Jurnal
Teknologi

Full Paper

A Novel Multimedia File Splitting Technique for Medical Data Grid Storage

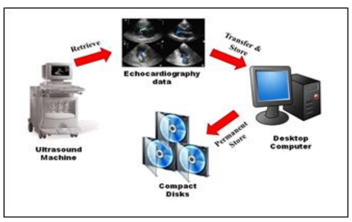
Mien May Chong[a], Rohaya Latip[a*], Masnida Hussin[a], Hamidah Ibrahim[b]

[a]Department of Communication Technology and Network, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia
[b]Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia

Article history
Received
3 December 2013
Received in revised form
2 July 2014
Accepted
25 November 2014

*Corresponding author
rohayalt@upm.edu.my

Graphical abstract



Abstract

Nowadays, videos and images dimensions are improved from 1D and 2D images into 3D or 4D (3 spatial dimension + time) images / videos. The improvement for these video data file were also increased the size of the data. Therefore, single and small storage is not enough for storing them. To overcome the problem, grid-based file storage service is introduced. However, the quality-of-service is very important for this applications and service. Therefore, several of quality-of-service requirements such as delay for data transmission, average CPU time per chunk used, and the total download time for a complete video file are need to be investigate. In this paper, our grid-based file storage test bed architecture, and the previous existing multimedia file splitting techniques will be discussed. Besides, our new multimedia file splitting technique – "Exponential-And-Uniform-based Splitting Technique" is also being proposed.

Keywords: Grid computing, file splitting technique, video file, data storage, medical video file

© 2015 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

In the early of 1990s, grid computing was a vision of sharing the computers to use for solving the large-scale of scientific, technical or business data or problems. It has emerged as a new technology that able to bring the high performance distributed infrastructures for users. [1] Some of the industries area are started to merge with the grid-based technologies such as Aerospace, Automotive, Agriculture, Financial and also the Medical area [2].

In medical area, there are many medical projects started to integrate with the grid computing. One of the projects is the Biomedical Informatics Research Network (BIRN) [10]. It is a national initiative that developed to assist the data sharing and online collaboration biomedical research. This project is focusing on the neuroimaging studies [6], [7], [9]. Besides, in Europe Union, there is a project called MammoGrid [7], [8], [11] was implemented, while in German, the project MedIGrid [7], [12] was introduced. MammoGrid is a project that aims to prove that Grid infrastructures to facilitate the collaboration between the researchers and clinicians across the Europe Union. MedIGrid project is used to investigate the application of Grid technologies for manipulating a large medical database.

Nowadays, the medical data such as medical videos and images all are improved their quality from 1D (e.g. cardiograms and encephalograms) and 2D (e.g. X-rays) images into 3D (e.g. tomography) or 4D (3 spatial dimension + time) images / videos. Those improvements have increased the size of these medical data. Thus, nowadays, a single, small storage is already not enough for used in storing them. By using the grid-based solution, large-scale of the medical data are stored using the grid storage. Unlike single, small storage, which just can be access by few users in a particular place, grid-based storage allows the entire hospital users such as doctors, nurses and physical technicians, that have the permission to access it for storing and retrieving all the related patient medical data under the same VO (virtual organization).

However, for a grid-based medical data storage service, the Quality-of–Service (QoS) is very important. If a service system has a better QoS, it will provide a better service for the users. Otherwise, it is a useless service. Thus, the Quality-of–Service (QoS) of a grid-based medical data storage service must be focused and studied [3].

2.0 EXISTING VIDEO SPLITTING TECHNIQUE

The volume for video of medical data is much bigger than the image data. Therefore, during the video data transfer to the grid storage, normally the large-scale volume of data will be split into few chunks and transfer to the storage for replication and storing. From the previous paper [3], they had investigated that the different type of media file splitting techniques will affect some of the Quality-of-Service requirements (such as initial provisioning delay for the video data streaming process, average CPU time used for parallel retrieving back the data, and also total time download for a complete data file) of a service system.

In previous papers [3], [4], [5], they stated some commons on splitting techniques that currently been used for splitting the media file data. One of the most common used splitting techniques is the uniform splitting technique and Fibonacci-based Splitting Technique.

2.1 Uniform Splitting Technique

This splitting technique, the entire video file will be splited into the same size of chunks. Below is the equation used for this splitting technique:

$$V^i = \frac{V}{n} \qquad (1)$$

Where:
V is assumed as the size of the whole video file,
V^i is the size of i^{th} chunk, with $i = 1,\ldots,n$,
n is the number of chunks.

2.2 Fibonacci-based Splitting Technique

In paper [3], Fibonacci-based Splitting technqiue was introduced. Fibonacci-based splitting is the splitting technique that based on the Fibonacci number, where the size of i^{th} chunk is assumed to be proportional to the i^{th} Fibonacci number. Equation (2) below is the formula for this techniuqe,

$$V^i = F(i) * \frac{V}{\sum_{i=1}^{n} F(i)} \qquad (2)$$

Where:

V is assumed as the size of the whole video file,
V^i is the size of i^{th} chunk, with $i = 1,\ldots,n$,
n is the number of chunks,
$F(i)$ indicates as i^{th} term of Fibonacci series.

2.3 Fibonacci-based Splitting With V1min Technique

However, there is a limitation, where the first size of the chunk has to be inferiorly limited; in fact, the value of the first chunk rapidly decreases when number of chunk increases, thus the values for the first chunk is not efficiently playable in multimedia streaming [3], where the size of the file are truncated which was enhanced from Fibonacci-based Splitting technique. This truncated version technique is named as "*Fibonacci-based Splitting With* V^1_{min} *technique*" [3].

This truncated version technique is also based on the Fibonacci number, which is similar to Fibonacci-based Splitting technique. However, the author decided to use only first $\bar{x}$ terms of fibonacci series and by forcing all the remaining terms to assume the F($\bar{x}$) value. Once the size of the file (V) and the number of chunks (n) are fixed, the value $\bar{x}$ can be obtained with respect to minimum value accepted for size of the first chunk, expressed as $V^1_{min.}$[3]. The $\bar{x}$ below is the formula for this truncated version technique,

$$\frac{V}{\sum_{i=1}^{\bar{x}} F(i) + \sum_{j=\bar{x}+1}^{n} F(\bar{x})} \geq V^1_{min} \qquad (3)$$

Using the following Fibonacci series property:

$$\sum_{i=1}^{\bar{x}} F(i) + 1 = F(\bar{x}+2) \qquad (4)$$

Relation (2) [3] can be expressed as:

$$F(x+2) + (n - x)\, F(x) \leq \frac{V}{V^1_{min}} + 1 \qquad (5)$$

that provides an easy way to identify the requested value of $\bar{x}$.

The dimension of each chunk (V^i) can be calculate as below:-

$$V^{i} = G(i) * \frac{V}{F(\bar{x}+2) + (n - \bar{x})\,F(\bar{x}) - 1} \qquad i = 1,\ldots,n \quad (6)$$

Where

$$G(i) = \begin{cases} F(i) & \text{if } 1 \le i \le \bar{x} \\ F(\bar{x}) & \text{if } \bar{x} \le i \le n \end{cases}$$

Among those existing techniques, the uniform splitting technique will gives a high initial delay for the first chunk downloading process. This is because the first chunk data size is bigger than the other splitting technique's first chunk data size. Thus, during the joining process, the start time for joining the video chunk back into a complete data file will increase the time due to the system had to wait several of time to retrieve back the first chunk from the grid storage.

For Fibonacci-based Splitting with V^{1}_{min} Technique, due to controlling the first chunk size by setting the minimum chunk size, it is giving the smallest initial delay among the other splitting techniques. However, it is not the best technique to used because the algorithm needs to check every possible first chunk size that is smaller than the minimum size, and choose the nearest first chunk size to split, thus this will increase the waiting time.

3.0 PROPOSED NEW "EXPONENTIAL-AND UNIFORM-BASED SPLITTING TECHNIQUE

3.1 Exponential-and-Uniform-based Splitting Technique (ExpoNUni)

To avoid the complexity of checking the most nearest minimum first chunk size, and also aimed to reduce the delay time from the existing techniques, a new splitting algorithm named – "*Exponential-And-Uniform-based (ExpoNUni) Splitting Technique*" is introduced by us. This proposed "*ExpoNUni*" technique is a splitting technique that is based on the exponential equation. The first four chunks will be following the exponential equation to split, while the remaining size of the video chunks file will split by forcing to use the F(3) value. Below is the formula for this technique,

$$v^{x} = G(x) * \frac{v}{\sum_{x=0}^{2} F(x) + \sum_{j=3}^{n} F(3)} \qquad (7)$$

Where:

$$G(x) = \begin{cases} F(x) & \text{if } 0 \le x < 3 \\ F(3) & \text{if } 3 \le x \le n \end{cases}$$

F(x) indicates as x^{th} term of Fibonacci series.

$f(x) = e^{x}$, where $x = 0, 1, 2.....$

V is assumed as the size of the whole video file,

V^{X} is the size of X^{th} chunk, with $X = 0,\ldots,n$,

n is the number of chunks.

In our algorithm, we use the first 4 terms of the Exponential series only, and forcing all the remaining terms to assume as the 4th term value. The reason that we use the first 4 terms value only is because starting from the fifth term value of the exponential series, all the remaining terms value are too big. Thus, they are not suitable to use for splitting.

From our algorithm, we found that the first chunk video size is smaller than the existing techniques' first chunk video size. Due to this reason, we believed that the "initial delay" evaluation metric in our splitting technique is lesser than the existing splitting techniques and also because of the reducing of initial delay, compare with the existing splitting techniques, our splitting technique will enhanced the joining of the video chunks together.

Thus, compare with the existing Fibonacci Splitting Technique, our splitting technique is giving by average 1.43% lesser time for downloading a complete video file to the user's pc from the grid storage, while compare to existing Uniform Splitting Technique, the average of 1.67% of downloading time is decreasing. For the file download time, in the averages of 1.43% (ExpoNUni vs Fibonacci) and 1.67% (ExpoNUni vs Uniform) of decreasing which we believed the reasons of exponential equation in our algorithm.

4.0 GRID-BASED VIDEO FILE STORAGE TEST BED IMPLEMENTATION

4.1 Case Study: Medical Data Grid Storage

In traditional medical data storage, the ultrasound machine will send the echocardiography data into the single, standalone desktop computer for viewing or temporary storing. Then, to avoid the missing of those echocardiography data and also for reducing the usage of the computers storage, commonly, the technician will using the compact disk (CD) or other device storage to physically store the echocardiography data as shown in Figure 1.

However, for long period, due to having more and more patients, there will be more and more compact disks used in physical storage of the echocardiography data. Thus, there will be more space need to use for storing these echocardiography data CD. Besides, the management on these CD is also a problem due to increasing of the CDs' volume. To avoid the missing of these disks, an idea on using the grid computing storage for storing these data is introduced.

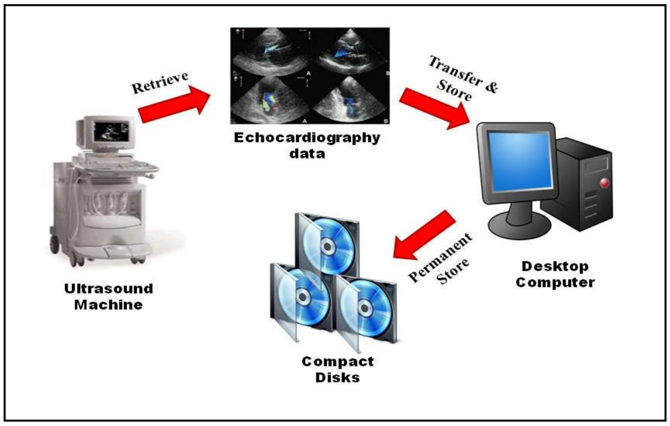


Figure 1 Process of retrieval, transferring and permanent store the echocardiography data in PPUKM

4.2 Our Grid Storage Test Bed Implementation

Due to avoid the missing of data, we proposed the grid storage to replace the compact disks (CD) to store all the related and important medical video data. In the test bed, the medical video data will be transferred into the computer and store in the grid storage through a grid portal that we developed as shown by Figure 2. During the data transmission, the data file will be split into various chunks for storing. At the final process, the evaluation metrics such as delay for the data transmission, average CPU time per chunk used, and the total download time for a complete video file are being measured.

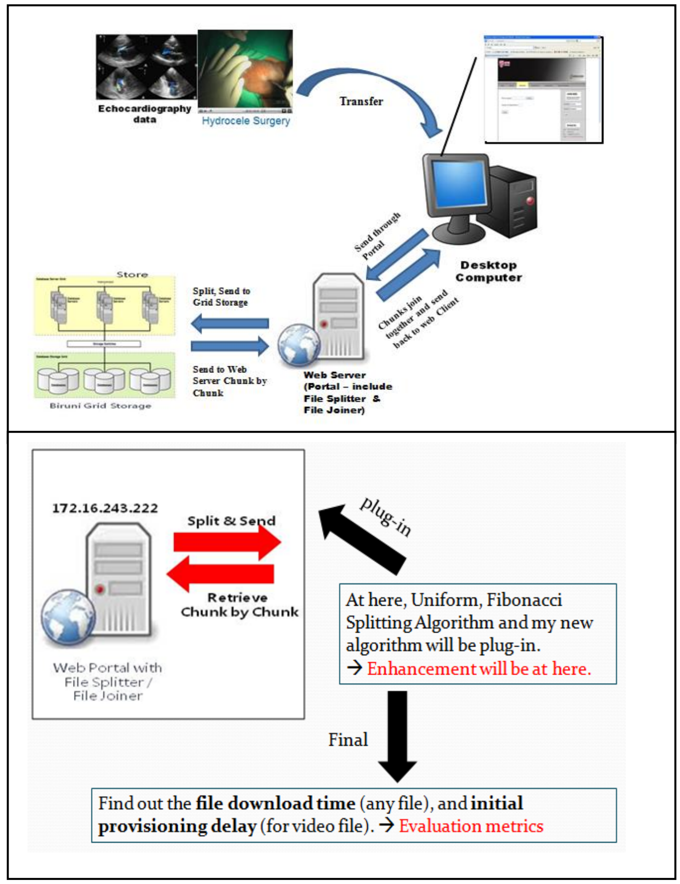


Figure 2 Our grid-based medical video file storage test bed

5.0 CONCLUSION

In this paper, we have discussed the current video file splitting techniques that is used for splitting the video data. We also introduced our new algorithm for the video file splitting technique that had overcome the download time of the video transmission by 1.43% of decrement for comparing with the Fibonacci Splitting Technique and 1.67% of decrement for comparing with the Uniform Splitting Technique. We have explained our grid-based file splitting test bed that used for testing and implementation of the algorithm.

From the discussions on those splitting techniques and our new technique, we believed that our Exponential-And-Uniform-based *(ExpoNUni)* Splitting Technique is the most appropriate technique for splitting the medical video data in terms of downloading time.

Acknowledgement

We would like to thank the InfoComm. Development Center (IDEC UPM), which have provided us with the grid infrastructure for us to do the research. We also would like to take this opportunity to thank the National University of Malaysia Medical Centre

(HUKM) and Hospital UiTM Sungai Buloh for giving us a chance to cooperate with them in improving the performances of their medical data storage.

References

[1] Foster, I., Kesselman, C. 1999. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco.

[2] Wikipedia. 2013. Data Grid Wikipedia. [Online]. From: http://en.wikipedia.org/wiki/Data_grid. [Accessed on 23 June 2013].

[3] Bruneo, D., Iellamo, G., Minutoli, G., Puliafito, A. 2009. GridVideo: A Practical Example of Nonscientific Application on the Grid. *IEEE Transactions on Knowledge and Data Engineering.* 21(5): 666-680.

[4] Nithya, L. M. and A. Shanmugam. 2010. A New Grid Architecture using JMF for Video-on-Demand Apllications. *International Journal of Computer Application.* 10(9): 34-39.

[5] Sidhu, J., S. Sarbjeet. 2010. Design, Implementation and Evaluation of Grid Environment for DV to MPEG4 Video Conversion. *International Journal of Computer Science and Communication*. 1(2): 151-155.

[6] 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*. 12(2): 162-172.

[7] 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.* 12(2): 154-161.

[8] Amendolia, S. R., Estrella, F., Hauer, T., Manset, D., McClatchey, R., Odeh, M., Reading, T., Rogulin, D., Schottlander, D., Solomonides, T. 2004. Grid Database for Shared Image Analysis in the MammoGrid Project. Proceedings 8th International Database Engineering and Applications Symposium. 302-311.

[9] Wikipedia. 2013. Biomedical Informatics Research Network Wikipedia. [Online]. From: http://en.wikipedia.org/wiki/Biomedical_Informatics_Research_Network. [Accessed on 15 July 2013].

[10] 2013. Biomedical Informatics Research Network (BIRN). [Online]. From: http://www.birncommunity.org/. [Accessed on 15 July 2013].

[11] Warren, R., A. E. Solomonides, C. del Frate, I. Warsi, J. Ding, M. Odeh, R. McClatchey, C. Tromans, M. Brady, R. Highnam, M. Cordell, F. Estrella, M. Bazzocchi, S.R. Amendolia. 2007. MammoGrid-A Prototype Distributed Mammographic Database for Europe. Clinical Radiology. 62(11): 1044-1051.

[12] 2013. MediGrid. [Online]. From: http://www.medigrid.de/. [Accessed on 15 July 2013].