Leveraging Globus Identity for the Grid

Suchandra Thapa
GlobusWorld, April 22, 2016
Chicago
Open Science Grid

Helps researchers speed up their research using high throughput computing methods
Helps campus HPC administrators share resources for multi-campus and national collaborative research
Last 30 days: 100M core-hours
Last 12 months: 200 Million jobs consumed 1 Billion hours of computing involving 1.5 Billion data transfers to move 223 Petabytes
Accomplished by federating 114 clusters providing 1h-100M hours each
OSG: 114 resource endpoints

... with campus users far and wide
... with X.509 Auth for virtual organizations

Replace the need for users to have X.509 proxies
Reduce time between sign-up and job submission
(hours not days)
→ OSG Connect
x.509 virtual organization validation occurs during wide area job distribution. 
trust relationship between the resource provider and the OSG VO. 
users not required to use x.509 certificates directly.
OSG and CI Connect

CI Connect provides a framework to create entry points to use OSG and campus resources
OSG Connect functions central entry point for campus-based users and individual PI’s
Access to resources using the OSG VO and glidein service
Potential integration with campus specific resources
An identity bridge: OSG Connect identity (Globus) • virtual organization roles (OSG)
CI Connect Growth

OSG Connect
ATLAS Connect
CMS Connect
Duke CI Connect
UChicago CI Connect

> 750 User Sign-ups

~20 new users/month
Establishing identity

How do we get from <campus researcher> to user=angus on OSG Connect? Globus Auth provides an answer.
IdM with Globus Auth

Customer-directed team management

SSH key upload (for login shell)
Python Client

- Globus Auth exposes a REST API
- Globus provides Python module
  - https://github.com/globus/globus-sdk-python
- Using the module, we:
  - search our group hierarchy for relevant changes
  - store these changes locally to track current state
  - provision user accounts into directory (nss_nis)
  - provision user files and directories in storage systems
  - populate local groups as a direct mirror of Globus Groups
  - define access rights to GridFTP and job submission
from nexus import GlobusOnlineRestClient

cfg = {
    'server': 'nexus.api.globusonline.org',
    'client': 'osgconnect',  # service account
    'client_secret': 'password',  # secret!
}

gc = GlobusOnlineRestClient(config=cfg)
headers, response = gc.get_group_members(groupuuid)
members = response['members']
members = [member for member in members if member and member['username']]
members.sort(lambda a, b: cmp(a['status'], b['status']) or cmp(a['username'], b['username']))
for member in members:
    print '%s (%s) %s' % (group, member['status'], member['username'])
    headers, profile = gc.get_user_profile(member['username'])
    if profile.has_key('credentials'):
        keys = sorted([cred['ssh_key'] for cred in prof['credentials'] if cred['credential_type'] == 'ssh2'])
        # store ssh keys into ~/.authorized_keys
Integrating with Login Hosts

Want to reduce the credentials users need to remember and use

Used curl C library and REST interface to authenticate logins

- Service integrated with Linux using PAM modules

```c
/* If authtok exists, try to authenticate with the old authtok. */
if (pass) {
    rc = globus_authenticate(NEXUSBASE, user, pass);
    if (rc == PAM_SUCCESS)
        return PAM_SUCCESS;
}
```
Summary: Data flow architecture

OSG Connect
• ~614 users
• ~104 projects
• ~120 campuses

Architecture extended to other campus integrations via CI Connect:
ATLAS
CMS
Duke University
UChicago
Future directions

Update infrastructure to use new Globus Auth API and interface

'Genericize' infrastructure and extend to support users and clusters in Virtual Clusters for Community Computation (VC3)

Further integration with Globus transfer
Thank you!

And our thanks to the Globus, CILogon and OSG teams. In particular:

Rachana Ananthakrishnan (Globus)
Mattias Lidman (Globus)
Stephen Rosen (Globus)
Kyle Chard (Globus)
Mats Rynge (OSG)
Further information

Open Science Grid

http://opensciencegrid.org/

OSG Connect

http://osgconnect.net/

Globus Python SDK

https://github.com/globus/globus-sdk-python