LIGO Authentication and Authorization 2.0

CILogon Fall 2009 Workshop

Urbana, IL – Sep 29, 2009

Scott Koranda & Warren Anderson

University of Wisconsin - Milwaukee

Who we are

- LIGO Laboratory
  - » CIT, MIT, LIGO detector sites in Hanford, WA and Livingston, LA
  - » ~ 250 people
- LIGO Scientific Collaboration
  - » 47+ institutions, groups, or organizations
  - » Intersects with the LIGO Laboratory
  - » Some organizations have subgroups
    - – GEO project is German/English collaboration with subgroups
  - » ~ 660 people
- Virgo collaborators
  - » Sister project
  - » Some use LIGO computing resources
  - » ~ 224 given LIGO.ORG Kerberos principal
- External collaborators
  - » Radio astronomers, numerical relativists, neutrino scientists, ...
  - » Just beginning to need to solve authorization issues

Where We Were

How to not scale in thousands of easy steps

- Systems evolved rather than being planned
- Coordination minimal or non-existant between services/sites/admins
- Membership in LIGO Scientific Collaboration not tracked
- Independent management of services and users:
  - Works OK for ~100 users on ~10 services
  - Does not scale to ~1000 users on ~100 services

Where We Were

The Mess We Made on the Web

- Web servers stood up by individuals and groups with private data
  - Each admin/scientist/application uses their own AuthN/Z scheme
  - Each admin administers AuthN/Z info independently of others
  - Usernames/passwords not coordinated between sites
- ~ 10 “wiki-like” applications with many instances each, including homespun ilog
  - Each has an internal accounting system which is not shared
- ~ 3 problem tracking systems with many instances each
- Shared “well-known” password used by all scientists for many sites
  - Found written down on whiteboards
  - Distributed in open emails
  - Posted on unprotected web pages by accident

Where We Were

The Mess We Made on the Grid/Shell

- LIGO Data Grid (LDG)
  - » We use slightly modified VDT for client/server distribution
  - » Users get X.509 certificates for authentication. Most use DOEGrids
    - – PI of group at each university/institute provides verbal verification for chain of trust, which sometimes takes weeks/months
    - – Cert request/retrieve/renew scripts have problems on some platforms
    - – Users make mistakes - some make many mistakes in a row
  - » Users request LDG account
    - – Successful account requests added into grid map file at each LDG site
    - – Grid map files maintained at each site independently by site admin
    - – No mechanism for removing/updating accounts automatically
- General computing and critical systems
  - » Only at laboratory sites at LHO, LLO, CIT, MIT
  - » Not coordinated between sites
  - » Use shared accounts for some critical systems

Where We Are Going

Requirements

- Easy for users
  - Single sign-on across as many services as possible
  - Minimize user management of credentials
- Easy for admins
  - Centralized management of user accounts
  - Hardened protocols/tools, widely used and well integrated into clients
  - Automate as much maintenance as possible
- Easy for collaboration management
  - Centralized management of personnel
  - High availability - network outages do not prevent work from happening at observatories or compute centers
  - Quick turn around - users can be added or removed everywhere in minutes

Where We Are Going

Pieces of the Puzzle

- my.ligo.org
  - » user database (mySQL) with PHP front end (in house development) to collect, manage and report personnel data
- LIGO.ORG Kerberos realm
  - » Authentication service
  - » One KDC per compute site for robustness
- LDAP
  - » Stores account and authorization information (user attributes)
  - » No secrets in the LDAP!
  - » One LDAP per compute site
- Grouper (I2)
  - » Create and manage groups for authorization, pushed to LDAP
- Sympa
  - » Mailing list engine to notify users/admins/managers of pending actions - draws mailing lists from groups in LDAP

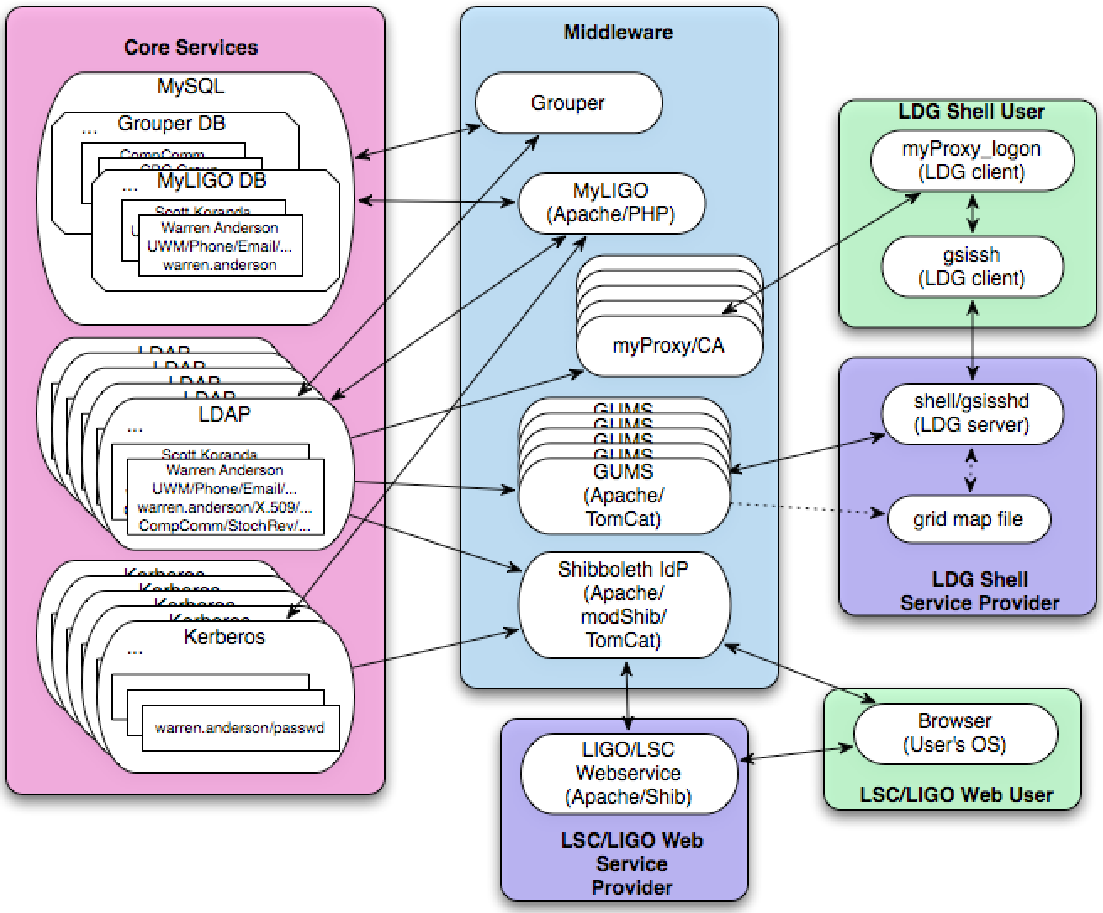
Where We Are Going

More Pieces of the Puzzle

- MyProxy (NCSA/Globus)
  - Exchange Kerberos ticket for short-lived X.509 credential/proxy
  - Embedded certificate authority (LIGO CA most likely)
  - One MyProxy server per compute site for robustness
- Shibboleth (I2)
  - Authentication via Kerberos (IdP)
  - Authorization via LDAP (IdP)
  - Web single sign-on
- GridShib (NCSA/Globus) ?
  - Ideally we would leverage a tool like GridShib for authorization to LDG services
  - But no Java services on LDG!
  - More on this later...

Where We Are Going

The puzzle



Where We Are Going

A day in the life - morning

- New member joins UWM LSC group
  - » Fills out myLIGO application with her information
  - » myLIGO notifies UWM group managers via Sympa
  - » UWM group manager approves application
    - – Kerberos credential created for user by myLIGO
    - – LDAP entry added for user by myLIGO
    - – User added to LSC members group
    - – Grid map files get updated with user certificate DN
  - » User downloads and executes “ligo-logon” script
    - – User gets Kerberos ticket and short-lived X.509/proxy
  - » User opens browser and goes to portal site to find where instructions and pipeline for running a pulsar search reside
    - – Shibboleth IdP authenticates her
    - – Portal site SP authorizes her because she is in LIGO members group

Where We Are Going

A day in the life - afternoon

- User goes to pulsar pipeline page
  - Pulsar pipeline page SP checks if she is member of pulsar analysis group
  - She is not, so access is politely denied with instructions to mail pulsar analysis group managers Sympa list to ask for access
- User mails pulsar analysis group managers and asks
  - Pulsar group manager goes to Grouper page, searches for her in Grouper/LDAP, adds her to pulsar analysis group, and emails to inform her
- User goes to pulsar pipeline page again, gets access, downloads pulsar pipeline and reads instructions
- User runs pipeline, which uses short-lived credential/proxy to launch job on CIT cluster
- User has gone from zero to hero in one day

Where We Are

What have you done for me lately?

- my.ligo.org is deployed
  - » Basic operations supported, development is ongoing
- Kerberos realm is operational
  - » Static web pages
  - » moin wikis, twikis, media wikis
  - » online document control system
  - » mod_auth_kerb used right now
- Git software repository
  - » Anonymous read via git protocol
  - » Write access tunneled through OpenSSH
    - – Kerberos (directly with ticket or via PAM)
    - – GSI-enabled OpenSSH (grid-mapfile managed by hand!)

Where We Are

What have you done for me lately?

- LDAP is operational
  - Twiki mapping from Kerberos principal to TwikiName via LDAP
  - Twiki authorization is done directly using LDAP
- Grouper 1.4 deployed
  - Back end to myLIGO, maintains all group membership
  - Only basic collaboration memberships supported
  - Still need to support most working groups

Where We Are

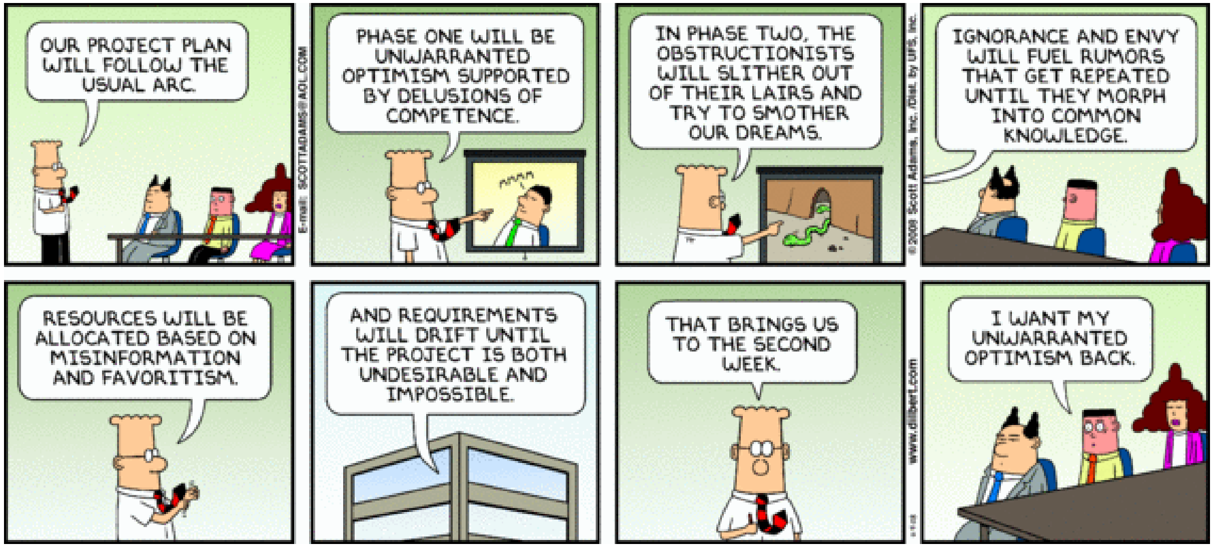
What have you done for me lately?

- Sympa deployed
  - No management interface supported yet, but in use for some groups
  - Group membership pulled from LDAP
  - Email addresses pulled from LDAP
  - LDAP group membership managed by Grouper
- Shibboleth, MyProxy, GridShib still on horizon
  - Shibboleth testbed operational
  - MyProxy thoroughly tested but not deployed
  - No work on GridShib so far (more later...)

Where We Are

Project Arc

- Our project is following the usual arc … we’ve reached panel 7 …


OUR PROJECT PLAN WILL FOLLOW THE USUAL ARC.
PHASE ONE WILL BE UNWARRANTED OPTIMISM SUPPORTED BY DELUSIONS OF COMPETENCE.
E-mail: SCOTTADAMS@AOL.COM
IN PHASE TWO, THE OBSTRUCTIONISTS WILL SLITHER OUT OF THEIR LAIRS AND TRY TO SMOTHER OUR DREAMS.
© 2008 Scott Adams, Inc./Dist. by UFS, Inc.
IGNORANCE AND ENVY WILL FUEL RUMORS THAT GET REPEATED UNTIL THEY MORPH INTO COMMON KNOWLEDGE.
RESOURCES WILL BE ALLOCATED BASED ON MISINFORMATION AND FAVORITISM.
AND REQUIREMENTS WILL DRIFT UNTIL THE PROJECT IS BOTH UNDESIRABLE AND IMPOSSIBLE.
THAT BRINGS US TO THE SECOND WEEK.
www.dilbert.com
I WANT MY UNWARRANTED OPTIMISM BACK.

- Project Contributors: Stuart Anderson, Warren Anderson, Phil Ehrens, Sam Finn, Jonathan Hanks, Scott Koranda, Diego Menéndez,Tom Nash, Shannon Roddy, Abe Singer,Hannah Williams (CIT, LHO, LLO, PSU, UWM)

What we need: GSI-OpenSSH

- A *critical* tool for LIGO
  - » No substitute for logging into a cluster to debug
  - » The “Grid vision” should not prevent...
- Adoption by major Linux distributions
  - » CentOS 5.3 and Debian Lenny are LIGO reference OS
  - » Go vote now!
  - » If primary OpenSSH developers won't adopt patch, should pursue Linux porters and/or specific Linux distributions?
- Specific logging for GSI authorizations
  - » Can this already be done with syslog(-ng)?
- Attribute-based authorization ala' GridShib
  - » eg. only members of “LIGO pulsar group” can login
  - » Note that OpenSSH is not coded in Java!

What we need: MyProxy

- Major functionality is all present and tested
  - » Used primarily to deliver short-lived credentials from embedded CA
- Attribute-based authorization and policy would be nice
  - » eg. scott.koranda@LIGO.ORG can obtain a short-lived credential with a 72 hour lifetime because he is member of group Communities:LVC:LSC:LSCAdminGroup
- Short-lived credential refresh
  - » Condor jobs that cache a short-lived credential/proxy should be able to use it to refresh and obtain a new short-lived credential/proxy
  - » Not uncommon for some LIGO jobs to run for a week
  - » Need to authenticate to publish results
- Can we cross the web-grid boundary?
  - » SPNEGO helps when going grid to web...
  - » If user authenticates first to IdP can we deliver grid credential?

What we need: GridShib

- Support for services NOT coded in Java!
- LIGO needs support for C and Python
  - GSI-enabled OpenSSH - C
  - globus-gridftp-server - C
  - Globus Replica Location Service (RLS) - C
  - LIGO Data Replicator (LDR) – Python under Apache httpd
  - LIGO Archival System (LARS) – Python under Apache httpd
  - Gravitational-wave candidate event database
    - (GraCEDb) – Python under Apache httpd
- Would like attribute based authorization via GridShib for all of these

What we need: all

- Regular and timely releases
  - » Prefer not tied to large Globus Toolkit releases
- Support for native packaging
  - » LIGO quickly moving to using only native packaging
  - » RPM/Yum for CentOS and Debian packages
- Transparent testing
  - » Prefer transparent nightly build tests on 64 bit CentOS 5.3 and Debian Lenny

What we offer

- Small but dedicated team with which to collaborate
  - Use LIGO/Condor and LIGO/Globus interactions as model
  - Bi-weekly (Condor) or monthly (Globus/CDIGS) calls
  - Responsible (we hope!) bug reporting and testing
- We prefer to consume rather than build our own
  - We don't do “not invented here” for computing infrastructure
  - If we can get what we need externally we will leverage it
- Production environment
  - We have lots of data
  - We have lots of scientists that need to analyze the data
  - This is not a drill!