A Sketch of the LIGO Identity Management System

Scott Koranda for the LIGO Scientific Collaboration

May 31, 2012
LIGO-T1200258-v3

Abstract

Harvesting the science content from the Laser Interferometer Gravitational Wave Observatory (LIGO) data is a collaborative effort between the instrumentalists, data analysts, modelers, and theorists from the LIGO Laboratory and the LIGO Scientific Collaboration (LSC). Realizing the full scientific potential of LIGO also requires collaboration between LIGO and other interferometer projects including the French and Italian Virgo project and the KAGRA project in Japan, as well as a full spectrum of other projects ranging from gamma-ray satellite experiments to numerical relativity groups.

Efficient collaboration begins with scalable and robust identity management infrastructure that can easily be leveraged and integrated with the wide spectrum of tools LIGO scientists use to collaborate and analyze the LIGO data. The LIGO Identity Management (IdM) project designs, deploys, operates, and supports the IdM infrastructure necessary to support the LIGO science mission.

1 LIGO and the LSC

The California Institute of Technology (Caltech) and the Massachusetts Institute of Technology (MIT) constructed and operate the Laser Interferometer Gravitational Wave Observatory (LIGO)[1] under a cooperative agreement with the United States National Science Foundation (NSF). LIGO operates two interferometers in the United States, one in Hanford, Washington and the second in Livingston, Louisiana. The combination of the Caltech and MIT LIGO staff with the LHO and LLO staff comprise the LIGO Laboratory with more than 200 current members.

Research groups at universities and other institutions sign memorandum of understanding (MOU) with the LIGO Laboratory to join the LIGO Scientific Collaboration (LSC)[2]. The LSC carries out the science mission of LIGO and is organized around three general areas of research: analysis of LIGO and other interferometer data searching for gravitational waves from astrophysical sources, detector operations and characterization, and development of future large scale gravitational wave detectors. The LSC was founded in 1997 and today includes more than 800 scientists from dozens of institutions and 13 countries worldwide.[3] Note that many, but not all, members of the LIGO Laboratory are members of the LSC.

The GEO project[4] is a German-British collaboration that built and operates the GEO600 interferometer in Hannover, Germany. Through an MOU signed with the LIGO Laboratory and the LSC GEO itself is a part of the LSC and all GEO members are members of the LSC. Likewise the Australian Consortium for Interferometric Gravitational Astronomy (ACIGA)[5] is wholly part of the LSC. More recently the Korean Gravitational-Wave Group (KGWG)[6] and the Indian Initiative in Gravitational-wave Observations (IndiGO)[7] consortiums have joined the LSC.

2 External collaborators

Scientists rely on analzying data from a world wide network of gravitational wave detectors to confidently detect and locate astrophysical sources of gravitational waves. The French and Italian Virgo[8] collaboration is composed of more than 200 scientists mainly from the Centre National de la Recherche Scientifique (CNRS)

and the Istituto Nazionale di Fisica Nucleare (INFN) laboratories and from the European Gravitational Observatory (EGO). The collaboration operates the Virgo gravitational wave antenna. The Kamioka Gravitational wave detector, a large-scale cryogenic gravitational wave telescope, is currently under construction in Japan by the KAGRA project[9].

Realizing the full science potential of LIGO and other gravitational wave detectors requires collaboration with scientists from other projects ranging from the NASA Swift Gamma-Ray Burst Mission[10] to the IceCube South Pole Neutrino Observatory[11] and the Numerical INJection Analysis (NINJA)[12] Project bringing together numerical relativity and gravitational wave data analysis. To date the LSC has signed MOUs with more than a dozen different projects for collaborative work to fully explore and realize the science potential of the LIGO and GEO interferometer data.

3 The LIGO Identity Management project

The LIGO Laboratory and the LSC (hereafter "LIGO") continue to grow and soon will include more than 1000 members from across four continents. Building and operating LIGO and analyzing interferometer data in such a collaboration requires a large number of working groups and committees, each with its own needs for electronic collaborative tools. The larger working groups or committees often require email lists, wikis or electronic notebooks, software version control repositories, and tools for audio and video conferencing. LIGO also operates collaboration-wide services used across many different groups such as the Document Control Center (DCC) and the LIGO Data Grid (LDG) used for computationally intensive data analysis. Other services include data streaming, detector characterization portals, and various metadata services necessary for data analysis.

Starting in 2007 the LIGO Identity Management project began to design, develop, deploy, operate, and support an IdM infrastructure capable of supporting the needs of the LIGO community and easing the burdens of working collaboratively across four continents. The primary goal of the LIGO IdM project is to enable each LIGO member to use a single electronic identity to efficiently access and consume all electronic resources, services, and tools necessary to carry out the LIGO science mission. Resources are categorized into three major types: web services, grid or cluster computing, and terminal or shell access.

4 Enrollment, identity, and group management

After formally joining LIGO either by signing an MOU with the LSC or becoming a member of the LIGO Laboratory each person enrolls electronically using the MyLIGO web portal[13].

4.1 MyLIGO

The current MyLIGO tool is a custom set of PHP code developed by and for LIGO. New members enroll using different workflows depending on how each is joining the collaboration. Enrollment flows for joining the LIGO Laboratory are more sophisticated than those for the LSC since the laboratory needs more closely resemble those of a classic structured enterprise rather than a distributed or virtual organization (VO) like the LSC.

After a successful enrollment the MyLIGO web portal stores information and collected attributes such as given name, family name, address, and telephone number in a MySQL relational database. Using given, family, and sometimes middle name(s) the portal creates for each user a unique identifier to serve as that person's single LIGO electronic identity. The IdM project has branded that LIGO electronic identity as the "albert.einstein" or "@LIGO.ORG" identity. The branding has played a critical role in the adoption and uptake of the IdM infrastructure by collaboration members.

4.2 LIGO electronic identifiers

LIGO electronic identities take the form `given.family@LIGO.ORG` and uniquely identify a single LIGO member. Identifiers are not reused. At the time the MyLIGO portal creates the LIGO identity or identifier it provisions that identity into the LIGO master Kerberos Key Distribution Center (KDC). Each LIGO identity is simultaneously a Kerberos principal in the `LIGO.ORG` realm. Associated with each identifier is a password or pass phrase. Users may reset or change their password using the MyLIGO portal. The portal uses the administrative interface to the master KDC to set the password for the Kerberos principal. The KDC is the only password store used for all LIGO identities and at no time are passwords stored in plain text or encrypted and stored in any other way besides the KDC.

4.3 LDAP

The MyLIGO portal further provisions an entry for each LIGO member into the LIGO master LDAP server. LIGO uses the OpenLDAP 2.4.x slapd server. Each LDAP record is an instance of the person, organizationalPerson, inetOrgPerson, eduPerson, posixAccount, krbPrincipalAux, eduMember, and qmailUser object classes and each includes the cn, eduPersonAffiliation, employeeNumber, employeeType, gidNumber, givenName, homeDirectory, krbPrincipalName, locality, mail, mailAlternateAddress, mailForwardingAddress, postalAddress, postalCode, sn, telephoneNumber, uid, and uidNumber attributes. Users may manage certain attributes such as mailForwardingAddress and postalAddress using the MyLIGO portal with changes provisioned into the MySQL relational database and LDAP server as necessary.

4.4 LIGO attributes and simple group management

Principal Investigators (PIs) for LSC MOU groups and LIGO Laboratory managers also use the MyLIGO portal to manage LIGO specific attributes and for simple group management tasks. For example, each PI for an LSC group must record the percent of a full time equivalent (%FTE) spent by the individual as a member of the group, the fraction of that time spent on research, and the fraction of available research time spent on LIGO. Those values are used according to collaboration bylaws to determine which LIGO members are authors on collaboration papers.

The PIs also use the MyLIGO portal as a simple front end for simple group management such as removing (de-enrolling) members from the group and for managing the representation of the MOU group on the LSC Council. The MyLIGO PHP code uses web services calls to drive LIGO's Grouper deployment, the data store and foundation for the majority of LIGO's group management.

4.5 Group management

LIGO leverages Grouper[14] from Internet2 for the majority of its group management needs. Only IdM project members directly use the Grouper administrative interface for group management. Simple LIGO-specific interfaces, including the MyLIGO group management front end, have been built to enable users to manage certain group memberships. For example many of the LIGO email lists are "opt-in" and managed by joining a particular group. A simple Javascript-based tool in the web browser that drives Grouper via its web services interface allows users to add or remove themselves from the email group(s).

The majority of groups and group memberships managed by Grouper within LIGO are provisioned or reflected into the LIGO LDAP master server using the Grouper `ldappc-ng` (renamed the Provisioning Service Provider or `PSP`) tool. At this time there is up to a 15 minute latency for a change in Grouper to be provisioned into LDAP, though after LIGO upgrades to Grouper 2.x and begins using the real-time provisioning PSP functionality the latency will be inconsequential.

4.6 COmanage and MyLIGO

Much of the functionality needed by LIGO for managing electronic identities including flexible enrollment workflows, attribute collection and provisioning, identifier management, and a front end for group man-

agement by Grouper is common to many scientific VOs, and more so for those VOs which are themselves federations of collaborative groups like the LSC. For this reason LIGO has partnered with Internet2 and the iPlant Collaborative[15] to develop COmanage[16], a platform for collaborative organization (CO) management. LIGO will build the next generation MyLIGO portal using the COmanage suite of tools as a base. Although LIGO is contributing a sizeable effort to the design and development of COmanage it will be a general (but customizable) tool able to support the needs of a wide spectrum of COs or VOs.

5 Authentication, authorization, and services

All authentication for services and tools supported by the LIGO IdM project use or will use Kerberos and the LIGO electronic identities (Kerberos principals) for authentication. No other credential store is used for authentication. To enable a robust authentication infrastructure available to a widely distributed set of services and resources the LIGO master KDC is replicated to a number of slave KDCs throughout the world.

To enable single sign-on for web services and tools LIGO has deployed the Shibboleth[17] Identity Provider (IdP) and a SAML2 based infrastructure. The production IdP currently delegates authentication to the Apache `mod_auth_kerb` module so that users logins and passwords are tested against the LIGO KDC. A future enhancement will use a dedicated Kerberos JAAS-based login handler more tightly integrated with the IdP to provide a more customized user experience and support extra functionality such as forced re-authentication for high risk services. As part of the SAML2 identity assertion the IdP queries the LIGO LDAP server network for user attributes including group memberships and asserts those attributes for consumption by services.

LIGO has deployed over 50 instances of the Shibboleth Service Provider (SP) to consume identity and attribute assertions from the IdP and manage access to web content like electronic notebooks, wikis, and data analysis results. Working together with the IdP the SPs support a single sign-on experience across the majority of LIGO web services. Using attribute assertions from the IdP the SPs manage authorization to services. Most authorization decisions are based on group memberships as asserted by the IdP, which retrieved them from LDAP where they had been provisioned by the Grouper suite of tools.

LIGO Data Grid users currently authenticate to LDG resources including Linux clusters used for data analysis using RFC 3820 proxy certificates (derived from X.509 certificates) and tools enhanced to support them using the Globus Grid Security Infrastructure (GSI)[18]. Today LIGO users in the United States request and receive X.509 certificates signed by the DOEGrids[19] certificate authority (CA) while users from other countries rely on regional or national CAs that are members of the International Grid Trust Federation (IGTF)[20]. Each user is responsible for managing her own certificate and private key and the encryption of the private key is neither managed by the LIGO IdM infrastructure nor related in any way to her LIGO identity (Kerberos principal).

Soon, however, LIGO users will begin to retrieve short-lived X.509 certificates from the CILogon[21] service after authenticating via the LIGO IdP using their LIGO identities (Kerberos principals). The short-lived X.509 certificates and RFC 3820 proxy certificates based on them can be used with the current set of GSI-enabled tools. This change will reduce the burden of users having to manage their own certificate and private key and directly tie LDG authentication to the single LIGO identity.

Authorization for access to LDG resources is currently managed via static access control lists or grid-mapfiles. Each static grid-mapfile lists the X.509 certificate subjects authorized to access the resource, and when necessary include a mapping to a local account needed by the service. After transitioning to using short-lived certificates issued by the CILogon service where the certificate subject name is directly tied to the LIGO identity or Kerberos principal authorization will be managed using grid-mapfiles derived programmatically from LDAP. The group of users that should be authorized to access a particular LDG service will be managed using Grouper with the group membership being provisioned into LDAP. The grid-mapfile generation tool will simply query LDAP with the name of a group(s) that should be authorized to obtain memberships and receive a list of X.509 subject names derived from the corresponding LIGO identities.

Shell and terminal access to general computing resources at the LHO site, as well as authenticated access to email services like IMAP, POP, and SMTP is managed using Kerberos for authentication and authorization

against LDAP groups. A particularly elegant design choice made by the IdM project architect at LHO is to use a local Kerberos realm just for the LHO site and enable cross realm trust so that LHO members may seamlessly use their @LIGO.ORG identities while preserving flexibility for the local LHO infrastructure. At this time the LDAP groups used for authorization are not managed using Grouper but that enhancement is planned for the near future. The other LIGO Laboratory sites are at various stages of transitioning their infrastructure to leverage a similar design.

6 Federation with external collaborators

As detailed above, LIGO scientists need to efficiently collaborate with researchers from a number of other projects to realize the full science potential of the LIGO data. Federated identity management and tools that consume federated identities streamline collaboration and lower the burdens on both users and administrators.

For web services LIGO has chosen to leverage its SAML2/Shibboleth infrastructure to enable federation. LIGO joined the InCommon[22] identity federation in the United States and is beginning to pursue federation with other SAML based identity federations in Europe, Japan, Australia, and Canada.

The LIGO InCommon administrator has injected metadata for the Compact Binary Coalescence (CBC)[23] wiki into the InCommon metadata to enable federation of that SP. As a first use case access to particular wiki pages necessary to support a working group was granted to a non-LIGO researcher from MIT who authenticated using his MIT identity and gained access to the resource. Work is ongoing to inject metadata for the Data Analysis Software Working Group (DASWG)[24] wiki and the main LIGO wiki served from wiki.ligo.org to enable access to those resources for other LIGO collaborators. The LIGO InCommon administrator will soon inject metadata for the LIGO IdP to enable the CILogon functionality detailed above in support of obtaining short-lived X.509 credentials for LDG access.

Work is also underway to directly federate the SP supporting the LIGO Document Control Center (DCC) with the University of Tokyo IdP to enable federated access to the DCC for a number of KAGRA scientists in support of a LIGO-KAGRA project. We expect that federation to be completed this summer.

Unfortunately not all scientists with which LIGO researchers need to collaborate have access to secure and well managed federated identities. For this reason LIGO has deployed an "identity provider of last resort". The LIGO Guest infrastructure[25] provisions electronic identities of the form `given.family@GUEST.LIGO.ORG` for collaborators with no access to federated identities but who need to access LIGO web resources. The LIGO Guest IdP is federated with the CBC wiki and has been used by two NASA collaborators to access certain wiki pages in support of some joint work.

Managing access to resources for federated identities brings with it its own challenges and adds a burden to LIGO IdM project administrators. To ease that burden LIGO will also leverage COmanage for managing external collaborations along with its own collaborative structure. The goal is to allow LIGO scientists themselves to quickly define working groups and manage authorizations to LIGO web resources based on them without the need to involve an IdM project administrator.

Federation within the grid space and federated access to LDG resources occurs through LIGO's reliance on X.509 credentials issued by CAs that are members of the IGTF. Likewise LIGO scientists may use their current X.509 certificates issued by the DOEGrids CA or other regional CAs to access non-LIGO grid resources. Federated access for LIGO scientists to non-LIGO grid resources will continue after LIGO transitions to using the CILogon service since the CILogon CA initially targeted for use is recognized by many grids. Note, however, that the initial CILogon CA that LIGO will use is not IGTF accredited. Use of the CILogon CA that is accredited by IGTF will require LIGO to achieve the InCommon Silver accreditation for its IdP. LIGO plans to take that action and assert InCommon Silver in the future.

7 Discussion

Below we address the specific workshop requirements regarding identity management infrastructure:

- *An example of an infrastructure component*: We detail above LIGO's current and planned use of a number of IdM infrastructure components including Kerberos, LDAP, Internet2 Grouper, Internet2 Shibboleth IdP and SP, CILogon, and COmanage.

- *An example of an outside infrastructure component utilized by the project*: LIGO currently leverages the InCommon identity federation infrastructure to enable federated access to particular LIGO web resources. In the near future LIGO will leverage the CILogon service through its membership in InCommon to support LIGO users obtaining short-lived X.509 credentials for grid access.

- *An example of an infrastructure component utilized by other projects*: The LIGO "identity provider of last resort" is currently being utilized by researchers from NASA to obtain access to certain LIGO resources since at this time NASA researchers do not have access to SAML2 federated identities (NASA plans to soon offer that functionality). As COmanage matures we expect other science projects to leverage the tool both by deploying instances of it and through a hosted service managed by Internet2.

- *An example of an infrastructure component where the project is pushing the boundaries of the current state of the practice*: COmanage is a research project intended to identity and address challenges in collaboration management for large distributed collaborative organizations. Other research and products in this space include SURFconext[26] from the Dutch SURF organization and various services from the Globus Online[27] project.

- *A particular security or social factor affecting the ability to utilize externally managed infrastructure components*: LIGO envisions leveraging externally managed infrastructure including the InCommon infrastructure, CILogon, and certain Globus Online services. With each, however, the questions regarding long term funding and availability of the service remain. InCommon through its membership fees appears to have reached a sustainable model. It is unclear how CILogon and Globus Online services will achieve sustainable funding and if subscription or membership fees are required if LIGO will be able to pay those fees given its own unique collaborative makeup. At this time no costs are incurred by LSC members to leverage the IdM infrastructure developed and supported using funding from a small set of LSC members. The brunt of the direct cost has been born only by the LIGO Laboratory with, for example, the InCommon membership fees as an example. It is unclear how the LIGO Laboratory could recover part of those costs from the LSC.

References

[1] B. Abbott, et al., *The Laser Interferometer Gravitational-Wave Observatory*, Rep. Prog. Phys. 72 (2009) 076901

[2] http://www.ligo.org

[3] http://roster.ligo.org

[4] H Grote (for the LIGO Scientific Collaboration), *The status of GEO 600*, Class. Quantum Grav. 25 No 11 (7 June 2008) 114043 (9pp)

[5] http://www.anu.edu.au/Physics/ACIGA

[6] http://kgwg.nims.re.kr

[7] http://www.gw-indigo.org

[8] https://wwwcascina.virgo.infn.it

[9] http://gwcenter.icrr.u-tokyo.ac.jp/en

[10] http://www.nasa.gov/mission_pages/swift/main/index.html

[11] http://icecube.wisc.edu/

[12] https://www.ninja-project.org/doku.php

[13] http://my.ligo.org

[14] http://www.internet2.edu/grouper/

[15] http://www.iplantcollaborative.org/

[16] https://spaces.internet2.edu/display/COmanage/Home

[17] http://shibboleth.net/

[18] http://www.globus.org/security/overview.html

[19] http://www.doegrids.org/

[20] http://www.igtf.net/

[21] http://www.cilogon.org/

[22] http://www.incommon.org/

[23] https://www.lsc-group.phys.uwm.edu/ligovirgo/cbc/

[24] https://www.lsc-group.phys.uwm.edu/daswg/

[25] https://guest.ligo.org/

[26] http://www.surfnet.nl/nl/Thema/coin/Pages/Default.aspx

[27] https://www.globusonline.org/