Int. Computer Symposium, Dec. 15-17, 2004, Taipei, Taiwan.

On Utilization of the Grid Computing Technology for Video Conversion

Chao-Tung Yang[1] Chuan-Lin Lai[1] Chin-Ming Chi[1] Kuan-Ching Li[2]

[1] *High-Performance Computing Laboratory*
Department of Computer Science and Information Engineering
Tunghai University
Taichung, 407 Taiwan, ROC
ctyang@mail.thu.edu.tw

[2] *Parallel and Distributed Processing Center*
Department of Computer Science and Information Management
Providence University
Shalu, Taichung, 433 Taiwan, ROC
kuancli@pu.edu.tw

Abstract- *In this paper, we investigate and perform the recent popular computing technique named Grid Computing executing the video conversion application. We have developed a resource broker called Phantom that runs on grid computing testbed. Phantom resource broker's main job is to query each computing node of a grid environment about its system information. Later, according to the computing requirement that was had been provided to the user initially, the goal is to find one or more computing sites that best fit to user's requirements and then re-assign the job, executing this given job in shortest time.*

Keywords: Grid Computing, Resource Broker, Video conversion.

1. Introduction

The video conversion application has been performed to show the potential practical use of the grid technology, as well as to give us an idea what we may face when run data intensive applications on grid platforms. Past researches of video conversion testbed had been constructed based on local PCs, clusters or super computers, so the problem faced was the computing power and storage space that was not insufficient. In grid computing environment, each computing node behaves as a PC, a SMP, a cluster, a super computer system that shares its resources, which can be listed as: CPU clock, memory, storage space, hardware device and others. When a video file conversion is processed, we just submit the conversion job to the grid system. Then, the resource broker reassigns the job to the most suitable computing site. In shortest time, the video conversion job that user had submitted is concluded. Just imagine the following situation: It is possible to store the content of complete DVDs on a CD-ROM without any noticeable loss of quality. This makes buying an expensive DVD burner with limited record capacity obsolete.

To copy a video with up to 9 GB from a DVD to a CD-ROM requires large amount of computing power and time. All the data volume must be reduced to about a 12th of its original size to accommodate the 700 MB of limited storage capacity of the CD-ROM. A data compression of this magnitude for digital video is only possible with the new video compression standard MPEG-4. Generally speaking, MPEG-4 is an extension of the MPEG-2 technology, but MPEG-4 can be used more universally, with additional extensions.

Generally, when we want to conversion a DVD title to a MPEG-4 format on a single PC, we should perform the following steps as shown in Figure 1.

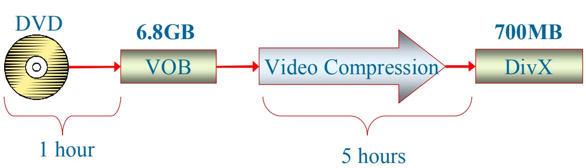

Figure 1. Video conversion – single stream

It is hard to reduce the transfer time from the DVD to a storage device, unless to upgrade the transfer bus to the SCSI or use a RAID storage system. The Video Conversion time is the key to reduce the total time. The sequential processing machine need a conversion time of 5 hours. If we divide the video file then submit the sub-file to the different conversion computing nodes, it's possible to reduce the video conversion time. It is performed the broker in the grid computing to find an available resource then to complete the video conversion job. After each node complete his conversion job, then send the result to the master node of the grid environment and then combine it. It can save quite a large amount of time. The references listed show this need and technique usage [2, 4, 5, 6, 7, 14, 15, 16, 17, 18, 19].

2. Background Review

2.1. Grid Computing

Grid computing enables the virtualization of distributed computing and data resources such as processing, network bandwidth and storage capacity to create a single system image, granting users and applications seamless access to vast IT capabilities. Just as an Internet user views a unified instance of content via the Web, a grid user essentially sees a single, large virtual computer.

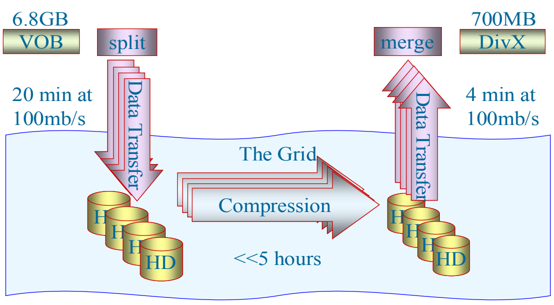

Figure 2. Video conversion – Grid Computing

At its core, grid computing is based on an open set of standards and protocols - e.g., Open Grid Services Architecture (OGSA) - that enable communication across heterogeneous, geographically dispersed environments. With grid computing, organizations can optimize computing and data resources, pool them for large capacity workloads, share them across networks and enable collaboration [2, 4, 5, 6, 7].

2.2. Globus Toolkit

The Globus Toolkit is an open architecture, open source software toolkit. It facilitates the creation of computational Grids. It provides software tools enabling coupling of people, computers, databases and instruments. With Globus, you can run job on two or more high-performance parallel machines at the same time, even the machines might be located far apart and owned by different organizations. Globus software helps scientists deal with very large datasets and complex remote collaborations. Globus software is used for large distributed computational jobs, remote instrumentation and remote data transfer.

Globus Toolkit, a set of services and software libraries to support Grids and Grid applications. The Toolkit includes software for security, information infrastructure, resource management, data management, communication, fault detection, and portability [1, 2, 4, 5, 9, 10, 11, 12, 13].

2.3. Video Format

	MPEG-1	MPEG-2	MPEG-4
Standard available since	1992	1995	1999
Max. video resolution	352 x 288	1920 x 1152	720 x 576
Default video resolution (NTSC)	352 x 288	640 x 480	640 x 480
Max. audio frequency range	48 kHz	96 kHz	96 kHz
Max. number of audio channels	2	8	8
Max. data rate	3 Mbit/sec	80 Mbit/sec	5 to 10 Mbit/sec.
Regular data rate used	1380 Kbit/s (352 x 288)	6500 Kbit/s (720 x 576)	880 Kbit/s (720 x 576)
Frames per sec. (NTSC)	30	30	30
Video quality	Satisfactory	Very good	Good to very good
Hardware requirements for encoding	Low	High	Very high
Hardware requirements for decoding	Very low	Medium	High

In general, if we want to backup a DVD title, we have four choices: VCD, MPEG4 (Divx), AVI and Video Cassette. Unfortunately, in these choices, AVI and Video Cassette are impossible in its usage. A raw AVI file is too big to store it. Raw AVI file per minutes needs about 207 Mbytes storage space to store it. In general, A DVD title has 135min. of playing time. If we want to backup it, we need about 27GBytes storage space. So, there is no reason to use raw AVI with larger size than DVD file to backup it. Video Cassette also has some drawbacks. First, if we want to backup a 135min. DVD title, then we must take 135min. (1:1 time) to back it up. It is too waste time for us to backup a DVD title. Second, When we success to backup a DVD title using video cassette, A video cassette depth of size about 2cm, it is thicker than a DVD with 0,1cm, so we need to find enough space to keep this cassette, if we want to backup more DVD title.

The history of MPEG goes back to the year 1987. MPEG stands for Motion Pictures Expert Group, a worldwide organization that develops manufacturer and platform independent standards for video compression. The first result was introduced as MPEG-1 in 1992. It was the basis for the less successful European Video-CD. Because of its limited resolution of 352 x 288 pixels, MPEG-1 is only suitable for the home environment, and the achievable video quality in relation to the data rate is rather low from today's point of view. MPEG-2 was introduced in 1995 and is mainly based on MPEG-1. The higher resolution with a maximum of 720 x 576 pixels is a major improvement enabling a significantly better video quality. The MPEG-4 was released by the MPEG group in December 1999.

MPEG-1 has small size of file but also has low quality, MPEG-2 has excellent video quality but with a large size of file. MPEG-4 have theirs advantage with "very good quality and small file size" [14].

A 130min movie	Source file size	Destination file size
DVD(MPEG-2)	4.3GB (VOB)	1.2GB (Divx)
VCD(MPEG-1)	1.2GB (DAT)	500MB (Divx)

Figure 3. DVD to Divx and VCD to Divx

3. Experiment Environment

3.1. Experimental Grid Environment

We built a grid computing testbed that includes four Linux PC clusters:

1. Site 1: 4 PCs with single Intel Celeron 1700 MHz processor, 256MB DDRAM, and 3Com 3c9051 interface,
2. Site 2: 4 PCs with Dual Intel Pentium3 866 MHz processor, 256MB SDRAM, and 3Com 3c9051 interface,
3. Site 3: 4 PCs with single Intel Pentium4 2.53GHz processor, 512MB DDRAM and Intel PRO100 VE interface,
4. Site 4: 4 PCs with single Intel Pentium4 2.4GHz processor, 256MB DDRAM and Accton EN-1216 interface.

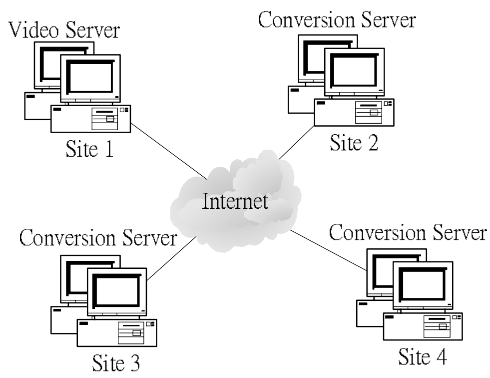

Figure 4. Our Grid computing testbed

Sites 1 to 3, it locates at different department and lab in Tunghai University, Taiwan. Site4 locates at NCHC (National Center for High-Performance Computing), Taiwan. We use general application to benchmark network traffic from Site 1 to Site4. Between Site (1, 2, 3) and Site (1, 2, 3), the average network latency is 3ms and the maximum transfer speed is 7600KBytes. Between Site (1, 2, 3) and Site4, the average network latency 5ms and the maximum transfer speed is 2000KBytes.

The main title (on the first page) should be centered, in Times 14-point, and boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two 12-point blank lines after the title.

3.2. Software for Video Compression

This section gives an introduction of DVD to DivX compression in video conversion grid. As shown in figure 5, the first step is to split VOB files into a number of chunks according to the number of nodes that the video conversion grid system has. The sizes of divided files are based on their information available in MDS (Figure 6).

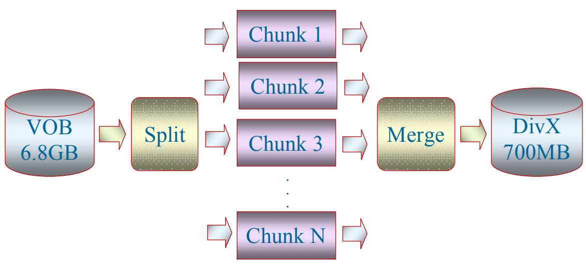

Figure 5. Split and Merge video files

By MDS, the client can gather useful information like computational capabilities, CPU loading, number of nodes, etc. Then the client can submit jobs represented by RSL with useful information through GRAM of Globus to remote servers. In this moment, numbers of divided VOB files transfer to NFS of each conversion server using GridFTP.

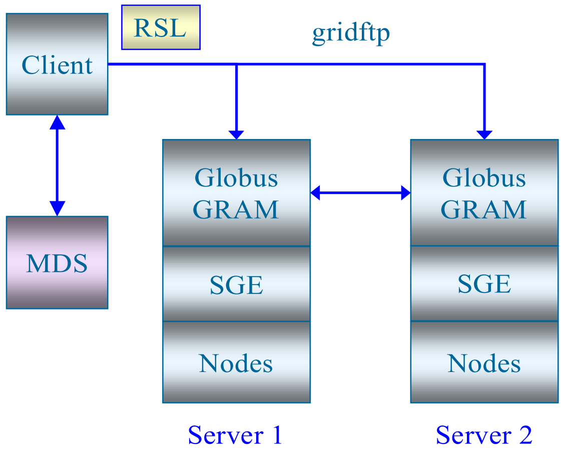

Figure 6. System Components

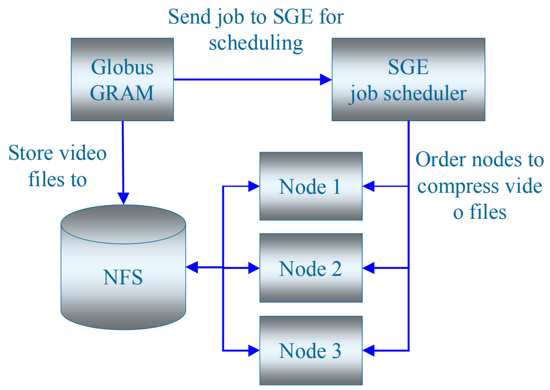

Figure 7. Component of the conversion server

After that, GRAM submits job to SGE for further scheduling on each cluster. As shown in Figure 7, SGE orders nodes belong to it to compress properly section of VOB file on its NFS. Each node submitted the job does compression and returns the DivX file after conversion to video server to be merged.

3.3. Phantom

We implement a Grid-Enabled Video converter: Phantom. Phantom's implementation is base on Java CoG and Globus API, as shown in Figure 8. Because of Java technology, Phantom can be executed on different platforms and OS and makes use of Globus services, such as resource allocation, information, data management services. Phantom gathers information such as where the idle system is and where enough storage spaces are.

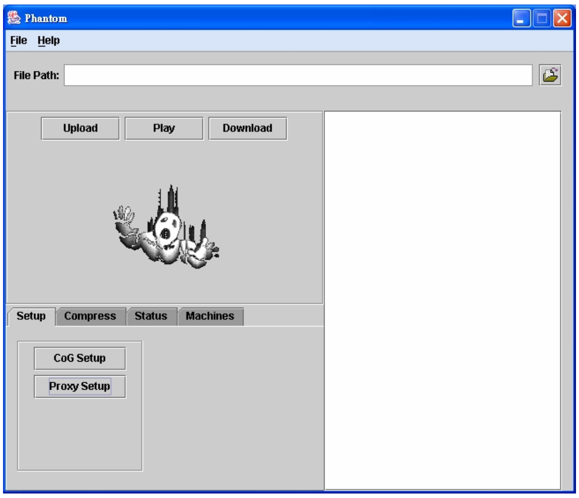

Figure 8. Phantom resource broker's main screen

When Phantom is performed, the video transfer time, video conversion time, storage system spaces will be predicted by obtaining system information using MDS. After deciding the best location for video conversion, Phantom will submits job to the master system. Figure 9 shows the input and output file format of Phantom.

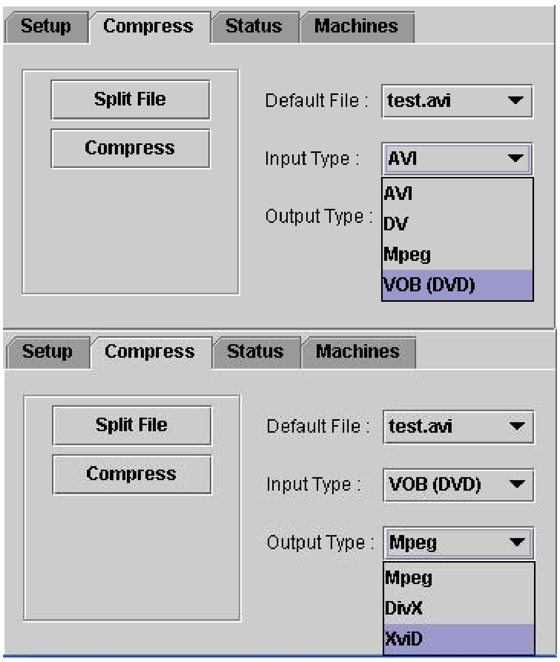

Figure 9. Phantom resource broker's input and output files format

The resource broker Phantom consists of following main components:

1. **Information Monitor:** This component monitors system factors such as video transfer time, CPU type, storage system spaces. MDS in Globus is used for gathering needed information.

2. **Location Finder:** Through the information monitor, location finder can evaluate and predict to find where the available system is.
3. **Data Transfer:** It transfers the video file to the destination node, where the system is selected by location finder. Currently, GridFTP is used for transfer video file.

Executer: When the video file transfer to the destination node, GRAM service in Globus is used for executing job on remote site.

3.4. Performance Results

The figures 10 and 11 demonstrate the performance results when using our Phantom system. From Figure 10, the conversion time is reduced form 2175.54s to 361.83s by using a grid system with 8 processors. Additionally from Figure 11, the conversion time is reduced respectively form 27.73s to 29.75s and 152.11s to 33.69s for MPEG to XviD and MPEG to DviX by using a Grid system with 8 processors.

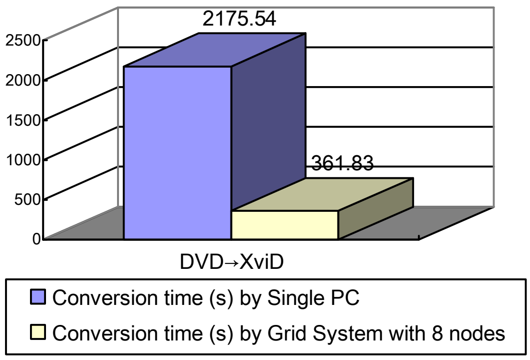

Figure 10. Comparison of conversion time from DVD to XviD of single PC and grid system

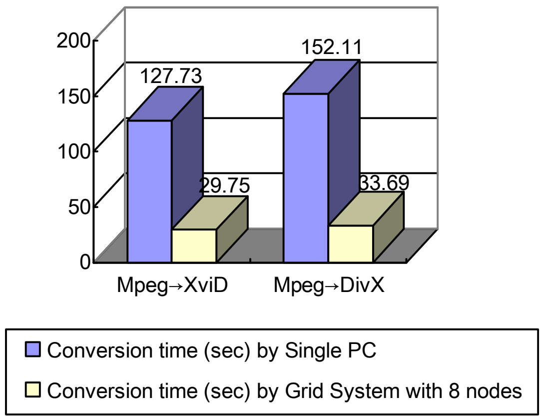

Figure 11. Comparison of conversion time from MPEG to XviD and MPEG to DivX of single PC and grid system

4. Conclusions and Future Works

This research intends to integrate the PCs, clusters and SMP machines in the campus or in the internet to form a computing farm by using grid computing technology, as to fully utilize the available or idle cycles present in these systems. We have developed the Phantom resource broker to assist us to find suitable resource on a grid system for video conversion without wasting unnecessary time. Our future work is to add the fault-detection policy, when a computing node fails. Phantom resource broker will be able to detect and re-assign the same job to other computing node that is listed by MDS service.

References

[1] Globus Toolkit, http://www.globus.org
[2] Global Grid Forum, http://www.ggf.org/
[3] Sun ONE Grid Engine, http://wwws.sun.com/software/gridware/
[4] IBM Redbooks, "Introduction to Grid Computing with Globus"
[5] IBM Redbooks, "Enabling Applications for Grid Computing with Globus"
[6] I. Foster, C. Kesselman, eds., "The Grid: Blueprint for a New Computing Infra-structure", Morgan Kaufmann; 1st edition (January 1999)
[7] I. Foster., "The Grid: A New Infrastructure for 21st Century Science", Physics Today, 55(2):42-47, 2002.
[8] Thomas Sterling, Gordon Bell, Janusz S. Kowalik, "Beowulf Cluster Comput-ing with Linux", MIT Press, Paperback, Published March 2002
[9] K. Czajkowski, I. Foster, N. Karonis, C. Kesselman, S. Martin, W. Smith, S. Tuecke., "A Resource Management Architecture for Metacomputing Systems", Proc. IPPS/SPDP '98 Workshop on Job Scheduling Strategies for Parallel Processing, pg. 62-82, 1998.
[10]K. Czajkowski, S. Fitzgerald, I. Foster, C. Kesselman., "Grid Information Ser-vices for Distributed Resource Sharing", Proceedings of the Tenth IEEE Inter-national Symposium on High-Performance Distributed Computing (HPDC-10), IEEE Press, August 2001.
[11]X. Zhang, J. Freschl, and J. Schopf., "A Performance Study of Monitoring and Information Services for Distributed Systems", Proceedings of HPDC, August 2003.
[12]W. Allcock, J. Bester, J. Bresnahan, A. Chervenak, L. Liming, S. Meder, S. Tuecke., "GridFTP Protocol Specification", GGF GridFTP Working Group Document, September 2002.

[13]B. Allcock, J. Bester, J. Bresnahan, A. L. Chervenak, I. Foster, C. Kesselman, S. Meder, V. Nefedova, D. Quesnal, S. Tuecke., “Data Management and Transfer in High Performance Computational Grid Environments”, Parallel Computing Journal, Vol. 28 (5), May 2002, pp. 749-771.
[14]MPEG-4 - Copying a DVD Video to CD-ROM, http://www.tomshardware.com/video/20000913/
[15]Tierney, B.L.; Johnston, W.E.; Herzog, H.; Hoo, G.; Guojun Jin; Lee, J.; Chen, L.T.; Rotem, D., “Using high speed networks to enable distributed parallel im-age server systems”, Supercomputing ‘94. Proceedings, 14-18 Nov. 1994.
[16]Lee, J.Y.B., “Parallel video servers: a tutorial”, IEEE Multimedia, Volume: 5 Issue: 2, April-June 1998.
[17]Yong He; Ahmad, T.; Liou, M.L., “MPEG-4 based interactive video using par-allel processing”, 1998. Proceedings. 1998 International Conference on Paral-lel Processing, 10-14 Aug. 1998.
[18]Yong He; Ahmad, T.; Liou, M.L., “Real-time interactive MPEG-4 system en-coder using a cluster of workstations”, IEEE Transactions on Multimedia, Vol-ume: 1 Issue: 2, June 1999.
[19]Gunawan, T.S.; Cai Wen Tong, “Parallel motion estimation on SMP system and cluster of SMPs”, 2002. APCCAS ‘02. 2002 Asia-Pacific Conference on Cir-cuits and Systems, Volume: 2, 28-31 Oct. 2002.