Authentication and Authorization Infrastructure for Data Storage and Access

Version	Date	Author	Comment
0.1	2012/03/18	Maarten Litmaath	Initial version
0.2	2012/03/19	Maarten Litmaath	ALICE approaches added, conclusions slightly adjusted
0.3	2012/03/26	Maarten Litmaath	Incorporated input from Paul Millar and Alessandro De Salvo
0.4	2012/04/02	Maarten Litmaath	Incorporated input from Jakob Blomer and Von Welch

Contents

Authentication and Authorization Infrastructure for Data Storage and Access 1
Abstract 4
Introduction 4
1. Status quo 4
2. Data protection 5
3. Issues with data ownership 6
4. Quotas 7
Conclusions 7

Abstract

In 2011 the WLCG Management Board mandated a set of Technical Evolution Groups (TEGs) to reassess the implementation and usage of the grid infrastructures that make up WLCG and document a strategy for technical evolution. Within the TEG on Security a working group was concerned with the Grid security aspects relevant to the storage and access of data.

Introduction

This document briefly describes the status quo concerning security in data storage and access, followed by reasons for and requirements on data protection. Next a number of issues related to data ownership are listed, followed by a brief overview of the area of quotas, to some extent related to security. Action items are listed in the conclusions.

1. Status quo

- SE and catalog configurations
    - Production data is protected against modification by unprivileged users
    - Some experiments prevent tape access by unprivileged users
    - User and group access is regulated by the experiment frameworks
        - Including quotas
        - An SE may be more permissive than desired → to be checked and fixed as needed
- Dealing with X509 overhead
    - Use bulk methods, sessions or trusted hosts as needed
        - Hosts should *at most* be trusted between services at the same site!
    - Cheap, short-lived tokens may become desirable
        - Their implementation may be expensive…
- For data operations ALICE are using detailed security *envelopes* instead of generic proxies
    - Minimize potential damage when a credential is stolen
- CASTOR still has backdoors for NS and RFIO → to be closed

2. Data protection

- Do different data classes need the same security model?
    - 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
- Read access regulation could be based on the confidentiality of the data
    - Would require careful management of distinct 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
        - Incur security handshake overhead also for large data volumes that could be made public
    - Most of the data is transferred over insecure channels!
        - Bulk encryption may remain too expensive
- Access audit trails are important for traceability
    - Both for security and performance investigations
- Protection is needed or desirable against:
    - Information leakage → not only the data itself, but also the file names matter
        - "Higgs-discovery.root"
        - Read access to directories and log files should be regulated
        - Encrypt selected, low-rate traffic?
    - Accidental commands → not everything should be writable for the whole VO
        - By default a user should not be able to impact someone else's data
        - An analysis group should not be able to impact data of another group
    - Malicious attacks
        - Outsiders
        - Insiders
            - Maintain as few privileges as possible per VO member
            - Use multi-person authorization ("two-man rule") for certain bulk data removals?
- Not only the data, but also the SE itself needs protection
    - Against illegal data → that the police would be interested in…
    - Against DoS

3. Issues with data ownership

- Missing concept: data owned by the VO or by a service
    - Use robot certificates for that?
        - ATLAS already do
- Mapping a person to/from a credential
    - DN changes may have consequences for data ownership
        - New certificate might indicate "formerly known as" → CA support would be needed
            - Would couple CAs and could make the infrastructure more fragile
            - Certificates should rather contain as little meta data to be vetted as possible
        - Could make use of VOMS nicknames or generic attributes instead of the DN → looks better
            - Link credentials in the authorization layer
        - ALICE LDAP service supports multiple DNs per AliEn user
    - X509 vs. Kerberos access
        - Is this only relevant for a few cases, e.g. EOS?
            - Try to avoid encumbering other SE implementations/deployments
        - How to map the Kerberos principal to a DN?
            - Yet another map-file?
            - Rather map DN to principal and use the latter for ownership?
            - Principal could be VOMS nickname (cf. ATLAS) or well-defined generic attribute
        - VOMS group membership could be determined from a map-file
            - Cf. "vo-role-mapfile" in dCache
- VO superuser concept desirable?
    - Avoid bothering SE admin for cleanups
    - Would be useful at least to ATLAS

4. Quotas

- Storage quotas
    - On the SE there may be a conflict with local replicas for performance/redundancy
    - Better handled by the experiment framework → current practice
    - Can still be useful to SE admin
    - Low priority, available on some SE types
- Quotas on other resources may be desirable to prevent DoS
    - Request rate
    - Bandwidth
    - Connections

Conclusions

The following areas would need attention in the near term:

1. Removal of backdoors from CASTOR
2. Checks of the actual permissions implemented by SEs
3. The issues with data ownership listed in Ch. 3