Network management challenges for user-to-user eResearch collaboration

Russell Butson[1], David Eyers[2]
[1]University of Otago, Dunedin, New Zealand, russell.butson@otago.ac.nz
[2]University of Otago, Dunedin, New Zealand, david.eyers@otago.ac.nz

INTRODUCTION

The KISS project is a healthcare study using eResearch infrastructure, particularly centralised storage, to effect collaboration across teams spanning multiple organisations within New Zealand. The project has been a catalyst for exploring access to eResearch infrastructure by non-technologists within the University of Otago, particularly regarding the use of national New Zealand eResearch infrastructure.

KISS researchers collect sensitive video and biometric data at field sites that are uploaded at the Eastern Institute of Technology, which is transported electronically for analysis at the University of Otago, and the subsequent, analysed data shared back to the members within the participating research organisations.

Previously, couriers transported unencrypted data, on external storage devices. The idea of a centralised, managed, electronic storage facility was perceived by the research team as a more efficient and secure approach to handling their sensitive raw data. The "DataFabric" had been developed in New Zealand to facilitate such use – similar in functionality (and software) to the ARCS Data Fabric service within Australia. The recently formed National eScience Infrastructure (NeSI) will maintain and extend the DataFabric within New Zealand. In both cases, the open source iRODS storage middleware [1] has been used, with Davis – an extension developed by ARCS – providing WebDAV functionality.

Moving the KISS project to managed, network-accessible storage has presented a number of challenges, although promising technological and educational solutions have been discovered. We discuss our findings regarding the conceptions and experiences of the researchers, the difficulties of ensuring that existing user software environments work well with the DataFabric, security requirements, and network quality of service and management policy.

SOFTWARE INTEGRATION

The researchers participating in the project are capable users of typical Microsoft Windows operating system installations, but were not trained or experienced in aspects of system administration. They could see the benefits to be gained by using the DataFabric, and were willing to be taught how to use the framework, but naturally could not be expected to endure changes beyond a certain point regarding the practices required to achieve their primary research objectives. This provided a particular challenge regarding software support: in many cases, users of eResearch infrastructure are technical experts, or have highly experienced technical support nearby (for example, for use of high performance computing facilities).

Given that WebDAV was supported by the operating systems in common use by the researchers, the initial plan was to use the available WebDAV interface to iRODS. It was rapidly discovered that the implementation of WebDAV, despite the protocol being a web standard, was far too inconsistent across the different operating systems to be of practical use. The decision was taken to install the BitKinex WebDAV client to connect to iRODS (through Davis). The multi-protocol BitKinex tool presents a significant amount of unnecessary information to users, and transferring files in a manner that was unfamiliar to them.

Using the WebDAV layer over iRODS also uncovered some subtle software bugs. Thanks to iRODS experts at the Centre for eResearch at the University of Auckland, at least one of these is now fixed: we determined a race condition that caused iRODS to fail intermittently when effecting certain file creation operations. The actual protocol traces that caused the problem represented fairly pathological behaviour to begin with, however, namely rapid sequences of duplicated requests. Due to resource shortages, the precise cause of the original problem has not been tracked down, but due to the iRODS fix, it is no longer breaking the system.

Further usability problems were encountered regarding how the DataFabric shows files regardless of whether transfers have actually completed or not. To operate safely on collections of files, out-of-band signalling was necessary between geographically distributed KISS project participants.

Although the workflow required by the KISS participants is quite straightforward, it is only a particular subset of the possible ways in which the (highly flexible) iRODS clients might be used. This highlights the desirability of software systems that can be configured (in particular limited) to meet the specific needs of given researcher workflows – while still providing important notifications regarding the success or failure of operations appropriately.

Next, browser-based tools were investigated. Moving back to the browser would allow interfaces to be customised to suit particular research projects, however there are two main problems. First, browser sessions need not to be interrupted for large file transfers to work successfully, but the security-sandboxed environment of typical web browsers precludes spawning surreptitious, long-lived background processes. Second, the capabilities for uploads using the HTTP(S) network protocol are awkward at best, in terms of the KISS project's simple requirement: to upload a tree of directories. Recent browsers were surveyed, and it was confirmed that none provide the functionality that we need, and that in the extreme cases, popular browsers failed with very obscure (to end users) error messages, or behaved incorrectly when faced with directory uploads, even within HTML5. A single archive file can be uploaded, but this is undesirable when a researcher needs to "babysit" a process to create that archive prior to upload: interfering with their working practice.

In parallel with the ongoing use of the NeSI DataFabric, Information Technology Services (ITS) at the University of Otago have been experimenting with offering an iRODS service themselves, which would hopefully be able to be federated with the NeSI infrastructure in the near future.

This local pilot has facilitated the use of GUI clients that speak iRODS network protocols directly, specifically the recently released iDrop tool. So far researchers have been able to use this software effectively, although there are many improvements to the environment that increase its ability to work in parallel with researchers. We are working with the software developers to achieve these ends.

The different protocols and clients also show up another area of network-based challenge: authentication. The previous generation of NZ eResearch authentication was highly arduous for access to infrastructure, involving arms-length management of short-lived certificates, and the difficulty of Shibbolethised interfaces not being available to organisations whose central IT staff were not yet able to implement their own identity providers.

NETWORK CONDITIONS

Multiple difficulties have occurred regarding network configuration. This is not surprising, as the KISS project is attempting to connect a number of comparatively unprivileged computers on independent organisational networks. For example, in early May 2012, a network card broke in a device close to one of the DataFabric servers. The first problem is that monitoring such pieces of infrastructure currently is not possible outside the hosting organisation's network, usually. In this case the only symptom observable from the researchers was that the transfers were proceeding unacceptably slowly. In total, more than ten network and system administrators were involved across five organisations in diagnosing the fault. Beyond being expensive in terms of time and human resources, there was the further complication that the organisations were not able to synchronise their diagnoses on demand. REANNZ, the organization who run the KAREN academic network, reported that minimal traffic was present on KAREN during KISS transfers – it turns out that the routing policy at Otago now load-balances between the unmetered, academic KAREN network and a commercial ISP, which avoided all of the optimisations available on KAREN. It would actually have been more useful in this case, for the service simply to have failed, rather than for routing to have done a fail-over. This represents desktop users' eResearch concerns impacting on organisations' core network routing management and policy.

We are investigating support for the "Science DMZ" concept to facilitate a degree of this type of network quality of service being effected for researcher-to-researcher connectivity, across local and national networks. This looks likely to be implemented using the OpenFlow Software Defined Networking protocols. Coupled with this is a deployment of perfSONAR, which should hopefully facilitate organisations choosing to expose some of their internal network and system monitoring details to create a shared picture of the network and service "weather map".

At the University of Otago we have further challenges in terms of network functionality, as high-level human resources policy has been pushed down into an unduly technical implementation of Internet censorship that is likely to disrupt eResearch traffic. Significant effort may be needed to correct this network micromanagement.

Finally, the security of the KISS data is paramount. We currently define the responsibility for security to remain with the research project, but ideally this would become a centrally managed property. At the moment, it is difficult within the eResearch infrastructure to get all organisations involved to agree on the service level agreements required.

REFERENCES

1. Rajasekar, A., Moore, R., Hou, C., et al., *iRODS Primer: Integrated Rule-Oriented Data System*. Synthesis Lectures on Information Concepts, Retrieval, and Services, 2010. 2(1): pp. 1–143.

ABOUT THE AUTHORS

Russell Butson is a senior lecturer in Higher Education at the University of Otago. His primary area of interest is the practice of collaboration, particularly within virtual research environments. The emerging nature of this field has meant that Russell has had to design and create a number of virtual research environments in order to advance his research area. While there is a technical component to this work, the actual research is more sociological than technical; with an emphasis on the nature of knowledge working through the analysis of the environments, communities, and practices involved in the practice of researcher collaborations. Russell has been in the eResearch domain for some years, promoting the uptake of ICT to support the practice of research both nationally and locally at the University of Otago. He has coordinated the adoption of eResearch infrastructure within the KISS project. He is a member of the University of Otago's eResearch Advisory Group and an advocate for eResearch generally.

David Eyers is a lecturer in Computer Science at the University of Otago. He does research into wide-area distributed systems, with particular interests in decentralized security and efficient network communication. Networking topics of relevance to this project include distributed access control systems, encrypted data management and techniques for integrated provenance tracking of sensitive information. David is interested in tools and technologies that can facilitate decentralized and collaborative network configuration and monitoring, particularly when spanning different administrative domains. He believes that academic networks need to differentiate themselves from typical commercial internet service providers by being able to support research-relevant quality of service from researcher to researcher across organisations. He is also interested in technical systems that can reflect security policy based on the high-level policies agreed in research project proposals that manage sensitive data. Undertaking eResearch projects facilitates the grounding of academic research topics in services that are of practical use to researchers.