Holiday computing using Vigyaan CD and network file system: application to molecular dynamics simulation

INSTITUTE OF CHEMICAL TECHNOLOGY PRAGUE

Vojtěch Spiwok, Petra Lipovová, Blanka Králová

Department of Biochemistry and Microbiology in Prague, Technicka 3, Prague 6, 166 28, Czech Republic
Tel: +420 224 353 028, Fax: +420 224 355 167, e-mail: spiwokv@vscht.cz

Abstract: Harnessing idle computers using grid technology became an attractive source of computer time in academic research as well as in pharmaceutical industry. We present a simple alternative to grid computing using a bootable Linux CD and network file system on LAN of the institute. Computers (nodes) of colleagues on holiday (not used for a week or more) were rebooted with Vigyaan CD (bootable, Knoppix-based biomolecular modelling workbench) in a CD drive. Then a file system of the master (a PC with stable Linux installation) was mounted on each node using network file system. Using this we can run computationally expensive applications on number of nodes while all input/output data are stored on master's hard disk. We present molecular dynamics simulation using this workbench in Gromacs molecular dynamics package. We discuss advantages (easy to setup, fully reversible solution, no installations required) and disadvantages (temporary, limited hardware support) of this solution.

Introduction

Success of grid computing in the field of protein folding [1] and protein-ligand docking [2] showed a high impact of exploitation of idle computer time for academic research as well as for drug development in pharmaceutical industry [3-4]. Several non-commercial [5] and commercial [6-8] as well as world-wide [9] grid environments are currently available. Despite a number of undeniable advantages there are also disadvantage of this solution e.g. the fact that grid software must be installed on each node, this software must be adapted for each platform etc.

Vigiaan CD [10] is a Linux distribution bootable from a CD based on Knoppix [11], which contains a number of bioinformatics and molecular modelling programs. Similarly to other bootable CD Linux distributions a complete operation system and software is installed on this CD and therefore no software installation on hard disk is required. This CD can be obtained in as an ISO image free of charge from its website [9]. The program Gromacs [12] - an open-source package for molecular dynamics simulations of biomacromolecules is also installed on the Vigiaan CD. Network file system (NFS) is a well established file system for sharing of disks in a network in Unix environment.

We decided to build a holiday alternative to a grid using Vigyaan CD and NFS. Idle computers (nodes) were kindly provided by our colleagues on holidays (see acknowledgments). Computers were booted with a Vigyaan CD. No movements of computers or network reorganisations were necessary. Secure shell (SSH) server was started on each node as soon as a Vigyaan environment was loaded. This enabled secure access to nodes from other computers. Than the filesystem of master (PC with a stable Linux installation) was mounted on each node using NFS. Gromacs calculations were then carried out on nodes while input/output data were loaded/stored on hard disk of master via NFS.

Advantages:
- easy, fast and flexible - each node can be added in five minutes
- full load - 100 % of CPU used, no interruptions
- reversible - data on a hard disk of node are not modified, hard disks of nodes are not used, low risk of data damage
- no installations necessary - all programs are installed on Vigyaan CD
- reproducibility of modelling data - same software is used as copies of Vigyaan CDs are identical
- good solution for high CPU but low RAM requirements - usually the case of molecular dynamics simulation
- provides a benchmarking of nodes (see Fig. 2)

Disadvantages:
- temporary solution - a limited time of usage
- no or limited access to data - data on hard drives of nodes can not be accessed by their users (or access is limited or complicated)
- hardware compatibility - only relatively new computers can be used, hardware must be compatible with Knoppix
- a single CD per node is required - number of CDs has to be burned
- possible security problems - NFS can be replaced by more secure file systems
- the operator of master has no holiday :-(

Future perspectives:
- parallel Gromacs run - mpi version of Gromacs announced in future version of Vigyaan CD
- secure version of network file system
- booting via LAN, Mosix, OpenMosix etc.
- other applications (protein-ligand docking, quantum chemistry etc.)

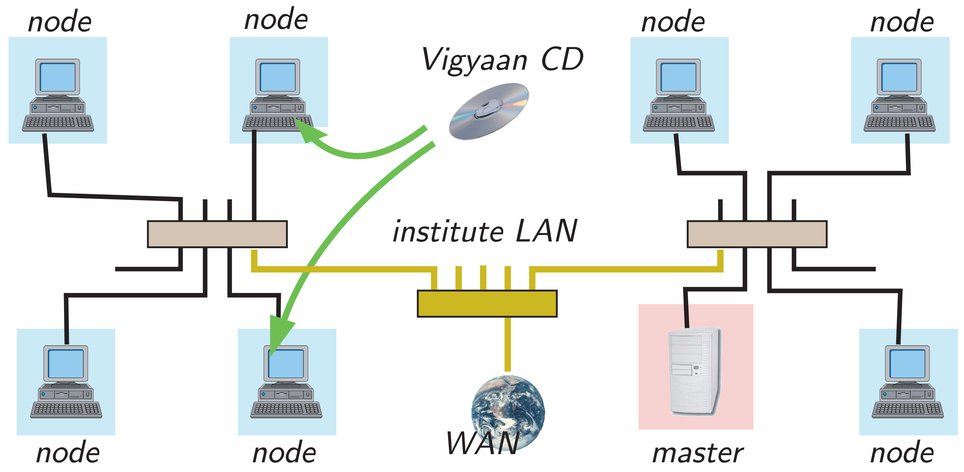

Fig1. Schematic view of setup of holiday computing.

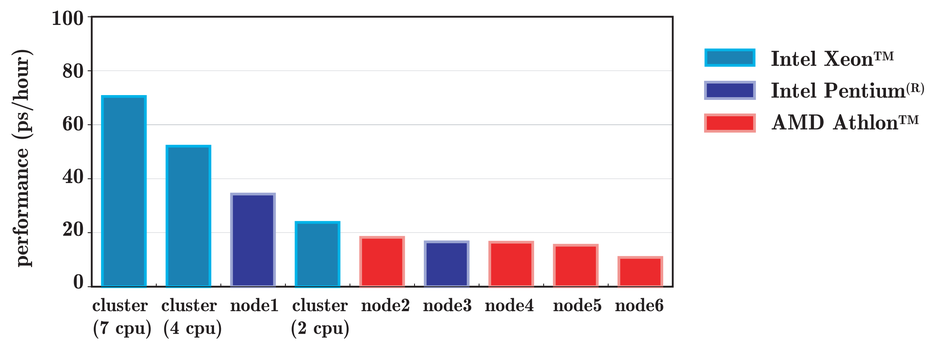

Fig2. Approximate benchmark results for computers involved in the study.

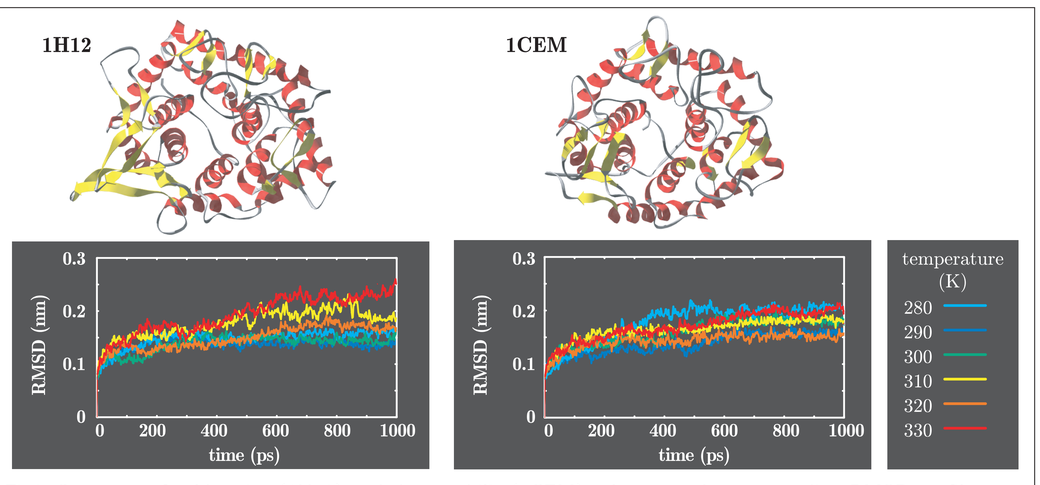

Fig3. Structure of cold-active (1H12) and thermophilic (1CEM) xylanase and coreesponding RMSD profiles.

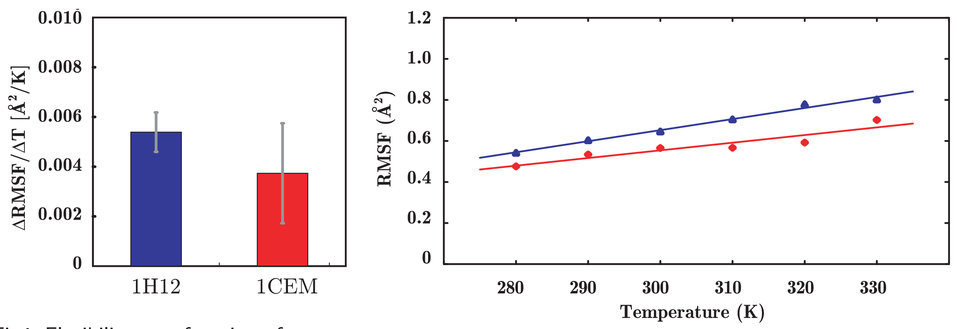

Fig4. Flexibility as a function of temperature.

Case study:

Program Gromacs was used for molecular dynamics simulation of two enzymes xylanases (EC. 3.2.1.8) - one from cold adapted Antarctic bacteria *Pseudoalteromonas haloplanctis* (PDB ID: 1H12) and one from bacteria *Clostridium thermocellum* adapted to high temperature (PDB ID: 1CEM). The aim of this study was elucidation of process of evolutionary adaptation of enzyme to function at low/high temperature. One nanosecond trajectory was calculated for each enzyme at six different temperatures (280 - 330 K). Trajectories monitored as RMSD (root-mean-square-deviation from the initial structure, Fig 3) and profile of flexibility (root-mean-square-fluctuation, RMSF, Fig. 4) as a function of temperature are illustrated. We pressent corresponding approximate benchmark data (Fig. 2) of PCs of contributors as well as a comparison with a Linux cluster.

You need:
- one computer (master) with Linux connected to institute LAN
- some space on hard drive on the master
- some Vigyaan CDs (v0.1)
- colleagues willing to share with you their computers (nodes) connected to institute LAN

Tutorial:

master.yourinstitute.cz - master
node1.yourinstitute.cz - one of nodes

Burn some Vigyaan CDs.

NODE
Put a Vigyaan CD into the CD drive of the node and restart. Make sure that the CD drive is the first in the booting sequence.
As soon as Vigyaan environment is loaded start SSH server in Knoppix menu. Set a password.

MASTER
login as root
make a directory e.g. "holiday"
```
[root@master vojta]$ mkdir holiday
```
store all input data into the directory holiday
start NFS:
```
[root@master vojta]$ /etc/rc.d/init.d/nfs start
```
information that NSF started should be printed
export the filesystem:
```
[root@master vojta]$ exportfs node1:/home/vojta/holiday -o rw
[root@master vojta]$ showmount -e
```
/home/vojta/holiday node1
exported directories with corresponding nodes should be listed

NODE
login to node as root (directly or using ssh)
make a directory "holiday":
```
[root@node1 knoppix]$ mkdir holiday
```
start portmapper
```
[root@node1 knoppix]$ /etc/init.d/portmap start
```
mount filesystem from the master
```
[root@node1 knoppix]$ mount master:/home/vojta/holiday holiday -o rw
[root@node1 knoppix]$ cd holiday
[root@node1 holiday]$
```
re-login as "knoppix" and perform all calculations

when finished relogin as root
```
[root@node1 holiday]$ cd ..
```
unmount filesystem
```
[root@node1 knoppix]$ umount holiday
[root@node1 knoppix]$ exit
```

MASTER
login as a root
stop export of filesystem:
```
[root@master vojta]$ exportfs -r
```
stop NFS
```
[root@master vojta]$ /etc/rc.d/init.d/nfs stop
[root@master vojta]$ exit
```

Troubleshooting:

Problem	**Solution**
booting:	
unsupported hardware	?
USB CD drive	look for some bootable floppy solution
running:	
file ownership confusion	use chown (as root)
problems with access permission to files and directories	use chmod (as root)
hostname confusion	contact your LAN administrator
problems with Gromacs	check Gromacs web [12], manual or mailing list

References:
1. M Shirts & VS Pande: COMPUTING: Screen Savers of the World Unite! *Science* (2000) **290**: 1903-1904.
2. WG Richards: Virtual screening using grid computing: the screensaver project. *Nat Rev Drug Discov* (2002) **1**: 551 -555.
3. A Chien, I Foster & D Goddette: Grid technologies empowering drug discovery. *Drug Discov Today* (2002) **7**(20 Suppl): S176-180.
4. Neue Zurcher Zeitung, 7. November 2003
5. www.cs.wisc.edu/condor/ or T Tannenbaum, D Wright, K Miller & M Livny: *Condor - A Distributed Job Scheduler* in Thomas Sterling, editor, *Beowulf Cluster Computing with Linux,* The MIT Press, 2002. ISBN: 0-262-69274-0.
6. www.entropia.com
7. www.turboworx.com
8. www.ud.com
9. www.globus.org
10. www.vigyaancd.org
11. www.knoppix.net
12. www.gromacs.org or HJC Berendsen, D van der Spoel, & R van Drunen: GROMACS: A message-passing parallel molecular dynamics implementation. *Comp Phys Comm* (1995) **91**: 43-56. E Lindahl, B Hess, & D van der Spoel: GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J Mol Mod* (2001) **7**: 306-317.

Acknowledgments:

Authors would like to gratefully acknowledge following contributors who allowed us to use their computers or for other help: Prof. Kateřina Demnerová, Pavel Jenč, Michaela Marková, Doc. Jarmila Pazlarová, Hynek Strnad, Prof. Olga Valentová (in alphabetical order). Comparison with Linux cluster was possible due to kind help of Prof. Leonardo Scapozza and his group on Department of Pharmaceutical Sciences of Swiss Federal Institute of Technology in Zurich, Switzerland. Work was supported by the Czech Science Foundation (GACR 204/02/0843).