

# Data Access and Management in Virtual Organisations

*Saleem Bhatti<sup>1</sup>, Soren-Aksel Sorensen<sup>2</sup>, Alan Dearle<sup>1</sup>, Graham Kirkby<sup>1</sup>*

*<sup>1</sup>School of Computer Science, University of St. Andrews*

*<sup>2</sup>Department of Computer Science, UCL*



# Talk Outline

- The problem space:
  - File management across Virtual Organisations (VOs)
- The solution space:
  - A consistent user viewpoint
  - Distributed file space
  - Network resource control
  - Security
  - Policy-based file management
- Next steps & Questions



- The problem space:
  - **File management across VOs**
- The solution space:
  - A consistent user viewpoint
  - Distributed file space
  - Network resource control
  - Security
  - Policy-based file management
- Next steps & Questions



# Managing files across sites [1]

- Multiple users, multiple sites
- Need to share files:
  - Files may be replicated at each site
  - Files may be updated by one or more sites
  - Consistency of file copies?
- Same file has different name at different sites:
  - Directory structure and naming of files
- Local administration and policy:
  - File management



# Managing files across sites [2]

- Different files versions:
  - ‘Current’ version
  - Older versions (archive)
  - Fall back to a previous version (checkpoint)
- No *consistent view* of file space:
  - Different users may see different views
  - Different directory structures
  - Multiple views of files for parties in a single VO



# Current file management tools

- Unix-based:
  - (S)FTP
  - rsync
  - rcp/scp
- Grid tools:
  - GridFTP
  - SRB
- Peer-to-peer systems:
  - File sharing systems (e.g. Chord, Pastry, JXTA)
- Others ...



# Current file management practice

- Manually intensive:
  - Much human interaction
- Across VOs:
  - Handling at each site?
  - User account juggling?
  - Security?
  - Administration policies?
- **Policy control:**
  - **External to system**



# Insights form ongoing work

- ASA - Autonomic Storage Architecture:
  - St. Andrews (EPSRC)
  - Work in progress
- NRS - Network Resource Scheduler:
  - UCL (EPSRC/e-Science)
- **VPFS - Virtual Policy-Based File-System**
  - St. Andrews and UCL



- The problem space:
  - File management across Virtual Organisations
- The solution space:
  - **A consistent user viewpoint**
  - **Distributed file space**
  - Network resource control
  - Security
  - Policy-based file management
- Next steps & Questions



# User viewpoint in VO

- A **single** viewpoint for a VO:
  - Across all sites
  - Across all users
  - Subject to policy
- Transparent co-ordination across sites:
  - Underlying system behaviour maintains consistency
  - “Self-managing” - **autonomic** system
- Responds (reacts) to user interaction



# The ASA

- <http://www-systems.dcs.st-andrews.ac.uk/wiki/ASA>
- “Home” ubiquitously available, irrespective of:
  - Machines and disks
  - Physical location
- Data highly durable - no need for backup
- Simple data sharing:
  - Accessed as a traditional file system
  - Existing programs and databases work without change
- Historical views:
  - Data never over-written



# ASA - Example

## Current Situation

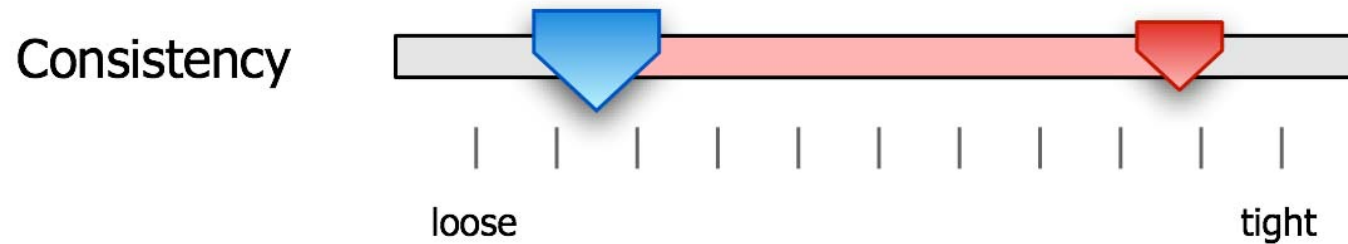
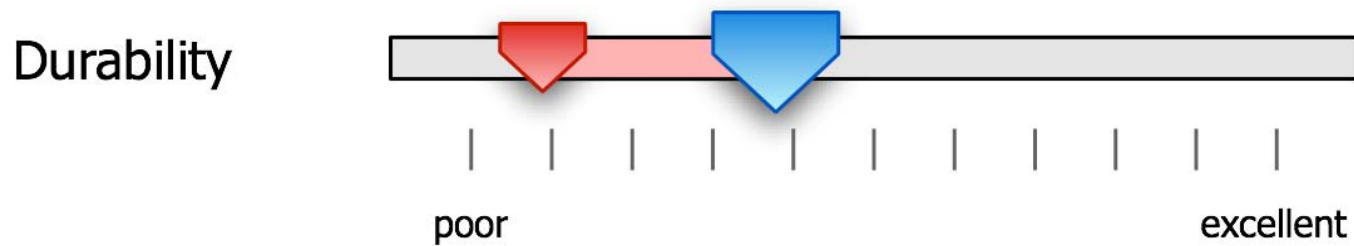
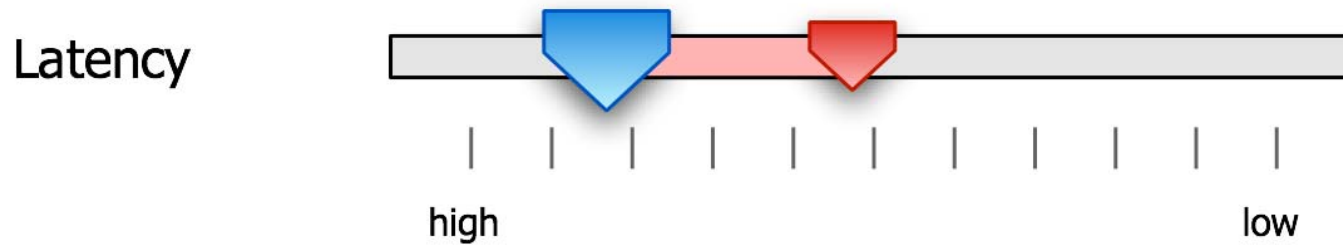
- Distributed environment - multiple storage sites
- Storage has to be managed explicitly:
  - Manual file copying - sftp, rsync, etc.
  - Non-trivial to ensure data is always accessible where it is needed, when it is needed
  - Additional problem with access from other locations

## Goal

- Groups of users have a logical view of one file system
  - Accessed from any location that can run the software



# ASA - User Control Example



- The problem space:
  - File management across Virtual Organisations
- The solution space:
  - A consistent user viewpoint
  - Distributed file space
  - **Network resource control**
  - Security
  - Policy-based file management
- Next steps & Questions



# Controlling the network

- When transferring files:
  - Files may be large
  - May have latency requirements
  - May have deadlines in a workflow
- Need to make sure network resources are available from file source(s) to file destination(s)
- Need network resource control

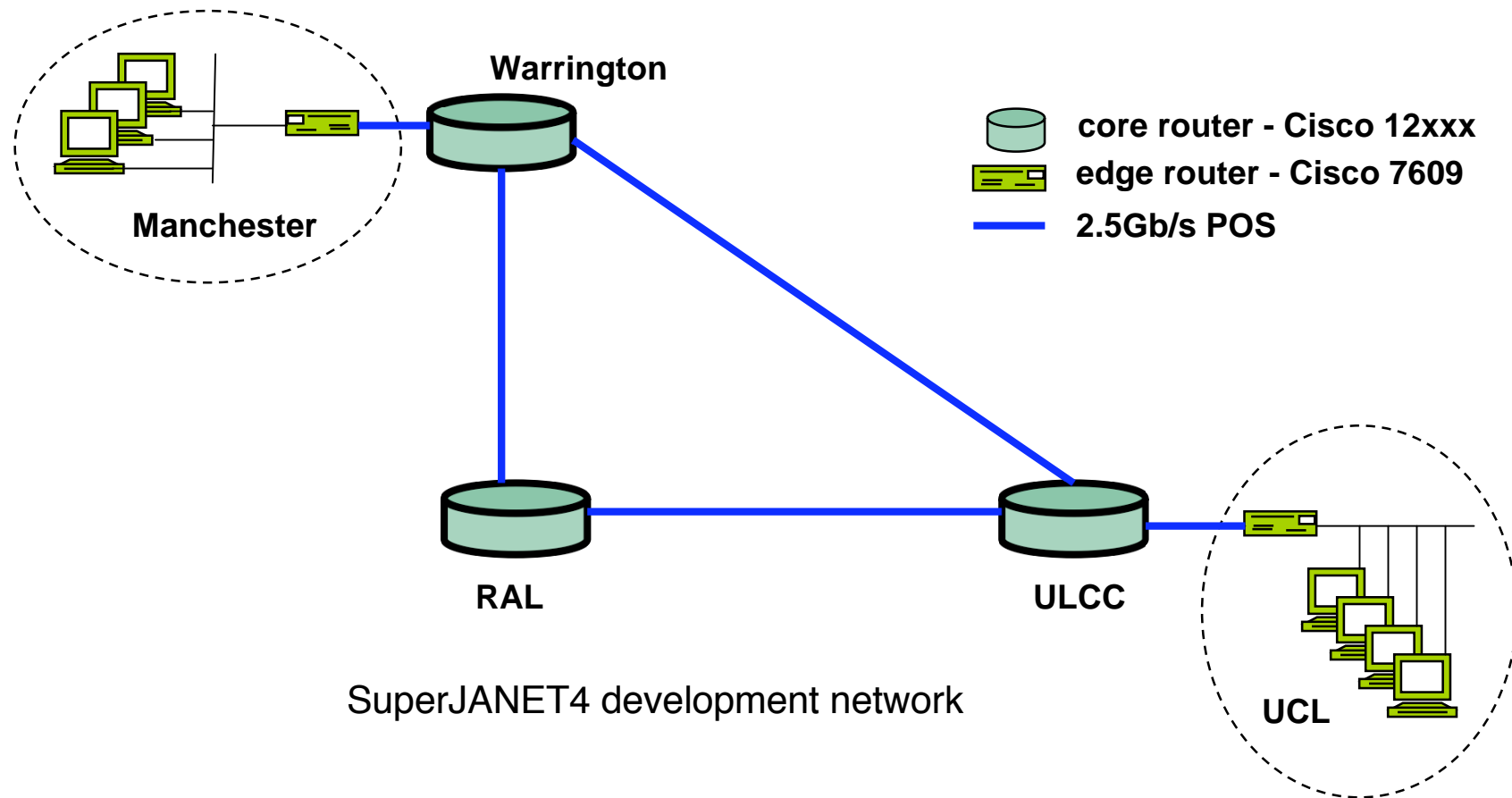


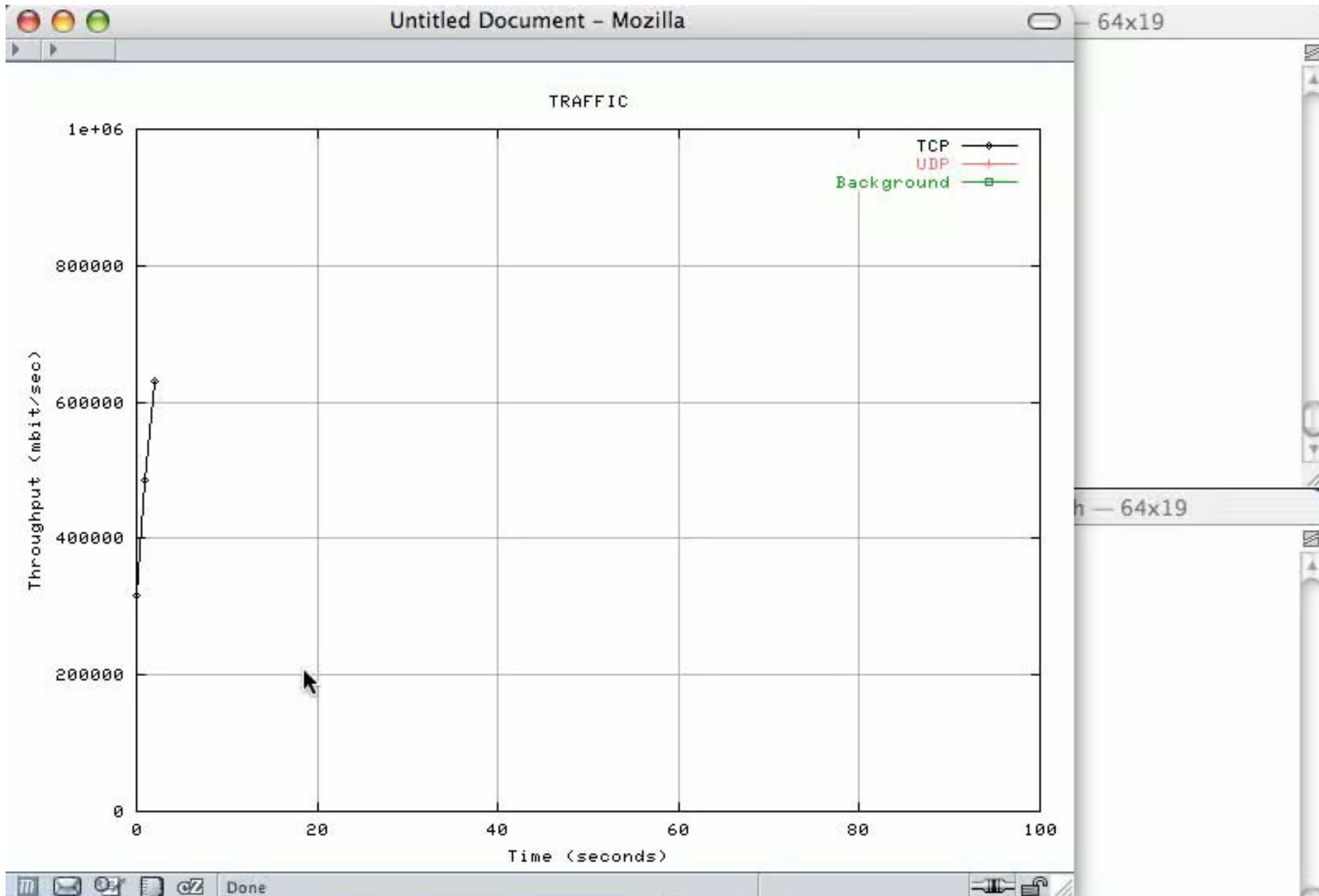
# NRS

- <http://www.cs.ucl.ac.uk/research/nrs/>
- Peer-to-peer network resource scheduling
- Provides inter-site network flow protection
- Accessible via API or management console
- Works using standard networking components
- Hierarchical security model
- Allows single-site-sign-on control:
  - Control between sites A & B from site C



# NRS demo





Done

xclock

Fri Apr 2 14:53:31 2004

- The problem space:
  - File management across Virtual Organisations
- The solution space:
  - A consistent user viewpoint
  - Distributed file space
  - Network resource control
  - **Security**
  - Policy-based file management
- Next steps & Questions



# Security

- On the wire & at storage/site location(s):
  - Privacy
  - Authencity
  - Integrity
- Security management:
  - Users
  - User groups
- Peer-to-peer, hierachical trust model (NRS):
  - Delegation of trust to site/locations



- The problem space:
  - File management across Virtual Organisations
- The solution space:
  - A consistent user viewpoint
  - Distributed file space
  - Network resource control
  - Security
  - **Policy-based file management**
- Next steps & Questions



# Policy-based requirements [1]

- Availability:
  - Robustness/resilience to node failures
- Consistency:
  - Always a single/multiple versions of the file only?
  - Allow multiple (per-user/per-site) versions?
- Security and access control:
  - On site, and on-the-wire
  - Per-group, per-site, per VO member
- Resource control and allocation:
  - Disc, network



# Policy-based requirements [2]

- These are **not** parameters of traditional file or storage systems!
  - Distinct from the file/storage system
  - Yet without policy, system is not useable for a VO
- Outside the tools:
  - Best practice
  - Agreed **policy** on file management
- Virtual Policy-based File System Required



- The problem space:
  - File management across Virtual Organisations
- The solution space:
  - A consistent user viewpoint
  - Distributed file space
  - Policy-based file management
  - Network resource control
  - Security
- **Next steps & Questions**



# Build it!

- We have many of the parts required:
  - Need to combine them
  - Some redesign required
  - Policy specification and control
- **VPFS - Virtual Policy-based File System:**
  - ASA + NRS + policy + security
  - Deadline based-scheduling (another talk 😊 )
- Integrate with CPU scheduling:
  - e.g. integrate with JYDE



# Questions

