OSG Needs and Desires for AAI

Brian Bockelman

Disclaimer!

- I am the “technology guy” for OSG, not necessarily the “security guy”.
    - However, I think this presentation is aligned with the security team.
        - Just want to apologize in advance for mistakes!
- Additionally, today I am at the kickoff meeting for another project, SciTokens, which operates in this space.
    - So, apologies for missing the working group meeting! Hopefully I have good things to report afterward!

OSG Needs

- OSG believes that **authentication** is best handled by the VO.
    - Historically, the VO ‘infrastructure’ has been the de-facto approver of the grid certificate.
        - Additionally, the certificate is worthless unless combined with a VOMS authorization.
        - Hence, the VO authenticated twice: once for the global identity (GSI cert) and once to join the VO.
- In ~2013, OSG made the decision that only one authentication is needed: the VO can approve the user once.
    - If the VO can demonstrate they authenticate users responsibly, then **no user certificate is needed**.
- This often works because OSG is composed of a small number of largish VOs.

OSG Needs

- Hence, OSG only needs users to authenticate with the VO:
    - **Transient trust** (OSG site trusts the VO; VO trusts the user) allows us to trust the user.
    - The *end-user identity thus exists only in the VO context*.
        - There is not a global `bbockelm` user; there is a `CMS:bbockelm` and `OSG:bbockelm`, however.
    - This may mean the user has multiple identities if he or she is in multiple groups.
- With regards to the above “needs”, the X509 ecosystem is overkill.
    - We have been working to reduce the user’s need for X509 identities for several years. Remaining use case: storage!

OSG Desires

- OSG does not want to handle identity management for users.
    - Identity management should be done through trustworthy VOs.
    - Strong preference to use growing federated identity infrastructure.
- Credential acquisition should be completely automated after logging in to VO-provided user interface (lxplus SSH terminal, Jupyter notebook). No separate “XYZ-proxy-init” commands anywhere!
    - There may be a need for a one-time grant of permissions. “Do you trust the CMS submit infrastructure with your CERN ID? If you click yes, CMS will have access to XYZ”
    - Used everywhere on the web; users seem to be comfortable with this model.

On Tracebility

- Traceability is an important (sometimes legal!) requirement from our sites.
- However, there may be two traceability use cases:
    - Traceability for **legal & auditing**: Accuracy is most important, even if process is slow.
    - Traceability for **debugging jobs**: Quick debugging is most important, even for slight reduction in accuracy.
- Important: the two use cases may be fulfilled by distinct mechanisms.
- It seems acceptable to have the VO use privacy-preserving identifiers that require the site to contact the VO to “map back” to a user.
    - However, this can significantly harm the site’s ability to help debug jobs.
    - It’s not quite clear how the privacy-vs-“debugability” question will play out!
- OSG today is satisfied if the VO can do tracing exercises similar to what the WLCG did in the summer: given a timeframe and a host identifier, can one determine all possible users who might have run on it?

Using SciTokens

- SciTokens (https://scitokens.org) is a recently funded NSF project to help particular science communities bootstrap an authorization-focused infrastructure, as opposed to an identity-focused infrastructure.
    - Its kickoff meeting is why I'm missing today's discussion!
- Goals include:
    - Define an OAuth2-inspired access token format (a "SciToken"). Can provide the bearer with authorizations to the VO's area in remote site storage. For scalability purposes, tokens can be verified in a distributed manner.
    - Develop software infrastructure (libraries, plugins) for these tokens to be utilized at service endpoints. Planning on a Python, C, and Java library. Plugins for (at least) Xrootd and CVMFS; likely will do something for FTS3.
    - Develop HTCondor integration so HTCondor can acquire, transport, and renew the tokens for running jobs. This integration aims to be generic - should be able to integrate with any OAuth2 ecosystem (such as Dropbox).
    - Improve the CILogon infrastructure so it can issue SciTokens.
    - Demonstrate/deploy usage of the tokens for our science stakeholders (LIGO/LSST).
- **NOTE**: Right now, this focuses primarily on the access to resources. Sidesteps how identity is established.
    - Intent is identity is established through OAuth, but working on demos to do this via Unix auth, CERN SSO, and X509 credentials.
- Done well, I believe this project can meet a number of OSG and WLCG needs.

Conclusions

- OSG takes a very VO-centric view of identity and is willing to trust (& verify) the VOs to comply with policies.

- Within this vantage point, the current ecosystem is wildly overkill; for non-WLCG *users*, we have been transitioning away for several years.

- We see the OAuth2 / OIDC ecosystem beginning to spin out many useful software products and integrations. We'd like to take advantage of these - and the SciTokens project will bring integrate several of these technologies with our everyday OSG-supported products (HTCondor, CILogon, Xrootd, CVMFS, FTS3).

    - Particularly, this solves the issues with accessing remote storage services using the same tried-and-tested model as ALICE.

- **In short**, we welcome and support more progress toward use of federated identity by WLCG!