Research Journal of Applied Sciences, Engineering and Technology 6(3): 512-514, 2013
DOI:10.19026/rjaset.6.4111
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: September 30, 2012 Accepted: December 13, 2012 Published: June 15, 2013

Research Article

# Research of Video Data Management Based on Cloud Storage

Silian Xie
Department of Communications and Control Engineering, Hunan Institute of Humanities Science and Technology, Loudi 417000, China

**Abstract:** This study introductions situation of video storage and the methods of video data management, then proposes a Scheduling Strategy algorithm based on Offset Cache (SSOC). According to the statistical information of data block buffer offset, through building a cloud environment and studying the storage layer's video data storage of the cloud storage, the algorithm schedules on mastering the supply and demand of data block comprehensively and accurately and solves the system's launch delay and the continuity of streaming media player. The method has applied in school's teaching resources.

**Keywords:** Cache offset, cloud storage, schedule, video data

## INTRODUCTION

With the development of science and technology, the Internet scale sharp expansion. In infrastructure scale, China's Internet bandwidth increased 16800 times from 1997 to 2007 and China telecom main bandwidth almost doubled every year, Google data center server number has reached a level. In the network service scale, China website reached 3 million in 2009 year, which is 5 times of 2005 years. In the user scale, the global have more than 1.7 billion Internet users by December 2009. Internet users have reached 380 million in China. Alibaba B2B registered enterprise has more than 5000; Taobao registered users has reached 150 million, Pay treasure registered users has reached 200 million; Tencent online communication user's has reached 100 million in peak. In data size, related research report shows that, the global have data quantity every 18 months will double; Amazon S3 object storage quantity reached 82 billion in 2009 and 4 billion growth of every month, Statistics show that YouTube video traffic reached billion times per minute of every day in 2009 in global and has more than 20 h of video upload quantity. Broadband rapid growth has laid a solid user base for the network video industry development and indicates the network video market good development trend. According to statistics, Internet users often use network services, online movie and TV to watch and download function accounted for 37.1% on the Internet in China.

Internet video traffic is explosive growth, which brought the infrastructure (for example, data center and backbone network) huge investment. According to Google data indicated that It construction investment has more than 7.4 billion us dollars in data center since 2006 and data center operating costs are higher than its construction cost, Internet video service provider's most concerned problems that it Can satisfy the dynamic change in the mass user diversified application experience, reduce the infrastructure investment and operation cost, balance the relationship between the scale and benefit.

We through the cloud storage technology to solve the problems of video service is very practical and also is feasible and therefore based on cloud storage of video storage technology research is of great theoretical value and broad application prospect and will become a trend of video service in future.

## ANALYSIS OF THE CURRENT RESEARCH METHOD

At present, there are lots of methods to study video data management, Such as: RAID, SAN, disk duplex, partition mirror, single disk array, disk array, Fault-tolerant software hot backup, replication, cluster technology, CDN technology, P2P technology, cloud storage technology and so on.

RAID5 technology (Muppalaneni and Gipinath, 2000) uses a block cross access to the data banding of distributed-check data, then writes to all disks in the parity. Data redundancy (Yoshitake *et al*., 1997) is provided by the parity information; data and parity information are placed in different disk among disk array. When a disk faults, the read performance will drop because the parity information needs to recover data; on the other hand, due to parity calculations, the use of block cross access to the data banding of distributed-check data is more effective than image-type.

Replication is at the expense of storage capacity to get the scalability of system. Due to increasing hardware storage capacity, the advantages of stripe exist no longer, Therefore, the video server cluster systems use replication technology to implement data distribution now.

By adding a new layer of network architecture to the existing Internet, CDN (Content Delivery Network) technology distributes center content to the nearest users and network "edge" node for the best service capacity; through the intelligent strategy, Which allows users to obtain the required contents, resolves Internet network congestion and improves user' response time of the web access.

With the (Decentralized) approach, P2P technology (Magharei *et al*., 2007) is such a new network computing technology that uses a large number of dynamic autonomy distribution resources[4] to accomplish specific functional. P2P technology combines with individual users into a network whose bandwidth and information can be shared with.

## SCHEDULING ALGORITHM BASED ON THE DATA OFFSET CACHE (SSOC)

In streaming model, the existing data scheduling strategy includes the following: Rarest-First strategy (Rarest-First, RF) (Agarwal and Rejaie, 2005). This strategy is conducive to speeding up the data block in the proliferation of network coverage and improving the overall throughput of the system and getting a better average playing continuity. But because of no taking into account the urgency degree of block elements, it results in a large start delay; Weighted Round-Robin (Weighted Round-Robin, WRR) (Zhou *et al*., 2007) strategy uses the size of the bandwidth, layer data block priority as the main basis for scheduling to get load balancing, but the chance of improving data is considered inadequately; Greedy strategy, known as the nearest deadline first (nearest deadline first) strategy (Chang *et al*., 2006), provides visually a good start late and playing continuity. But after testing and analysis, this strategy gets poor performance because Greedy strategy will likely lead to decline the shared probability of data blocks gradually and result in the performance reduce for overall system, eventually it makes each node of the playback not be achieved; Random strategy selects an ongoing request randomly from the partners node, this strategy's performances is not very stable, especially in the heterogeneous network environment. For the problems of scheduling algorithms and the cloud storage system, we propose a Scheduling Strategy algorithm based on Offset Cache (SSOC). According to the statistical information of data block buffer offset, the algorithm schedules on mastering the supply and demand of data block comprehensively and accurately. The experimental results show that SSOC algorithm has further improved in playing continuity and launch delay than the minimum priority scheduling algorithm.

**Algorithm description:**

- According to descending order of scheduling priority of data block sequence is $A_1, A_2, ..., A_m$.
- The corresponding sequence of data blocks has a collection of nodes are $S_1, S_2, ..., S_m$.
- Sending rate of all the neighbor nodes are R (1), R (2), ..., R (n).
- In the current scheduling cycle, the expected send wait time of each neighbor is Wait (1), Wait (2), ..., Wait (n), the initial values are 0.

Output Each data block has a sending node sender $A_i$.

**Algorithm:** max get ← min (m, τ×I); //Current scheduling cycle can receive the maximum number of data blocks
for i = 1 to $get_{max}$ do
$t_{min} = \infty$; // initialize for the first time $A_i$
for j = 1 to k do
$t_{deliver} = \frac{lenght}{R_j^i}$ // obtained $A_i$ expected time from the $S_j$
if $t_{deliver}$ + Wait ( $S_j^i$ ) <$t_{min}$ and $t_{deliver}$ + Wait ( $S_j^i$ ) <τ
t min ← $t_{deliver}$ + Wait ( $S_j^i$ ) $sender_i \leftarrow S_j^i$ ;
end for j;
if $sender_i \neq$ null
Wait ($sender_i$) ← $t_{min}$;
end for I;

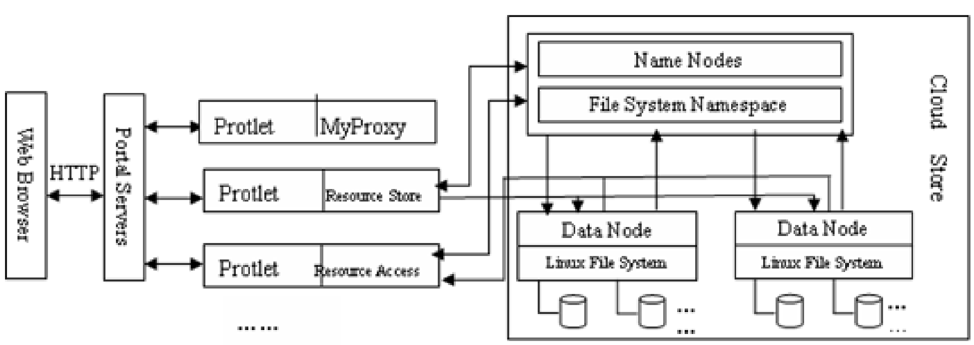


Fig. 1: Implementation model of cloud storage

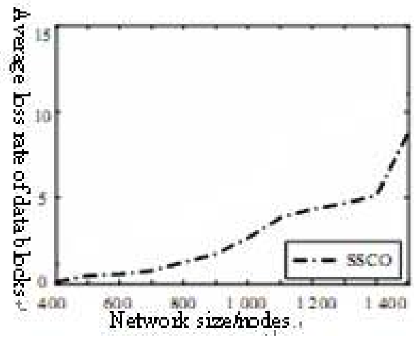


Fig. 2: Average loss rate of data blocks

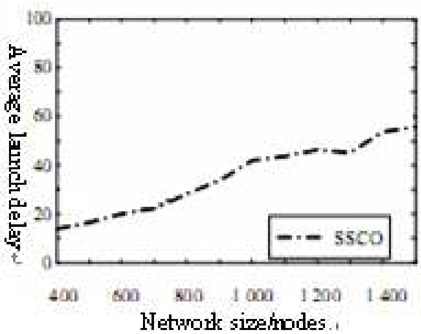


Fig. 3: Average launch delay

**Model implementation:** We use the general business machines as the underlying storage devices and Hadoop as a platform of cloud storage environment, managing the underlying business unit. We construct a virtual file system called HDFS. In HDFS, we develop application services modules of integrating teaching resources and video resources, including user management, directory management, resource management, system management and so on. By using the Liferay Portal as a container, we package each service mode in a Portlet. The process about teaching resources store and access in the cloud storage environment is shown in Fig. 1.

## EXPERIMENT AND RESULT ANALYSIS

We use NS2 as a simulator and the output bandwidth of the user nodes is [200, 600] Kb/sec, access bandwidth is [400, 1000] Kb/sec, server output bandwidth is 100 Mb/sec. Streaming media file playback time is 30 min in the data nodes of cloud store. Users join the network node shows Poisson distribution. Streaming media file playback speed is 400 Kb/sec. The simulation results are shown in Fig. 2 and 3.

We can be seen indicators curve growth is slow and smooth from the above of the experimental, It is indicating that data of scarcity degrees and high urgent data blocks will be scheduling priority in the same bandwidth conditions and can effectively reduce the loss rate, improve the system Continuous playback. System startup launch is correspondingly reduced.

## ACKNOWLEDGMENT

This study was partially supported by the Scientific Research Fund of Hunan Provincial Education Department under (Grant No. 10B053) and the Construct Program of the Key Discipline in Hunan Province.

## REFERENCES

Agarwal, V. and R. Rejaie, 2005. Adaptive multi-source streaming in heterogeneous peer-to-peer networks [C]. Proceedings of the 12th Annual Multimedia Computing and Networking. ACM Press, pp: 102-109.

Chang, F., J. Dean, S. Ghemawat W.C. Hsieh and D.A.W.M. Burrows, 2006. Bigtable: A distributed storage system for structured data [C]. Proceeding of the 7th USENIX Symposium on Operating Systems Design and Implementation (OSDI 06), 7: 15.

Magharei, N., R. Rejaie and Y. Guo, 2007. Mesh or multiple-tree: A comparative study of P2P live streaming services [C]. Proceeding of 26th IEEE International Conference on Computer Communications (INFOCOM 2007), pp: 1424-1432.

Muppalaneni, N. and K. Gipinath, 2000. Amulets-tier RAID storage system with RAID1 and RAID5. Proceedings of 14th International Parallel and Distributed Processing Symposium (IPDPS'00). Cncum, Mexico, May.

Yoshitake, S., M. Tetstutaro and Y. Naomi 1997. Software RAID technology for cluster environments. Syst. J., Retrieved from: Http:// www.storageconference.org/2001/2001CD/Pobshink.pdf.

Zhou, Y.P., D. Chiu and M. Lui, 2007. A simple model for analyzing P2P streaming protocols [C]. Proceedings of IEEE International Conference on Network Protocols (ICNP'07) [S.1.]. Beijing, pp: 226-235.