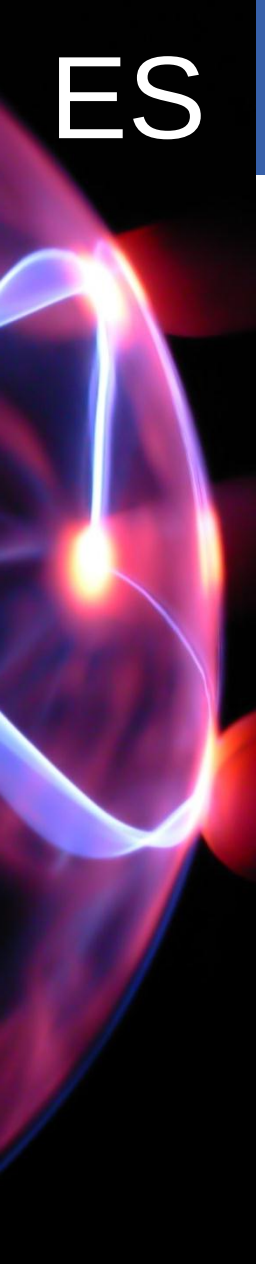


Authentication and Authorization (AAI)

issues concerning

Storage Systems and Data Access

Pre-GDB, 2013-02-12

Maarten Litmaath

History

- Security TEG WG “AAI on Storage Systems”
  - Jointly with Data & Storage Management TEGs
  - https://twiki.cern.ch/twiki/bin/view/LCG/AAIOnStorageSystems
- Summary of status quo and issues presented in pre-GDB of Feb 7, 2012
- Some further discussion on Security/DM/SM TEG mailing lists
- Summary document at v0.4 since Apr 2, 2012
  - Attached to the Twiki page

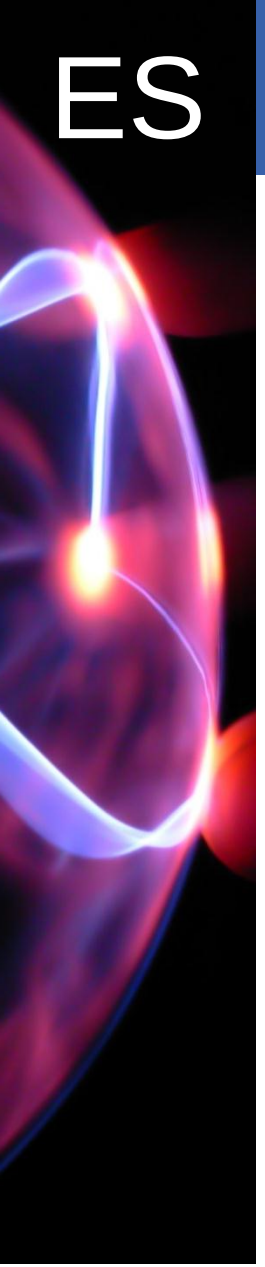
Status quo (1)

- Storage element and catalog configurations
  - Production/group data is protected against modification by unprivileged users
  - Tape access by unprivileged users is essentially prevented
  - User and group access are regulated by the experiment frameworks
    - Including quotas
    - An SE may in fact be more permissive than desired → to be checked and fixed as needed

Status quo (2)

- Dealing with X509 overhead
  - Use bulk methods, sessions or trusted hosts where needed
    - Hosts should at most be trusted within sites
  - Look into cheap, short-lived session tokens?
    - Implementation may be expensive
- ALICE use security *envelopes* for data operations
  - Minimal potential damage when stolen
- CASTOR NS and RFIO backdoors are being closed
  - Insecure legacy clients being phased out

Data protection issues (1)



- Do different data classes need the same protection?
  - Custodial → highest security to protect against tampering
  - Cached → can be replaced, but can the security therefore be relaxed?
  - User → highest security to protect against tampering and reading by others

Data protection issues (2)

- Read access based on data confidentiality?
  - Requires careful management of categories → expensive and/or bound to fail?
    - At least access to user data should be managed carefully
  - By default none of the data of one VO shall be readable by another
    - Thus incur security overhead also for large volumes that could be made public
      - Use bulk methods, sessions, trusted hosts as needed
  - Most of the data is transferred over insecure channels!
    - Bulk encryption may remain too expensive

Data protection issues (3)

- Access audit trails are important for traceability
  - For security
    - Tampering with existing data
    - Upload of illegal data
    - Access to confidential data
    - DoS attack
  - For investigations of performance issues
    - Which DN / VO is overloading my service?
- Traceability questionnaire:
  - https://twiki.cern.ch/twiki/bin/view/LCG/WLCGSecTraceability
  - Does your service record the necessary things, store the information in the right place with the right retention period and access control?

Data protection issues (4)

- Protection is needed or desirable against:
  - Information leakage → not only the data, but also the file names matter
    - “Higgs-discovery.root”
    - Read access to directories and logs to be regulated
    - Encrypt selected low-rate traffic?
  - Accidental commands → restrict write access as much as possible
    - Protect generic production data
    - Confine each analysis group to its own area
    - Protect user data
  - Attacks
    - See next page

Data protection issues (5)

- Protection against attacks
  - By outsiders exploiting a bug or doing a DoS
  - By insiders, or outsiders using stolen credentials
    - Maintain as few privileges as possible per VO member
      - May also help prevent accidents…
    - Use multi-person authorization (“two-man rule”) for bulk data removal of VO-wide or group data?
      - Ditto…
- Also the SE itself needs protection
  - Against upload of illegal data
  - Against DoS

Issues with data ownership (1)

- Missing concept: data owned by a VO, subgroup or service.
  - Historically the DNs of production managers were and are still being used as owners
    - What if such a person moves on?
      - See next page
    - What if such a person also works on private analysis?
      - Privileged access to data may remain
        - » Wrong proxy used by mistake for analysis
        - » Storage element may couple privileges to DN
  - Use robot certificates instead?
    - At least ATLAS do that already

Issues with data ownership (2)

- Mapping a person to/from a credential
  - DN changes may affect data ownership
    - Ownership could be by the user's VOMS nickname or a generic attribute instead of the DN
      - A big change, but maybe well worth the investment?
      - Can we do this per SE type?
      - Can we do this per SE instance?
      - Natural nickname candidate: CERN account name
        - » Each user is in the CERN HR DB → can have an account
      - ATLAS already use nicknames in some of their services
    - ALICE: LDAP server supports multiple DNs per user
      - Might such an approach be less cumbersome?

Issues with data ownership (3)

- Mapping a person to/from a credential
  - X509 vs. Kerberos access
    - Is this only relevant for a few cases, e.g. EOS?
      - Try to avoid encumbering other SE types or instances
    - How to map Kerberos principal to/from DN?
      - Yet another map-file?
      - Use principal for ownership, map DN to it?
      - Use principal as VOMS nickname or generic attribute?
        - » See previous page
    - VOMS groups could be determined from map-file
      - Cf. dCache “vo-role-mapfile”
      - For writing a primary group may need to be decided
        - » It could be taken from the directory instead

Issues with data ownership (4)

- VO superuser concept desirable?
  - Avoid bothering SE admin for cleanups
  - Useful at least to ATLAS
  - How can different SE types implement this?
    - Maybe more relevant for certain types?

Cloud storage

- Do we need to be concerned about cloud storage in the context of these discussions?
  - Can cloud storage technologies suggest answers to some of the questions?
  - Do we just need to be concerned with a grid layer on top of such storage?
    - Cloud storage might be just another back-end