

# Exploitation of Switched Lightpaths for e-Science Applications: an e-Health mini-project

Lee Momtahan and Andrew Simpson

Oxford University Computing Laboratory  
Wolfson Building, Parks Road, Oxford OX1 3QD

## Abstract

The Exploitation of Switch Lightpaths for e-Science Applications (ESLEA) project aims to pioneer the use of the UKLight network with application from a number of e-Science domains, with one of these domains being the field of e-Health. In this paper, we provide a motivation for the use of a network such as UKLight within the healthcare domain and describe the work being undertaken within the ESLEA project in this respect.

## 1. Introduction

The future of networking is moving towards switching in the optical domain. In this respect, several leading national network organizations are in the process of creating an international experimental test-bed to pilot this new paradigm. These include StarLight in the USA, SURFNET in the Netherlands (through the NetherLight project), the Canadian academic network CANARIE, and CERN in Geneva. Most recently the Czech Republic and NORDUnet have joined the international switched optical network community.

The UK research community is able to participate in this global initiative through UKLight. The UKLight network was initiated through a consortium of UK researchers in collaboration with UKERNA, the providers of SuperJANET. The researchers span a wide interest base from high-demand e-Science applications, through to network systems groups and the optical communications engineering research community. The UKLight initiative was given backing by the Joint Information Systems Committee (JISC) and funded via HEFCE.

The Exploitation of Switch Lightpaths for e-Science Applications (ESLEA) project [1] aims to pioneer the use of the UKLight network with applications from a number of e-Science domains that involve the transfer large volumes of data across Wide Area Networks (WANs). The project commenced in February 2005, and is a collaboration between six UK universities, more than a dozen overseas partners and two industrial partners (Cisco Systems and British Telecom). The ESLEA project aims to

demonstrate the potential of circuit-switched optical networking to the UK e-Science community via a series of 'mini-projects' aimed at demonstrating potential benefits. One of the mini-project exemplars involves demonstrating the relevance for e-Health applications, and it is this aspect of the project that the current authors have primary responsibility for.

The e-Health mini-project will take a two-pronged approach. This is in recognition of the fact that Grid-based e-Health applications are some distance from being deployed "for real" within the NHS: there are many technical, social, legal and ethical issues that must first be overcome before such deployment can happen.

The first aspect of the approach is the consideration and documentation of the above generic issues. This analysis will be carried out in conjunction with the GIMI consortium [2], which is concerned with tackling the interoperability and security issues associated with the deployment of a product-level health grid to support collaborative research and healthcare delivery. The project has three application areas through which the middleware will be validated: the self-management of chronic conditions; the linking of images associated with colorectal cancer; and the development of a training system for radiologists. The project builds upon the experiences of the e-DiaMoND project [5], and should provide fertile ground for investigating the potential of hybrid optical networks as two of three applications require the secure, timely transfer of large data sets of high-resolution medical images.

The second aspect of the approach will be the utilisation of UKLight to provide tangible

benefits to an e-Health project that is currently ongoing: the Integrative Biology project (IB) [3]. As well as developing an infrastructure to facilitate distributed, collaborative research, IB is developing detailed computational models in order to study heart disease and cancer tumours.

The IB project utilises computational resources in number of High Performance Computing sites distributed across the UK. As the level of detail modelled and number of simulation runs is ramped up in the quest for greater accuracy, the use of hybrid optical networks to deal with resultant quantities of data becomes not only attractive, but inevitable. Beyond the simple transfer of simulation data, real time visualisation will be used to demonstrate the benefits of UKLight. Furthermore the potential to run a model on heterogeneous computational resources, coupled by UKLight, will be explored.

Each of the use cases being considered feature non-functional concerns relating to security, performance and quality of service. These concerns are typical within the e-Health domain, and, as such, this is an appropriate candidate for validating the potential of hybrid optical networks.

## 2. Switched lightpaths

Research networks now have to accommodate a wide range of users. Emerging from the fields of physics, astronomy, biology and medical research are a set of users engaged in high-profile cutting-edge research, and most likely using Grid technologies. Although small in number, this set of users accounts for a significant proportion of the traffic flow through the network.

These users require the network to provide high-throughput transfers, often with tight quality of service constraints, between a well-defined set of nodes. For economic reasons it makes sense to use circuit switching technologies to accommodate these users.

When all the packets in a large data flow are going to the same destination there is a very high cost in examining each packet to determine where it should be routed, as is the case with traditional packet switched networks. In principle, it is far simpler to provide switched paths that switch protocol independent flows between end-points, as is done by optical networks. New technologies allow many paths

to be easily multiplexed/de-multiplexed to/from the fibre as individual wavelengths of “lambdas”, creating the possibility of low-cost dynamic switching suitable for research networks. The resulting lightpaths are very high capacity, avoid the expense of routing, have very predictable delay properties and minimal jitter, and, as they carry a unique flow, avoid problems such as contention, congestion, packet loss and packet reordering. In short, such technology is ideal for many e-Science applications.

## 3. The generic: issues and use cases

In considering the generic issues pertaining to the utilisation of high-speed networks within e-Health applications, the wider e-Health community will be engaged to capture potential use cases. Initially, the use cases will be derived from the DTI-funded GIMI (Generic Infrastructure for Medical Informatics) project.

The primary aim of GIMI is to develop a service-based infrastructure to support healthcare research, and, in the longer term, healthcare delivery. The key technical challenges that the GIMI project is seeking to solve are interoperability and security: GIMI must be interoperable in the sense that the system aims to interface with systems deployed within the NHS; it must be secure in the sense that the data owner determines exactly under what circumstances their data may be accessed. A generic medical informatics infrastructure capable of interfacing with technology deployed within the NHS will, amongst other things, be in the spirit of the government's goal of joined-up healthcare; aid the self-management of diseases for those suffering from long-term illnesses; and form the basis for clinical training applications.

A typical application builds upon the work of e-DiaMoND [5] and is associated with the development of a tool to support mammography radiologist training. A requirement of the tool is that it should mimic as much as possible the clinical environment. In this respect, a mammography radiologist will typically read, analyse, and make a decision concerning the actions to be taken for a patient, in approximately thirty seconds. This means that, in general, a radiologist will perform in the region of 100 readings per one-hour session. This amounts to approximately 100GB of data per reading session (assuming full field digital mammography). A typical screening centre will screen approximately 100 women per day. The

average woman will be having her fourth scan, so will have three previous screening sessions' worth of data (plus any other investigative data). In reality, if a clinic were to subcontract screening work, we can expect that it would do so on a batch basis, i.e., a day's work, which is 100 patients' worth of data. On average, this would mean three previous sets of four images plus one set of four images, i.e., 16 images, for each patient. If each image was fully digital, this would result in  $16 * 75M = 1.2GB$  of data to be transferred.

The investigation of the potential for other uses in this domain is ongoing. Of particular relevance in this regard, is the NeuroGrid project [10], the principal objective of which is to enhance collaboration within and between clinical researchers in different domains, and between clinical researchers and e-scientists.

#### **4. The specific: Integrative Biology**

The Integrative Biology project, which started in February 2004 is managed by the University of Oxford, and has a project team that is spread across six Higher Education Institutions in the UK (Oxford, Nottingham, Leeds, Birmingham and Sheffield Universities, and University College, London), the Central Laboratory of the Research Councils (CCLRC), and IBM UK. In addition, international collaborators are drawn from the Universities of Auckland in New Zealand, Tulane, UCSD and UCLA in the USA, Graz in Austria and Utrecht in the Netherlands.

IB aims to develop a customised virtual research environment that is capable of supporting research scientists in the integration of biological information across multiple spatiotemporal scales. Such integration should provide comprehensive descriptions at the system level, which can aid in the determination of biological function in both normal- and patho-physiology.

The IB framework is being built upon middleware developed in earlier e-Science projects, including <sup>m</sup>yGrid [4], e-DiaMoND [5], BioSimGrid [6], RealityGrid [7], gViz [8], and GEODISE [9].

It is intended that the tools developed by IB will support and accelerate the work of clinical and physiological researchers. The long-term beneficiaries of the work should be patients with heart disease and cancer, which together cause 60% of all deaths within the UK.

To this end, IB's end-users are drawn from the bio-mathematical modelling, cardiac physiology, and clinical oncology communities. These groups all wish to make use of mathematical modelling and HPC-enabled simulation to investigate multi-scale and multi-physics phenomena in the modelling of the heart or cancer. All of this work has the following generic characteristics: it is grounded in experimental data; the mathematical models are complex, and usually involve the solution of non-linear systems of coupled ordinary and partial differential equations (or equivalent stochastic representations of the underlying processes on very small spatial scales) in complex, deformable three-dimensional geometries; very large quantities of simulation data are generated which must be validated against experiment, curated, synthesised, analysed, and visualised; efficient solution of the equations on HPC facilities requires assistance in developing MPI versions of codes, together with computational steering and performance control tools, and will be enhanced by provision of suitable workflow support tools.

To provide some quantification of the scale of activity that IB is looking to support, consider the following. Current models of action-potential propagation within whole-heart models can involve finite element meshes with tens of millions of nodes, and up 50 solution variables. Solutions in real-time for just a single heartbeat require (at least) millisecond temporal accuracy, resulting in data generation in the 10-100 TeraByte range from a single simulation.

#### **5. ESLEA and IB**

The Integrative Biology project will use a variety of computational and storage resources distributed throughout the UK, including HPCx at Daresbury, CSAR at Manchester and NGS Data Clusters at RAL. Linking these with UKLight will provide significant benefits to the end-users.

In the first phase, we plan to set up a link between Daresbury and RAL. As simulation model runs ramp up, TeraBytes of simulation data will need to be transferred between the main computation site at Daresbury, and the main storage site at RAL. UKLight can provide a dedicated multi-Gigabit per second link between the two.

In the second phase, we plan to utilise the same link for visualisation. Data being streamed for storage will also be passed to visualisation servers based at RAL. This processing will reduce the high bandwidth stream of raw simulation results into a relatively low bandwidth stream of geometry information. This can then be shipped to the end-user for rendering on their desktop over normal low bandwidth networks. When used for computational steering, the visualisation needs to occur at near real-time speeds. This phase will test the ability of UKLight to provision high bandwidth dedicated circuits with reliable quality of service.

Our third phase is more ambitious, and involves the execution of a heart model across heterogeneous compute resources. In modelling heart mechanics, the equations that model the mechanical pumping action are coupled to the electrical activity generating the muscular contractions via a large system of ordinary differential equations which model the ion flows through the cell membrane. A potentially attractive approach is to solve the ODE systems on a massively parallel processing (MPP) architecture whilst simultaneously solving the mechanical problem on a symmetric multi processing (SMP) architecture. To make this approach feasible would require very rapid communications between the HPC resources, with minimal latency, jitter and packet loss. We therefore plan to use UKLight to couple the MPP and SMP resources (respectively the HPCx and CSAR facilities), to provide an optimal architecture on which to execute the heart model.

This phase will not only stress aforementioned quality of service characteristics of the switched lightpath, but will also provide challenges in the co-allocation of UKLight, HPCx and CSAR resources. The extent to which the benefits of this approach outweigh the cost is a key research question that we aim to provide the answer to.

## 6. Conclusion

The primary motivation for the ELSEA project is the demonstration of the potential of switched lightpaths for e-Science applications, and within the project the authors are tasked with demonstrating the potential for e-Health applications. We aim to demonstrate this potential in two ways: via engagement with, and

delivering benefits for, a specific project – Integrative Biology; and by soliciting other potential use cases from other projects within the domain. By taking this two-pronged approach, we hope to convince the wider e-Health community of the potential value of UKLight.

## References

1. ESLEA home page.  
[www.eslea.uklight.ac.uk](http://www.eslea.uklight.ac.uk)
2. GIMI home page.  
[www.gimi.ox.ac.uk](http://www.gimi.ox.ac.uk)
3. Integrative Biology home page.  
[www.integrativebiology.ox.ac.uk](http://www.integrativebiology.ox.ac.uk)
4. myGrid home page.  
[www.mygrid.org.uk](http://www.mygrid.org.uk)
5. e-DiaMoND home page  
[www.ediamond.ox.ac.uk](http://www.ediamond.ox.ac.uk)
6. BioSimGrid home page.  
[www.biosimgrid.org](http://www.biosimgrid.org)
7. RealityGrid home page.  
[www.realitygrid.org](http://www.realitygrid.org)
8. gViz home page.  
[www.comp.leeds.ac.uk/vis/gviz](http://www.comp.leeds.ac.uk/vis/gviz)
9. GEODISE home page.  
[www.geodise.org](http://www.geodise.org)
10. NeuroGrid home page  
[www.neurogrid.ac.uk](http://www.neurogrid.ac.uk)