

# **UNIVERSITI PUTRA MALAYSIA**

FAST DATA TRANSFER TECHNIQUE FOR LARGE DATA IN GRID ENVIRONMENT

**CHONG MIEN MAY** 

**FSKTM 2015 3** 



# FAST DATA TRANSFER TECHNIQUE FOR LARGE DATA IN GRID ENVIRONMENT



By

**CHONG MIEN MAY** 

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

April 2015

# COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia



Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the Degree of Master of Science

# FAST DATA TRANSFER TECHNIQUE FOR LARGE DATA IN GRID ENVIRONMENT

By

# CHONG MIEN MAY

April 2015

# Chairman: Rohaya Latip, PhDFaculty: Computer Science and Information Technology

Nowadays, multimedia files such as videos and images are being improved in terms of their quality, where the dimension of the video or image is improving from two dimensional (2D) to become three or four dimensional (3D or 4D). These quality improvements have also increased the size of the multimedia data. Thus, the large size of video data has used a lot of time to transmit through the network. To solve the problem, in our research, we proposed a new video splitting technique, which is the "Exponential-and-Uniform-based (ExpoNUni) Splitting Technique" to improve the transmission time and maintain the quality of the data.

Our "ExpoNUni" Splitting Technique has performed a better result compared to the existing techniques such as Fibonacci-based Splitting with V1min Technique and Uniform-based Splitting Technique. For the Initial Delay, overall, our splitting technique by average has 26.23 % lesser initial delay time for comparing with the existing Fibonacci-based Splitting with V1min Technique. It shows an 11.62% average of initial delay time lesser comparing to Uniform-based Splitting Technique.For download time, the overall result that compared with the existing Fibonacci-based Splitting technique by average 1.43% lesser time for downloading a complete video file to the user's PC from the grid storage, while the Uniform-based Splitting Technique, it shows 1.67% decrement of average downloading time.

To prove the Exponential-and-Uniform-based splitter component is able to reduce the transmission time in a Grid-based Video Storage Framework, we also compare the uploading time and downloading time between the framework without the splitter component and the framework with our splitter component. The results show that the framework with our splitter component proofs it takes lesser data transmission time, where overall with an average of 9.86% lesser time for the uploading time and overall with an average of 12.24% lesser time for the downloading time. Besides, the

MATLAB tool proves that our splitting technique did not affect the quality of the retrieved data



Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Sarjana Sains

# TEKNIK PEMINDAHAN DATA YANG CEPAT UNTUK DATA YANG BESAR DALAM ALAM PERSEKITARAN GRID

Oleh

# CHONG MIEN MAY

April 2015

# Pengerusi : Rohaya Latip, PhD Fakulti : Sains Komputer dan Teknologi Maklumat

Pada masa kini, fail multimedia seperti video dan imej sedang meningkat tahap kualiti mereka, di mana dimensi video atau imej telah bertukar daripada dua dimensi (2D) kepada tiga atau empat dimensi (3D atau 4D). Peningkatan kualiti ini juga telah meningkat saiz data multimedia. Oleh itu, video yang bersize besar telah menggunakan banyak masa semasa dihantar melalui rangkaian internet. Untuk menyelesaikan masalah ini, dalam kajian kami, kami mencadangkan teknik pemisahan video yang baru bernama "Eksponen-dan-seragam berasaskan (ExpoNUni) Pemisahan Teknik". Teknik ini digunakan untuk meningkatkan masa penghantaran data dalam rangkaian internet dan juga digunakan untuk mengekalkan kualiti data.

Teknik Pemisahan "ExpoNUni" kami telah memberikan keputusan yang lebih baik berbanding dengan Teknik Pemisahan berasaskan Fibonacci V1min dan Teknik Pemisahan berasaskan Uniform. Secara keseluruhannya, Teknik Pemishan "ExpoNUni" kami telah berjaya menurunkan purata 26.23% masa kelewatan berbandingan Teknik Pemisahan berasaskan Fibonacci V1min dan menurunkan purata 11.62% masa kelewatan berbandingan Teknik Pemisahan berasaskan Fibonacci V1min, Bagi masa muat turun, berbandingan Teknik Pemisahan berasaskan Fibonacci V1min, teknik pemisahan kami telah berjaya menurunkan purata 1.43% masa muat turun, manakala berbandingan Teknik Pemisahan berasaskan Uniform, teknik kami berjaya menurunkan purata 1.67% masa muat turun.

G

Untuk membuktikan komponen splitter berdasarkan "ExpoNUni" mampu mengurangkan masa penghantaran data dalam Rangka Kerja Penyimpanan Video berasaskan alam sekitar Grid, kami juga membandingkan masa muat naik dan masa muat turun di antara rangka kerja tanpa komponen splitter dan rangka kerja dengan komponen splitter kami. Keputusan menunjukkan bahawa rangka kerja yang membawa komponen splitter kami telah mencapaikan purata kekurangan 9.86% untuk masa muat naik dan purata kekurangan 12.24% untuk masa memuat turun. Selain itu, alat

MATLAB juga membuktikan bahawa teknik pemisahan kami tidak memberi sebarang kesan negative kepada video data.



# ACKNOWLEDGEMENTS

First of all, I wish to express my sincere gratitude to my supervisor, Dr. Rohaya Latip, for her guidance, encouragement, dedication, support and invaluable advice throughout these semesters. I would like to thank also my Co-supervisor, Prof. Dr. Hamidah Ibrahim for her fruitful discussion and valuable suggestions. Besides, I would like to thank the InfoComm. Development Center (IDEC UPM), which have provided the grid infrastructure for me to do the research. I also would like to take this opportunity to thank the Hospital UiTM Sungai Buloh National and University of Malaysia Medical Centre (HUKM) for giving me a chance to cooperate with them in improving the performances of their medical data storage.

And, for my final words, I would to deeply thank my lovely parents, sister and brother for everything that they've done for me. Thank you for their encouragement, fully support and love throughout my study. Lastly, I acknowledge that my achievement is a result of cooperation and help of my lecturers, friends, course mates and those whom I have not mentioned above.

THANK YOU TO ALL ~

Chong Mien May 21<sup>ST</sup> JULY 2015

I certify that a Thesis Examination Committee has met on 15 April 2015 to conduct the final examination of Chong Mien May on her thesis entitled "Fast Data Transfer Technique For Large Data In Grid Environment" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science.

Members of the Thesis Examination Committee were as follows:

# Shamala A/P K. Subramaniam, PhD

Professor Faculty of Computer Science and Information Technology Universiti Putra Malaysia (Chairman)

# Azizol B. Haji Abdullah, PhD

Faculty of Computer Science and Information Technology Universiti Putra Malaysia (Internal Examiner)

# Mustafa Mat Deris, PhD

Professor Software Engineering Department Faculty of Computer Science and Information Technology Universiti Teknologi Tun Hussein Onn Malaysia (External Examiner)

### ZULKARNAIN ZAINAL, PhD

Professor and Deputy Dean School of Graduate Studies Universiti Putra Malaysia

Date: 12 August 2015

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

# Rohaya Latip, PhD

Senior Lecturer Faculty of Computer Science and Information Technology Universiti Putra Malaysia (Chairman)

# Hamidah Ibrahim, PhD

Professor Faculty of Computer Science and Information Technology Universiti Putra Malaysia (Member)

### BUJANG BIN KIM HUAT, PhD Professor and Dean School of Graduate Studies

School of Graduate Studies Universiti Putra Malaysia

Date:

# **Declaration by graduate student**

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: \_\_\_\_\_ Date: \_\_\_\_

Name and Matric No.: Chong Mien May (GS28355)

# **Declaration by Members of Supervisory Committee**

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature:	
Name of	
Chairman of	
Supervisory	
Committee: Rohava Latip, PhD	

Signature: \_\_\_\_\_ Name of Member of Supervisory Committee: <u>Hamidah Ibrahim, PhD</u>

# TABLE OF CONTENTS

Page

	C
ABSTRACT	i
ABSTRAK	iii
ACKNOWLEDGEMENTS	v
APPROVAL	vi
DECLARATION	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS	xvi

# CHAPTER

1	INTROD	UCTION	
	1.1 Resear	ch Background	1
	1.2 Problem	m Statement	2
	1.3 Resear	ch Objectives	4
	1.4 Resear	ch Scope and Limitation	4
	1.5 Organi	zation of Thesis	5
2	LITERAT	CURE REVIEW	
	2.1 Introdu	iction	6
	2.2 Grid C	omputing	6
	2.3 Data T	ransmission	8
	2.4 Video	Compression Technique	10
	2.5 Existin	g Video Splitting Techniques	12
	2.5.1	Uniform-based Splitting Technique	12
	2.5.2	Fibonacci-based Splitting Technique	13
	2.5.3	Fibonacci-based Splitting with V <sup>1</sup> <sub>min</sub> Technique	14
	2.6 Curren	t Related Grid-based Data Storage Frameworks	16
	2.6.1	Grid-based E-learning System Framework	
		(Educational Video)	17
	2.6.2	MammoGrid - Distributed Mammographic Database	18
	2.6.3	Clustered-based Video Storage Framework	19
	2.7 Chapte	r Summary	21
3	METHOI	DOLOGY	
	3.1 Introdu	iction	23
	3.2 Define	Study Phase	23
	3.2.1	Input Variables	24
	3.2.2	Output Variables	24
	3.2.3	Assumptions	24
	3.3 Initializ	zation of Variable Phase	24
	3.4 Design	and Development Phase	25
	3.4.1	Design New Video Splitting Technique	25
	3.4.2	Design and Development of New framework	25
	3.5 Evalua	te and Experimental Phase	27

	3.6 Analyze and Validate Phase	28
	3.7 Chapter Summary	28
4	EXPONENTIAL-AND-UNIFORM-BASED ( <i>ExpoNUni</i> ) SPLITTIN TECHNIQUE AND GRID-BASED VIDEO STORAGE FRAMEWORK	١G
	4.1 Introduction	29
	4.2 Exponential-and-Uniform-based ( <i>ExpoNUni</i> ) Splitting Technique	29
	4.2.1 Definition of Exponential Function	29
	4.2.2 Equation	30
	4.2.3 Algorithm	31
	4.3 Development of the Framework	32
	4.3.1 Implementation on Grid Testbed	38
	4.3.2 Real Implementation in Hospital	43
	4.4 Chapter Summary	48
5	RESULTS AND DISCUSSIONS	
	5.1 Introduction	49
	5.2 Determination of the Equation Value for the Splitting Technique	49
	5.3 Initial Delay Time	50
	5.4 Video Uploading Time	52
	5.5 Video Download Time	53
	5.6 Quality of Split Video	54
	5.7 Chapter Summary	57
6	CONCLUSIONS AND FUTURE WORKS	
Ŭ	6.1 Introduction	58
	6.2 Conclusions and Contributions	58
	6.3 Future Works	59
RI	EFERENCES	60
A	PPENDIX A	66
BI	IODATA OF STUDENT	70
LI	IST OF PUBLICATIONS / AWARDS	71

G

# LIST OF TABLES

,

 $\bigcirc$ 

Fable		Page
2.1	History of Video Compression Standards	10
3.1	Research Parameters	25
5.1	Size of First Chunk for 600MB	50
5.2	Initial Delay (s) Versus Chunks	51



# LIST OF FIGURES

Figure		Page
1.1	Academic Grid Environment	4
2.1	Example of Virtual Organization Concept (Tsiknakis et al., 2007)	7
2.2	Layers in Grid Architecture	8
2.3	Taxonomy of Data Transmission.	9
2.4	Block-matching Motion Estimation Compression (Ahmed et al., 2011)	11
2.5	Motion Compensation Technique (Ahmed et al., 2011)	12
2.6	Diagram of Uniform-based Splitting Technique	13
2.7	Diagram of Fibonacci-based Splitting Technique	14
2.8	Diagram of Fibonacci-based Splitting with V <sup>1</sup> <sub>min</sub> Technique	16
2.9	E-Learning System Framework (Mitra et al., 2015)	17
2.10	Multi-layered Framework of E-Learning System. (Roy et al., 2014)	17
2.11	Figure 2.11. Layered architecture of E-Learning Grid. (Roy et al., 2014)	18
2.12	Schematic of MammoGrid Project. (Warren et al., 2007)	19
2.13	Video Splitting Scheme in Clustered-based Video Storage Framework (Liao & Jin, 2005)	20
2.14	Architecture of one Cluster Video Server (Liao & Jin, 2005)	20
2.15	Architecture of one Cluster Video Server Storage System. (Liao & Jin, 2005)	21
3.1	Phases of the Research	23
3.2	Design of the Grid-based Video Storage Framework	26

6

4.1	Diagram of Exponential-And-Uniform-based Splitting Technique	30
4.2	Algorithm of Exponential-and-Uniform-based Splitting Technique	31
4.3	Infrastructure of Grid-based Video Storage Framework	33
4.4	Process Flow of the Framework	33
4.5	Flow of Grid-based Video Storage Framework	34
4.6	The flowchart of upload process	35
4.7	The flowchart of splitting process	36
4.8	The flowchart of the download process	37
4.9	Invoke the verification query inside the Academic Grid VOMS	38
4.10	Interface example for the Upload Page of the framework	39
4.11	Invoke the Storing Process	39
4.12	Query for creating the folders.	40
4.13	Query for storing the video chunks	40
4.14	Interface example for the Download Page of the framework	40
4.15	File List's Query	41
4.16	Query for downloading the chunks.	41
4.17	PHP code for downloading the chunks	42
4.18	Retrieval and Storing Echocardiography Data Process in Hospital UKM or Hospital UiTM Cardiology Department	43
4.19	Current PAC System in Hospital UiTM (Malaysia) Cardiology Department	44
4.20	Integration of the Framework in Hospital UiTM (Malaysia).	45
4.21	Diagram of the Process for the Web Portal	46

4.22	Interface for Searching Patient's Data	47
4.23	Interface for the Selected Patient's Data List	47
4.24	Interface for viewing the Echocardiography without downloading	48
5.1	Analysis of the Exponential Series	49
5.2	Analysis of graph to determine the Exponential Series in the equation of Exponential-and-Uniform- based Splitting Technique	50
5.3	The Initial Delay Time with Different Number of Video Chunks	51
5.4	Upload Time with Different Size of Video Data	52
5.5	The Total Download Time with different number of chunks	53
5.6	The Total Download Time with Different Video Size	54
5.7	Sample 1	55
5.8	Sample 2	56

C

# LIST OF ABBREVIATIONS

AVC	Advanced Video Coding
BIRN	Biomedical Informatics Research Network
CDs	Compact Disks
СТ	Computed Tomography
ExpoNUni	Exponential-and-Uniform-based
HPC	High Performance Computing
HUKM	Hospital Universiti Kebangsaan Malaysia
IS	Information System
MAD	Mean Absolute Differences
MAP	Maximum A Posteriori
MBs	Macro-Blocks
MediGrid	Medicine Grid
MPEG	Motion Picture Experts Group
MRF	Markov Random Field
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
MV	Motion Vector
NM	Nuclear Medicine
PACS	Picture Archiving Communication System

C

PC	Personal Computer
PHP	Hypertext Preprocessor
RAM	Random Access Memory
SAD	Sum of Absolute Differences
SE	Storage Element
UiTM	Universiti Teknologi MARA
UPM	Universiti Putra Malaysia
US	UltraSound
VO	Virtual Organization
VOMS	Virtual Organization Management System
VOs	Virtual Organizations
WMS	Workload Management System



## **CHAPTER 1**

### INTRODUCTION

### 1.1 Research Background

Grid computing is the service-oriented distributed systems that generally are with the high performance computational computers. It normally composed of heterogeneous resources (Hussin & Latip, 2013). Today, it has become a new technology that able to bring the high performance distributed infrastructure for the community (Bruneo et al., 2009). Grid computing is giving a best way for the community to solve a number problems.

Thus, nowadays, there are a lot of applications integrated with the grid computing technology. These applications includes drug diagnostics applications (Lee & Park, 2014; Redolfi et al., 2009), protein folding applications (Desell et al., 2010), financial modeling applications (DAWSON et al., 2014; Jacob et al., 2005), earthquake simulation applications (Cinquini et al., 2014; Cossu et al., 2010; Dabas & Arya, 2013; Fern ández-Quiruelas et al., 2011; Renard et al., 2009), climate or weather modeling applications (De & Nationale, 2014; Sper de Almeida, 2012) and other applications. In medical area, there are a lot of the projects running under the grid environment, such as BIRN Projects (Astakhov et al., 2005; Glover et al., 2012; Grethe et al., 2005; Keator et al., 2008), MammoGrid Projects (S. R. Amendolia et al., 2004; S. Amendolia et al., 2004; Warren et al., 2007), MediGrid Projects (Bertero et al., 2003; Boccia & Guarracino, 2005; Krefting et al., 2009; Weisbecker, Falkner, & Rienhoff, 2009) and other projects (Kumar et al., 2008).

Since there are a number of applications using the grid computing to execute their works, there is a need to study on the issue of the grid service's quality. As what's presented in the paper of Bruneo et al., 2009, to provide a better service in the grid environment, the requirements for the Quality-of-Service (QoS) must be focused and studied, such as the delay perceived by the user, the time for uploading or receiving a data from the grid storage. Thus, it is important to study on the issue of data transmission.

Text, audio, video, and image graphics are the data that normally transmit in the network. Among these data, video has become one of the important mediums for the area of communications and entertainment (Thampi, 2013). Like what have been presented in the paper of Toharia et al., 2008, many fields such as the news broadcasting (news, shows, series, etc.), advertising, and medical applications, are using the large amount of video. Thus, to enhance the video clarity and quality, nowadays, the video has been improved from the two dimensions (2D) video images into three dimensional (3D) or four dimensional (4D) video images.

However, the improvement of the video dimension also increased the size of the video. Thus, video has become the data that takes up a lot of space and time during the transmission and storage session. To reduce the data transmission time for a large multimedia data, such as video, normally there are using two ways to reduce the size of the video before transmit through the network, such as the video compression technique or video splitting technique.

There researchers that worked on the video compression techniques are included Ahmed et al., 2011, Anand Deshpande & Rao, 2014, Ismail et al., 2012, D. Kim et al., 2006; N.-J. Kim et al., 2009, Okade & Biswas, 2013; Shen, 2005; Shenolikar & Narote, 2009; Xiang et al., 2013. Although video compression is able to reduce the size of the video, however, this technique still facing some limitations, such as the video quality problem (noise problem) and also the algorithm computation time problem.

In the papers of Bruneo et al., 2009; Jin et al., 2014; Nithya & Shanmugam, 2010; Shenoy & Vin, 1997; Sidhu & Singh, 2010, they are focusing on the video splitting technique. They introduced some splitting techniques, such as the Uniform-based Splitting Technique, Fibonacci-based Splitting Technique, and Fibonacci-based Splitting Technique, Fibonacci-based Splitting Technique, By using the Uniform-based Splitting Technique, Fibonacci-based Splitting Technique, and Fibonacci-based Splitting technique, they were found that the transmission time for large video is able to reduce.

By studying the video splitting techniques, we found that unlike the video compression techniques, video splitting techniques not only reduce transmission time, they also maintain the quality of the video. In short, after studying some previous papers, we decided to introduce a new video splitting technique that is capable of reducing video data transmission time and maintaining the quality of the video data. Besides, to test the video splitting technique, the Video Storage Framework was also introduced under the Academic Grid Environment. The proposed splitting technique was applied in the framework.

# **1.2 Problem Statement**

In a network, transmitting large data on a limited bandwidth is very critical. There are two common techniques used to reduce the video data transmission time, which are the video compression techniques and video splitting techniques. Video compression techniques can reduce the size of the video, however, this techniques also reducing the quality of the video, such as removing the pixels or the detail of the video images. Thus, video compression technique is not the best solution to reduce the size for some critical and important video such as the medical video data because medical data must be accurate for diagnosing the patient's illness.

To prevent the pixels removed from the video images, we choose to study on the video splitting technique. However, from the previous Uniform-based Splitting Technique, and Fibonacci-based Splitting with  $V^lmin$  Technique, we found that the techniques

have some limitations. In Uniform-based Splitting Technique (Bruneo et al., 2009; Jin et al., 2014; Shenoy & Vin, 1997; Sidhu & Singh, 2010), the delay time is high; while for the Fibonacci-based Splitting with  $V^lmin$  Technique (Bruneo et al., 2009), it used a lot of time on the complex mathematical algorithm where searching the first chunk size is needed before starting the splitting process. Thus, Fibonacci-based Splitting with  $V^lmin$  Technique also increasing the data transmission time. To reduce the data transmission time and maintain the quality of the video images, a new splitting technique should be proposed.

In paper of Toharia et al., 2008; Toharia et al., 2010, presented that many fields such as the news broadcasting (news, shows, series, etc.), advertising, medical applications, educational field and others fields, all requires large video storage to store a large amount of video data. However, due to the size of video, the storage space on the server for storing the large video data has become an issue.

To solve the storage problem, there are some Grid-based Videos and Images Storage Frameworks developed, such as the Distributed Mammographic Database Framework in the MammoGrid Project (Warren et al., 2007), the Grid-based E-learning System Framework for the educational video purpose (Mitra, Das, & Roy, 2015), and others framework. By using the grid computing storage, the space storage can be more manageable and could overcome the limited storage of the server. Besides, the video images data is more reliable and available due to the replication function inside the grid storage. However, there is a limitation for these frameworks, where their framework do not have a splitter function at the server that can reducing the data transmission time in the grid environment.

In paper of Jin et al., 2014 and Liao & Jin, 2005, they worked on the clustered video servers. They using the video splitting and distributed placement scheme for their clustered video servers, where the scheme is dividing the movie into several parts with the same time length of same file size (Uniform-based Splitting Technique) based on the requirements of the clients. However, in first problem statement, we found that the Uniform-based Splitting Technique is having a high delay time. Thus, the delay time problem is becoming a limitation for this framework. Besides, the other limitation is the framework is working on the cluster based technology and not the grid computing technology. Since grid computing sharing the computer power and data storage capacity over the Internet, thus, the grid computing can provide more storage than the cluster computing.

Based on the limitations found in the previous studies, in this research, a new Gridbased Video Storage Framework with the Exponential-and-Uniform-based Video Splitter Server was developed.

3

# 1.3 Research Objectives

First, the main objective of this research is to propose a new video splitting technique for reducing the data transmission time in grid environment. The new video splitting technique must be able to reduce the initial delay time and total download time for the Video Download Service in the grid environment. Besides, it should maintain the quality of the video after the splitting and reconstruction process have been done.

The secondary objective of this research is to develop a framework for managing the large video data inside the grid environment. The new technique, named as Exponential-and-Uniform based Splitting Technique will be embedded in the framework. This framework runs on the Academic Grid testbed to store all the large video data.

# 1.4 Research Scope and Limitation

The scope of this research is to propose a new video splitting technique, which can reduce the initial delay time and the download time for a video download from the grid storage. Besides, the Grid-based Video Storage Framework is built and run on the Academic Grid testbed. Academic Grid is known as the A-Grid, which is a learning and discovery grid. Figure 1.1 shows the diagram of the Academic Grid Environment. There are more than three terabytes of storage spaces are made available to grid users to store their data in the Academic Grid Storage. The members of the Academic Grid Malaysia include the Biruni Grid of Universiti Putra Malaysia, the Crystal Grid of Universiti Malaysia, and the Grid of Universiti Teknologi Malaysia.



Figure 1.1. Academic Grid Environment

Besides, the research is focused mainly on the 600MB video file as the dataset. This is because of the previously studies (Bruneo et al., 2009; Nithya & Shanmugam, 2010), are using 600MB video data for testing. Thus, to have an accurate comparison analysis with the previous techniques, a 600MB video file is chosen as our dataset. There are five 600MB video files with the same video format but different video contents were used in the experiments.

## 1.5 Organization of Thesis

# **CHAPTER 1: INTRODUCTION**

In this chapter, a brief description of the research area is presented, followed by a presentation of the problem statement and the objectives of the research. The outline of the thesis is as follows.

### **CHAPTER 2: LITERATURE REVIEW**

In Chapter Two, an overall assessment of different theories connected to the research is presented. It included the Grid Computing's Background, and also the review of the existing video splitting techniques. Besides, this chapter also discussed on the data transmission's issues

# **CHAPTER 3: METHODOLOGY**

This chapter covers the test bed design and the planning for the implementation processes. This section explained the work and process flow of the research as well as the methods and the techniques used in the research.

# CHAPTER 4: EXPONENTIAL-AND-UNIFORM-BASED (ExpoNUni) SPLITTING TECHNIQUE AND GRID-BASED VIDEO STORAGE FRAMEWORK

This chapter covers the details of the Exponential-and-Uniform-based (*ExpoNUni*) Splitting Technique and development and design of the framework. In this section, all the technical process of the research and the framework will be explained.

# **CHAPTER 5: RESULTS AND DISCUSSIONS**

In this chapter, it covers the details of the test bed implementation, and also the experiments' results in the real test bed.

# $\bigcirc$

# **CHAPTER 6: CONCLUSIONS AND FUTURE WORKS**

A discussion and conclusion of the research are found in Chapter Six. In this chapter too, there are some future works has been discussed too.

### REFERENCES

- Ahmed, Z., Hussain, A. J., & Al-Jumeily, D. (2011). Fast Computations of Full Search Block Matching Motion Estimation (FCFS). *Proceedings of PGNeT Conference*.
- Amendolia, S., Estrella, F., Hassan, W., Hauer, T., Manset, D., McClatchey, R., Rogulin, D., Solomonides, T. (2004). MammoGrid: A Service Oriented Architecture Based Medical Grid Application. In H. Jin, Y. Pan, N. Xiao, & J. Sun (Eds.), *Grid and Cooperative Computing - GCC 2004 SE - 143* (Vol. 3251, pp. 939–942). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-30208-7\_143
- Amendolia, S. R., Estrella, F., Hauer, T., Manset, D., Mcclatchey, R., Odeh, M., Reading, T., Rogulin, D., Schottlander, D., Solomonides, T. (2004). Grid Databases for Shared Image Analysis in the MammoGrid Project.
- Anand Deshpande, P. P., & Rao, D. H. (2014). Motion Estimation Algorithms in Video Super-Resolution.
- Ansary, M. A. A., Chowdhury, R. R., Ehsan, M. R., Anisuzzaman, M., Khan, H. A., & Ahmed, K. I. (2013). Comparison between time domain and frequency domain strain estimation in elastography. In *Advances in Electrical Engineering (ICAEE)*, 2013 International Conference on (pp. 257–262). http://doi.org/10.1109/ICAEE.2013.6750343
- Astakhov, V., Gupta, A., Santini, S., & Grethe, J. (2005). Data Integration in the Biomedical Informatics Research Network (BIRN). In B. Ludäscher & L. Raschid (Eds.), *Data Integration in the Life Sciences SE* - 31 (Vol. 3615, pp. 317–320). Springer Berlin Heidelberg. http://doi.org/10.1007/11530084\_31
- Babafemi, O., Sanjay, M., & Adigun, M. (2013). Towards Developing Grid-based Portals for e-Commerce on-Demand Services on a Utility Computing Platform. *{IERI}* Procedia, 4(0), 81–87. http://doi.org/http://dx.doi.org/10.1016/j.ieri.2013.11.013
- Bertero, M., Bonetto, P., Carracciuolo, L., D'Amore, L., Formiconi, A., Guarracino, M. R., Laccetti, G., Murli, A., Oliva, G. (2003). MedIGrid: a medical imaging application for computational Grids. In *Parallel and Distributed Processing Symposium*, 2003. Proceedings. International (p. 8 pp.-). http://doi.org/10.1109/IPDPS.2003.1213457
- Boccia, V., & Guarracino, M. R. (2005). A Grid Enabled PSE for Medical Imaging: Experiences on MedIGrid. 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05), 529–536. http://doi.org/10.1109/CBMS.2005.9
- Bruneo, D., Iellamo, G., Minutoli, G., & Puliafito, A. (2009). GridVideo: A Practical Example of Nonscientific Application on the Grid. *IEEE Trans. on Knowl. and Data Eng.*, 21(5), 666–680. http://doi.org/10.1109/TKDE.2008.191

- Cabo, M., Possani, E., Potts, C. N., & Song, X. (2015). Split-merge: Using exponential neighborhood search for scheduling a batching machine. *Computers & Operations Research*. http://doi.org/10.1016/j.cor.2015.04.017
- Cinquini, L., Crichton, D., Mattmann, C., Harney, J., Shipman, G., Wang, F., Ananthakrishnan, R., Miller, N., Denvil, S., Morgan, M., Pobre, Z., Bell, G., Schweitzer, R. (2014). The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems*, 36, 400–417. http://doi.org/10.1016/j.future.2013.07.002
- Cossu, R., Petitdidier, M., Linford, J., Badoux, V., Fusco, L., Gotab, B., Hluchy, L., Lecca, G., Murgia, F., Plevier, C., Renard, P., Schwichtenberg, H., de Cerff, W., Tran, V., Vetois, G. (2010). A roadmap for a dedicated Earth Science Grid platform. *Earth Science Informatics*, 3(3), 135–148. http://doi.org/10.1007/s12145-010-0045-4
- Dabas, P., & Arya, A. (2013). Grid Computing : An Introduction. International Journal of Advanced Research in Computer Science and Software Engineering, 3(3), 466–470.
- DAWSON, C. J., Rick, A. H. I. I., Joseph, J., & Seaman, J. W. (2014). Grid computing accounting and statistics management system. Google Patents. Retrieved from http://www.google.com/patents/US8713179
- De, D., & Nationale, M. (2014). A Test Case on the Usage of Grid Infrastructure in Regional Climate Simulation Centre National pour la Recherche Scientifique et Technique , Rabat , Morocco, 19(6), 769–774. http://doi.org/10.5829/idosi.mejsr.2014.19.6.12533
- De Haan, G., Biezen, P. W. A. C., Huijgen, H., & Ojo, O. A. (1993). True-motion estimation with 3-D recursive search block matching. *Circuits and Systems for Video Technology, IEEE Transactions on, 3*(5), 368–379,388. http://doi.org/10.1109/76.246088
- Desell, T., Waters, A., Magdon-ismail, M., Szymanski, B. K., Varela, C. A., Newby, M., Anderson, D. (2010). Accelerating the MilkyWay @ Home Volunteer Computing Project with GPUs, 276–288.
- Fern ández-Quiruelas, V., Fern ández, J., Cofi ño, A. S., Fita, L., & Guti érez, J. M. (2011). Benefits and requirements of grid computing for climate applications. An example with the community atmospheric model. *Environmental Modelling & Software*, 26(9), 1057–1069. http://doi.org/http://dx.doi.org/10.1016/j.envsoft.2011.03.006
- Field, E. H., Gupta, V., Gupta, N., Maechling, P., & Jordan, T. H. (2005). Hazard map calculations using grid computing. *Seismological Research Letters*, 76(5), 565– 573.

Foster, I., & Kesselman, C. (2002). What is the Grid. A Three Point Checklist, 20.

- Foster, I., & Kesselman, C. (2003). The Grid 2: Blueprint for a new computing infrastructure. Elsevier.
- Glover, G. H., Mueller, B. A., Turner, J. A., van Erp, T. G. M., Liu, T. T., Greve, D. N., Voyvodic, J., Rasmussen, J., Brown, G., Keator, D., Calhoun, V., Lee, H. J., Ford, J. M., Mathalon, D., Diaz, M., O'Leary. D. S., Gadde, S., Preda, A., Lim, K. O., Wible, C. G., Stern, H. S., Belger, A., McCarthy, G., Ozyurt, B., Potkin, S. G. (2012). Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. *Journal of Magnetic Resonance Imaging*, *36*(1), 39–54. http://doi.org/10.1002/jmri.23572
- Grethe, J. S., Baru, C., Gupta, A., James, M., Ludaescher, B., Martone, M. E., Papadopoulos, P. M., Peltier, S. T., Rajasekar, A., Santini, S., Zaslavsky, I. N., Ellisman, M. H. (2005). From Grid to Healthgrid: Proceedings of Healthgrid 2005. *Studies in Health Technology and Informatics*. Retrieved from http://iospress.metapress.com/content/LGPQ69GTY5TTRKGW
- Hardy, G. H., & Wright, E. M. (1979). An introduction to the theory of numbers. Oxford University Press.
- Hussin, M., & Latip, R. (2013). Adaptive Resource Control Mechanism Through Reputation-Based Scheduling in Heterogeneous Distributed Systems. *Journal of Computer* Science, 9(12), 1661–1668. http://doi.org/10.3844/jcssp.2013.1661.1668
- Ismail, Y., McNeely, J. B., Shaaban, M., Mahmoud, H., & Bayoumi, M. A. (2012). Fast motion estimation system using dynamic models for H. 264/AVC video coding. Circuits and Systems for Video Technology, IEEE Transactions on, 22(1), 28–42.
- Jacob, B., Brown, M., Fukui, K., & Trivedi, N. (2005). *Introduction to grid computing*. IBM, International Technical Support Organization.
- Jin, H., Liao, X. F., Hu, Q. H., & Pang, L. P. (2014). Video splitting and distributed placement scheme for clustered video servers. Google Patents.
- Keator, D. B., Grethe, J. S., Marcus, D., Ozyurt, B., Gadde, S., Murphy, S., Pieper, S., Greve, D., Notestine, R., Bockholt, H. J., Papadopoulos, P. (2008). A national human neuroimaging collaboratory enabled by the Biomedical Informatics Research Network (BIRN). *IEEE Transactions on Information Technology in Biomedicine : A Publication of the IEEE Engineering in Medicine and Biology Society*, *12*(2), 162–72. http://doi.org/10.1109/TITB.2008.917893
- Kim, D., Lee, J., Jeon, K., & Jeong, J. (2006). Macroblock Mode Decision Scheme for Fast Encoding in H.264/AVC. In Y. Zhuang, S.-Q. Yang, Y. Rui, & Q. He (Eds.),

*Advances in Multimedia Information Processing - PCM 2006 SE - 43* (Vol. 4261, pp. 365–374). Springer Berlin Heidelberg. http://doi.org/10.1007/11922162\_43

- Kim, N.-J., Erturk, S., & Lee, H.-J. (2009). Two-bit transform based block motion estimation using second derivatives. *Consumer Electronics, IEEE Transactions* on, 55(2), 902–910. http://doi.org/10.1109/TCE.2009.5174473
- Krefting, D., Bart, J., Beronov, K., Dzhimova, O., Falkner, J., Hartung, M., Hoheisel, A., Knoch, T. A., Lingner, T., Mohammed, Y., Peter, K., Rahm, E., Sax, U., Sommerfeld, D., Steinke, T., Tolxdorff, T., Vossberg, M., Viezens, F., Weisbecker, A. (2009). MediGRID: Towards a user friendly secured grid infrastructure. *Future Generation Computer Systems*, 25(3), 326–336. http://doi.org/http://dx.doi.org/10.1016/j.future.2008.05.005
- Kumar, V. S., Rutt, B., Kurc, T., Catalyurek, U. V, Pan, T. C., Chow, S., Lamont, S., Martone, M., Saltz, J. H. (2008). Large-scale biomedical image analysis in grid environments. *IEEE Transactions on Information Technology in Biomedicine : A Publication of the IEEE Engineering in Medicine and Biology Society*, 12(2), 154–61. http://doi.org/10.1109/TITB.2007.908466
- Lecca, G., Petitdidier, M., Hluchy, L., Ivanovic, M., Kussul, N., Ray, N., & Thieron, V. (2011). Grid computing technology for hydrological applications. *Journal of Hydrology*, 403(1–2), 186–199. http://doi.org/http://dx.doi.org/10.1016/j.jhydrol.2011.04.003
- Lee, H., & Park, D. (2014). A Grid Service-Based Virtual Screening System. *The Computer Journal*, 57(2), 302–307. http://doi.org/10.1093/comjnl/bxt015
- Liao, X., & Jin, H. (2005). A new distributed storage scheme for cluster video server. Journal of Systems Architecture, 51(2), 79–94.
- Liu, Q., Sclabassi, R. J., Li, C.-C., & Sun, M. (2005). An Application of MAP-MRF to Change Detection in Image Sequence Based on Mean Field Theory. *EURASIP J. Appl. Signal Process.*, 2005, 1956–1968. http://doi.org/10.1155/ASP.2005.1956
- Luther, A., Buyya, R., Ranjan, R., & Venugopal, S. (2005). Alchemi: A. NET-based Enterprise Grid Computing System. In *International Conference on Internet Computing* (pp. 269–278).
- Mitra, S., Das, A., & Roy, S. (2015). Development of E-Learning System in Grid Environment. In J. K. Mandal, S. C. Satapathy, M. Kumar Sanyal, P. P. Sarkar, & A. Mukhopadhyay (Eds.), *Information Systems Design and Intelligent Applications SE - 15* (Vol. 340, pp. 133–141). Springer India. http://doi.org/10.1007/978-81-322-2247-7\_15
- Moreno, R., & Alonso-Conde, A. B. (2004). Job scheduling and resource management techniques in economic grid environments. In *Grid Computing* (pp. 25–32).

- Nithya, L. M., & Shanmugam, A. (2010). A New Grid Architecture using JMF for Video-on-Demand Applications. *Technology*, 10(9).
- Okade, M., & Biswas, P. K. (2013). A novel motion vector outlier removal technique based on adaptive weighted vector median filtering for global motion estimation. 2013 Annual IEEE India Conference (INDICON), 1–5. http://doi.org/10.1109/INDCON.2013.6726121
- Philip, J. T., Samuvel, B., Pradeesh, K., & Nimmi, N. K. (2014). A comparative study of block matching and optical flow motion estimation algorithms. In *Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD), 2014 Annual International Conference on (pp. 1–6).* http://doi.org/10.1109/AICERA.2014.6908204
- Redolfi, A., McClatchey, R., Anjum, A., Zijdenbos, A., Manset, D., Barkhof, F., Spenger, C., Legré, Y., Wahlund, L., di San Pietro, C. B. (2009). Grid infrastructures for computational neuroscience: the neuGRID example. *Future Neurology*, 4(6), 703–722.
- Renard, P., Badoux, V., Petitdidier, M., & Cossu, R. (2009). Grid Computing for Earth Science. *Eos, Transactions American Geophysical Union*, 90(14), 117–119. http://doi.org/10.1029/2009EO140002
- Roy, S., De Sarkar, A., & Mukherjee, N. (2014). An Agent Based E-Learning Framework for Grid Environment. In *E-Learning Paradigms and Applications* (pp. 121–144). Springer.
- Shen, B. (2005). Submacroblock motion compensation for fast down-scale transcoding of compressed video. *Circuits and Systems for Video Technology, IEEE Transactions on, 15*(10), 1291–1302. http://doi.org/10.1109/TCSVT.2005.854216
- Shenolikar, P. C., & Narote, S. P. (2009). Motion estimation on DWT based image sequence. International Journal of Recent Trends in Engineering, 2(4), 168–170.
- Shenoy, P. J., & Vin, H. M. (1997). Efficient striping techniques for multimedia file servers. In Network and Operating System Support for Digital Audio and Video, 1997., Proceedings of the IEEE 7th International Workshop on (pp. 25–36).
- Sidhu, J., & Singh, S. (2010). Design, Implementation and Evaluation of Grid Environment for DV to MPEG4 Video Conversion, *1*(2), 151–155.
- Sper de Almeida, E. (2012). Reducing Time Delays in Computing Numerical Weather Models at Regional and Local Levels : A Grid-Based Approach. *International Journal of Grid Computing & Applications*, 3(4), 1–17. http://doi.org/10.5121/ijgca.2012.3401

- Thampi, S. M. (2013). A Review on P2P Video Streaming. arXiv Preprint arXiv:1304.1235.
- Toharia, P., Sánchez, A., Bosque, J. L., & Robles, O. D. (2008). Efficient grid-based video storage and retrieval. In On the Move to Meaningful Internet Systems: OTM 2008 (pp. 833–851). Springer.
- Toharia, P., Sánchez, A., Bosque, J. L., & Robles, O. D. (2010). GCViR: grid contentbased video retrieval with work allocation brokering. *Concurrency and Computation: Practice and Experience*, 22(11), 1450–1475. http://doi.org/10.1002/cpe.1492
- Tsiknakis, M., Rueping, S., Martin, L., Sfakianakis, S., Bucur, a, Sengstag, T., Brochhausen, M., Pucaski, J., Graf, N. (2007). Developing a European grid infrastructure for cancer research: vision, architecture and services. *Ecancermedicalscience*, 1, 56. http://doi.org/10.3332/ecms.2007.56
- Wang, H., Sun, C., & Li, H. (2008). A Novel Macroblock-Level Rate-Distortion Optimization Scheme for H. 264/AVC. In Advances in Visual Computing (pp. 762–771). Springer.
- Warren, R., Solomonides, A. E., del Frate, C., Warsi, I., Ding, J., Odeh, M., McClatchey, R., Tromans, C., Brady, M., Highnam, R., Cordell, M., Estrella, F., Bazzocchi, M., Amendolia, S. R. (2007). MammoGrid — a prototype distributed mammographic database for Europe. *Clinical Radiology*, 62(11), 1044–1051. http://doi.org/http://dx.doi.org/10.1016/j.crad.2006.09.032
- Weisbecker, A., Falkner, J., & Rienhoff, O. (2009). MediGRID Grid Computing For Medicine and Life Sciences. In S. Lin & E. Yen (Eds.), *Grid Computing SE - 5* (pp. 57–65). Springer US. http://doi.org/10.1007/978-0-387-78417-5\_5
- Wozak, F. (2007). Medical Data Grids as a Base-Architecture for Interregional Shared Electronic Health Records. Institute for Health Information Systems, University for Health Sciences, Medical Informatics and Technology (UMIT), Doctoral Thesis.
- Xiang, T., Yu, C., & Chen, F. (2013). Fast Encryption of JPEG 2000 Images in Wireless Multimedia Sensor Networks. In K. Ren, X. Liu, W. Liang, M. Xu, X. Jia, & K. Xing (Eds.), Wireless Algorithms, Systems, and Applications SE - 17 (Vol. 7992, pp. 196–205). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-39701-1\_17

