Identity management and distributed computing: What LIGO wants from Condor

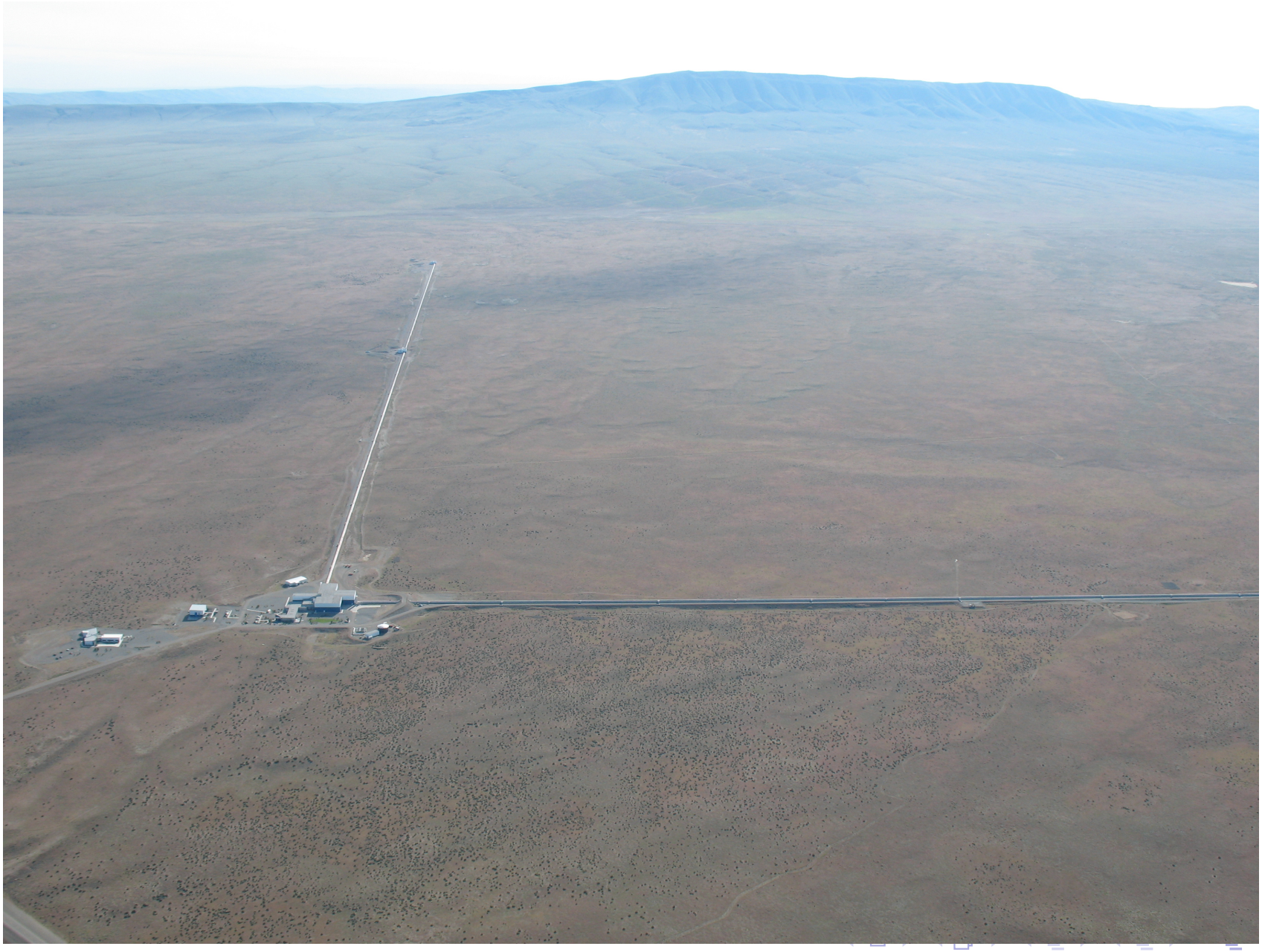
Scott Koranda for LIGO

LIGO and University of Wisconsin-Milwaukee

April 28, 2010
LIGO-XXXXXXX-v1

(Because LIGO has never asked Condor for anything before, right?)

LIGO Hanford, WA

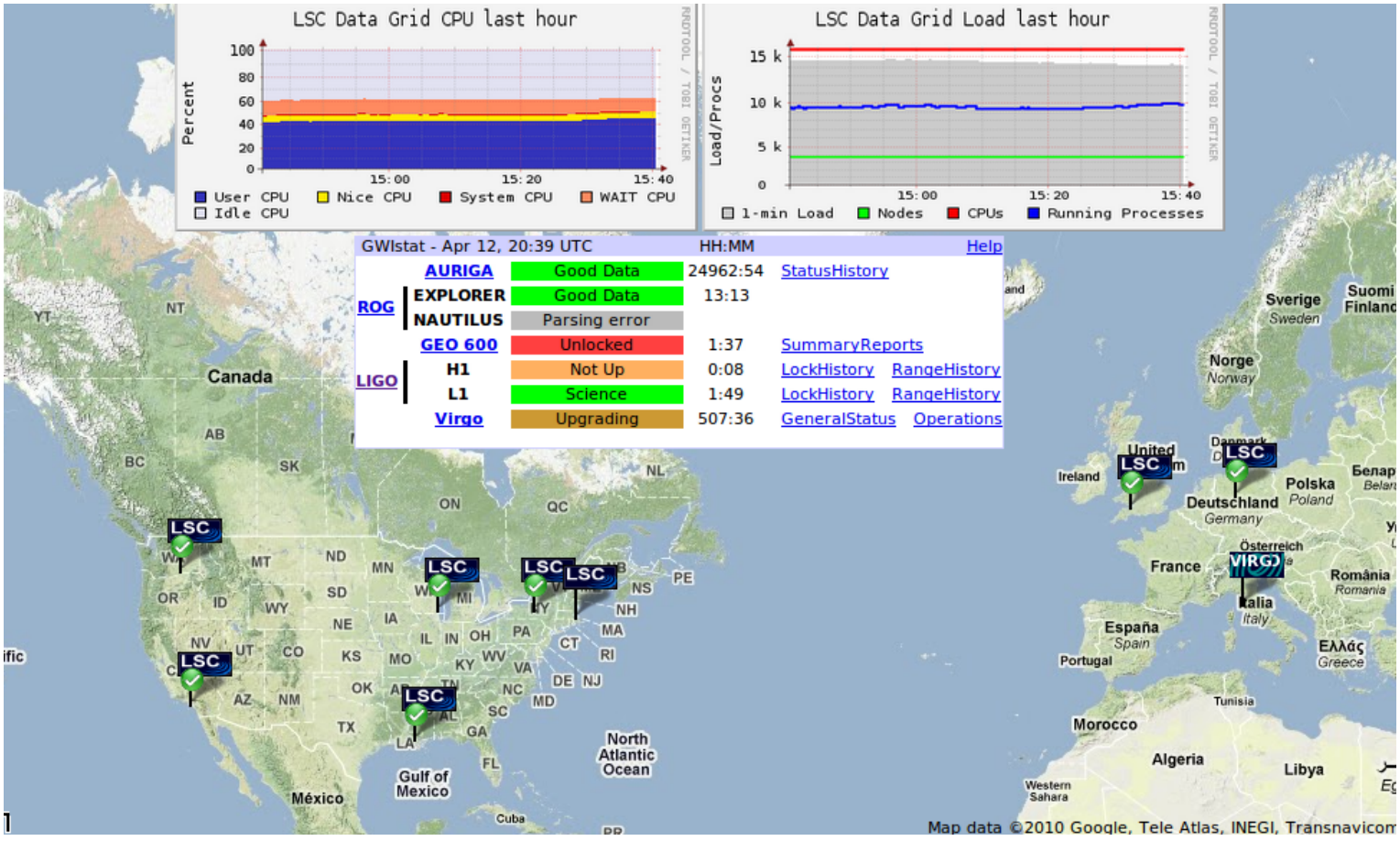


LIGO Livingston, LA

GEO600 interferometer, Hannover, Germany

LIGO Scientific Collaboration

LIGO Data Grid



- 15000+ cores, all Condor
- 10 sites
- Many flavors of data and metadata services
- $>$ 200 users

Condor: Success for LIGO

- 1000s of DAGs managed by Pegasus and DAGman
- Up to 5M jobs in über-DAG/DAX
- Single user/workflow keep cluster busy for week
- All production computing done with Condor

Running large workflows easy

Great. So what's next?

Isn't it the LIGO Data *Grid*?

It's taken a while, but we finally have traction...

- glideinWMS deployed at UWM and Caltech
- can glide into OSG (Nebraska)
- corralWMS as front end is being pursued
- scientists more interested than ever...

But it took *years* to go from cluster to grid.

Why?

Long road from clusters to grids

Just easier to login to multiple clusters and run Condor

```
$ grid-proxy-init
Your identity: /DC=org/DC=doegrids/OU=People/CN=Scott Koranda 212488
Enter GRID pass phrase for this identity:
Creating proxy ..................................... Done
Your proxy is valid until: Thu May  5 11:42:49 2011
$ gsissh ldas-grid.ligo.caltech.edu
$ condor_submit_dag analyze-this.dag
```

Note: users love the single sign-on, hate X.509 certificates

Data is analyzed. Now what?

A funny thing happened on the way to publishing the paper...

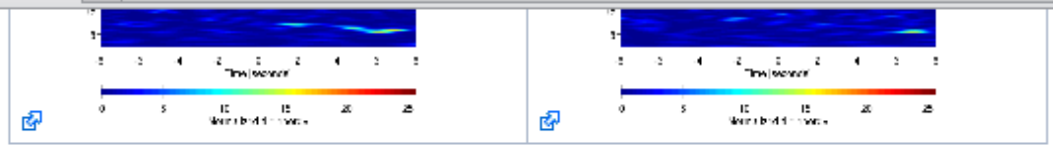
- Group of 10 to 60 people analyzed the data...
- Generated TBs of ouput...
- Now need to dig out the science...
- Collaboratively!

How do 60 physicists across 3 continents collaboratively sift through TBs of data spread around 10 computing clusters?

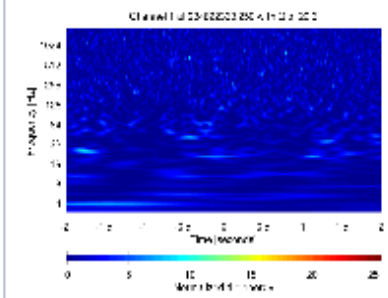
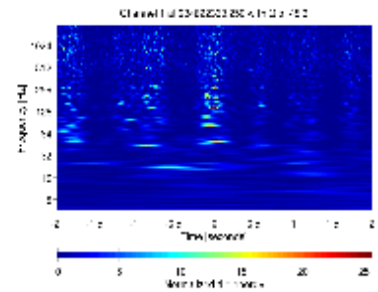
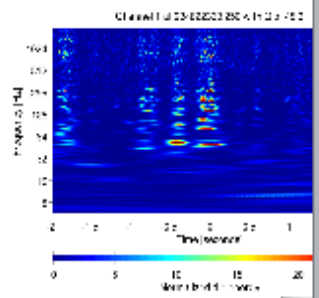
Collaborative science through a web browser(?)

File Edit View History Bookmarks Tools Help

S6Plan/100105082151S6Plan...

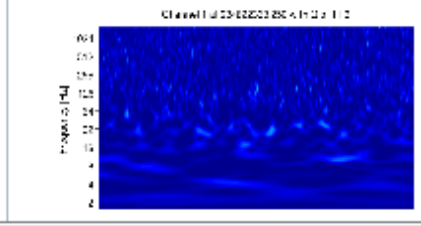
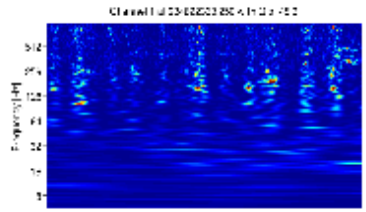


L1:OMC-QPD1_Y_OUT_DAQ and L1:OMC-QPD2_Y_OUT_DAQ have some noise at higher frequencies:

LSC-DARM_ERR	OMC-QPD1_Y_OUT_DAQ	OMC-QPD2_Y_OUT_DAQ
		

H1: There is a H1:DMT-PRE_LOCKLOSS_1800_SEC flag.

Looks like some H1:OMC-QPD1 channels had some noise at higher frequencies:

LSC-DARM_ERR	H1:OMC-QPD1_Y_OUT_DAQ
	

Seemed like a good idea, but...

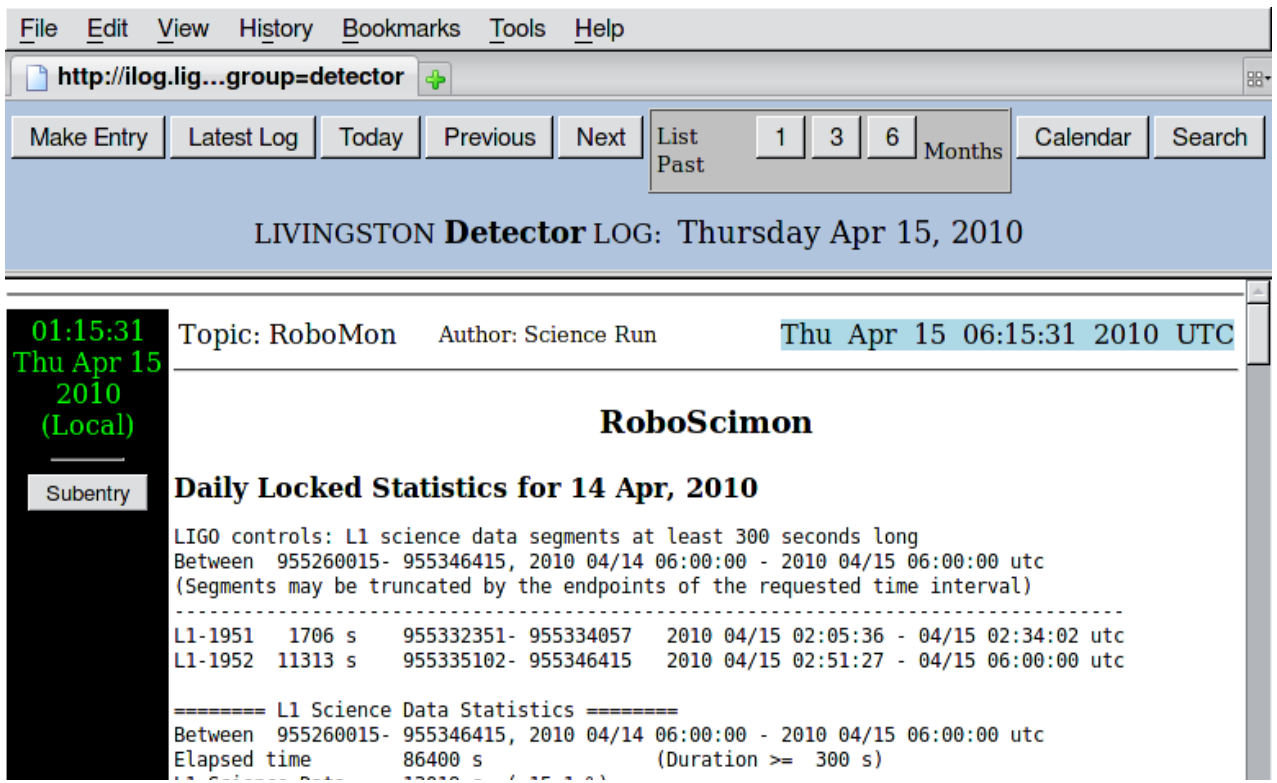
things were bad.

The mess we made on the Web

- Multiple sites deploying web tools
  - Moin, Twiki/Foswiki, Docuwiki, MediaWiki,...
  - GNATS, Bugzilla, Redmine, Trac, Gitorious?
  - Each requiring new login/password for user

The mess we made on the Web

- Specific example: eLogs at the detector sites
  - Web based electronic notebooks
  - Email "the" admin for access (hopefully he knows you)
  - Unique accounts, but...
  - All accounts use the same password
  - Loop not closed when people leave collaboration



The mess we made on the Web

Users frustrated

First response is “well known login/password”

- shared login and password collaboration wide
- used for protecting “low risk” information
- who monitors what is low risk?
- found login/password in the wild

The mess we made on the web

As the number of web tools and services grew we knew we had a problem...

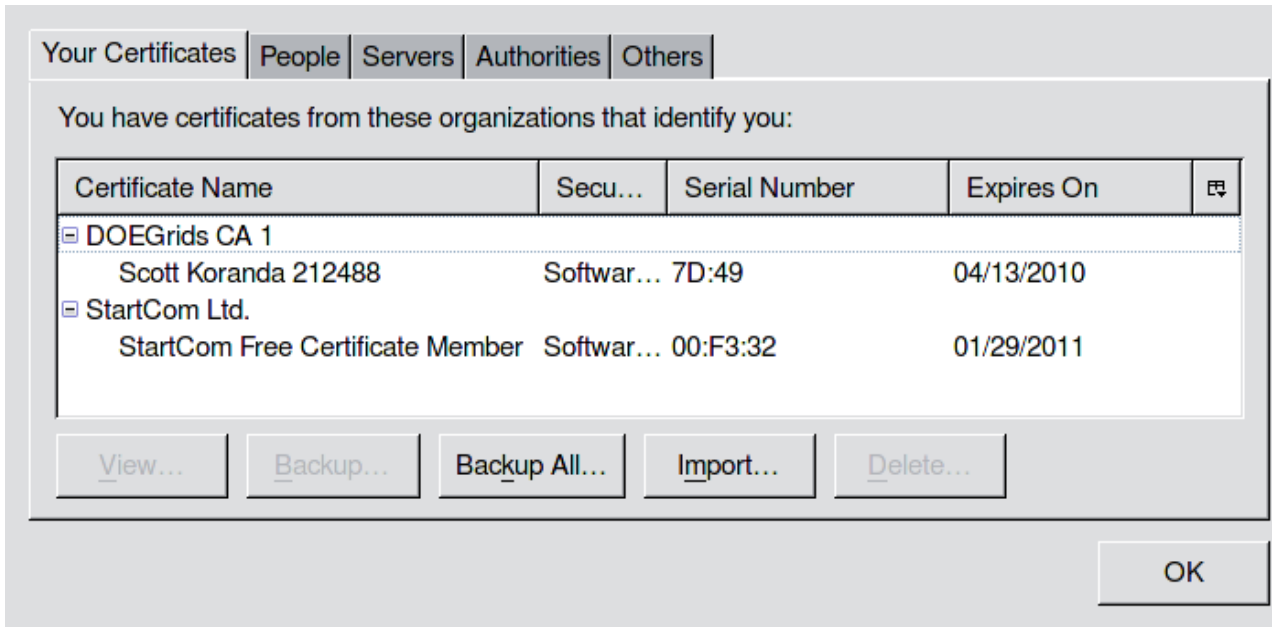
...but we were in production, busy doing science, and didn't take the hint...

The mess we made on the Grid

- LDG emerged in 2001
- Sought single sign-on and promise of Grid utopia
- Most Grid tools require PKI and GSI

The mess we made on the Grid

- User must request, retrieve, manage X.509 cert
  - Not all web browsers do PKI well
  - Grid tools require PEM but web browsers write PKCS12
  - "17, but steps 6) and 9) have 12 or 13 subitems each"
  - Turns out Ph.D. physicists on average cannot do this
  - Command line tools don't help much



Hi Scott,
The e-mail is attached below. When I click on the "import your certificate", it returns a "Add Certificates" pop-up that asks whether we want to add certificates to a key chain. The keychain options are: login, Microsoft_Intermediate_Certificates, System and X509Anchors. It also opens a panel as attached below. I am not certain how the import is happening in this system. I do not see any .p12 file in my directories and hence the subsequent export commands do not work. Sorry for bothering you. If you have any directions, please let me know. Thanks very much in advance,

The mess we made on the Grid

- No roster of who is/is not member of LIGO
- Each cert request must be vetted
  - Requires "secure communication" with each group PI
  - Getting attention of PIs can be difficult
  - SMIME email difficult for most PIs
  - Loop not closed when people leave group

The mess we made on the Grid

After X.509 cert issued user must be authorized

- Cumbersome
  - Each user added by hand to ACL file(s) at each site
  - Only grid-specific solutions available for managing ACLs
- Not uncommon for new member to wait weeks for credentials and access to LDG resources

The mess we made on the Grid

Managing access to LDG was one of the first hints we needed better identity management...

...we didn't take the hint...

The mess we made on the command line

Version control repositories

- CVS, SVN, git
- Distributed across multiple sites
- Each requiring yet another login/password
- People leave collaboration but still have access

Same issues for other command line tools

The mess we made on the command line

Managing access for hundreds of people to multiple code repositories was a nightmare...we knew we had a problem...

..but we were in production, busy doing science, and couldn't take the hint...

We had a mess

Yes, we can analyze the data with Condor...
But collaborating with the available tools is tedious.

Need a new approach to "logging in"
(identity management or IdM)

No single event precipitated new approach
It really came down to two things:

1. Sustained whining from frustrated users
2. Chatting with Ken Klingenstein (I2) over drinks

LIGO Identity Management Project

Knit together existing technologies and tools

Goals:

- Single identity for each LIGO person
- Single credential for each LIGO person
- SSO across web, grid, command-line
- Single source of privilege assertions

LIGO Identity Management Project

Found we had two building blocks:

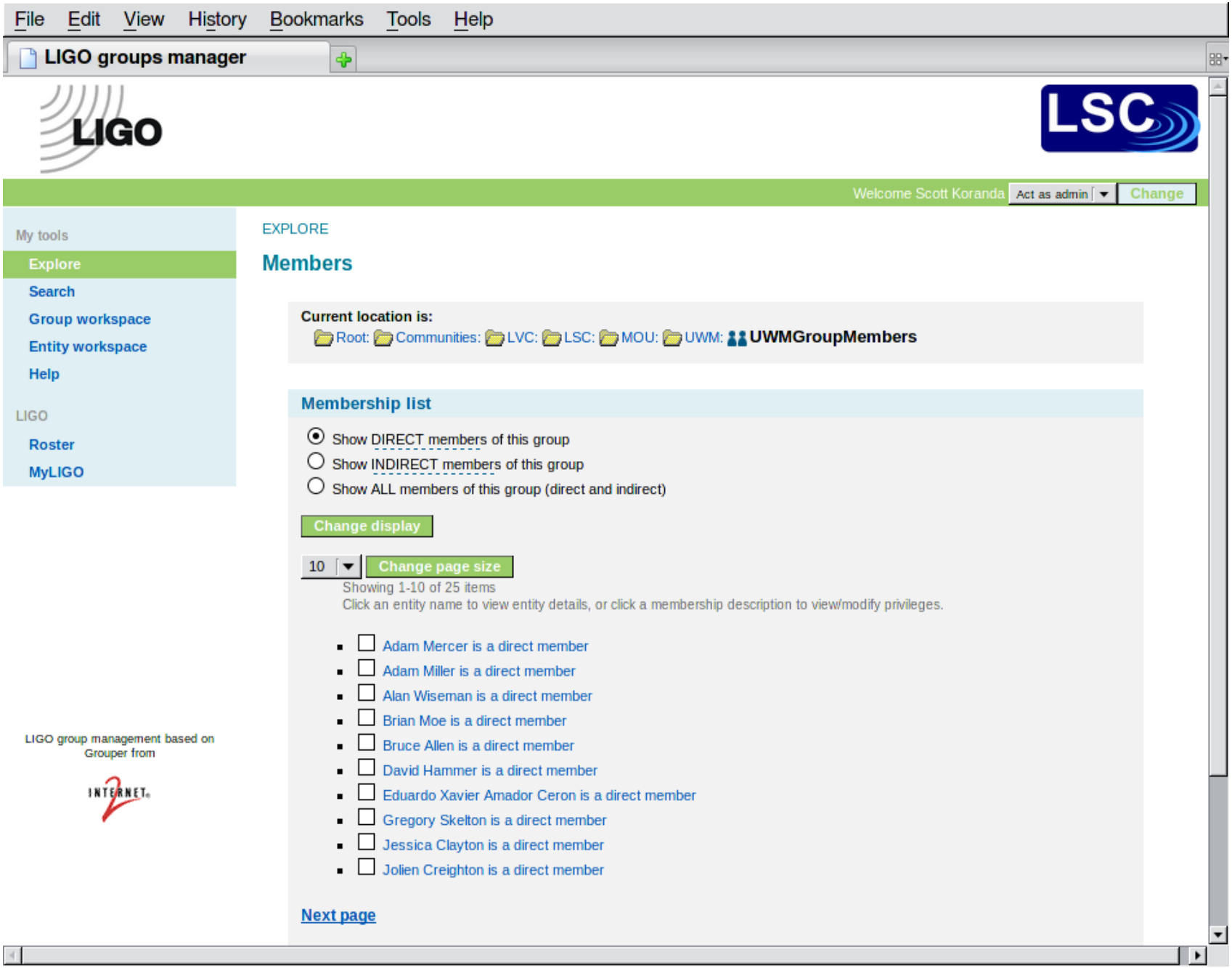
1. The nascent "LIGO Roster" project
   - PHP + Apache + MySQL
2. Kerberos principal for each LIGO member
   - unused at the time
   - `scott.koranda@LIGO.ORG`
   - users call it their "at LIGO.ORG login"
   - also known as their "albert.einstein" login
   - roster drives creation of principal for each member
   - roster pushes principal and details into LDAP

Single authoritative source of privilege

Primarily focus on group membership

Decided to leverage Grouper from I2

- Flexible enough to reflect community structure
- Ready-to-use web front-end
- SOAP and RESTful WS APIs
- Privilege and RBAC support
- Reflect into LDAP


File Edit View History Bookmarks Tools Help
LIGO groups manager
LIGO
LSC
Welcome Scott Koranda
Act as admin
Change
My tools
Explore
Search
Group workspace
Entity workspace
Help
LIGO
Roster
MyLIGO
EXPLORE
Members
Current location is:
Root: Communities: LVC: LSC: MOU: UWM: UWMGroupMembers
Membership list
Show DIRECT members of this group
Show INDIRECT members of this group
Show ALL members of this group (direct and indirect)
Change display
10
Change page size
Showing 1-10 of 25 items
Click an entity name to view entity details, or click a membership description to view/modify privileges.
Adam Mercer is a direct member
Adam Miller is a direct member
Alan Wiseman is a direct member
Brian Moe is a direct member
Bruce Allen is a direct member
David Hammer is a direct member
Eduardo Xavier Amador Ceron is a direct member
Gregory Skelton is a direct member
Jessica Clayton is a direct member
Jolien Creighton is a direct member
Next page
LIGO group management based on Grouper from
INTERNET2

```
[root@oregano ~]# ldapsearch -LLL -b "ou=people,dc=ligo,dc=org"
-H ldap://ldap.ligo.caltech.edu -x '(cn=Scott Koranda)'
isMemberOf
dn: employeeNumber=882,ou=people,dc=ligo,dc=org
isMemberOf: Communities:LVC:LSC:MOU:UWM:UWMGroupMembers
isMemberOf: Communities:LVC:LSC:MOU:UWM:UWMGroupManagers
isMemberOf: Communities:LVC:LSC:LSCGroupMembers
isMemberOf: Communities:LVC:LSC:CompComm:CompCommGroupMembers
```

Single identity and authoritative membership is key

LIGO Roster, Grouper, and Kerberos a powerful combination

- Kerb principal enables single identity
- Roster enables management of those identities
- Grouper enables management of memberships

With this foundation we could tackle web, grid, and command line spaces...

Single sign-on for LIGO web space

Deploy I2 Shibboleth System

- Single sign-on across LIGO web tools/pages
- LIGO Identity Provider (IdP)
  - Authenticate via `REMOTE_USER` and `mod_auth_kerb`
  - Attributes pulled from LDAP master server
  - Focus mainly on `IsMemberOf` (via Grouper)
- Look to federate in future
  - InCommon for many U.S. institutions
  - European federations (UK, DFN-AAI)
  - LCGT project in Japan

MyProxy, GridShib, CILogon integrate LIGO Data Grid

- MyProxy exchanges Kerb ticket for X.509 cert
- GridShib/CILogon exchanges SAML2 for X.509 cert
- User "sees" @LIGO.ORG cred required for both
- X.509 certs are "short-lived"
- Can also be converted to RFC 3820 proxy cert

LIGO Data Grid Authorization

Grid authorization driven by Grouper and LDAP

- ACL files derived from `IsMemberOf`
- Simple LDAP query with local caching
- X.509 DN and login pulled from LDAP

```
# isMemberOf Communities:LVC:LSC:LSCGroupMembers
"/DC=org/DC=ligo/O=LIGO/OU=People/CN=scott.koranda@LIGO.ORG" skoranda
"/DC=org/DC=ligo/O=LIGO/OU=People/CN=sergey.klimenko@LIGO.ORG" klimenko
"/DC=org/DC=ligo/O=LIGO/OU=People/CN=sukanta.bose@LIGO.ORG" bose
"/DC=org/DC=ligo/O=LIGO/OU=People/CN=szabolcs.marka@LIGO.ORG" smarka
"/DC=org/DC=ligo/O=LIGO/OU=People/CN=teviet.creighton@LIGO.ORG" teviet
```

Integrating the command line

CVS, SVN, git tunnel through SSH

- Most Linux OpenSSH sshd GSS-API + Kerberos
- Grid-enabled OpenSSH also deployed
- NCSA "mechglue" enables Kerb + GSI
- PAM also work with Kerberos

This pattern same for other command line tools

(note that `curl` works well with SAML2/Shib)

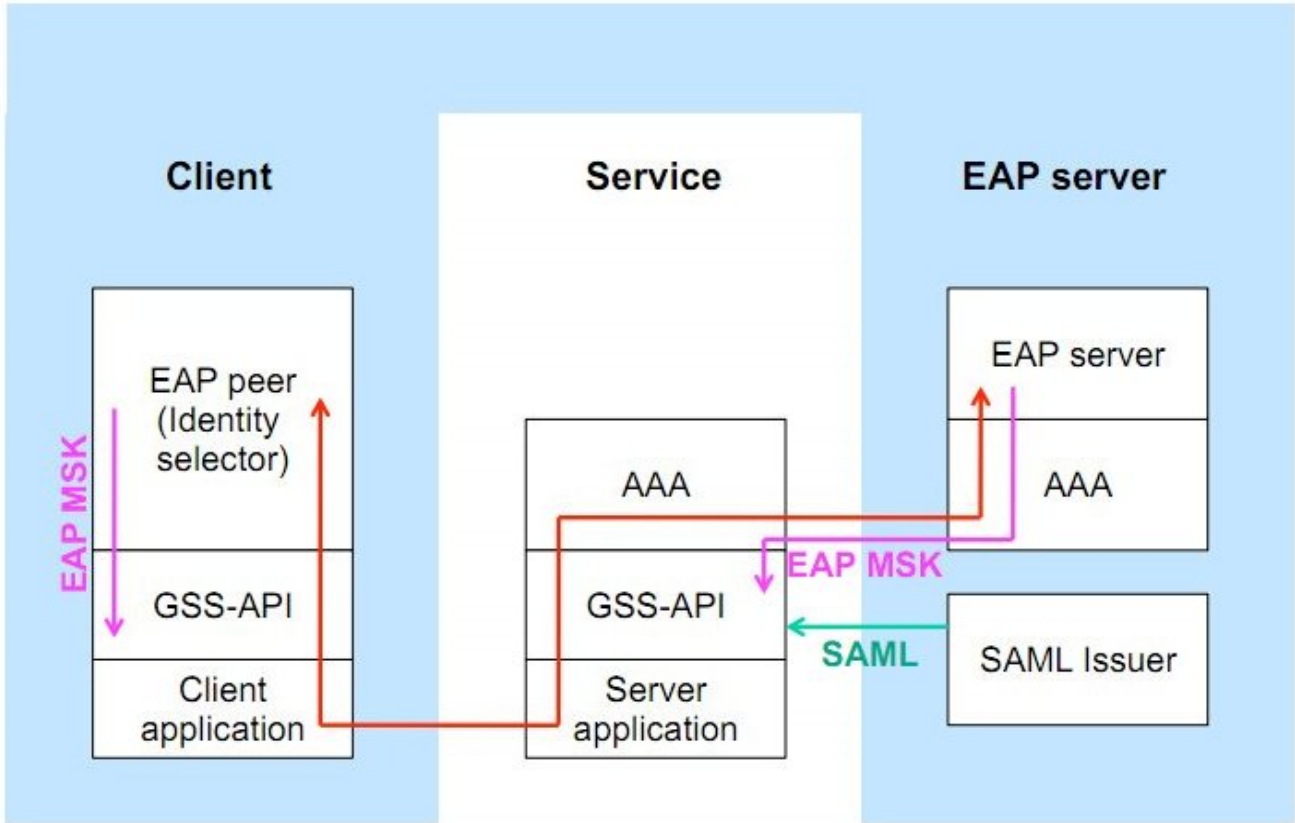
Other approach for grid and command line?

Project Moonshot

http://www.project-moonshot.org/

Project Moonshot is a JANET(UK)-led initiative, in partnership with the GEANT project and others, to develop a single unifying technology for extending the benefits of federated identity to a broad range of non-Web services, including Cloud infrastructures, High Performance Computing & Grid infrastructures and other commonly deployed services including mail, file store, remote access and instant messaging.

Moonshot


Client
Service
EAP server
EAP peer (Identity selector)
GSS-API
Client application
AAA
GSS-API
Server application
EAP server
AAA
SAML Issuer
EAP MSK
EAP MSK
SAML

Putting it all together

Within 15 minutes of joining LIGO Albert Einstein using his albert.einstein@LIGO.ORG credential can...

1. Access LIGO wikis to find HOWTOs
2. Download and install client tools
3. Login to cluster
4. Checkout code from git repository
5. Email analysis discussion list for help
6. Build code, submit analysis jobs

From 0 to science with one @LIGO.ORG credential

So everything is perfect?

The problem of single sign-on

- Compromised @LIGO.ORG credential yields much access
  - Users don't do great job protecting credential
- Do we want to enable access all way to instruments?
  - Probably require N-factor auth closer to instrument
- Federation: should the same ID faculty use for managing grants and grading be used to access LIGO resources?
  - We will have to get better at levels of assurance

Some ideas for Condor

Consume privilege assertions from external sources

- LDAP resolution of privilege groups?
- Query out to SAML Attribute Service (IdP)?
- Moonshot approach leverging existing GSS-API code?

(Of course push assertions into classads...)

Some ideas for Condor

Jobs pulling/pushing to/from web resources

- Happy to see Zach's libcurl work!
- Need to manage session cookies
- Existing tokens (X.509, ...) for auth
- (SAML keyword is "ECP")

Some ideas for Condor

Management of tokens/credentials at DAG level

- Job needing credential submitted at *end* of workflow
- Need Condor to keep credential available
- May need to refresh credentials
- (with proper delegation)

Distinction between web and grid is fading

- Scientists just want to use tools
- Don't care if "web" or "grid" or "cloud"
- Typical use case:
  - Submit large workflow to grid
  - Jobs run for week analyzing data
  - Workflow generates 1000's of summary images
  - Need to POST summary into analysis wiki
- Seamless IdM across "grid", "web", "cloud"
- Need HigherEd and Grid communities to build together