PROCEEDINGS OF SCIENCE

Using lambda networks to enhance performance of interactive large simulations *

Matthew J Harvey

Imperial College London

Shantenu Jha†

Louisiana State University and University College London

Mary-Ann Thyveetil

University College London

Peter Coveney

University College London

The ability to use a visualisation tool to steer large simulations provides innovative and novel usage scenarios, for example, the ability to use new algorithms for the computation of free energy profiles along a nanopore [1]. However, we find that the performance of interactive simulations is sensitive to the quality of service of the network with latency and packet loss in particular having a detrimental effect. The use of dedicated networks (provisioned in this case as a circuit-switched, point-to-point optical lightpath or *lambda*) can lead to significant (50% or more) performance enhancement. When running on say 128 or 256 processors of a high-end supercomputer this saving has a significant value. We discuss the results of experiments performed to understand the impact of network characteristics on the performance of a large parallel classical molecular dynamics simulation when coupled interactively to a remote visualisation tool.

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

*This talk is a condensed version of work originally presented in Ref. [9].

†Speaker.

PoS(ESLEA)014

1. Introduction

Lambda networking involves using different wavelengths (lambdas) of light in fibres for separate connections. Lambda networks provide high-levels of Quality of Service (QoS) by giving applications and user communities dedicated lambdas on a shared fibre infrastructure. The implementation requires Dense Wavelength Division Multiplexing (DWDM) to accommodate many wavelengths on a fibre, optical switches, and other optical networking equipment. Grid computing applications have so far mostly made use of best-effort, shared TCP/IP networks, *i.e.* the network has simply been the glue that holds the middleware-enabled computational resources together. In contrast, by using lambdas the networks themselves are schedulable "first class" grid resources. These deterministic lambda networks, carrying one or more lambdas of data, form on-demand, end-to-end dedicated networks, often called lightpaths; lightpaths form the basis of the next generation of network-centric applications.

Lightpaths have the ability to meet the needs of very demanding e-science applications as a consequence of their ability to provide several features that are not possible using regular, production best-effort networks. These include providing higher bandwidth connections (*e.g.* [2]), user-defined networks [3], implementation of novel protocols [4] and provide essentially contention-free and high quality-of-service links.

Most applications however, have tended to use lambdas for their high bandwidth alone. For example, an important class of applications driving the development and research of lambdas are visualisation of large and complex data sets. Here we report on one of the first uses of lambdas to couple interactive, steered visualisation with "active" simulations. This work was conducted as part of the SPICE (Simulated Pore Interactive Computing Environment) project details of which have been discussed elsewhere [1, 9] . In the next section, we describe the project's motivating scientific problem and the technical solutions adopted.

2. Simulated Pore Interactive Computing Environment

The transport of bio-molecules like DNA, RNA and poly-peptides across protein membrane channels is of primary significance in a variety of areas. Although there has been a flurry of recent activity, both theoretical and experimental [5, 6], aimed at understanding this crucial process, many aspects remain unclear.

Of the possible computational approaches, classical molecular dynamics (MD) simulations of bio-molecular systems have the ability to provide insight into specific aspects of a biological system at a level of detail not possible with other simulation techniques. MD simulations can be used to study details of a phenomenon that are often not accessible experimentally [7] and would certainly not be available from simple theoretical approaches. However, the ability to provide such detailed information comes at a price: MD simulations are extremely computationally intensive – prohibitively so in many cases. As was discussed in Ref. [1], advances in both the algorithmic and the computational approaches are imperative to overcome such barriers.

SPICE, the Simulated Pore Interactive Computing Environment project [1], implements a method, henceforth referred to as SMD-JE, to compute the free energy profile (FEP) along the vertical axis of the protein pore. This method reduces the computational requirement for the problem

of interest by a factor of at least 50-100, at the expense of introducing two new variable parameters, with a corresponding uncertainty in the choice of their values. Fortunately, the computational advantages can be recovered by performing a set of "preprocessing simulations" which, along with a series of interactive simulations, help inform an appropriate choice of the parameters. To benefit from the advantages of the SMD-JE approach and to facilitate its implementation at all levels – interactive simulations of large systems, the pre-processing simulations and finally the production simulation set – we use the infrastructure of a federated trans-Atlantic grid [8].

Interactive simulations involve using the visualiser as a steerer, *e.g.* to apply a force to a subset of atoms, Figure 1(b), and requires bi-directional communication – there is a steady-state flow from the simulation to the visualiser as well as the visualiser to the simulation. As a consequence of requiring geographically distributed resources, high-end interactive simulations are dependent on the performance of the network between the scientist (visualiser) and the simulation. Unreliable communication leads not only to a possible loss of interactivity, but equally seriously, a significant slowdown of the simulation as it waits for data from the visualiser.

On switching traffic flow from the production network to a lambda network, we found an improvement in the performance of around 50%. The simulation performance over the production network varied, *i.e.* was apparently sensitive to prevailing network conditions. Interactive MD simulations thus require high quality-of-service – as defined by low latency, jitter and packet loss – networks to ensure reliable bi-directional communication. This leads to the interesting situation where large-scale interactive computations require both computational and visualisation resources to be co-allocated with networks of sufficient QoS [8].

3. Experiment

In order to quantify the impact of network performance characteristics on the efficiency of an interactive MD simulation a series of measurements were made under controlled conditions. The full details of our methodology are described in [9] and are presented here in outline.

The two compute resources on which the molecular dynamics simulation (using NAMD[10]) and visualiser were run (VMD[11]), were connected via a dedicated circuit on the UKLight[12] optical network, provisioned at 300Mbps. UKLight provides the user with sole-use of dedicated, manually-configured SDH circuits which hence have similar properties to dedicated lambda networks.

To control the characteristics of the network, a third system was introduced between the visualiser and the UKLight link. This system acted as an IP bridge and employed the NISTnet [13] package to modify the traffic flow. The wall-time per simulation timestep (t_s) was measured for interactive simulations over a range of network characteristics controlled by NISTnet. The parameters varied were 1) Packet transmission delay (α), to simulate different latency network paths, 2) Packet Loss (β), to simulate packet loss due to network congestion.

3.1 Quality of Service: Latency

The effective latency of the UKLight link was varied to emulate different paths, service times and congestion – all characteristic of best effort networks between two given end points.

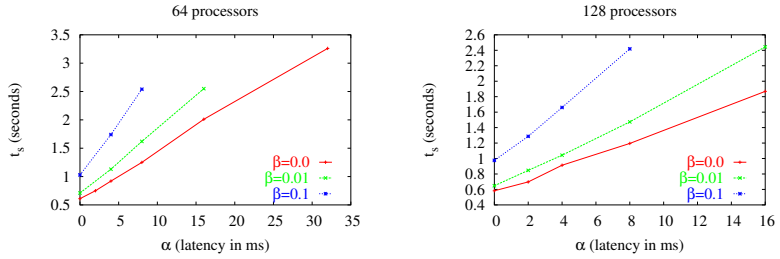

Figure 1: Plots showing the effect of latency on the performance (t_s). An increase in latency leads to a linear increase in the wall-time taken per simulation timestep, independent of the number of processors used. The performance degradation remains linear for different values of β.

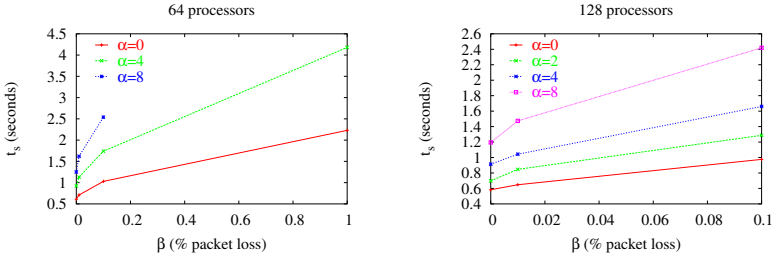

Figure 2: Plots showing the dependence of performance (t_s) on the packet loss (β) for different values of α. The qualitative characteristics remain the same for 64 and 128 processors, *i.e.* at a fixed value of α, the ratio of t_s for a pair of β values is similar for the two processor counts. For example for α = 4, the ratio of t_s at β = 0.1 and 0.01, is 1.54 and 1.59 for 64 and 128 respectively. At the same values of β but for α=8 the values of the ratio are 1.56 and 1.62 respectively.

Not surprisingly the time taken for each timestep increases linearly with greater latency. Our results are plotted in Fig. 1. It is interesting to note, that although the average wall-time taken per simulation time-step is of the order of hundreds of milli-seconds, introducing latencies of a few milli-seconds has a significant effect. This is attributed to the fact that a significant fraction of the simulation time is spent waiting for I/O operations to complete. Thus we conclude that reducing any avoidable latency is a good performance enhancing strategy.

3.2 Packet Loss Effects

Packet loss is interpreted by the TCP protocol as an indication of congestion and causes the window size to be immediately reduced to a minimum size (4096 bytes in this case) and then renegotiated up. The effect of increasing β (as shown in Fig. 2 is to increase the frequency of window size reduction (reducing the average size over time) and consequently reducing effective throughput. We observe that, in general, default settings for TCP window size parameters are unsuited to networks with high bandwidth-delay products.

4. Conclusion

It can be argued that with significant effort, a highly optimised I/O mechanism could be implemented within the NAMD code to withstand performance degradation arising from production networks. Whereas we do not contest that this in principle is possible, doing so would require significant re-factoring of a very complex code which has been developed by the community over many years (we estimate the number of person-years effort to be easily a hundred). Equally important, it is impractical to aim to introduce special-purpose code for every unique usage scenario; thus it is highly desirable to be able to use the same general-purpose code over a wide range of scientific problems and usage scenarios. Our efforts to quantify the advantages of lightpaths need to be understood in the above mentioned context.

Not only can grids use lightpaths to more closely integrate distributed environments, but they *must* use lightpaths to couple distributed environments to overcome some of the bottlenecks of traditional programming methodologies of high performance codes as well as problem solving approaches for challenging scientific problems. SPICE provides an example of a large-scale problem that depends on using algorithms amenable to distributed computing techniques and then implementing them on grids. In order to effectively utilise these algorithms, interactive simulations on large computers are required.

In order to enable meaningful interactive exploration, the responses must be computed in reasonable times. Thus as larger systems are studied – MD simulations of a million atoms are now just appearing in the literature [15] – not only will larger computers be required, but the need for efficient and reliable communication will also grow.

5. Acknowledgements

This work has been supported by EPSRC grant number GR/T04465/01 (ESLEA), by EPSRC Grant EP/D500028 (SPICE) and the EPSRC-funded RealityGrid project (GR/R67699 and EP/C536452/1).

References

[1] S. Jha, P. V. Coveney, M. J. Harvey, and R. Pinning, "SPICE: Simulated Pore Interactive Computing Environment," *Proceedings of the 2005 ACM/IEEE conference on Supercomputing*, p. 70, 2005, dx.doi.org/10.1109/SC.2005.65.

[2] 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 Generation Computer Systems*, vol. 22, pp. 876–883, 2006.

[3] J. Mambretti, R. Gold, F. Yeh, and J. Chen, "Amroeba: Computational astrophysics modeling enabled by dynamic lambda switching," *Future Generation Computer Systems*, vol. 22, pp. 949–954, 2006.

[4] R. L. Grossman, Y. Gu, D. Hanley, M. Sabala, J. Mambretti, A. Szalay, A. Thakar, K. Kumazoe, O. Yuji, and M. Lee, "Data mining middleware for wide-area high-performance networks," *Future Generation Computer Systems*, vol. 22, pp. 940–948, 2006.

[5] D. K. Lubensky and D. R. Nelson. *Phys. Rev E*, 31917 (65), 1999; Ralf Metzler and Joseph Klafter. *Biophysical Journal*, 2776 (85), 2003; Stefan Howorka and Hagan Bayley, *Biophysical Journal*, 3202 (83), 2002.

[6] A. Meller *et al*, *Phys. Rev. Lett.*, 3435 (86) 2003; A. F. Sauer-Budge *et al. Phys. Rev. Lett.* 90(23), 238101, 2003.

[7] M. Karplus and J. A. McCammon, "Molecular Dynamics Simulations of Biomolecules," *Nature Structural Biology*, vol. 9, no. 9, pp. 646–652, 2002.

[8] B. Boghosian, P. Coveney, S. Dong, L. Finn, S. Jha, G. Karniadakis, and N. Karonis, "Nektar, SPICE and Vortonics – Using Federated Grids for Large Scale Scientific Applications," in *Proceedings of Challenges of Large Applications in Distributed Environments (CLADE) 2006*, vol. IEEE Catalog Number: 06EX13197, Paris, June 2006, pp. 32–42, iSBN 1-4244-0420-7.

[9] M. J. Harvey, S. Jha, M. A. Thyveetil, and P. V. Coveney, "Using lambda networks to enhance performance of interactive large simulations," *2nd IEEE International Conference on e-Science and Grid Computing, 4-6 December 2006, Amsterdam*, 2006.

[10] J. C. Philips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schilten, "Scalable molecular dynamics with NAMD," *Journal of Computational Chemistry*, vol. 26, pp. 1781–1802, 2005.

[11] W. Humphrey, A. Dalke, and K. Schulten, "VMD – Visual Molecular Dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.

[12] JISC, "UKLight Switched Optical Lightpath Network," http://www.uklight.ac.uk

[13] M. Carson and D. Santay, "NISTNet - A Linux-based Network Emulation Tool," *Computer Communication Review*, vol. 6, 2003.

[14] U. o. S. C. Information Sciences Institute, 1981, rFC 793: Transmission Control Protocol `http://rfc.net/rfc793.html`.

[15] K. Y. Sanbonmatsu, S. Joseph, and C.-S. Tung, "Simulating movement of tRNA into the ribosome during decoding," *PNAS*, vol. 102, no. 44, pp. 15 854–15 859, 2005. [Online]. Available: