XrootD and GridFTP development plans at RAL

Alastair Dewhurst, George Vasilakakos, Ian Johnson
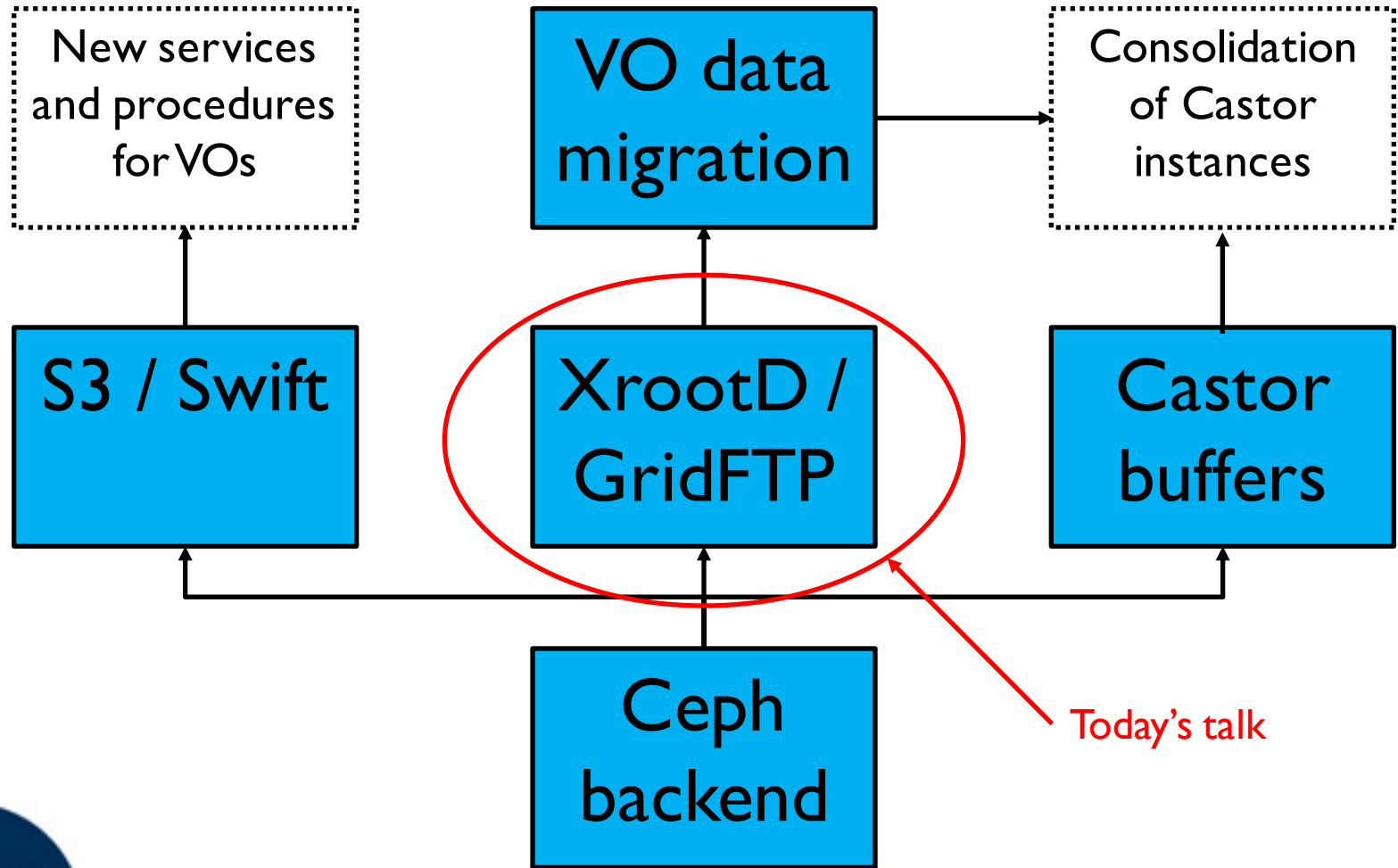
Alastair Dewhurst, 9th May 2016

Outline

- Motivation

- XrootD authorization

- GridFTP

  - Authorization

  - Performance

- Leading /

- Summary

The Echo Project

New services and procedures for VOs

VO data migration

Consolidation of Castor instances

S3 / Swift

XrootD / GridFTP

Castor buffers

Ceph backend

Today's talk

Alastair Dewhurst, 9th May 2016

Motivation

- The LHC VO are not currently able to run all their workflows through an SE providing only S3/Swift access.

- Minimum requirement to provide GridFTP and XrootD access to storage.

  - Aim to create thinnest layer on top of Ceph (why CephFS + GridFTP servers are not favoured).

  - Only intend to provide needed functionality for existing use cases.

- Relying heavily on work done by Sebastien Ponce and Brian Bockelman.
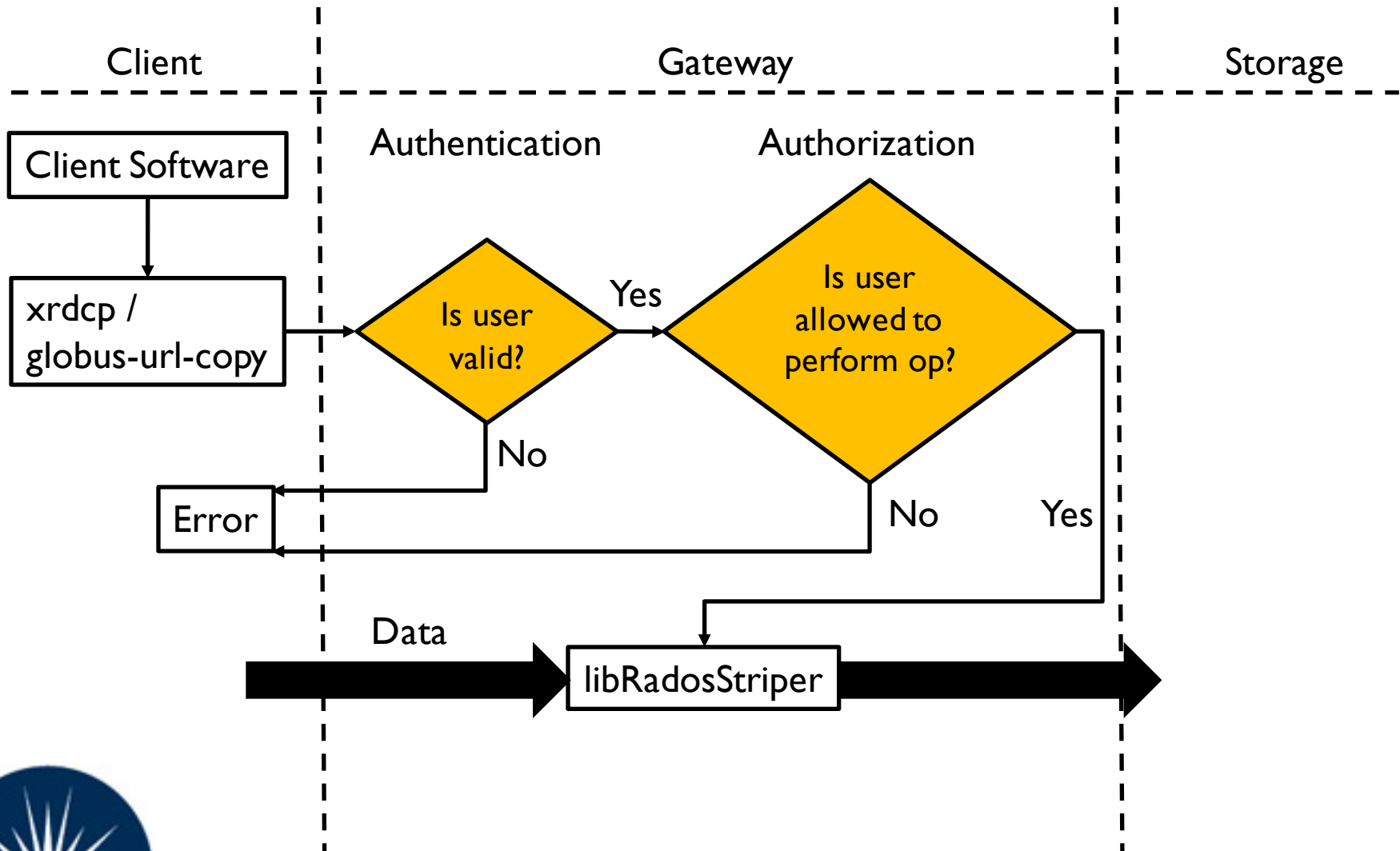
For XrootD the Gateway could be the WN
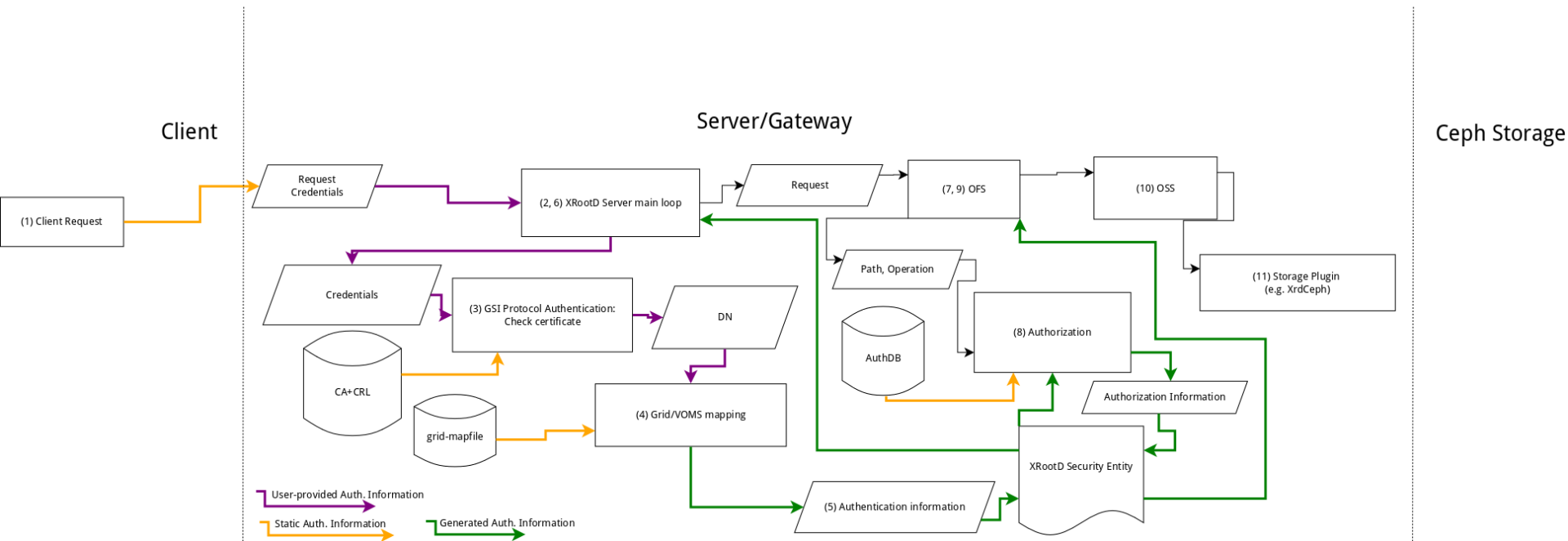
Alastair Dewhurst, 9th May 2016

XrootD authorization (1)

- "Just" need to develop authorization code for XrootD.

- Grid-map file + AuthDB are existing components we are trying to reuse.

XrootD authorization (2)

Xrootd.conf
all.export *?
all.export /

authDB
u atlasprod /test a ◀—— Created pool called /test

grid-mapfile
"/C=UK/O=eScience/OU=CLRC/L=RAL/CN=alastair dewhurst" atlasprod

-bash-4.1$ xrdcp 128m root://ceph-gw1.gridpp.rl.ac.uk/test:/128m

[0B/0B][100%][===][0B/s]

Run: [ERROR] Server responded with an error: [3010] Unable to create test:/128m; permission denied

-bash-4.1$ xrdcp 128m root://ceph-gw1.gridpp.rl.ac.uk//test:128m

[0B/0B][100%][===][0B/s]
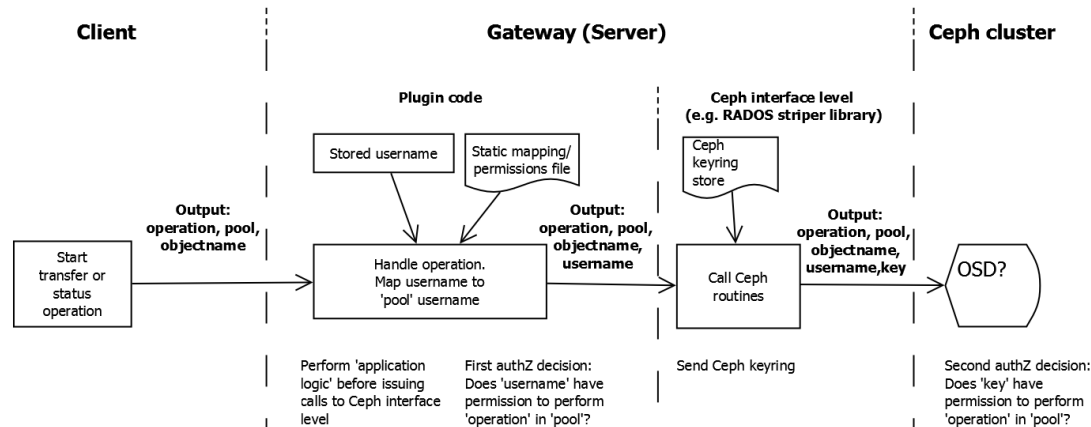
Run: [ERROR] Server responded with an error: [3005] Unable to open /test:128m; invalid argument

Alastair Dewhurst, 9th May 2016

GridFTP authorization

- Will need to build GridFTP authorization.

  - If possible re-use code.

  - Must share Grid-map file and authDB.

- GridFTP authorization development depends on how we get XrootD authorization to work.

2b. Proposed authorization - to be performed for every operation

Client	Gateway (Server)	Ceph cluster

Plugin code

Ceph interface level (e.g. RADOS striper library)

Stored username

Static mapping/ permissions file

Ceph keyring store

Output: operation, pool, objectname

Output: operation, pool, objectname, username

Output: operation, pool, objectname, username, key

Start transfer or status operation

Handle operation. Map username to 'pool' username

Call Ceph routines

OSD?

Perform 'application logic' before issuing calls to Ceph interface level

First authZ decision: Does 'username' have permission to perform 'operation' in 'pool'?

Send Ceph keyring

Second authZ decision: Does 'key' have permission to perform 'operation' in 'pool'?

Updated plan: Only use one Ceph keyring

Alastair Dewhurst, 9th May 2016

GridFTP Performance (1)

	Rados get/put	Globus-url-copy	FTS transfers
Reads	~80MB/s	~80MB/s	6-7MB/s
Writes	~80MB/s	~80MB/s	6-7MB/s

- Current plugin based on proof of concept work by Sebastien.

- Acceptable performance if using Globus-url-copy directly (Streaming).

- Poor performance using FTS (block mode).

  - Tried to get larger chunks to be sent but this is not viable over WAN.

Alastair Dewhurst, 9th May 2016

GridFTP Performance (2)

```
┌──────────┐      ┌──────────┐      ┌──────────┐
│   FTS    │ ──▪──▶│          │      │   Ceph   │
│  Client  │──▪──▶ │ Gateway  │─▪▪▪▪─▶│ Backend  │
│          │──▪──▶ │          │      │          │
│          │──▪──▶ │          │      │          │
│          │──▪──▶ │          │      │          │
└──────────┘      └──────────┘      └──────────┘
```

- Ian Johnson is working on this currently.

- Based on Brian Bockelman's GridFTP code for HDFS

  - https://github.com/bbockelm/gridftp_hdfs/tree/master/src

Alastair Dewhurst, 9th May 2016

Leading /

- There are several different issues with leading /.

1. Unable to delete paths (with XrootD) not starting with a /

    - https://github.com/xrootd/xrootd/issues/314

2. XrootD authDB requires / else path is considered a template.

    - Work around means creating a pool starting with a /.

3. GridFTP plugin adds a leading /.

    - Ian Johnson has removed this from plugin.

Alastair Dewhurst, 9th May 2016

Conclusion

- A lot of effort has gone into understanding XrootD

  - Still trying to decide on best solution.

- GridFTP plugin development ongoing.

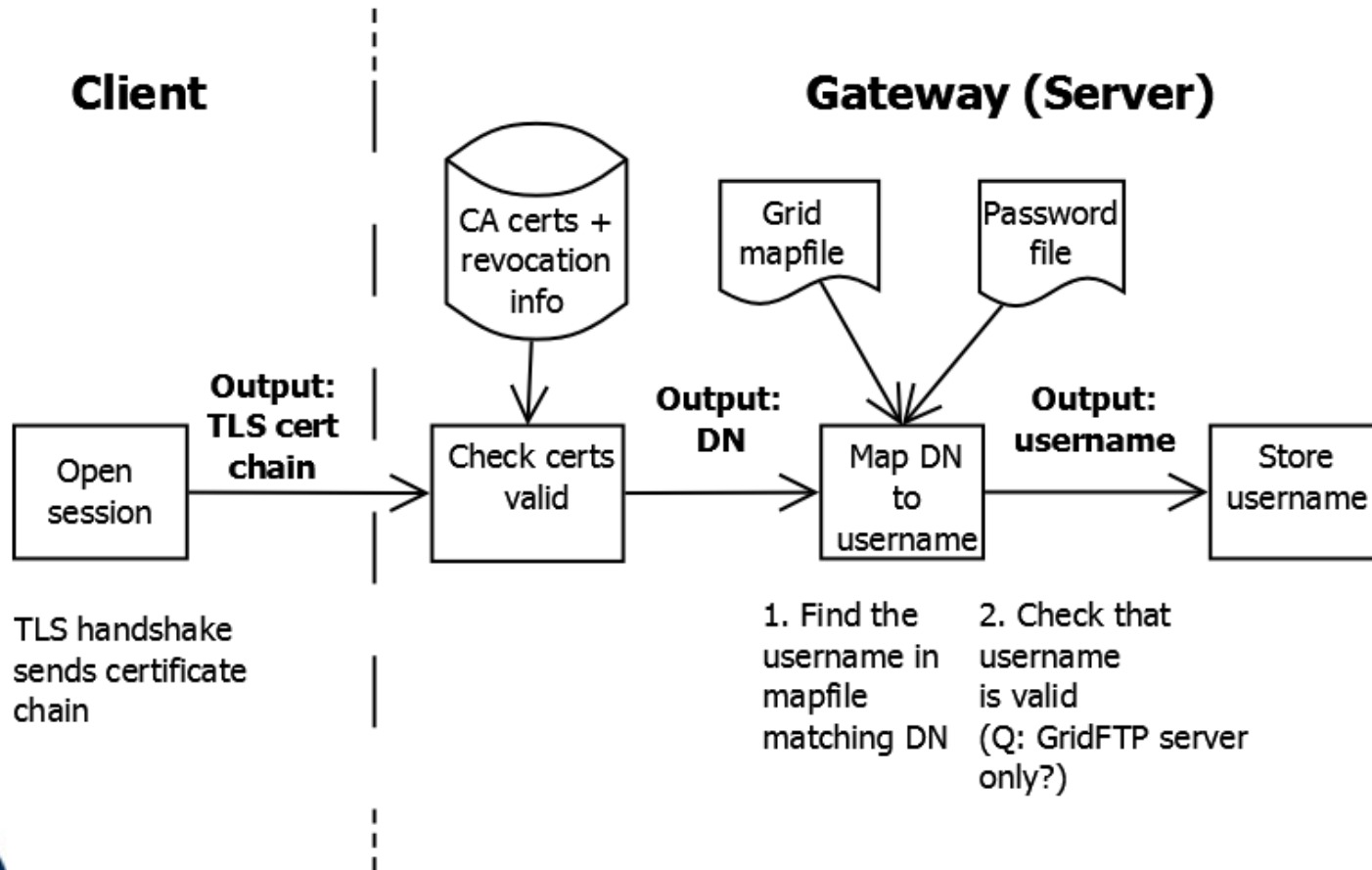- Concern about leading / preventing objects being accessible via both protocols.

Backup

Alastair Dewhurst, 9th May 2016

Backup

1. GSI Authentication - performed at session start-up

Client

Gateway (Server)

CA certs + revocation info

Grid mapfile

Password file

Open session

Output: TLS cert chain

Check certs valid

Output: DN

Map DN to username

Output: username

Store username

TLS handshake sends certificate chain

1. Find the username in mapfile matching DN

2. Check that username is valid (Q: GridFTP server only?)

Alastair Dewhurst, 9th May 2016

Backup

2a. Current GridFTP authorization - performed for every operation

Alastair Dewhurst, 9th May 2016