SSO Questions

Request for 10 minute overview/demo on Grouper software package set-up (IT-oriented)

- I can demo production LIGO Grouper (1.6.3)
- I can demo sandbox Grouper (2.1.3)
- Grouper has 4 main components
  1. Core/API–Java libraries manipulate database state
  2. User Interface (UI)
     - Run as Java JSP in container (Tomcat for LIGO)
     - Admin UI
     - “Lite” UI
     - New UI in development
  3. Web Services (WS)–REST and SOAP
  4. Provisioning Service Provider (PSP)
- Show LIGO custom “simple” UI page(s)
Discuss issues with Linux vs IIS setup/installation

Sorry, LIGO is a Linux shop. I have no practical IIS experience.

I can only talk about what I know from community.
SSO Questions

Do you have a centralized sign-up process for a LIGO identity and how does it work (quick demo)?

Yes. We have a service called “My LIGO”

- Quick demo
- New version in development that will be based on COmanage
  - Configurable enrollment flows
  - Quickly spin up new enrollments
  - Notifications
  - Provisioning into LDAP and Grouper and ?
  - Collaboration with Internet2 and iPlant
  - Would love to have STScI involved
SSO Questions

Do you separate your "internal" users from your "external" users?

Yes (and no)

- Did not do this initially—mistake
- Have deployed “LIGO Guest”
  - Separate Kerberos, LDAP, Grouper, IdP
  - Will be separate COmanage for enrollment
  - Don’t see problem with one LDAP for STScI
Did all the LIGO-participating organizations have to "buy into" the same set of software tools for SSO to work (i.e. Shibboleth)?

- LIGO is leading its collaborators so they are following
- LIGO has done everything in compliant ways, so Shib not required per se, just SAML2
- Is this a SAML2 versus OpenID question?
Can you set up web-service to non-web-service Authentication passing of information?

Yes

- Demo command line tool ligo-proxy-init
- Talk about SAML2 ECP
  - Enhanced Client or Proxy profile
  - Intended for non-browser clients
  - Also used for delegation (though most use OAuth)
- Talk about “Kitten: SAML Enhanced Client SASL and GSS-API Mechanisms”
- Talk about Project Moonshot

^1 tools.ietf.org/html/draft-cantor-ietf-kitten-saml-ec-01
^2 https://community.ja.net/groups/moonshot/wiki/start-here
Can you tell us about your experiences with Single Logout?

LIGO does not have SLO in place
- Waiting for community solutions
- SLO in federated context is not clear cut
- Use SAML2 forced re-authentication for high risk services
- Version 2.4.0 of Shib IdP implements a type of logout
  - End session with initiating SP
  - End session with IdP
  - Leaves other SP sessions in place
- Version 3.x Shib IdP roadmap includes SLO
SSO Questions

Have there been any major security issues since you started working with SSO?

- No, no major security issues
- Central authorization service has helped a few times
  - Revoke all privileges quickly
- Single credential also helped a few times
  - Turn off authentication as “show employee the door”
- Just beginning push to better password entropy
- Also implementing 2FA
  - Tokens for the highest risk (instrument access)
  - Phone for medium risk
Service Provider (SP) Questions

Request for 10 minute overview/demo on Grouper software package usage (SP-oriented)

- I can demo production LIGO Grouper (1.6.3)
- I can demo sandbox Grouper (2.1.3)
- Go over proposed LIGO authorization scheme
  - eligible subjects eligible for the privilege
  - locked subjects locked out of privilege
  - closure subjects losing privilege routinely
  - denied union of locked and closure
  - authorized eligible minus denied

Services look at authorized
Service Provider (SP) Questions

Did you have to unify identities from various projects into one "master list"?

- No, not really
- Got people to give up other identifiers, focus on new identity
- Used LIGO “killer app”
  - Access to the paper drafts
  - All thought leaders need to read the paper drafts
  - Let their influence bring others along
  - Not long before users pushing us
- Holdout are Linux cluster logins
  - Users embed logins in things they shouldn’t
  - Using a “carrot first, stick later” approach
Do you have an SSO alternative for folks who use `.htaccess` and `.ftaccess` files for AuthN/AuthZ?

- Use `.htaccess` files but do authorization by groups/roles/privileges
- Use centralized group/role/privilege management (Grouper/COmanage)
- Leverage delegated management of group/role/privileges

```plaintext
AuthType shibboleth
ShibRequestSetting requireSession 1
Require STScI_PrivilegeA
```
Other Questions

Request for 10 minute overview/demo on COmanage software package

Bad timing!

- Main demo box being used for webinar today (Tuesday)
- Development sandbox down with bad fan
- See what we can do with timing...
Other Questions

Request for 10 minute overview of Lessons Learned from LIGO’s SSO experience

- Branding of identity more important than I thought
- Separate authentication and authorization
- Enrollment/onboarding policy is time sink
- Offboarding as important as onboarding
- Privacy issues king for NREN IdPs
  - Will have to deal with opaque identifiers
- Be conservative with SAML2 protocols
  - Use HTTP-Redirect and HTTP-Post
  - Limit use of Artifact Resolution if you can
  - IdPs prefer stateless protocols
Other Questions

*How do you join a Shibboleth Federation? Is there something we need to do now in order to make this an easier process in the future?*

- Some type of contract (InCommon) or MOU (IDEM)
- Prepare privacy policy ahead of time and publish
- Do not treat InCommon as a “vendor”
  - You are not buying a service or product
  - You are joining an organization
  - It’s about shared trust and risk
  - Help your legal team understand in advance
- Helps if your services fall under one domain scope
  - eg. @ligo.org
  - Not strictly necessary, but expedites things
- Think about how you will manage metadata
- Show InCommon federation manager
Other Questions

*Can we easily Federate with an organization that has X.509 certificates? What about with OpenID AuthN?*

Work required depends on what level is necessary

- One application?
- Multiple service providers?
- Entire infrastructure?
- One-way or bilateral?

Ideas, things to consider

- Let application do the work if it can
- Shibboleth SP has useful features
  - SAML artifact spoofing
  - External authentication handler
- Social2SAML gateways for OpenID popular
  - Much support in Higher Ed for this
- At IdP level X.509 and OpenID login handlers
Other Questions

What are the major hurdles regarding setting-up/joining International Federations and how have you managed to work those out?

- Opaque identifiers and lack of attributes from IdPs
  - Collect other attributes at enrollment (COmanage)
  - Bind attributes using attribute authority
    - SAML2 attribute authority
    - LDAP
- Different/multiple signing keys for each federation
  - Due to misunderstanding of trust model
  - Issue slowly going away
  - In meantime play games with relying party configurations
- Show report for CTSC
I am curious about how LIGO uses Condor, which problems they encountered, how they worked around those etc.

- Show LIGO Data Grid monitoring pages
- Show vulture plots

Major problems and mitigations
- Configuration complexity
  “There’s a knob for that”
  - Close relationship with Condor team
  - Involvement in community
  - Sharing among all administrators
Condor Questions

Major problems and mitigations? What are their biggest problems/issues with the Condor environment at their site?

- Naive users have too much power
- Too easy to expose weakness in storage and network
  - Set memory limits on processes
  - Separate I/O paths for interactive/non-interactive
- Building complex workflows is...too complex
  - DAGMan syntax is essentially “assembler level”
  - Build application-specific DAGMan wrapping scripts
  - Leverage Pegasus workflow planning tool
- Users have trouble debugging
  - Educate, educate, educate
  - Build application-specific tools as necessary
If they were to re-build their system from scratch, would they still choose Condor?

Yes, absolutely.

- Flexibility is king
  - Support wide spectrum of analysis workflows
  - Support wide spectrum of hardware configurations
Condor Questions

How do they do workflow-level logging? Are they happy with Condor text log files? Do they ingest them in a DB?

More complex workflows leverage Pegasus

- pegasus.isi.edu
- pegasus-state for state of workflow
- pegasus-analyzer for querying success/failure

Simple workflows use text files
Condor Questions

*How do they do step-level logging? Do they use Condor Job Hooks? Do they have the logging logic in the workflow steps themselves?*
Condor Questions

What do they use for GUls?

Sorry, we do not use GUls. All workflows are submitted on the command line.

Field is relatively new, only hints that community is ready to think about GUls.
Condor Questions

How long does a typical workflow step take in their normal data processing? Do they have/had issues with Condor latency in scheduling jobs?

- Varies widely—from a few minutes to a few days
- Flagship analysis has many stages
  - Early stages mostly Python scripts—short
  - Main analysis jobs (FFT’s and matched filtering) in C—hours
  - Post processing and plotting mostly Python—short
- Latency worked around
  - Pegasus “bundles” some of the short jobs
  - Short jobs run in local universe on head node
  - Users encouraged to rework pipelines
Condor Questions

Do they have a local processing cluster with machines dedicate for their use? If so, how many and their types? (mixed platforms or uniform, like RedHat Linux?)

- 20K cores across 7 main clusters
- CIT, MIT, LHO, LLO, SYR, UWM, HAN
- India coming on line
- Cardiff coming on line
- Mix of Intel and AMD (mostly Intel now)
- All run 64bit SL 6.1 or Debian Squeeze (5)
- Reference OS committee meets every 2 years
Condor Questions

- New hardware has many cores/slots per CPU
- Many cores per unit memory
- Begin leveraging dynamic partionable slots
  - Driven by large memory jobs
  - Driven by hugely threaded (> 1024) jobs
Condor Questions

*Do they ”steal cycles” from user/desktop machines?*

- Not with Condor
- Einstein@Home is Boinc powered
- Run E@H as backfill on clusters
Condor Questions

*Do they ever go offsite to obtain more compute power (cloud/grid/whatever)?* If so, *do they have partnerships with organizations that provide them compute resources?*

LIGO’s involvement in Open Science Grid a spectacular failure

- Much work done by LIGO infrastructure people
- But never a top priority for LIGO
- Scientists simply would not pick it up
- Just too easy to SSH into cluster and work
- Will not even use our own clusters as “grid”
- Scientists would rather wait then learn to use OSG
- No culture of leveraging shared resources
- No formal OSG/LIGO arrangement now
Condor Questions

Do they copy input/output data files to/from the worker nodes or use a central storage server to serve the data to them?

- All clusters use shared central storage
- All storage NFS
  - Most simple NFS from multiple servers
  - Caltech uses NFS and QFS
- Need for good parallel file system is evident
- Problem is lack of institutional investment in the technology