A GridFTP interface for iRODS

Shunde Zhang

shunde.zhang@arcs.org.au

The Motivation

- Communication with the Grid
- Data moving to/from other data sources
- Transporting large data sets
- Performance

GridFTP

- The de facto standard for the Grid
- Encrypted control channel (and possible data channel)
- Advanced features to increase transfer rate
  - Extended block mode, partial download
  - Parallel/Striped transfer
  - UDP
  - Checksum
  - Third party transfer

Griffin

- A GridFTP interface
  - GridFTP version1
  - Compatible with existing GridFTP tools, e.g. guc, Grisu, Hermes, FTS, DataMINX DTS, Globus SaaS
- Abstract layer to hook up different file systems
  - Protocol converter

Parallel Transfer
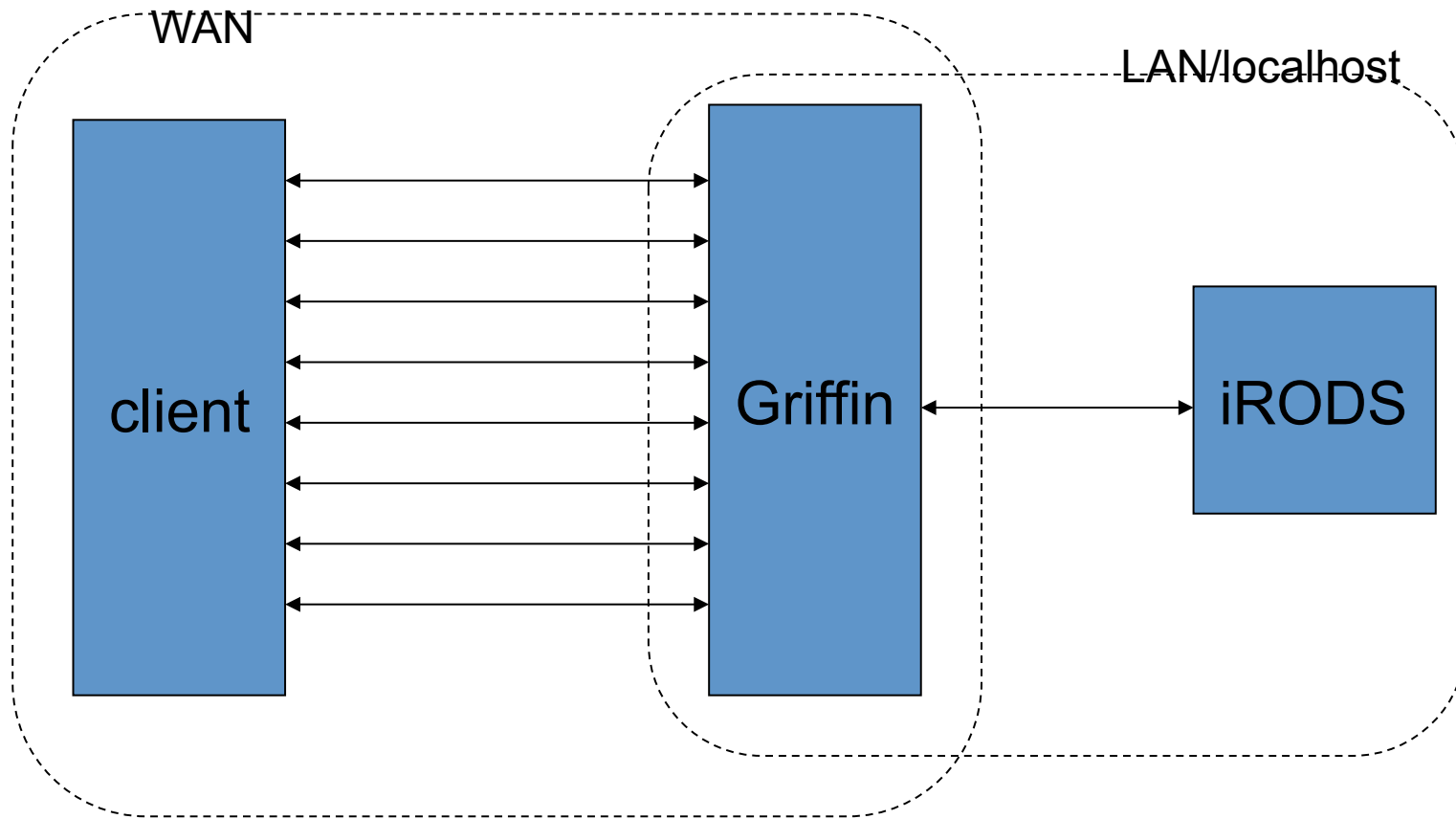
The implementation

- Java-based
  - OS independent
  - Easy to install, easy to run
  - JNLP, start from the web (potential)
- Modular design, with spring framework

Deployment

- A service for CentOS

- Control channel port

- Data channel ports

- irods-mapfile for slave iRODS servers (no need for masters)

- Can specify a default resource (otherwise it will be chosen by the rules)
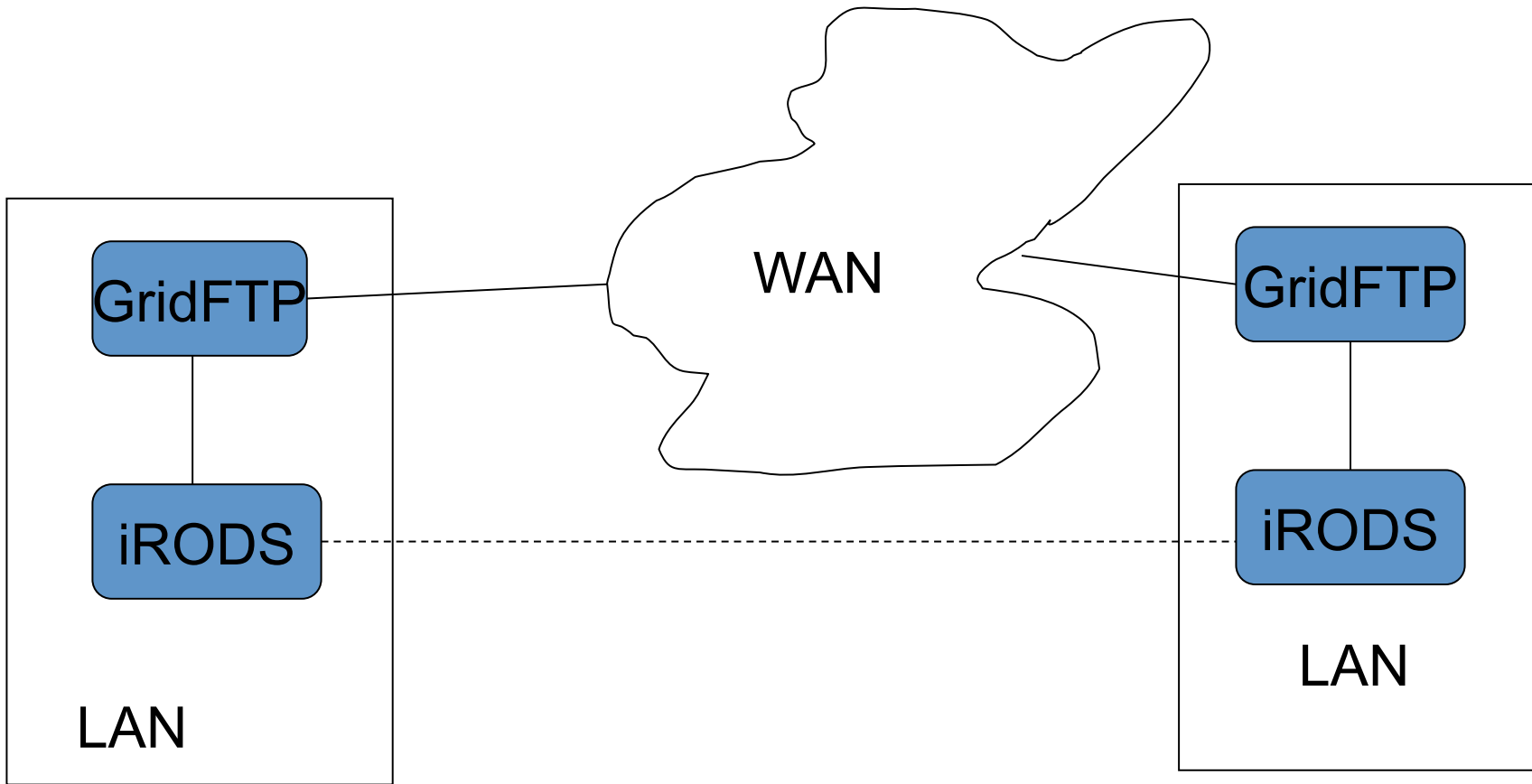
Some tests

transferring twenty-one 320M files (totally 6.7G) from Hobart to Melbourne (310Mbps connection)

Test	Time
Globus GridFTP 5 on disk (UDT, 2 FTP connections, 2 threads on each)	10.5 mins
Globus GridFTP 5 on disk (TCP, 2 FTP connections, 2 threads on each)	15 mins
Griffin to iRODS (TCP, 2 FTP connections, 2 threads on each)	14 mins
iput	13 mins

Use Case 1
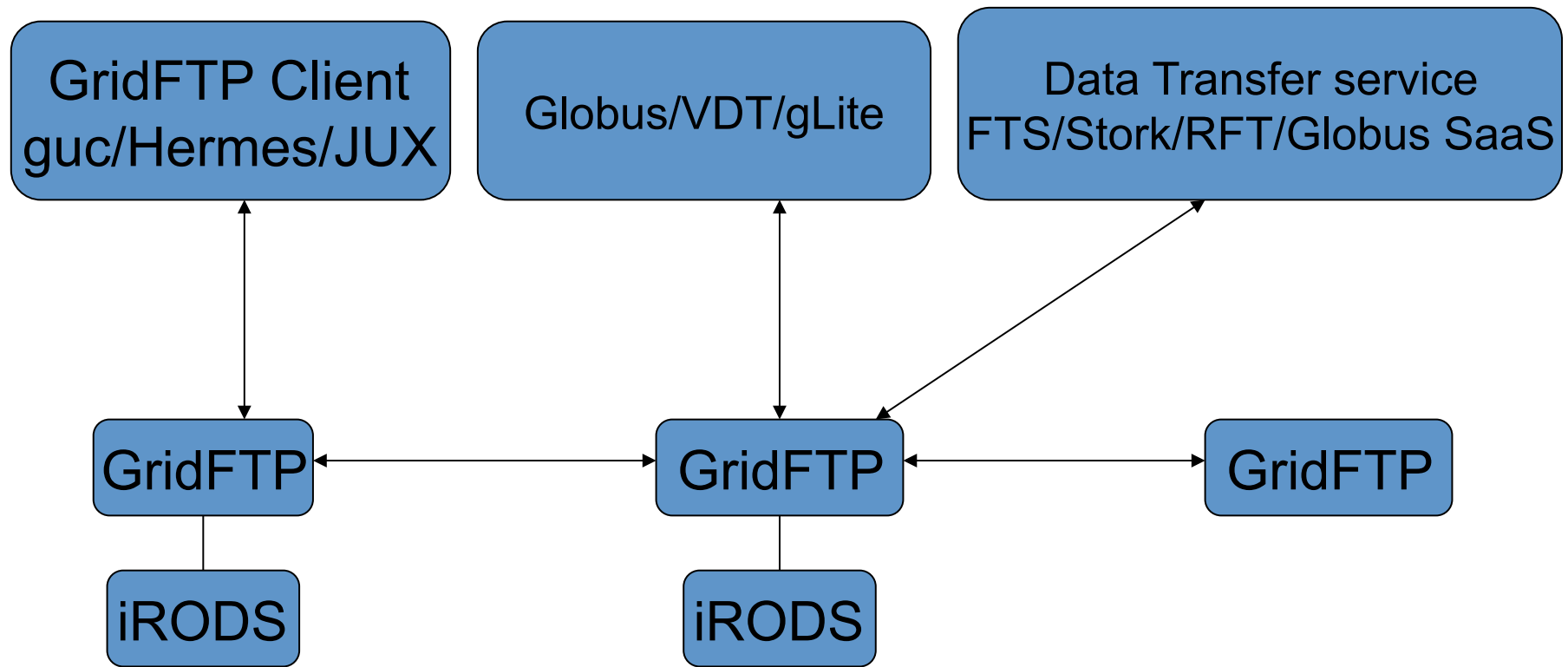
Use Case 2

Future work

- Performance/stress tests

- UDP

- SSHFTP

- GridFTP v2
  - checksum

Thank you!

- Where to get it
  - https://projects.arcs.org.au/trac/griffin
- Questions?
- Email: shunde.zhang@arcs.org.au