Encrypted Data Storage.

John White (Helsinki Institute of Physics)
Joni Hahkala (HIP)
Kalle Happonen (HIP)
Ákos Frohner (ex-CERN)

www.eu-egee.org

Contents

- **Background.**
- **Hydra.**
- **Medical Data Management.**
- **Support and Deployment plan.**

Background

- In some research and user communities data security is imperative.
  - Medical data
  - Financial data
  - Personal data
- eg. **In the medical research field:**
  - Data files should be anonymized.
  - Data files should be encrypted:
    - On storage elements.
    - During transfer.
  - Metadata and Data files controlled by ACLs.
  - These files should be protected from “local attack” also.

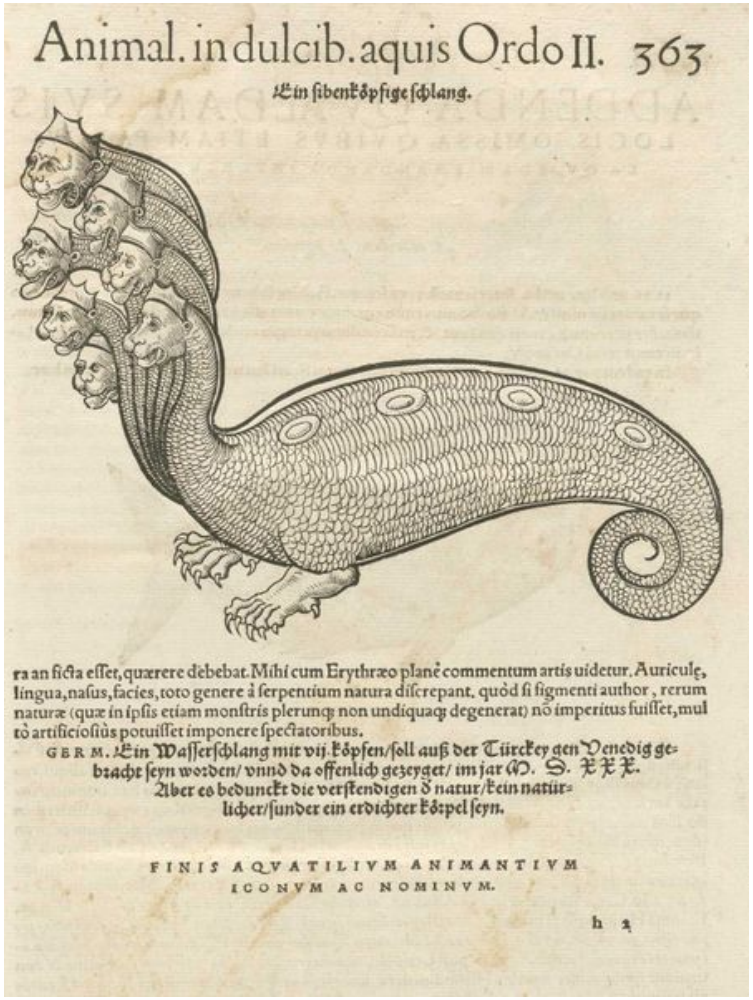
Encryption

- Strong encryption of files gives more flexibility:
  - Unauthorized access to the encrypted file is not catastrophic.
- But how to encrypt?
  - Decryption a requirement:
    - By the owner.
    - By a group defined at file-level granularity.
- Authentication certificates are not suitable for long-term.
  - Renewal each year → files lost?
  - Revocation or lost pass phrase → files lost?
  - No group access.
- Symmetric keys can be hard to manage
  - Large number of keys needed for high granularity.
  - User-managed keys → lost keys → lost files.

Key Management

- Central key storage is not so good.
  - All eggs in one basket.
  - Single security breach gives away all keys.
  - Rogue root-privilege user could access every key (file).
- Distributed key storage a better strategy.
  - Secret is shared across separate keystores (DB or other).
  - Ideally on administratively different heterogeneous servers.
  - Not enough to hack one server.
  - Coordinated attack needed to access the keys.

Hydra


Animal. in dulcib. aquis Ordo II. 363
Ein ſibenköpfige ſchlang.
ra an ficta eſſet, quærere debebat. Mihi cum Erythræo planè commentum artis uidetur. Auriculę, lingua, naſus, facies, toto genere à ſerpentium natura diſcrepant. quòd ſi figmenti author, rerum naturæ (quæ in ipſis etiam monſtris plerunq; non undiquaq; degenerat) nō imperitus fuiſſet, multò artificioſiùs potuiſſet imponere ſpectatoribus.
GERM. Ein Waſſerſchlang mit vij. köpfen/ſoll auß der Türckey gen Venedig gebracht ſeyn worden/ vnnd da offenlich gezeyget/ im jar M. D. XXX.
Aber es beduncket die verſtendigen d natur/ ein natürlicher/ ſunder ein erdichter köpel ſeyn.
FINIS AQVATILIVM ANIMANTIVM ICONVM AC NOMINVM.
h 2

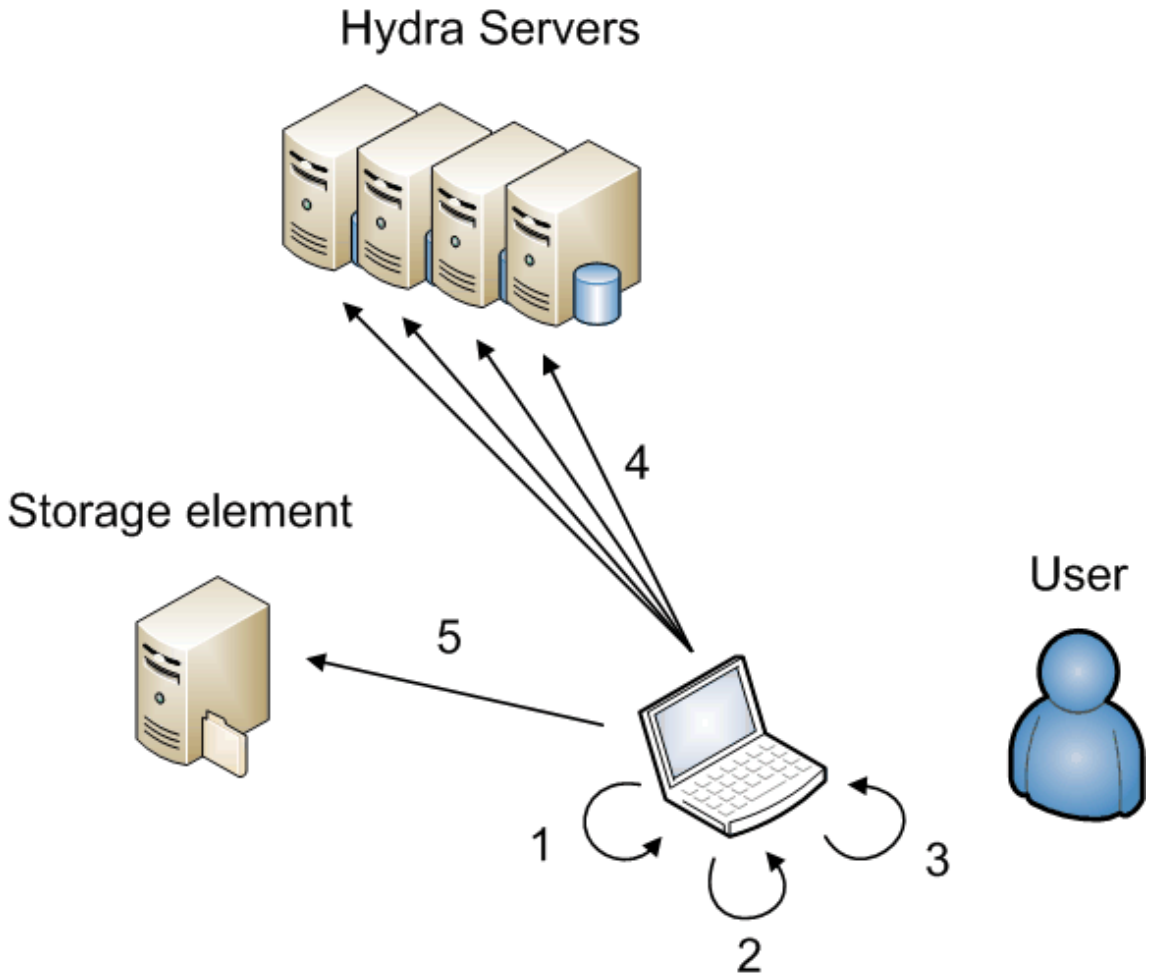
Hydra is a distributed key storage solution.

- File encryption key generated, split and distributed to keystores (databases)
- Splitting based on **Shamir Secret Sharing Scheme**.
  - (m, n) threshold scheme. With $n > m$.
  - $m - 1$ degree polynomial with n points calculated.
  - Need m points to reconstruct the polynomial.
  - $m - 1$ fragments do not help to determine the constant, which is the secret.
- This scheme provides:
  - Security: m fragments required to reconstruct key.
  - Fault Tolerance: $m - n$ keystores unavailable: key recoverable.
  - eg. Splitting key using $(3, 5)$ scheme:
    Key "lost" if 3 servers attacked.
    Any 2 servers may be offline.

Hydra

- A key per file:
  - Distributed to *n* (preferably) geographically and managerially separated servers.
  - Key fragment access controlled with access control list.
  - Access for only owner or list of users.
  - Access to list of groups (VOMS).
- **Caveats:**
  - **Still vulnerable to user credential compromise.**
    - Proxies are in many places.
    - Secure token service could help.
  - **Decrypted files eventually are used:**
    - A WN will have access with user credentials when computing, thus a compromise there reveals the contents.

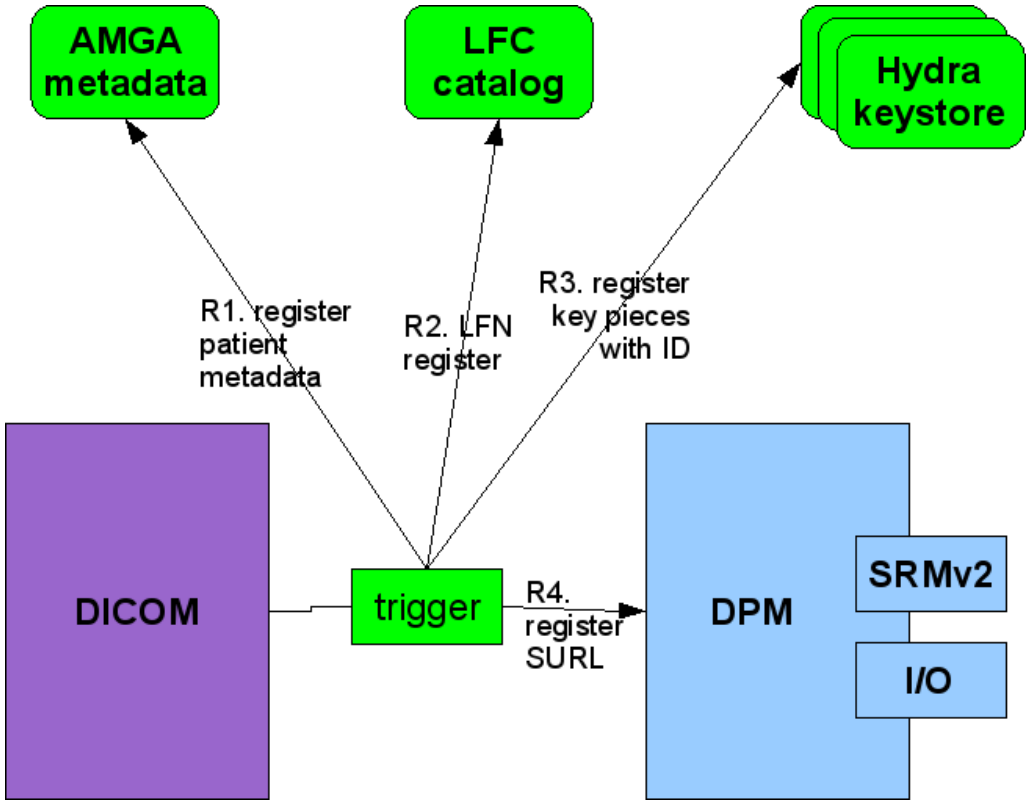
Hydra



Hydra in EGEE/gLite

- Hydra integrated to produce **Medical Data Manager** (MDM).
- Medical data: Managed by DICOM storage.
  - **D**igital **I**maging and **Co**mmunications in **M**edicine (**DICOM**): standard.
  - DICOM designed for internal hospital usage.
  - Should not be exposed to general Grid environment.
- EGEE solution: extension of Data Management tools.
  - DICOM to DPM trigger.
  - Encryption/decryption of data on the fly.
  - Metadata registration of metadata to catalog (AMGA), file logical info to LFC
  - Storage of key fragments to Hydra using LFC GUID.
  - SRMv2, GFAL, gridFTP for SURLs, TURLs and transfers.

File Registration



- DPM-DICOM trigger:
  - Registered to AMGA, LFC.
  - Uses LFC (or DICOM) GUID for ID.
  - Registers keys under ID to Hydra.
  - Stores encrypted file to DPM.
- GUID important (retrieval).

```
dpm-dicom-trigger <DICOM file name>

MDM-register <DICOM file name>
```

Status and Deployment.

- Installation/Documentation/Notes continuous updating:
  - To be updated with more focus on end-user tasks.
  - Technical/development details to be (re)moved.
- Certified for SL4. SL5/Debian ongoing.
  - Hydra clients in the general gLite UI.
  - Hydra service released separately (client wishes).
- Security Audit:
  - Done. Code review issues raised.
  - Code being updated.
- **Future work depends on the needs of clients.**

Status and Deployment.

- Installation/Documentation/Notes continuous updating:
  - To be updated with more focus on end-user tasks.
  - Technical/development details to be (re)moved.
- Certified for SL4. SL5/Debian ongoing.
  - Hydra clients in the general gLite UI.
  - Hydra service released separately (client wishes).
- Security Audit:
  - Done. Code review issues raised.
  - Code being updated.
- **Future work depends on the needs of clients.**
- **Hydra part of EMI. Continued HIP support.**

Status and Deployment.

- Installation/Documentation/Notes continuous updating:
  - To be updated with more focus on end-user tasks.
  - Technical/development details to be (re)moved.
- Certified for SL4. SL5/Debian ongoing.
  - Hydra clients in the general gLite UI.
  - Hydra service released separately (client wishes).
- Security Audit:
  - Done. Code review issues raised.
  - Code being updated.
- **Future work depends on the needs of clients.**
- **Hydra part of EMI. Continued HIP support.**
- **Under discussions with external company.**
  - Front end for generic cloud storage.
  - Prototype Java interface. .NET integration after.

`https://twiki.cern.ch/twiki/bin/view/EGEE/DMEDS`

Questions?

p.s. There's a LaTeX template available:

https://edms.cern.ch/document/964536/1