Improving Grid User's Privacy with gLite Pseudonymity Service

Henri Mikkonen, Joni Hahkala and John White

5th EGEE User Forum

12-16 April 2010

Uppsala, Sweden

www.eu-egee.org

e-infrastructure

Introduction

- **The user activity in the Grid can be hidden from the other Grid users and network observers...**
  - ...but the resource administrators are able to figure out the used resources and applications.
- **Information creep is of serious concern to applications in areas of highly competitive research, such as biomedicine**
  - The following requirement was already described in the EGEE Global Security Architecture document (DJRA 3.1)
    - "...an outsider should not be able to deduce a particular user's activities, such as how much of the resources the user consumes or what applications are run..."
- **Solution: users should be able to hide their real identity behind a pseudo-anonymous identity**

Terminology [Pfitzmann2009]

- **A pseudonym refers to an identifier of a subject (holder) other than one of the subject's real names**
  - A subject itself is pseudonymous if a pseudonym is used as identifier instead of its real names
- **For the linking between a pseudonym and its holder, three kinds of pseudonyms are defined:**
  - Public pseudonym: publicly known from the very beginning
  - Initially non-public pseudonym: the linking may be known by certain parties, but is not public at least initially
  - Initially unlinked pseudonym: the linking is not initially known to anybody with the possible exception of the holder himself

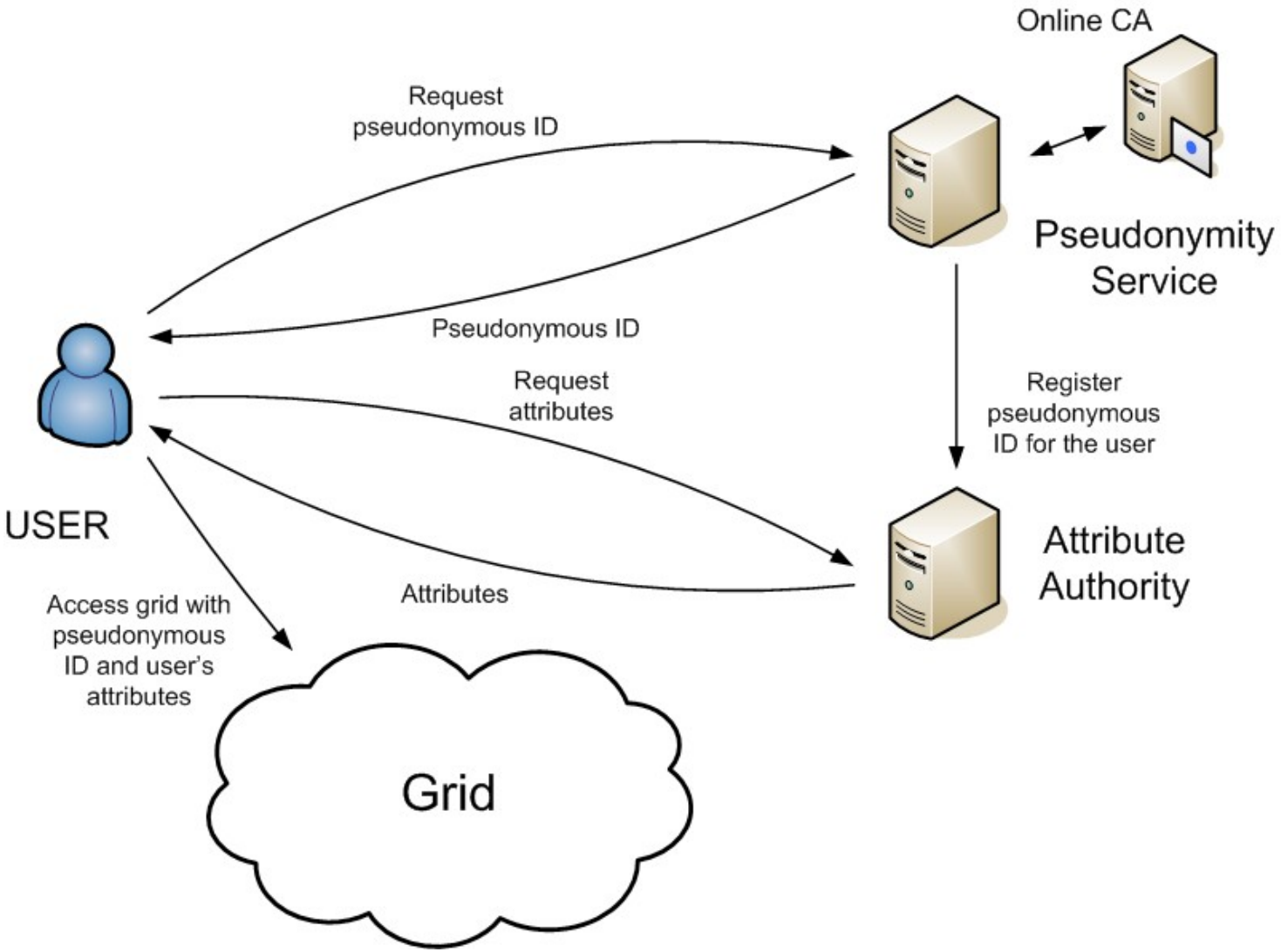
Use Case

- **A user initializes a pseudonymous Grid identity (proxy) with his real Grid credentials**
  - The pseudonymous proxy has an anonymized DN and a set of user's real VO attributes
- **The user utilizes the pseudonymous proxy for submitting a job to the Grid**

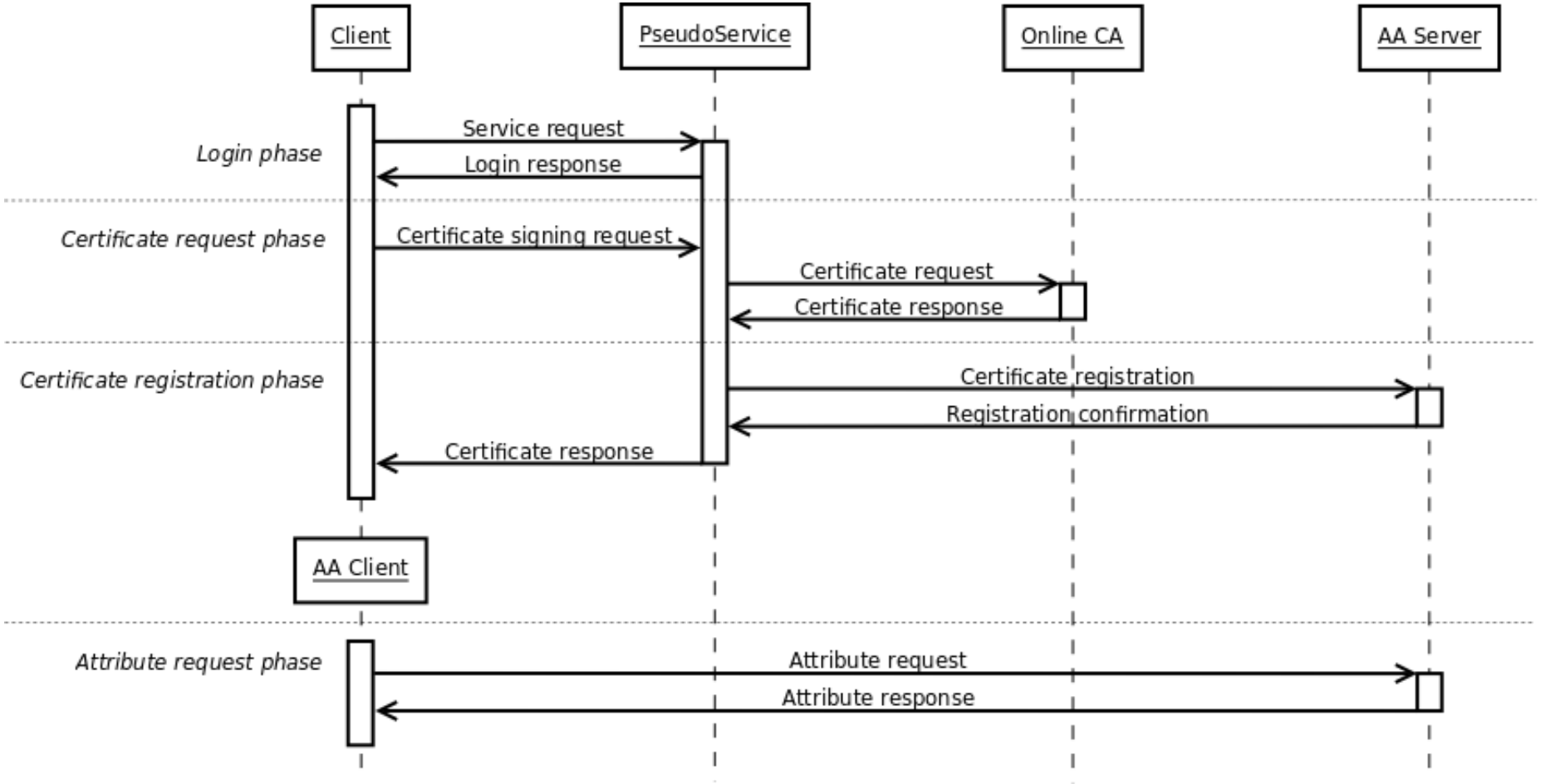
Use Case

- **A user initializes a pseudonymous Grid identity (proxy) with his real Grid credentials**
  - The pseudonymous proxy has an anonymized DN and a set of user's real VO attributes
- **The user utilizes the pseudonymous proxy for submitting a job to the Grid**
- **After the original pseudonymous proxy is expired, the user must initialize a new one with his real credentials**
  - The pseudonymous proxy has a new anonymized DN, but the similar set of user's real VO attributes
- **The user can access the results for the job submitted with the original pseudonymous proxy**

System Overview



Sequence


Client
PseudoService
Online CA
AA Server
Login phase
Service request
Login response
Certificate request phase
Certificate signing request
Certificate request
Certificate response
Certificate registration phase
Certificate registration
Registration confirmation
Certificate response
AA Client
Attribute request phase
Attribute request
Attribute response

Server-Side Requirements

- **Apache Tomcat container with gLite Trustmanager**
  - The VOMS proxy is used for user authentication & authorization
- **An RDBMS instance for storing the relationships between real and pseudonymous identities**
  - Any system supported by Hibernate can be used
- **VOMS-server and VOMS-Admin 2.5+ as the Attribute Authority**
  - The support of multiple certificates for each user is required
  - The pseudonymity service must be given a VO admin role
- **An Online CA instance supporting the CMC or CMP protocols**
  - Interoperability with an open source CA software called EJBCA is tested via the CMP protocol

Client-Side Requirements

- **The client tool is a standalone Java software**
  - Java 5 is required, interoperability tested with SUN Java
  - No other binary dependencies
- **voms-proxy-init is required for initializing VOMS proxies**

Grid Middleware Requirements

- **The middleware security must be based on PKI certificates**
  - Pseudonymous credentials are modeled as X.509 certificates
- **The middleware policy must not require the use of same subject DN from the users**
  - Pseudonyms should be short-lived, ideally one-time, identities with unique identifiers
    - Otherwise vulnerability for data correlation attacks would rise
  - Both authentication and authorization must be based on the attributes instead of the DNs

Conclusions

- **gLite Pseudonymity Service can be used for obtaining initially non-public pseudonymous identity for Grids**
  - The relationships between the real and pseudonymous identities can be revealed by authorized parties eg. in the cases of misuse
- **The users can hide their identity from the outsiders, including resource administrators**
  - However, other VO members with similar attributes can access the same data in the Grid
- **The work is scheduled to continue in the European Middleware Initiative (EMI)**

References

- **A. Pfitzmann and M. Hansen, *"A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management*", version 0.32, December 2009.**
  - http://dud.inf.tu-dresden.de/Anon_Terminology.shtml
- **J. Hahkala, H. Mikkonen, M. Silander and J. White, "*Requirements and Initial Design of a Grid Pseudonymity System*", in Proceedings of HPCS 2008, Nicosia, Cyprus, June 2008.**
  - http://www.scs-europe.net/conf/ecms2008/ecms2008%20CD/hpcs2008%20pdf/hpcs08w1-6.pdf
- **J. Hahkala, H. Mikkonen, M. Silander and J. White, "*A Pseudonymity System for Grids*", in Journal of Computers, Volume 4, Issue 5, May 2009.**
  - http://www.academypublisher.com/jcp/vol04/no05/jcp0405415422.pdf

- **Thank you for your attention!**
- **Any questions?**
- **Contact: firstname.lastname@cern.ch**