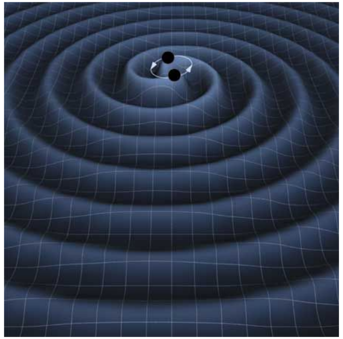
SciTokens in LIGO

James Alexander Clark (LIGO lab / Caltech) & Ron Tapia (PSU ICDS / IGC)

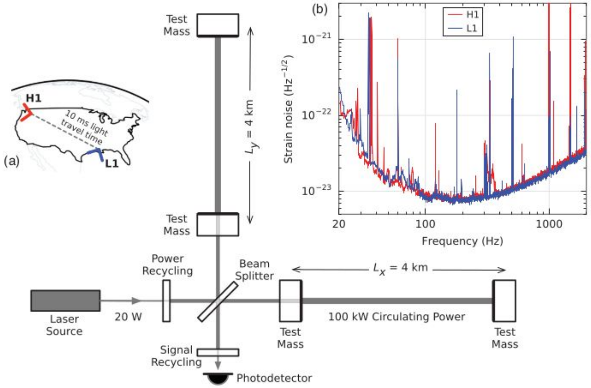
WoTBAn&Az 2021: Oct 18, 2021

Gravitational Waves

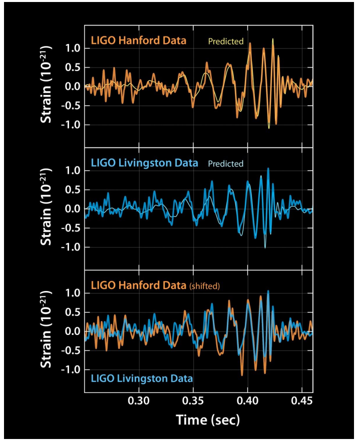
- Perturbations of the space-time metric produced by rapid changes in shape and orientation of massive objects.
- Gravitational waves carry information from the coherent, relativistic motion of large masses



Artist's impression of gravitational waves from two orbiting black holes. [Image: T. Carnahan (NASA GSFC)]

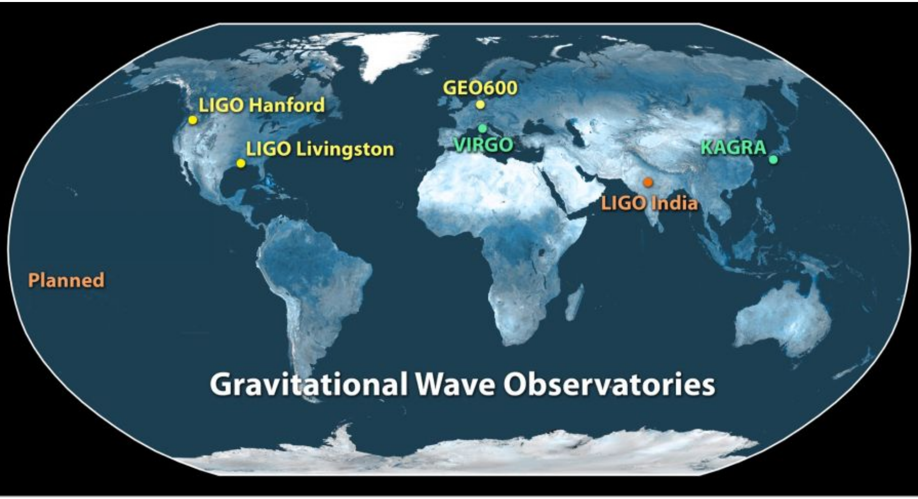


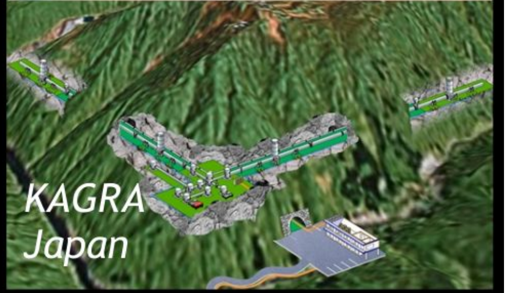
https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.116.061102



- Passage induces stretching / squeezing tidal strain, h ~ Delta L / L
- A “strong” gravitational wave: displacements (Delta L) on the order of 10^{-18} meters
- Detection: multiple large laser interferometric detectors & digital signal processing

International Gravitational Wave observatory Network (IGWN)





LIGO Hanford (WA)

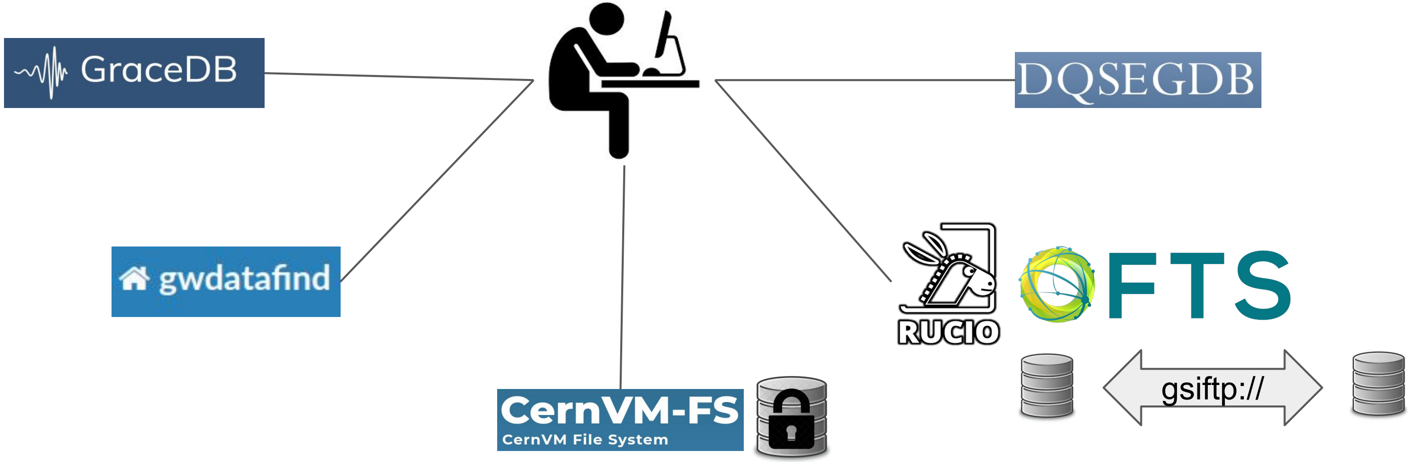
LIGO Livingston (LA)

Virgo (Italy)

GEO600 (Germany)

LIGO Computing & An/Az

- LIGO operates & uses a variety of mission critical services with specific auth. requirements



Currently use X.509 for authentication:

- LIGO & Virgo users receive a short-lived proxy certificate from CILogon via CLI tool ligo-proxy-init
- Services use proxy certs generated from robot certificates

Services using X509

- **CVMFS:**
  - LIGO hosts embargoed instrument data in CVMFS for distributed HTC (e.g. OSG) workflows; embargo is currently enforced via X509 credentials, with proxy certs passed along with HTCondor jobs
- **GWDataFind**:
  - instrument data-discovery utility; users to query for the location of files containing gravitational-wave detector data for consumption by data analysis pipelines.
- **DQSegDB**:
  - data quality segment database service & client package used to store, access instrument status metadata
- **GraceDB**:
  - Gravitational-Wave Candidate Event Database, provides a centralized location for aggregating and retrieving information about candidate gravitational-wave events
- **Rucio / FTS**:
  - Bulk archival data management and replication; operator authentication currently through SSH & transfers between GridFTP end points authenticated with delegated X509 proxy

X.509 for Authentication

- X.509 certificates for authentication
- Authorization via group membership (LDAP)
- Possession of cert implies identity
- Each service responsible for authorization configuration
- Possessor of certificate is entitled to *all* capabilities granted to identity
- Services using X.509: XRootD, CVMFS, DQSegDB, GWDataFind, GraceDB, GridFTP

Some statistics from the last 15 days:

- Unique robot entity that have invoked x509 credentials —> 35.
- Average number of times a robot x509 credentials called daily —> 54
- Unique people who have have invoked x509 credentials —> 333.
- Average number of times a personal x509 credentials are called daily —> 230

SciTokens Motivation

- OSG plans to retire Grid Community Toolkit (Jan 2022)
  - Implications for CVMFS, GSI OpenSSH, Grid FTP
- Improved security:
  - Capabilities based authorization vs identity based authorization
- LIGO observing run O4 early start date: June 2022

SciToken Goals

- Replace X.509 certificates with SciTokens
  - Retire ligo-proxy-init
- Replace grid-mapfile authorization with capabilities-based authorization
  - Grid map files are used by sites to associate X.509 distinguished names with a local users
- Migrate to federated identity
  - Remove reliance on LIGO.ORG kerberos
  - Kerberos supported but not required

SciTokens Use Cases

- HTCondor Jobs
  - Access data: XRootD, CVMFS, StashCache
  - Access GraceDB
- CLI Tools on cluster submit nodes
  - DCC, GraceDB, DQSegDB, GWDataFind
- Robots:
  - Cron jobs accessing DQSegDB
  - CI jobs (GitLab)
- The researcher’s laptop
  - Only 64-bit Linux *must* be supported
  - Other operating systems are supported as best-effort

The LIGO SciTokens Team

- System administrators and programmers
- Mostly, not IAM people
- Mostly IAM adjacent
- Mostly concerned with running services that require authn/authz

Consult Experts

- Jim Basney (NCSA) and the SciTokens/SciAuth projects
    - https://scitokens.org/
    - https://sciauth.org/
- Dave Dykstra (FermiLab) help with HashiCorp Vault/CILogon
- Brian Bockelman (Morgridge Institute for Research) XRootD
- Bi-weekly working meetings
- OSG Slack

HTCondor Local Issuer

- SciTokens generated by HTCondor credmon
- `iss` in the token set to https://scitokens.org/ligo
- Static website based on https://github.com/scitokens/ligo
- Private key configured into HTCondor
- Public keys manually added to https://scitokens.org/ligo/oauth2/certs
- OSG XRootD configured to trust https://scitokens.org/ligo
- OSG XRootD configured to map scopes to file system paths

Local Issuer Lessons

- **Issuer** is an overloaded term
  - The value of `iss` in the SciToken payload (serves public key)
  - The generator of SciTokens (uses private key to sign token)
- Easy Condor-only solution
- Not easily adaptable to non-HTCondor use cases

Vault + CILogon

- SciTokens served by HashICorp Vault server
  - vault.ligo.org
  - CLI client: `htgettokens`
- Vault configured to use CILogon:
  - cilogon:/client_id/caltech/ligo/test
  - cilogin:/client_id/caltech/ligo/prod
- `iss` in SciToken set to: https://cilogon.org/ligo
- JWKS discovery: https://cilogon.org/ligo/.well-known/openid-configuration
- JWKS (certs): https://cilogon.org/oauth2/certs
- XRootD configured to trust https://cilogon.org/ligo

Vault + CILogon Features

- htgettoken paper by Dave Dykstra
  - https://github.com/fermitools/htgettoken/files/6063416/CHEP21_Paper_Htgettoken.pdf
- HTCondor integration
- Kerberos support & convenient CLI → easy integration & adoption with existing workflows
- Support for long-lived processes/robots
  - Method 1: User stores refresh token in a vault path accessible by a Kerberos credential
  - Method 2: Vault admin gives user an indefinitely renewable vault token
- Supports *researcher laptop* use case
- Direct line & support from developers :)

HTCondor with Vault

- *Using Vault as the OAuth client* in HTCondor Admin Manual
- Install `condor-credmon-vault` from the HTCondor yum repository
- Vault config based on: https://github.com/fermitools/htvault-config
- Tokens fetched using https://github.com/fermitools/htgettoken
- `htgettoken` is available in OSG 3.5 yum repository

Current vault deployment:

- vault.ligo.org: single VM (2 CPUs, 2G RAM, 20G HDD), hosted by LIGO lab @ CIT
- Expect to exploit native HA support for production, machine specs TBD

Current State

- Development/testing server vault.ligo.org
- Researcher Laptop use case supported for CVMFS (note: Kagra)
- HTCondor use case supported for CVMFS
- Robot use case supported
- OIDC workflow an alternative to kerberos
- Path forward for other services that use X.509

Next steps

- Finalize namespaces:
  - Requires coordination
    - CILogon configuration
    - LDAP/Grouper groups
    - Service configuration/behavior
  - Audience values
  - Scope values

Outstanding Issues

- Can use SciToken to access XRootD metadata, but not data.
    - This is a known issue and being worked on
- Tracing/auditing
    - Services need to be able to associate access via a token with a responsible party
    - Identity is encoded in the sub claim.
    - Use LIGO username as sub claim `albert.einstein`
    - Honor system to not use sub for authorization
    - GDPR implications for VIRGO users?

Timeline for a rolling transition (i.e., X.509 → mix of X.509 & scitokens → scitokens):

- ~End of 2021 / start of 2022: Production-level Vault service and CILogon configuration, integration/support for authenticated LIGO data in CVMFS
- ~June 2022: Transition all services by start of next observing run (O4)

Acknowledgements

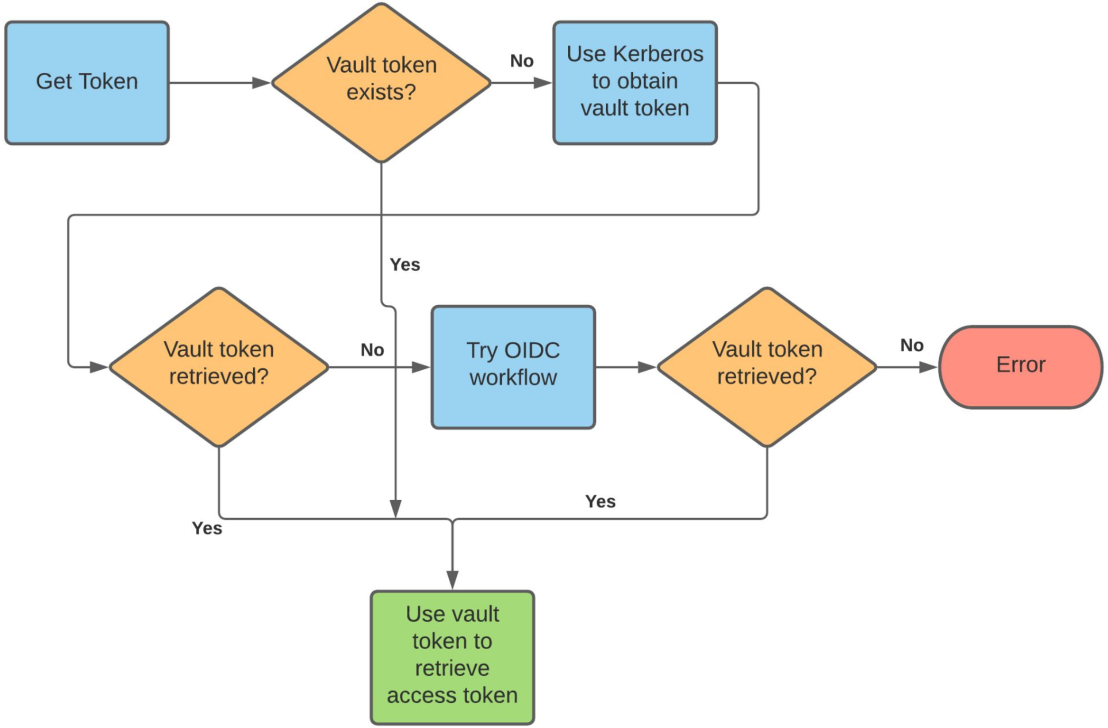
- This work was supported by NSF Grants PHY-1764464, PHY-1700765 and PHY-2110594
- This material is based upon work supported by OAC-1738962

extra

HashiCorp Vault

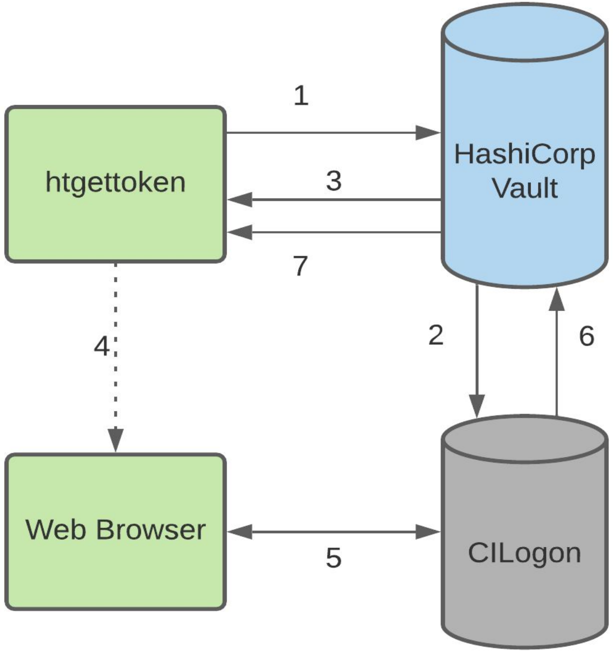
- Secure path-secret storage
- Secrets can be dynamic
- Supports authentication plugins (`vault-plugin-auth-jwt`)
- Supports secret backend plugins (`vault-plugin-secrets-oauthapp`)
- Supports kerberos authentication

htgettoken Flowchart



OIDC Workflow

Only done if kerberos fails.



1. htgettoken contacts vault server.
2. Vault contacts CILogon to start transaction.
3. Vault responds with a URL and then htgettoken asks the user to use a browser to complete the workflow.
4. The user uses a browser to complete the workflow.
5. The user is redirected to a CILogon URL, where the user selects an identity provider and authenticates.
6. After successful authentication, CILogon contacts vault and sends vault refresh and access tokens.
7. Vault responds to htgettoken with a success or failure message. Upon success it sends a vault token and an access token.

OIDC Workflow Video

Permission denied at 1:08. Use gear to set quality to 1080p.