Secure Cloud Data Storage for Galaxy Workflows

Carsten Friedrich, Catherine Wise, Richard Sinnott

March 2014
The Problem

• Galaxy workflows often deal with sensitive data

• It is attractive to execute workflows in the cloud

• Clouds are usually not highly trusted

How can we secure sensitive data in untrusted environments for processing in Galaxy?
TrustStore – Secure Data Storage in the Cloud

TrustStore Architecture

- TrustStore Client
- Key Management Service
- Integrity Management Service
- Collaboration/Profile Service
- Amazon S3
- NeCTAR
- MS Azure
- ...
TrustStore – Store Manager

![TrustStore - Store Manager](image)

<table>
<thead>
<tr>
<th>Name</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient X</td>
<td>carsten@csiro</td>
</tr>
<tr>
<td>ISDN1</td>
<td>carsten@csiro</td>
</tr>
<tr>
<td>galaxy-test</td>
<td>carsten@csiro</td>
</tr>
<tr>
<td>galaxy-store2</td>
<td>galaxy-user</td>
</tr>
<tr>
<td>test-uploads</td>
<td>galaxy-user</td>
</tr>
<tr>
<td>galaxy-store</td>
<td>galaxy-user</td>
</tr>
</tbody>
</table>
TrustStore –
Uploading files

![Image of TrustStore interface with files and progress bars](image_url)
TrustStore – Windows Client
TrustStore – Changing Store Access
TrustStore – Secure Data Storage For Galaxy Workflows
TrustStore –
Client Data Workflow

Cryptographic Library

TrustStore Client

- Store Profile
- Fragment / Merge
- Encrypt / Decrypt
- Integrity Management
- Upload / Download

Collaboration/Profile Management Service

Key Management Service

Integrity Management Service

Cloud Storage
TrustStore – Collaboration Workflow

Store Profile

Cryptographic Library

TrustStore Client

- Get Access List
- Get Public Keys of Collaborators
- Encrypt profile with keys (S/MIME)
- Upload Encrypted Profile

Collaboration/Profile Management Service
Lessons learned

It does work. It does the job. There is room for improvement:

• Galaxy can’t keep a secret (It’s not even trying):
  • Can’t hide passwords in forms.
  • Writes passwords into logs / databases.
  • No streaming support.
  • Can’t protect temporary storage.

• Writing Galaxy modules is tricky:
  • Sparse documentation.
  • Sparse error messages.
  • Very fickle library dependency management.

• Implementing as an external data source provider may have been better in hindsight.
Thank you

Carsten Friedrich
Research Team Leader
IE Lab, CSIRO Computational Informatics
Carsten.Friedrich@csiro.au
+61 2 6126 7019