Networking and Information Technology
Research and Development

The government seeks individual input; attendees/participants may provide individual advice only.

Middleware and Grid Interagency Coordination (MAGIC) Meeting Minutes
April 6, 2022, 12-2 pm ET

Virtual

Participants

Alex Vinson (NASA)	Mallory Hinks (NCO)
Arjun Shankar (ORNL/OLCF)	Marcy Collinson (Oracle)
Bill Miller (DOE)	Mark Day (NERSC)
Cory Snavely (NERSC)	Martin Halbert (NSF OAC)
Douglas Lattman (ORNL)	Michael Corn (University of California San Diego)
Donald Petravick (University of Illinois)	Michael Shapiro (University of Illinois)
Eric Blanco (NCO)	Mike Heroux (Sandia)
Eric Lancon (BNL)	Miron Livny (Wisconsin)
Florence Hudson (Columbia University)	Rich Carlson (DOE/SC)
Hall Finkel (DOE/SC)	Robert Bohn (NIST)
Jeff Conklin (NCO)	Roland Haas (NCSA)
Jim Basney (University of Illinois)	Sharon Broude Geva (U of Michigan)
Juan Jenny Li (NSF/OAC)	Terry Fleury (University of Illinois)
Kevin Thompson (NSF)	Tevfik Kosar (NSF)

Introductions: This meeting was chaired by Rich Carlson (DOE/SC) and Tevfik Kosar (NSF)

CILogon: Enabling Federated Identity and Access Management for Scientific Collaborations

Jim Basney, Principal Research Scientist, Cybersecurity Division, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

- Jim provided an overview of the tool.
    - CILogon enables researchers to log on to cyberinfrastructure (CI). CILogon provides an integrated open-source identity and access management platform for research collaborations, combining federated identity management (Shibboleth, InCommon) with collaborative organization management (COmanage). Federated identity management enables researchers to use their home organization identities to access research applications, rather than requiring yet another username and password to log on. Collaborative organization management enables research projects to define user groups for authorization to collaboration platforms (e.g., wikis, mailing lists, and domain applications). CILogon implements the AARC Blueprint Architecture.

    - CILogon: 10+ year sustained effort to enable secure logon to scientific cyberinfrastructure (CI)
    - for seamless identity and access management (IAM) using federated identities (SAML, OIDC, OAuth, JWT, X.509, LDAP, SSH, etc.) so researchers log on with their existing credentials from their home organization
    - supporting 17,500+ active users from 450+ organizations around the world with onboarding/offboarding/attributes/groups/roles
    - managed consistently across multiple applications supporting access to science applications on HPC clusters, in Jupyter notebooks, using Globus, via REST APIs, and many other interfaces
    - using existing identity providers from the researcher's home organization (SAML/ADFS) or external sources (Google, ORCID, GitHub)
- Realizing our vision
    - align with InCommon Trusted Access Platform (https://www.incommon.org/trusted-access/) Shibboleth, COmanage, Grouper
    - provide hosted services common IAM platform across many collaborations growing CILogon operations (since 2010) reliability / sustainability
- Open Source
    - CILogon (https://github.com/cilogon) OpenID Connect, OAuth, X.509
    - InCommon (https://www.incommon.org/trusted-access/) Shibboleth, COmanage, Grouper
    - IdentityPython (https://idpy.org/) pyFF, SATOSA
    - SciTokens (https://scitokens.org/)
    - OpenLDAP with voPerson (https://voperson.org)
- Sustainability
    - development supported by NSF/DOE
    - operational support from XSEDE
    - non-profit subscription model administered by NCSA/UIUC
    - supports long-term sustainability
    - provides contracted SLAs
    - CILogon remains open source and focused on research & scholarship needs
    - CI Logon has 10+ year proven history
    - From 2016 to present number of monthly users has grown 1,000 users to almost 28,000 users
- Campus and Researcher IDs
    - 4,000+ identity providers available via eduGAIN including CERN, NCSA, LIGO, XSEDE, ...
    - OAuth-based identity providers ORCID GitHub Google Microsoft
    - supporting researcher mobility
    - supporting researchers w/o campus IdPs
- managing 1000s of IdPs
    - SAML Metadata Query Protocol
        - https://datatracker.ietf.org/doc/draft-young-md-query/

        - https://spaces.at.internet2.edu/display/MDQ
    - monitor daily IdP changes
        - https://groups.google.com/a/cilogon.org/g/idp-updates
        - https://cilogon.org/idplist/
- Baseline Federations Program – InCommon activity
    - Federation support of research and scholarly communities
    - sets a baseline in terms of trust and operational security practices across the Federation
    - aligned with international baseline expectations Program standards
    - Protects collaboration resources
    - Shared federation legal, policies and practices help collaborations manage riskurity
    - Sec Incident Response Standard
    - https://www.incommon.org/federation/baseline/
- AARC Blueprint Architecture
    - a standard model for describing how research projects need have a proxy service to handle all the different types of researcher authentication.
    - authentication and authorization for research collaborations project in Europe.
    - Adopted this in CILogon
- Federation Proxy
    - apps don't need to handle the complexities of federation in isolation
    - many apps can't handle 1000s of identity providers (e.g., AWS)
    - a federation proxy service can handle federation for many (related) apps
    - a federation proxy can handle targeted user identifiers consistently
    - open-source software for operating your own proxy
    - CILogon operates these proxies because they have to – applications should be able to drive federated authentication and federated identity at scale directly and query for attributes from attribute providers, but many applications just don't have that power. That's what drives the requirement for the proxy services.
- Tokens for science – different types of tokens needed for services – some examples:
    - OpenID Connect (OIDC) ID Tokens (e.g., SCiMMA) containing user attributes and group memberships from the research community (via COmanage) and from the researcher's home institution (via InCommon)
    - SciTokens (e.g., LIGO) containing authorization scope values determined by per client/subscriber policy
    - WLCG Tokens (e.g., Fermilab) support for wlcg.groups and storage.*|compute.* scopes
    - GA4GH Passports (e.g., Australian BioCommons) support for AffiliationAndRole, AcceptedTermsAndPolicies, ResearcherStatus, ControlledAccessGrants, and LinkedIdentities
- Jim presented a slide focusing on token standards.
- Jim presented a slide on the token transition timeline and more detail can be found at https://opensciencegrid.org/docs/security/tokens/overview/

- Jim displayed a slide illustration the projects using CIlogon and Jim focused on the Cloudbank project as an example.
    - Cloudbank mission is to facilitate access to cloud resources for NSF researchers.
    - Jim walked the audience through the login page using CIlogon, and moved through the consent page and then the dashboard. Jim talked about the actual steps CILogin goes through once the user clicks the login button.

Q&A

- Douglas Lattman wanted to know aside from the RFD standards called out in slide 16 if Jim's group also addressed the NIST standards.
    - Jim stated "the identity assurance program maps to the NIST identity assurance levels and give us a standard way of asserting those levels of assurance in the saml assertion and also in the in the json web token. That's something that is able to pass through the proxy so that level of assurance that we get from campus in the saml assertion can then go into the ID token that's issued by the CIlogon proxy. If the application has a requirement for a multifactor authentication or identity proofing it can check for that in the different types of assertions that they get from us".
- Miron Livny stated that there seems to be too much identity and not enough capability. He asked should we try to do less identity in our infrastructure because it seems we are moving away from the gsi that was very identity heavy.
    - Jim stated that for the cloud it is not your driver's license that gains you access to your resources, it's instead your credit card. But we still need some kind of user authentication to get access to resources to generate some kind of credential. That credential needs very little identity information for a lot of use cases. Just need some sort of ID for researcher accessibility without going through a lot of identity proofing steps for access to the resources.
- Arjun Shankar had a question about sustainability and how the Europeans thinking about this in terms of risk exposure for attacks. How are the designers making this offering robust in this context?
    - Jim stated that the key aspects for this offering is to operate a highly reliable service that operates with other service providers like the services in Europe. Jim gave a few examples of these European providers and stated that they are long-lived organizations with good sustainability models. There are ongoing discussions on what are priorities for customers like off hours support, but the goal is to have multiple token providers, multiple identity providers that they can trust from a security perspective, from reliability perspective from a sustainability perspective.
- Douglas Lattman had two more questions. The first question he asked if Jim worked with any identity providers that can issue identity cards like Evan IML three level(sic) where they provide who they are and who they are working for, like with identity cert where they are recognized across boundaries? The second question is do you meet any of the FEDramp requirements in how you are hosting the data for organizations that are contractually required to be FEDramp?

    - Jim answered the first question by stating that they don't do credentialing themselves but work with service providers that are doing high level credentialing. He stated the are moving away from identity for a lot of their application, but the infrastructure is still there for it.
    - For the second question Jim stated they don't operate under a federally certified regime and FEDramp has not been a requirement for their subscribers so far. They have a good security policy but conceded that maybe FEDramp could be in their future but at the moment it is not something they are doing.
- Rich Carlson asked what are some of the new trends coming out of the new RCS that we should be looking for?
    - Jim answered that federation at scale in the JSON web token OAUTH open ID connect space is very active research right now, a next generation of what we currently do on the saml side for federating with identity providers at scale. We also have OAUTH 2.1 in our future and so there's a lot of operational security improvements happening in in the protocols that are important for us to track.
- Rich Carlson had a follow-up question. He was interested in understanding what's happening in the automation space for workflows and how they can process and use somebody authorization or identity.
    - Jim stated that their pilot job model is really key for supporting workflows and a number of these environments like in LIGO and having specific authorization to spin up the pilot framework that accepts jobs from the experiment is part of the transition work that we're doing right now getting the right authorizations into those tokens that are used by the pilot framework is what enables that in a least privilege way.
- Doug Lattman had a final question around FEDramp and frameworks – Does Jim's group work on implementing anything like 62443 or ISO 27000 or some security framework that can be audited and verified?
    - Jim stated that some of their practices map to a framework, like the NIST framework, but the primary framework that they operate under is the trusted CI framework that gives them the four pillars of their cybersecurity program. They haven't done an ISO 27,000 audit, for example.

The Path to Cloud Federation via Standards

Dr. Robert Bohn, Cloud Computing Program Manager, CTL/NIST

- Background
    - Cost & Efficiency drivers - US IT Budget ~ $80B/year:
    - Federal Cloud Computing Strategy (Cloud First – Feb 2011)
    - NIST's Goal – To accelerate the federal government's adoption of cloud computing
        - Build a USG Cloud Computing Technology Roadmap
        - Lead efforts to develop standards and guidelines
    - Starting Material – NIST Definition of Cloud Computing (SP 800-145)
    - Develop a Reference Architecture for Cloud Computing
        - Determine the "What" of Cloud Computing, not the "How"

- USG Cloud Computing Technology Roadmap Requirements (NIST SP 500-293)
    - International voluntary consensus-based standards
    - Solutions for High-priority Security Requirements, technically de-coupled from organizational policy decisions
    - Technical specifications to enable development of consistent, high-quality Service-Level Agreements
    - Clearly and consistently categorized cloud services
    - Frameworks to support seamless implementation of federated community cloud environments
    - Updated Organization Policy that reflects the Cloud Computing Business and Technology model
    - Defined unique government regulatory requirements and solutions
    - Collaborative parallel strategic "future cloud" development initiatives
    - Defined and implemented reliability design goals
    - Defined and implemented cloud service metrics
- NIST/IEEE Collaboration
    - NIST recognized the importance of "Frameworks to support seamless implementation of federated community cloud environments" in USG Cloud Computing Roadmap (NIST SP 500-293).
    - Collaboration between NIST and IEEE P2302 will help build consensus on creating an Intercloud - an open, transparent infrastructure amongst cloud providers to support evolving technological and business models and the growing demand for standards that address Intercloud interoperability.
- IEEE SIIF Objectives
    - Purpose: This standard creates an economy amongst cloud providers that is transparent to users and applications, which provides for a dynamic infrastructure that can support evolving business models.
    - Scope: To define topology, functions, and governance for cloud-to-cloud interoperability and federation.
    - Support Transparent Infrastructure
        - Like the Internet
        - Like the Phone Network
    - Cloud Implementation Independent
        - Like the Internet Router
        - Like the Phone Network CO Switch
        - Based on Standards
    - Simple Protocol Set, Easy to Join
        - Like an ISP, simple IP based protocols enough to get started
        - Supports Regional Governance
    - Support for Generalized Resource Federation
        - Not Just VM's – IaaS, PaaS, *aaS
        - Extensibility to Any Describable Resource Type
        - Communities can Add Resource Types

    - Support for Multiple (Open or Proprietary) Federation Topologies
        - Network Abstraction
    - Global Scale Capable
- NIST PWGFC/IEEE P2302 Goals & Outputs
    - The NIST PWGFC will develop a cloud federation vocabulary and conceptual model based on the Scope and Purpose.
        - The PWGFC interim outputs will be contributed to the IEEE P2302 Working Group in real-time.
        - The PWGFC output will be a NIST Special Publication.
    - The IEEE P2302 Intercloud Working Group will develop a cloud federation standard based on the Scope and Purpose.
        - The PWGFC interim contributions will serve as input.
        - Feedback on PWGFC vocabulary and conceptual architecture contributions will be provided to the PWGFC in real-time.
        - The P2302 initial output will be an IEEE Standard.
        - Plan to contribute the P2302 Standard to ISO/JTC1 to create an International Standard.
- NIST Cloud Federation Reference Architecture
    - Sites decide to collaborate and establish trust
    - Federation Managers (FMs) are deployed
    - FMs configured to communicate based on trust relationship
    - FedAdmin(s) create one or more federation instances
    - A federation is a Virtual Administrative Domain
    - Resource Owner register services for a specific federation
    - Federation members can invoke federation services
- IEEE 2302-2021 is a Collaborative NIST / IEEE Effort
    - IEEE 2302-2021 formalizes the CFRA model as a Federation Hosting Service (FHS)
        - Codifies an Architectural Approach and API for Federation Instances
    - A federation instance has:
        - Members, including Service Owners and at least one Fed Admin
        - Federation-specific roles and attributes
        - Federation members can discover and use registered services regardless of location
    - Fed instances are hosted on one or more FHSs
        - A set of FHSs constitute a distributed API Gateway
        - As with ordinary API Gateways, an FHS can "back-end" an external IdP
        - IEEE 2302 does not mandate a specific type of Identity Provider or identity credential
    - A set of FHSs are the “railroad” on which many federation instances can run
        - FHSs operators establish trust and secure communication among themselves
        - Federation members establish trust among themselves
- The IEEE 2302-2021 Federation Hosting Service (FHS) Model

    - The FHS model is essentially a set of communicating API Gateways
    - Three APIs in OpenAPI 3.0:
        - FHSOp API
        - FHSMember API
        - FHS-FHS API
    - Four Federation Capability Levels
        - L1: Core
        - L2: Accounting, Billing, Auditing
        - L3: Legal/Compliance Agreements
        - L4: Automation
    - IEEE 2302-2021 defines L1 Core API
    - There are three pre-defined roles:
        - FedAdmin
        - ServiceOwner
        - Member
    - FHS configurable to use various IdPs
        - OIDC, PKI, Verifiable Credentials, etc.
- Range of Deployment and Governance Configurations
    - Centralized (and by definition, Homogeneous)
        - Single FHS operated by a trusted operator
    - Decentralized and Homogeneous
        - Set of communicating FHSs with the same type of IdP
        - FHSs could be operated by different organizations
    - Decentralized and Heterogeneous
        - Set of communicating FHSs with different IdPs
        - Differences in credential formats could be address using token translation
            - Similar to universal token translation approach of the Authentication and Authorisation for Research Collaborations (AARC) project
        - To reiterate: attribute semantics and release are not issues since attributes are federation-specific
- The User's Perspective
    - After authentication, a federation member has a consistent view and ability to interact with federated resources
    - Member can use services and access data based on their authorizations, regardless of who owns the resource
    - Resource owners retain unilateral control over the discovery and access policies of their resources
    - Different user-friendly front-end interfaces possible. For example:
        - Web portals
        - Virtual desktops
        - Jupyter notebooks
- Building Out the Capability Levels

    - Accounting & Billing
        - Resource usage cost structure applied to monitored usage data
    - Regulatory Compliance
        - Federation credentials have sufficient identity and authorization information to properly enforce well-defined policies
    - Trust Domain Management
        - Simple federations can manually manage trust “out-of-band”
        - OIDC Federation Specification could be used to establish trusted communication between (OIDC-based) FHSs that have a common trust anchor
        - Trustmark Frameworks could be used where users build a Trustmark Interop Profile
    - Automation
        - Commercial federation providers could operate fleets of on-demand FHSs
        - Operation similar to current cloud-based Content Distribution Networks
- Comparison with Existing Approaches and Systems
    - IEEE 2302-2021 architectural approach is similar to:
        - Dataverse (similar proxy approach)
        - OpenStack Keystone service (“Federate In, Federate Out”)
        - DARPA Security Enclaves (research work published in 2000)
    - InCommon, in contrast, provides a single, monolithic, trust environment
        - Defined by the metadata file of trusted IdPs and SPs
        - Has created a “trust ecosystem” wherein additional services have been developed to provide enhanced capabilities and flexibility, e.g., CILogon, Grouper, COmanage
    - Open Science Pool: OSG Data Federation and OSG Compute Federation
        - Access Points (APoints) provide workflow management, including input/output data staging
        - Execution Points (XPoints) enable sites to contribute server capacity to an OSPool
    - NIH Cloud Interoperability approaches
        - “Find data”, “Authorization to use data”, “Set up place to do analysis”
        - GA4GH Passports and Data Repository Services
        - Researcher Auth Service (RAS) utilizes OAuth/OIDC
    - REFEDS Federation 2.0 WG chartered to make recommendations for future academic interfederations
        - “Drive innovative technical architectures, standards and policies”
        - Evolve academic interfederations by leveraging technologies and standards from across the federation landscape
- Summary
    - The need for collaboration that is inherently distributed is fundamental
        - Application domains exist across academia, industry and government
        - Widely adopted common practices and standardized tooling are needed

- The 2302-2021 model and API is a "Ford Model T" with lots of development needed.
  - Supports range of deployment and governance
  - Small-scale deployments
  - Individual projects and organizations can own and operate their own FHSs
  - Large-scale, commercial, deployments
  - Commercial federation providers could operate a fleet of FHSs across their data centers to provide on-demand federations with a potentially global footprint
  - Similar to provisioning of Content Distribution Networks
- Provides a possible evolutionary path for current federation environments
  - Architectural approach provides flexibility that ad hoc federation approaches can't
- Economic market development is the future challenge

Questions

- Rich Carlson had a comment and a question - we've talked a lot today about the identity systems and the back end so you can assert an identity and share that out what's the state of the art with being able to have a resource provider consume those identities, or those the attributes that are being shipped around?
  - Robert Bohn deferred to Jim Basney and Jim answered the question – Many are using site tokens lots of web applications right now are using it as well. It's well supported in various identity platforms that people are building their applications with. We are all involved in this move from X.509 and SAML over to the open ID connect and JSON web token space and with JSON tokens really defining the state of the art.
  - Rich followed with - what about having the ability for resource providers to create policies and have those policies implemented?
  - Jim Basney - So far, CIlogon uses the OAUTH client management API to allow the service owners to register policies with us as a token issuer. Those policies can be applied when the tokens are being issued based on the attributes of your authentication and the requests that are part of your current session. You can get a capability allocated to you that implements the policies, both for your virtual organization and research collaboration and the specific application that you're accessing. That's a JSON policy language that we're using, that standard OAUTH client management API to provision into. We are still working out some of the standards for how those policies are reflected in the token itself.
- Mark Day had some comments for Jim Basney about the model he showed in his presentation and about multi-factor authentication (MFA) in particular. He spoke about signaling and MFA and in the end asked if there was something to be done or was his group ahead of the curve on this one.
  - Jim answered by saying that his group was ahead of the curve on this one. It's known in the federation that if something is not required, they are not going to do it. Because it's a minimization principle we're not going to share that information,

we're not going to require the user to do it or have the user do it unless it's absolutely required for access to the service. So, this notion of "your experience with the service will be better if you do MFA as part of your login, but we can handle it if you don't" is still working its way through the Federation. I think you're feeling some of the pain of being ahead of the curve there. For CIlogon right now we don't offer a step-up service so we're not feeling that pain yet. In some back-and-forth Jim mentioned the proxies make this more complicated as they don't pass on the MFA preferences.

- Arjun Shankar asked if in some of these answers was policy and mechanisms were getting mixed together, mixing scalability issues and policy issues.
    - Jim stated that the proxy is a multi-tool for sure, it's doing token translation where it's pulling in attributes from multiple sources and applying policies potentially for multiple sources. It's convenient to do that, all in one place, we could have multiple layers in the system. We certainly have cases where we've got a proxy talking to a proxy talking to a proxy. Some of that policy can be pushed all the way down to the application or are pushed all the way up to the IDP. We can introduce layers, but you know every layer also add some complexity, so we've got some real practical trade-offs there.

Roundtable

- Mike Heroux: On Thursday April 21 there will be a Leadership Scientific Software (LSSW) townhall meeting 3 - 4:30 pm ET entitled "Expanding the Scope of What is Reusable: A panel discussion"

Next Meeting

May 4 (12 pm ET)