authorization**

**IP Address Recognition**
- Device IP address in institution namespace
- Asserting that the user of the device is authorized user

**Remote IP Recognition Options**
- VPN
- Proxy Servers

**Individual Authentication**

**Publisher-specific accounts**
- Typically for personalized features only
- Typically do not provide authorization

**Federated Authentication via SAML:** Security Assertion Markup Language
- Validates the end user, not the device
- Provides authorization
- Shibboleth is a specific implementation of the SAML protocol
The Scholarly Authentication/Authorization Ecosystem

### Device Authorization

- **IP Address Recognition**
  - Device IP address in institution namespace
  - Asserting that the user of the device is an authorized user
- **Remote IP Recognition Options**
  - VPN
  - Proxy Servers

### Individual Authentication

- **Publisher-specific accounts**
  - Typically for personalized features only
  - Do not provide authorization
- **Federated Authentication via SAML:** Security Assertion Markup Language
  - Validates the end user, not the device
  - Provides authorization
  - Shibboleth is a specific implementation of the SAML protocol

**Library May be Using Any and/or All of These**
The Scholarly Authentication/Authorization Ecosystem

Device Authorization

IP Address Recognition
- Device IP address in institution namespace
- Asserting that the user of the device is authorized user

Remote IP Recognition Options
- VPN
- Proxy Servers

Individual Authentication

Publisher-specific accounts
- Typically for personalized features only
- Do not provide authorization

Federated Authentication via SAML: Security Assertion Markup Language
- Validates the end user, not the device
- Provides authorization
- Shibboleth is a specific implementation of the SAML protocol

Today’s Focus
Federated Authentication/SAML/Shibboleth

- Built upon system of trust between identity providers (e.g. academic institutions) and service providers (e.g. publishers)
- Identity provider controls the information (i.e. attributes) about its users that gets sent to service providers
What is (was) RA21?

RA21 Initiative
• Resource Access in the 21st Century
• Sponsored by NISO and STM
• Began in 2016
• Pilot projects to test various technical solutions and user experiences, including a pilot with corporate librarians from pharma companies
• Ended June 2019

Outcome
• NISO Recommended Practice, published in June 2019
What is SeamlessAccess?

A coalition of 5 founding organizations:

Community-driven effort to support seamless access to information resources, scholarly collaboration tools, and shared research infrastructure

SeamlessAccess is creating an operational service based upon the RA21 recommended practices
SeamlessAccess Provides an Improved User Experience
Makes SAML “Where Are You From” Easier to Navigate for End User

Information about user’s institution(s) saved in the local browser

Trusted service providers are able to access this institution data

Reduces the number of times a user must search for their institution
Toward A More Seamless World
What does SeamlessAccess do, specifically?

Enables the ability for sites to display the “Access through <<your institution>>” button by using data about the user’s institution(s) via browser storage.
What does SeamlessAccess not do?

• SeamlessAccess is not involved in the actual federated authentication process.
• Service providers (e.g. publishers) and identity providers (e.g. academic institutions) must already have an existing relationship that supports SAML-based federated authentication for SeamlessAccess to work.

• Example:
  • A campus running Shibboleth is a member of an identity federation like InCommon
  • A publisher is also a member of InCommon and supports SAML federated authentication
  • The InCommon federation governs the policy framework and facilitates the exchange of SAML configuration details between all parties in a federation
Dependencies for SeamlessAccess to Work:

• Institution must have SAML service.
• Library must request SAML be enabled for publisher platform.
• Publisher must enable SAML on platform.
• Publisher must implement Seamless Access service.

• User must authenticate through SAML on a Seamless Access enabled publisher platform.

• If user is using the same browser on the same device, then Seamless Access will work on all enabled platforms.

• User authenticates.
Who is SeamlessAccess?

A governance committee with representatives from across the stakeholder groups

An implementation team including:
- Project Director
- Publisher Outreach
- Technical support from SUNET (with funding from GÉANT)
- Library Outreach
- User Experience

An outreach committee including representatives from 6 libraries

An attribute release working group

A contract language working group (coming soon)
What is the current status of SeamlessAccess?

- Piloting Phase RA21: December 2016 – June 2019
- SeamlessAccess Beta Phase (Supported by SUNET, the Swedish Federation; Full 24x7 Operational Service): July 2019 – December 2020
- Full Service (Anticipated Support from GÉANT and Internet2): January 2021 - onwards
Where is SeamlessAccess Implemented?

**In Progress**
- CCC
- DARIAH
- Elsevier
- O’Reilly
- Silverchair platform
- Taylor & Francis
- Wiley

**Integrated**
- ACS
- Atypon platform
- Springer Nature
- SUNET
- TENET
COVID-19: Expanding Federated Access

February 2020:

“As you may know, a new outbreak of coronavirus has recently occurred in China. ... all students are required to study online at home. However, now traditional remote access faces significant pressure by the limitation of VPN capacity ... CERNET is actively promoting service providers to join CARSI (CERNET Authentication and Resource Sharing Infrastructure) to help users access ... CARSI bases on the Shibboleth system, it’s among the world’s most widely deployed federated identity solutions”

March 2020:

[Image of a flyer showing online classes and breaking news about colleges suspending in-person classes]
ACS Publications’ Experience

March/April saw a significant increase in federated authentications, likely due to:

• Shift in demand for remote access due to campus closures

• Deployment of the SeamlessAccess user experience on March 1, 2020

• Expansion of institutions enabled for federated access:
  - For existing federations, associating an IdP’s EntityID with subscription record enables WAYF discovery
  - Also joining many new federations
Lessons Learned – Attributes

Many IdPs are asserting personally identifiable attributes even though ACS is not requesting them

• A bug on the ACS site displayed the user’s first name if supplied via attributes:

• Now corrected even if personally identifiable attributes are received:

• SeamlessAccess is sponsoring an Entity Categories and Attribute Bundles Working Group to help IdPs select the appropriate attribute set for a given SP
  • Authorization Only (no attributes), Anonymous, and Pseudonymous
  • Publisher SPs likely will request Pseudonymous to assist campuses with credential theft
Lessons Learned – IdPs for University Systems

• Some university systems run a single IdP for multiple campuses
  • E.g. Indiana University system, University of Alaska system

• No standard way to associate users with their affiliated campuses using attributes
  • Important both for authorization and for usage reporting

• Discussing a concept with David Bantz (Chair, InCommon Community Trust and Assurance Board)
  • IdP asserts both primary affiliation and list of all affiliated campuses
  • SP would check authorization at primary affiliation, then others as needed
Lessons Learned – User Experience

User consent for attribute release is sorely needed, e.g.:

Use of SP’s MDUI elements (i.e. DisplayName, Description, Logo) can help users understand the federated authentication workflow.
What is GetFTR (Get Full Text Research)?

• A new service under development by the scholarly publishing community that also leverages federated authentication.

• Provides on-the-fly verification of a user’s entitlement rights to a research article based on the user’s institutional affiliation.

• Works with publisher platforms to determine entitlement status using two pieces of information:
  • Unique identifier for a document (DOI)
  • Unique identifier for an institution (EntityID)

• Compatible with (all of) today’s research discovery tools, scientific collaboration networks, library resource management systems, etc.
How Does GetFTR Work?

• Returns “smart links” (WAYFless URLs) customized for the user’s institution

• These links work regardless of the user’s physical location.
  • If connected to a campus network with a recognized IP address, institutional access is provided based upon IP address recognition.
  • If away from a campus network, institutional access is provided via federated authentication.
    • GetFTR plays no role in the federated authentication workflow between the user and the publisher.
Final Thoughts

• Campus IT and campus libraries need to work together.
  • Determine appropriate attribute release policies for SPs procured via the library.
  • Implement an attribute release user experience for your campus.
  • Determine contract language, where appropriate.

• Findings from recent survey conducted by SeamlessAccess indicates some areas for improvement between campus IT and campus libraries:
  • 72% of IT respondents said they had implemented federated authentication for library resources, while only 46% of library respondents said they had.
  • 88% of library respondents who had implemented federated authentication said they worked with IT.
    • 94% of IT respondents indicated the library was minimally involved; only 20% said library was very involved.

Survey responses: 159 library, 131 IT
Final Thoughts (cont.)

• Send publisher SPs pseudonymous identifiers for users.
  • Stop sending personally identifiable attributes.
  • Pseudonymous attributes (e.g. eduPersonTargetedID) can greatly assist campuses to identify compromised user credentials without compromising privacy.

• Expansion of federated authentication is underway and will likely continue.
  • Logging for SeamlessAccess service is very limited by design to protect user privacy
  • Taking a conservative approach, in April 2020 SeamlessAccess observed between 1 -2 million user flows, mostly from ACS and nature.com.

• A new divide is emerging between institutions that already support federated authentication and those that do not.
  • U.S. Department of Education lists more than 4,000 degree-granting academic institutions, compared to 558 IdP members of InCommon.
  • Large institutions are more likely to support federated authentication. Better support is needed to help smaller institutions to learn how to get started.
Comments?
Questions?
Discussion?
IAM Online - Special Edition - How do YOU use eduroam?
May 18, 2020 - 2 pm ET | 1 pm CT | Noon MT | 11 am PT

A look at how three organizations have leveraged eduroam to meet new and emerging needs for online learning.

June IAM Online - Hiring for Identity and Access Management
June 10, 2020 - 2 pm ET | 1 pm CT | Noon MT | 11 am PT

Challenges and opportunities for both CIOs and identity management professionals in higher education.
# InCommon Trusted Access Platform Training

[https://incommon.org/academy/software-training/](https://incommon.org/academy/software-training/)

<table>
<thead>
<tr>
<th>Software Component</th>
<th>Virtual Training Dates</th>
<th>Early-Bird Rate Deadline</th>
</tr>
</thead>
<tbody>
<tr>
<td>COmanage</td>
<td>May 19-20, 2020</td>
<td>Passed</td>
</tr>
<tr>
<td>Grouper</td>
<td>June 2-3, 2020</td>
<td>Passed</td>
</tr>
<tr>
<td>midPoint</td>
<td>June 16-18, 2020</td>
<td>May 15, 2020</td>
</tr>
</tbody>
</table>
InCommon Virtual BaseCAMP
July 20-24, 2020 (Noon - 4 pm ET each day)

- An introduction to identity and access management basics
- An introduction to the InCommon Federation
- An overview of the community-built InCommon Trusted Access Platform services and software

Program and registration information: https://meetings.internet2.edu/2020-basecamp/
IAM Online Evaluation

https://www.surveymonkey.com/r/IAMOnline-May-2020