Security and Trust in an Industrial Grid Project

ISGC 2011 (23 March 2011, 臺北市, Taiwan)

Andreas Schreiber <Andreas.Schreiber@dlr.de>
German Aerospace Center (DLR), Cologne
http://www.dlr.de/sc
Abstract

In usual Grid security infrastructures based on personal Grid certificates, it is possible for users (i.e., employees), to first copy data (or software) to a Grid resource using their personal certificate and then copy this from another security domain to some other place. In the D-Grid project AeroGrid, which provides a Grid infrastructure and client tools for an industrial application from the aerospace domain, the industrial partner is a large turbine manufacturer with high demands on security. It is an important requirement that employees are not able to copy any data outside the security domain of the company. Within the project, a security policy for solving this problem has been defined. The basic strategy for a solution is as follows: The policies and the administrators of the company must forbid and enforce that employees can take the private key that belongs to the Grid certificate with them outside the company. Then the Grid certificate would be not usable for accessing data stored on some Grid resources. For the implementation of this strategy, a company-internal Grid Certificate Authority is deployed and a policy for handling certificates and private key is defined.

A second industrial requirement is reliability of data arising from complex processes. To have a reliable documentation of the individual steps performed in engineering calculations, it’s important to trace all processing steps, i.e. the complete Provenance of the process that led’s to a certain result. Within the project, a Service-Oriented Provenance architecture for recording Provenance information (such as user interactions in the graphical user interface or execution of numerical codes) has been provided.

This talk presents the security and the Provenance infrastructure of the AeroGrid project as well as details on the implementation and deployment of the security solution.

Outline

- AeroGrid
- Industrial Background
- Trust
  - Security
  - Provenance
AeroGrid

Grid-based cooperation between
- industry,
- research centers, and
- universities
in aerospace engineering

http://www.aero-grid.de
Project Partner

**Industry**
- MTU Aero Engines GmbH
- T-Systems Solutions for Research GmbH

**Research**
- German Aerospace Center (DLR)
  - Institute for Propulsion Technology
  - Simulation and Software Technology (Coord.)
- Deutsches Zentrum für Luft- und Raumfahrt e.V.

**University**
- University of the Armed Forces, Munich
  - Institute for Jet Propulsion
AeroGrid
Use Case and Project Goals

Usage Case

→ Collaboration in designing engine components

Project goals

→ Allow cooperation in research and development projects
→ Use of up-to-date program versions, data, and compute resources across all locations
→ Detailed documentation of history of a computational process that leads to a certain result ("Provenance")
AeroGrid Architecture

Site A (e.g., D-Grid)
- UNICORE6 Gateway
- Liferay Server
- WebDAV Server
- Resources
- Data/ Metadata
- CPU Resources

Site B (Service provider)
- UNICORE6 Gateway
- Liferay Server
- WebDAV Server
- Resources
- Data/ Metadata
- CPU Resources

User
- Code- Developer
- Simulation User

DataFinder

Web-Portal

ePROTAS
AeroGrid Deployment

MTU Aero Engines

Unicore Gateway
Monitoring and Discovering Service MDS
Unicore/X TSI
Vsite-TS

GridSphere Portal
Provenance Service
Unicore Gateway

Metadata Store (WebDAV)
CFD Developer
Provenance Store
XUUD8
Unicore/X TSI
Vsite-AT

LSF
Provenance Service
Provenance Store
EPROTAS
CFD User

CFD User
DataFinder
AeroGrid -Portal

Background: Turbo Machinery Simulation Tasks

Simulation of turbine components

→ Design (variants)
→ Optimization
→ Aero elasticity
→ Aero acoustics
→ Cooling
→ Complex geometries
→ Multistage components
Workflow Turbine Simulation

Start

<<Pre-Processing>>
GMC

<<Simulation>>
TRACE

<<Monitoring>>
GNUPlot

Problems?
Yes
<<Break>>
Stop Simulation

No

<<Post-Processing>>
TRACE-POST

<<Visualization>>
Tecplot

Optimum reached?
Yes
End

No
Trust in Industrial Context

- **Employees** are not trusted
  - Need for protecting confidential and classified data
  - Solution is a suitable security policy

- **Results** are not trusted
  - Need for traceable workflows and reliable documented results
  - Solution is recording of process and data Provenance
Industrial Security Concerns

- UNICORE 6 security model is based on client and server certificates

- With personal user certificates:
  - Users can access resources from within secure company network
  - Users can access resources from other locations
  - Users can access data or software from other locations

- Security solution must forbid to copy confidential or classified data or software
Security Solution

- Certificate Authority
  - Internal Grid-CA, trusted by D-Grid resources
  - Definition of policy for this CA

- Assure that private personal keys cannot leave the company
  - Security policies of the companies already forbids to copy any data

- Prevent active misuse of the security policy
  - Wrapper for Grid client software
    - Users cannot read and copy their private personal keys
    - Only the Grid client software can read the key
Security Solution

User

Simulation User

ePROTAS

Wrapper

UNICORE Client

Certificate

Remote Site (e.g., D-Grid Resource)

UNICORE6 Gateway

Liferay Server

WebDAV Server

UNICORE 6

Resources

Data/Metadata

CPU Resources

 UNICORE / WebDAV
Trust for Results
About “Provenance”

- lat.: provenire, “to come from“
  - Synonym: “Lineage”

- In Art Curation:
  - Source, Origin: The history of ownership or location of a piece of art
  - Actions performed on artifacts

- Data Provenance: What for?
  - Question of Item Identity
  - Question of Product Quality
  - Question of Production Efficiency
  - Question of Production Error Sources
  - Question of Repeatability
  - Question of Trust
Provenance in Computer Science

The Provenance of a piece of information is the history of its creation.

- What details can be documented about data production processes?
  - Input Parameters of Tools and Workflows
  - Used Resources (Computers, Other Data)
  - Responsible Contacts (User Sessions)
  - Produced Files
  - and Relationships among each other

Difference to classic Logging!
Example Provenance Use Case Questions in Engineering Applications

- Which simulation produced a certain file?
- Which simulation calculated a certain model?
- In which simulation a certain parameter was used?
- What monitoring data was recorded in a simulation with parameter == x?
- Which simulations were run with a certain numeric or model configuration?
- Has all data stayed within the company network during a confidential calculation?
Process Documentation of Complex Simulations

- What is recorded in complex simulations?
  - Model Parameters
  - Tools (Versions, Path, Origin of the Binaries)
  - Used Libraries and Compilers (Versions, Parameters)
  - Used Resources (e.g., Data, Computers, OS Environment, …)
  - Produced Files
  - User Session References
  - Execution Dependencies and Causal Chains

- Benefits:
  - Detailed Trace of the Emergence of Results
  - Clear Documentation of Distributed Workflows
  - Possibility to “Re-run“ Simulations
  - Formalities Compliance Checks
Specific AeroGrid Use Cases

- Resource Search
  - Which users and resources were involved in the production of a certain result?

- Error Search
  - Find successful simulation runs with a certain parameter configuration

- Expert Search
  - Who has used a certain configuration of parameters already?
AeroGrid Workflow: Provenance-Aware Applications

- Stand-alone GUI Applications
- Command Line Executables
- Script Integrated Tools
- Component-Based Software
  - e.g. Objects
- Interactions through
  - Shell Executions
  - Event Handling
  - File System
  - ...

- Start
- <<Pre-Processing>> GMC
- <<Simulation>> TRACE
- <<Monitoring>> GNUPlot
- <<Post-Processing>> TRACE-POST
- <<Visualization>> Tecplot
- <<Front End>> DataFinder
- <<Break>> Stop Simulation
- Optimum reached?
- No
- Problems?
- Yes
- End
Structure of Provenance Documentation

Example: Login and create a collection
InteractionRecord of Login Operation

Sending Actor: DataFinder GUI
Receiving Actor: DataFinder Meta Data Management
InteractionRecord of Collection Creation

Sending Actor: DataFinder GUI
Receiving Actor: DataFinder Meta Data Management
Questions?