PROCEEDINGS OF SCIENCE

Use of UKLight as a Fast Network for Data Transport from Grid Infrastructures

M.-A. Thyveetil*, **S. Manos, J. L. Suter and P. V. Coveney**
Centre for Computational Science, Department of Chemistry, University College London,
20 Gordon Street, London, United Kingdom, WC1H 0AJ.
E-mail: m.thyveetil@ucl.ac.uk

Large-scale molecular dynamics simulations run on high-end supercomputing facilities can generate large quantities of data. We simulate mineral systems up to 10 million atoms in size in order to extract materials properties which are otherwise difficult to obtain through existing experimental techniques. Simulating clay systems this large can generate large files up to 50GB in size. These simulations were carried out on remote sites across the US TeraGrid and UK's HPCx supercomputer. Using the dedicated network, UKLight, connected to these high-end supercomputing resources has significantly reduced the time taken to transport large quantities of data generated from our simulations. UKLight provides excellent quality of service, with reduced packet loss and latency. This enhanced data transfer method paves the way for faster communication between two coupled applications, such as in the case of real-time visualisation or computational steering.

Lighting the Blue Touchpaper for UK e-Science - Closing Conference of ESLEA Project
March 26-28 2007
The George Hotel, Edinburgh, UK

*Speaker.

1. Introduction

Computational grids [1, 2, 3] provide an attractive environment in which to undertake the intensive compute tasks required for large-scale molecular dynamics (MD) simulation. Large-scale atomistic simulations, which we define as containing more than 100,000 atoms, provide a bridge between atomistic and mesoscopic scale simulations [4]. Operating over at least tens of thousands of atoms, emergent mesoscopic properties are observed in full atomistic detail. Large-scale simulations of clay nanocomposites reveal long wavelength, low amplitude thermal undulations [5]. Small simulation sizes implicitly inhibit long wavelength clay sheet flexibility due to the periodic boundaries used in condensed matter molecular dynamics, which effectively pin the clay sheet at the edges of the simulation cell. It is often difficult to predict how nanocomposites will behave from theories of conventional composite behavior due to the disparity of dimensions; hence the need for large-scale molecular simulation to sample all possible length scales[6, 7, 8, 9]. We perform many simulations at various system sizes; the largest approaches that of realistic clay platelets and contains upwards of a million atoms. Using computational grid resources allows the turnaround on the large number of simulations required to be on a feasible timescale. We utilise Grid resources on the US TeraGrid and the UK's flagship machine HPCx. Jobs are launched remotely using the Application Hosting Environment [10][11].

With increasingly large simulation sizes now possible, new problems appear as our largest simulations can generate files up to 50GB in size. These files contain important atomistic data which need to be retrieved from the remote Grid resource to a local machine, but this can become time consuming. An answer to the problems of slow and unreliable networks is to use switched-circuit networks. In a switched-circuit network the user has sole use of a dedicated network connection, thus eliminating contention with other traffic and providing excellent, predictable, Quality of Service (QoS). Switched-circuit networks can be implemented in various ways, though there has been much recent work on allocating users or groups sole use of individual wavelengths (lambdas) in multi-wavelength optical fibres[12]. In the UK dedicated connections are available via the UKLight network which uses manually-configured SDH circuits. In this paper we present network tests intended to optimise the use of local machines at UCL connected by UKLight to external Grid resources. We conclude with a summary of our findings and future plans we have with UKLight.

2. Performance Testing of UKLight

Currently, a lambda network operates in the UK called UKLight[1]. It provides a fast connection between UCL and various supercomputing resources such as EPCC's HPCx[2] and the TeraGrid[3]. We carried out a series of tests on the performance of the link between UCL and HPCx, as well as UCL and the TeraGrid. Preliminary results showed that the link was not as fast as expected. Two methods could have been used in order to improve the bandwidth of the link. The first was to use GridFTP in order to use multiple streams over one link. The problem with this method is that it can be difficult to implement on a network such as UKLight. This led us to use common software

[1] http://www.uklight.ac.uk

[2] http://www.hpcx.ac.uk

[3] http://www.teragrid.org

such as SSH, along with network tuning to achieve maximum bandwidth over a single stream. Specifically, we needed to tune the TCP window size in order to achieve maximum bandwidth. This section describes the methodology we used to test the system and the results we obtained once network parameters were tuned.

2.1 Methodology

Currently high-performance dedicated network connections are provided in the UK by the UKLight[4] network. The first connection studied was between a Linux box connected to the same UKLight switch as UCL's SGI Prism and a box connected to the UKLight switch of HPCx. The second connection was between UCL and the TeraGrid's IA-64 Linux cluster NCSA. NCSA's network parameters were already tuned, therefore a separate machine was not needed for testing.

Iperf is a network tool which we used to test UDP and TCP bandwidth between networked computers[5]. Iperf UDP tests provided the maximum bandwidth of the connection before packet loss is seen. This test was carried out in both the production network and UKLight. In all UDP tests the Grid resource acted as the client while the UCL Linux machine was the server. The result of the UDP tests can be used to calculate the bandwidth delay product (BDP) of the link, which is found by: bandwidth × round trip time. The round trip time is the time elapsed for a message to travel to a remote place and back again. The BDP helps determine the maximum window size for TCP communication.

In order to get the best performance from a network, the TCP window size defined on the kernel and application side, needs to be tuned. The Linux kernel parameters which we needed to adjust were as follows:

$/proc/sys/net/core/wmem\_max$
$/proc/sys/net/core/rmem\_max$
$/proc/sys/net/ipv4/tcp\_rmem$
$/proc/sys/net/ipv4/tcp\_wmem$

In addition, we used the application called High Performance Enabled SSH/SCP (HPN-SSH)[6]; this is a patch for recent OpenSSH releases which allows adjustment of the TCP window size within the application. With these changes in place, we then tested the performance of SCP on all connections, including the production network, UKLight with untuned network parameters and also with tuned network parameters.

2.2 Results

The UDP tests showed that the connection between UCL and other grid resources could be as large as 40MB/s, as summarised in Table 1. The TCP tests showed that a maximum of 34.4MB/s could be achieved, as shown in Table 2. Using these parameters, the Linux kernel parameters were tuned and HPN-SSH was used to compare the data transfer rates for the production network, as well

[4] http://www.uklight.ac.uk

[5] http://dast.nlanr.net/Projects/Iperf

[6] http://www.psc.edu/networking/projects/hpn-ssh

as the untuned and tuned UKLight network. As summarised in Table 3, a massive improvement can be seen using tuned network parameters. The academic Super Janet production network operates at 4.5MB/s for connections within the UK and 600KB/s from UCL to TeraGrid's NCSA, whilst with tuned network parameters, the bandwidth of the connection goes up to 16MB/s for the NCSA UKLight link and 28MB/s for the HPCx link.

Grid resource	Maximum Bandwidth (MB/s)	Round trip time	Bandwidth delay product
NCSA	40 MB/s	92ms	3.2MB
HPCx Linux	32MB/s	8ms	300KB

Table 1: Results for UDP tests of UKLight connection between UCL and Grid resources. The Grid resource acted as the Iperf client while the UCL Linux machine was the server.

Grid resource	Maximum Bandwidth (MB/s)	Maximum Window Size (MB)
NCSA	34.3 MB/s	4MB
HPCx Linux	34.4MB/s	3MB

Table 2: Results for TCP tests of UKLight connection between UCL and Grid resources. The Grid resource acted as the Iperf client while the UCL Linux machine was the server.

	Maximum bandwidth (MB/s)		
Grid resource	Janet Network	UKLight (untuned)	UKLight (tuned)
NCSA	0.6 MB/s	0.7 MB/s	16 MB/s
HPCx Linux	4.5 MB/s	8 MB/s	28 MB/s

Table 3: Comparison of networks using the maximum bandwidth obtained from SCP. A great improvement is seen using UKLight over the academic Super Janet network.

3. Conclusion

We conclude that UKLight makes a significant impact when transporting large files from Grid resources. In order to achieve the maximum bandwidth of this link over a single stream network parameters must be tuned. The results also highlight the fact that UKLight provides an efficient way to transport data for more complicated uses such as real-time visualisation and computational steering[13]. Visualisation of molecular simulations as presented in this study is very important but with system sizes greater than 1 million atoms, most visualisation is currently carried out only after the simulation has completed. Ideally we would like to be able to carry out real-time visualisation in order to see the evolution of the system while it is running. In addition to this, computational steering provides a way to interact with and monitor a simulation.

Real-time visualisation and computational steering are processes which cannot be carried out on batch systems, used by most high performance computing facilities. This means that advanced reservation and co-scheduling of the resources are needed, so that the user is able to determine when their simulation has started. In the future we hope to be able to use UKLight in order to carry out real-time visualisation and steering of our clay nanocomposite simulations across UKLight with the aid of co-scheduling.

References

[1] P. V. Coveney, editor, "Scientific Grid Computing", *Phil. Trans. R Soc. A* **7** (2005) 24–32.

[2] I. Foster, C. Kesselman and S. Tuecke, "The anatomy of the grid: Enabling scalable virtual organizations", *Intl J. Supercomp. Appl.* **15** (2001) 3–23.

[3] B. Boghosian and P. V. Coveney, "Scientific applications of grid computing", *Comp. Sci. and Eng.* **7** (2005) 10–13.

[4] P. Boulet, P. V. Coveney and S. Stackhouse, "Simulation of hydrated Li^{+}-, Na^{+}- and K^{+}-montmorillonite/polymer nanocomposites using large-scale molecular dynamics", *Chem. Phys. Lett.* **389** (2004) 261–267.

[5] J. L. Suter, P. V. Coveney, H. C. Greenwell, and M.-A. Thyveetil, "Large-Scale Molecular Dynamics Study of Montmorillonite Clay: Emergence of Undulatory Fluctuations and Determination of Material Properties", *J. Phys. Chem. C*, *in press* (2007).

[6] H. C. Greenwell, W. Jones, P. V. Coveney, and S. Stackhouse, "On the application of computer simulation techniques to anionic and cationic clays: A materials chemistry perspective", *J. Mater. Chem.* **16** (2006) 708–723.

[7] H. C. Greenwell, A. A. Bowden, B. Q. Chen, P. Boulet,, J. P. G. Evans, P. V. Coveney, and A. Whiting, "Intercalation and in situ polymerization of poly(alkylene oxide) derivatives within M+-montmorillonite (M = Li, Na, K)", *J. Mater. Chem.* **16** (2006) 1082–1094.

[8] E. S. Boek, P. V. Coveney, and N. T. Skipper, "Monte Carlo molecular modeling studies of hydrated Li-, Na-, and K-smectites: Understanding the role of potassium as a clay swelling inhibitor", *J. Am. Chem. Soc.* **117** (1995) 12608–12617.

[9] H. C. Greenwell, M. J. Harvey, P. Boulet, A. A. Bowden, P. V. Coveney, and A. Whiting, "Interlayer structure and bonding in nonswelling primary amine intercalated clays", *Macromolecules* **38** (2005) 6189–6200.

[10] P. V. Coveney, R. Saksena, S. J. Zasada, M. McKeown and S. Pickles, "The Application Hosting Environment: Lightweight Middleware for Grid-Based Computational Science", *Comp. Phys. Comm.* **176** (2007) 406–418.

[11] S. Zasada, "The Application Hosting Environment: Lightweight Middleware for Grid Based Computational Science", *Lighting the Blue Touchpaper for UK e-Science - Closing Conference of ESLEA Project*, (2007).

[12] A. Hirano, L. Renambot, B. Jeong, J. Leigh, A. Verlo, V. Vishwanath, R. Singh, J. Aguilera, A. Johnson, and T. A. DeFanti, "The first functional demonstration of optical virtual concatenation as a technique for achieving terabit networking", *Future Gener. Comput. Syst.*, 22 (2006) 876–883.

[13] M. Harvey, S., Jha, M.-A. Thyveetil and P. V. Coveney, "Using Lambda Networks to Enhance Performance of Interactive Large Simulations", *Second IEEE International Conference on e-Science and Grid Computing (e-Science'06)*, (2006) 40.