International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.2 (2014), pp.153-164
http://dx.doi.org/10.14257/ijsip.2014.7.2.15

Model Research of p2p VoD System based on Fluency

Yubin Wang[1], Jingyi bo[1] and Weili Xu[2]

[1]*College of Math and Information Science & Technology, Hebei Normal University of Science& Technology, Qinhuangdao, Hebei, 066004, China,*
[2]*College of Business Administration, Hebei Normal University of Science& Technology, Qinhuangdao, Hebei, 066004, China,*
[1]*qhdwyb@126.com*, [2]*qhdbjy@126.com*

Abstract

In the peer to peer (p2p) VoD system, there are many factors to influence the fluency of video, such as watching behavior, distribution of segments, download rates, data transmission, and so on. Most research concentrates on single factor without considering the interrelation of all the factors. In the paper, factors would be researched separately to find the relation with fluency of video playing. The mechanism can improve the download ratio with condition of keep watching smoothly. The model proposes a precision method to understand users' watching behavior and can help to optimize design in other kinds of p2p application.

Keywords: *peer-to-peer, VoD, model, fluency*

1. Introduction

The internet develops rapidly with mass customizable of broad band connection. More and more videos have been appeared on the internet. Video online playing is more and more popular. There are two categories of these systems: tree-based and mesh-based. A tree-based system constructs one or more spanning trees, whose structure is static making the system easy to analyze, measure and control. But as the heterogeneity, dynamicity of the Internet, tree-based systems are not very suitable. A mesh-based system doesn't form a fixed overlay. Peers exchange their buffer maps, which presents the available chunks to upload, to get the information of their neighbors and most commercial P2P streaming systems belong to this type [1]. Some typical commercial p2p streaming systems work well. Gridmedia has the record to support 220000 users in the same time to watch high-quality Internet TV by one server [2].

Due to the vast users and limit bandwidth, online smoothly playing is often difficulty. Even, waiting time is more than watching time. The bandwidth is also mainly occupied by this service. In order to resolve the problem, many tools and algorithms have been developed to help to keep the watching fluency. For example, crawler was developed to measure the global characteristics and estimate the video playback quality, batching [3-4], patching [5], and other tight delay methods are used to reduce time delay in online watching [6-9]. In addition model research has also get great improvement [10-13]. In order to cooperate tools, algorithms and models, cache size has also been studied. Mostly, research is focused on the minimize size of cache memory [14].

VoD is different from traditional Web application. It has characteristics of high bandwidth occupation and time delay sensitivity and these make VoD bad scalability. If all the multimedia files are stored in one server, no matter how wide the bandwidth is, the access

ISSN: 2005-4254 IJSIP

would be over loaded with the development of digital technology and increasing of users. So, the multimedia files can store in different servers. Each user can get wanted video from the nearest server. CDN (content distribution networks) algorithm has been developed [15-17]. This algorithm improves the scalability of VoD but take a new problem in. How to place and preserve the servers would be very difficult. But with the development of computer technology, personal computer can also play the role of server. Then, p2p technology would have been greatly developed.

Due to the limit of application, p2p has been researched for a long time according to simulation. After the VoD becomes universal based on p2p, many more studies are still going on. The research field still includes many aspects, such as users' behavior [18-19], download speed, download progress, and so on.

The paper would adopt p2p flow crawler to collect useful information [20]. The actual data collected would be used to verify the non-negligible effect between watching and download progress. A model is established to describe the relation between watching and download progress. The model provides a more accurate method.

In fact, users' watching behavior is greatly influenced by location of resource. The location of resources also have effect on the download speed of each user. If a user watches video while downloading the chunks, the download speed would have effect on the watching fluency. So, how to deal with the relation between downloading and watching properly is quite an important work. In p2p VoD system, watching fluency is effected by location of chunks, download speed, and download mechanism.

The major contribution in the paper would be showed as the following:

(1) According to the data collected, mechanism of data transmission is analyzed.

(2) Based on analysis data collected in multi channels, download speed is connected with location of segments of video. Relation between download speed and location of segments would reflect the nature property.

(3) Developing a mathematics model to describe the relation of watching fluency and download speed. Simulation method is used to verify the accuracy of the model. On one hand, it can acquire the record of users' watching history according to the model; on the other hand, the model can pre-assess watching fluency, this can help to understand the relation of watching fluency and download progress, watching progress, download mechanism, download speed.

(4) A new model is developed to analyze the data flow in channels and verify the non-negligible effect between watching and download progress.

The remainder of the paper is organized as follows. P2p VoD system is described in Section 2. Relationship between download speed and watching fluency is shown in Section 3. Simulation and results are shown in Section 5 and the followed section 5 is conclusion.

2. P2p VoD

P2p VoD system initially uses server to distribute multimedia objects. Segments dividing, nodes joining, process of resource requiring are the same as other p2p system; P2p VoD has its special characteristics. Characteristics of different p2p systems are showed in Figure 1 to Figure 3 as the following.

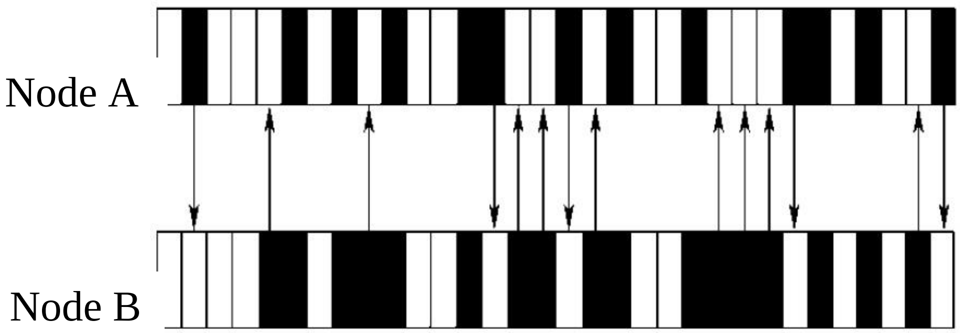

Figure 1. Chunks Exchange between Two Nodes in BitTorrent Mode

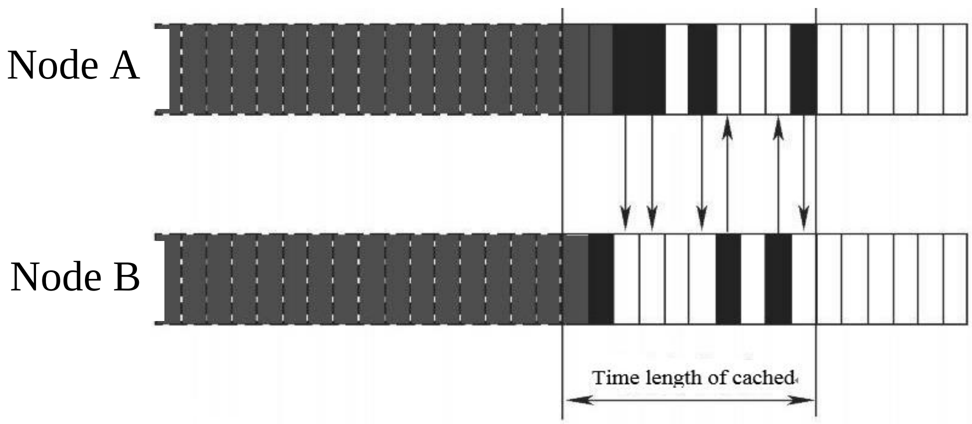

Figure 2. Chunks Exchange between Two Nodes in Live Mode

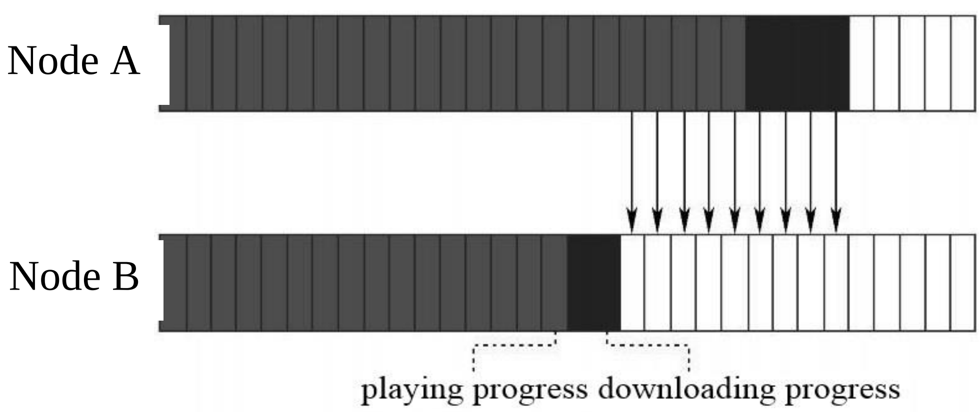

Figure 3. Chunks Exchange between Two Nodes in VoD

According to figures above, we can get some conclusions:

(1) In Bit Torrent system, whole media should be download completely before watching. So, in order to prevent too little chunks to download in the network. Mechanism of random or lowest first would be adopted [21]. Time of sharing resource is relative with time of completely downloading. But in VoD system, users usually watch as they download the video.

So, it is very important to keep fluency playing. User would download the needed chunks from the server or other users who watch ahead of him. Time of user download the video is up to time of user watching. So, users' watching habit would also influence the distribution of chunks and distribution influences the watching fluency.

(2) VoD system is different from video live. Live is more like TV. Users can just choose watching or not watching. Live has almost no interaction with customers. VoD is like DVD and users can choose the time to play. While in playing, customers can also pause, play and drag the video. In same channel, users' watching process is almost the same in live mode while it is quite different in VoD system. That is to say, in VoD system, resource share is quite limit. So, VoD system is more challenging than live. Characteristics of live and VoD are showed in Table 1.

Table 1. Characteristics of Live and VoD

	Video live	VoD
Request concurrency	High concurrency	Low concurrency, asynchronous, distributed
End to end delay	As short as possible	Not many demand
Number of channels	Usually small, users' attention is on several channels	Usually big, users often distribute in many channels
Demand of VCR	No interaction	User can have interaction with video playing, such as pause, replay, drag, and so on
Difficulty degree of data sharing	Little difference between users, easy to share	Difference between users, difficult to share
Memory size	Small	Big

3. Relationship between Download Speed and Watching Fluency

3.1. System Introduction

A typical VoD system is selected which develops by one company. There are nearly 40,000 channels and maximum number of dozens of thousands online numbers. Crawler is used to collect 20h data. The data refers 200 channels and ten thousand independent IP addresses.

According to analysis of the data, continuing download strategy is used in the system. The continuing download strategy is chunks downloading following the chronological order from server or other computers. Usually, in order to keep playing fluency, mean download speed is higher than play speed. So, downloaded chunks without playing increase with time going on. At the same time, chunks downloading speed is influenced by chunk number in the network. So, relationship between downloading speed and playing speed is not linear. It is not accuracy with downloading speed instead of playing speed to study the users' play behavior.

3.2. Relationship between Downloading and Playing Speed

According to existing research, most users watch from begin part of one video. If users drag in watching, they usually drag once [22]. After the drag, they would watch continuously. So, influence of drag can be neglected. The model is also based on the continuous watching and it really has enough accuracy. Table 2 shows main parameters in relation between downloading speed and watching speed and meaning of parameters. Playing progress represents number of chunks have been watched. Analysis of this progress can get users' watching behavior; downloading progress means the number of chunks downloaded when it is playing chunk x .

Table 2. Main Parameters used in Relation of Download and Play Progress

Symbol	Meaning
r	Playing speed
x	Playing progress
$l(x)$	Downloading progress
L	Total number of chunks
$u(c)$	Downloading speed of chunk c
$P_w(x)$	Probability distribution of playing progress in channel
$P_d(l)$	Probability distribution of downloading progress in channel
$P_w(x)$	Distribution function of playing progress in channel
$F_d(l)$	Distribution function of downloading progress in channel

The whole algorithm is described as the following.

Take a node as a research object. If one user watches video continuously, the watching progress after time length of t is:

$$x = r \times t \tag{1}$$

Because of different number and location of chunks, total downloading speed is showed as the following:

$$l = \int_0^t u(c)dt \tag{2}$$

We know $t = x / r$, if we set $y = rt$, then

$$l(x) = \int_0^{x/r} u(l(rt))dt = \frac{1}{r}\int_0^x u(l(y))dy \tag{3}$$

Downloading speed of each chunk can be deducted. When downloading chunk l , total time consumption t can be deducted as the following.

$$\frac{dl}{dt} = u(c), \quad dt = \frac{dl}{u(c)}, \quad t = \int_0^l \frac{dl}{u(c)} \tag{4}$$

So,

$$x(l) = r\int_0^l \frac{dc}{u(c)} \tag{5}$$

Equation (3) and (5) give the relation of playing progress and downloading progress at time t .

One chunk would be played only if it is completely downloaded. So, to keep the fluency, downloading speed should be higher than playing speed. If define x^* as the playing progress when video downloads completely. x^* can be given as the following:

$$x^{*} = r\int_{0}^{L} \frac{dc}{u(c)} \quad (6)$$

After x^{*}, watching progress can't be deducted by downloading progress. But, it has no effect on watching fluency. As the same reason, watching behavior is not influenced by distribution of chunks after x^{*}. Combined with actual condition, equation (3) and (5) can be modified as equation (7) and (8).

$$l(x) = \min\{ \frac{1}{r}\int_{0}^{x} u(l(y))\,dy ,\ L\} \quad (7)$$

$$x(l) = \min\{ r\int_{0}^{l} \frac{dc}{u(c)} ,\ r\int_{0}^{L} \frac{dc}{u(c)}\} = \min\{ r\int_{0}^{l} \frac{dc}{u(c)} ,\ x^{*}\} \quad (8)$$

Equation (7) and (8) show the relation between watching progress and downloading progress. Further, the conclusion can spread to the whole system to analyze the relation between distribution of downloading progress and watching progress.

According to bitmap information of all nodes in the channel, distribution of node download progress probability $P_d(l)$ and distribution of user download progress function $F_d(l)$ can be acquired. According to equation (5), distribution function $F_w(x)$ would be acquired.

$$F_w(x) = F_w(x(l)) = F_d(l(x)) = F_d(l) \quad (9)$$

Take actual condition into consideration, relation between download progress and watching progress of nodes can be described as the following:

$$F_d(l) = \begin{cases} F_w(x(l)) & l < L \\ 1 & l = L \end{cases} \quad (10)$$

$$F_d(l) = \begin{cases} F_w(x(l)) & l < L \\ F_w(x^{*}) & l = L \end{cases} \quad (11)$$

$$F_d(l) = \begin{cases} F_d(l(x)) & x < x^{*} \\ 1 & x = x^{*} \end{cases} \quad (12)$$

$$F_w(l) = \begin{cases} F_d(l(x)) & x < x^{*} \\ P_d(L) & x = x^{*} \end{cases} \quad (13)$$

Difference between downloading and watching progress is mainly caused by downloading speed, because playing speed is constant. So, number of resources and bandwidth are the key factors to influence download speed [23-25].

Figure 4 and 5 show the number of chunks and downloading speed in an ordinary channel in the VoD system. For the reason of different users have different interest, nearly all the video would play for minutes. Once users have no interest on the video, he would change another one. So, number of chunks in start is bigger than that of in end. Usually, number variation of chunks would become smoothly with chunk ID increases. That is because if one can watch a video for dozens of minutes, he would watch it continuously with less probability to change. Longer time user to watch, less probability to change. If bandwidth is constant, the downloading speed is in proportion with number of chunks. So, two curves keep same trends with each other. With the results from analysis, downloading speed is deducted as the following:

$$u(x) = \phi(a, nF_d(l)) = \phi(a, nF_d(l(x))) = \phi(a, nF_w(l(x))) \quad (14)$$

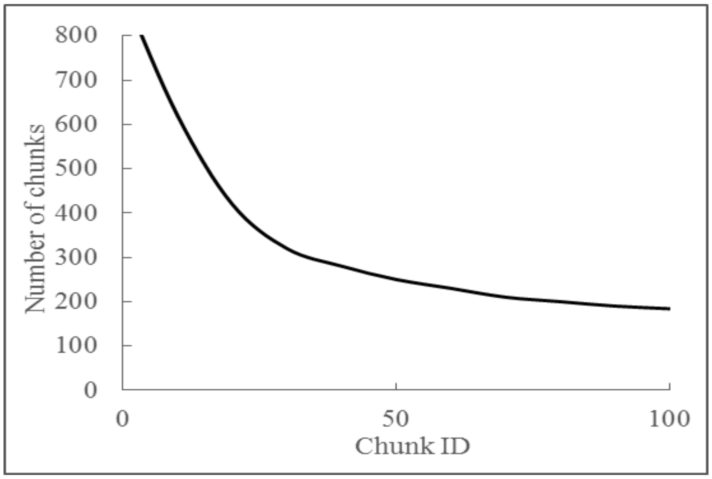

Figure 4. Number of Chunks in Different Positions of Channel

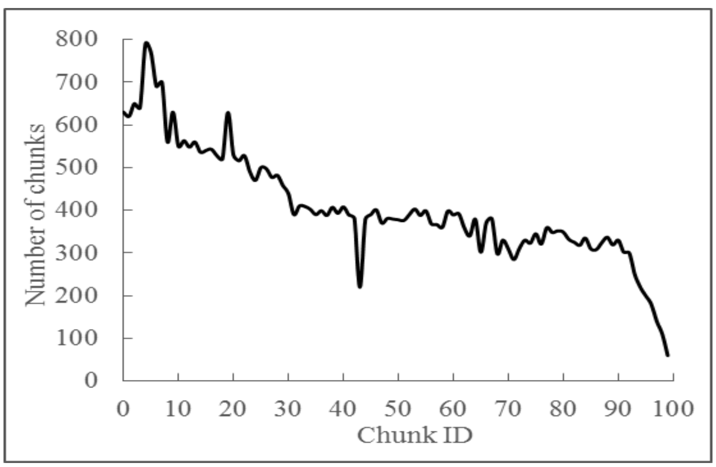

Figure 5. Download Speed in Different Positions of Channel

Where, a is mean value of statistics of downloading speed in system; n is total number of nodes; $nF_w(l(x))$ means number of chunks in location l. According to equation (14), (7) and (8). Downloading progress and watching progress can be described as:

$$l(x) = \min\{\int_0^x \phi(a, nF_w(y))dy, L\} \quad (15)$$

$$x(l) = \min\{ \int_0^l \frac{dc}{\phi(a, nF_d(c))}, x^*\} \quad (16)$$

Where,

$$x^* = \int_0^L \frac{dc}{\phi(a, nF_d(c))} \quad (17)$$

So, when chunks distribution $F_d(l)$ has been acquired in the channel. We can get watching progress distribution $F_w(x)$ according to equation (10).

$$F_d(l) = \begin{cases} F_w\left[\int_0^l \frac{dc}{\phi(a, nF_d(c))}\right] & l < L \\ 1 & l = L \end{cases} \quad (18)$$

Where, $F_d(l)$, $F_w(x)$ and n are mean values of statistics at steady state. Chunks distribution would be deducted as the same method:

$$F_w(x) = \begin{cases} F_d\left[\int_0^x \phi(a, nF_w(y))dy\right] & x < x^* \\ 1 & x = x^* \end{cases} \quad (19)$$

The method can be used to assess effect of variation of downloading bandwidth on system. It also can be used to assess performance of different downloading mechanism.

4. Simulation

4.1. Environment Introduction

At the beginning, there is only one server in network. Other nodes can take part in or quit freely. Nodes life time is used to describe the behavior of nodes and obey exponential distribution. Range of life time is in 1s to 10 000s.

The whole video is divided into 250 chunks with size of each chunk is 2 MB. The playing speed is 400 kbit/s. the mean value of maximum downloading bandwidth is 1 Mbit/s. The system uses second as time unit.

4.2. Results of Simulation

Statistics method is firstly used to calculate the accumulated distribution of downloading speed and watching speed at a certain time; then, model with equation (18) and (19) would be used to calculate and compare with collected data.

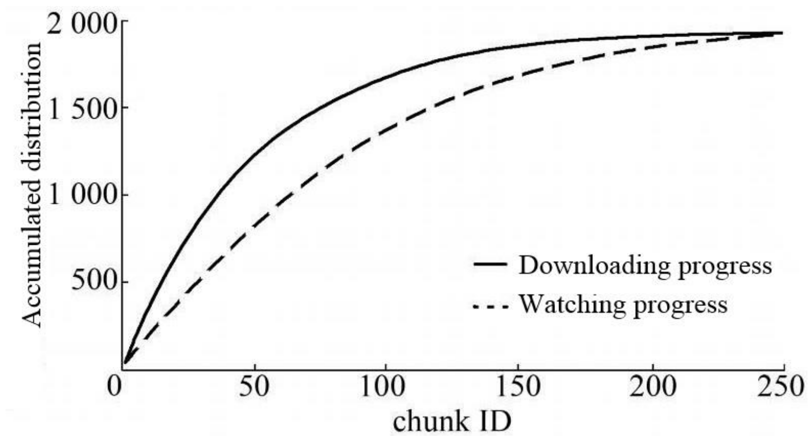

Figure 6. Accumulated Distribution of Watching and Downloading Progress

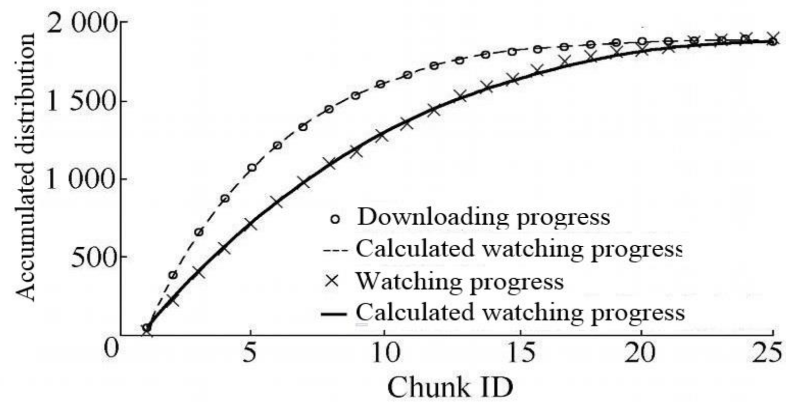

Figure 7. Calculation Results

Where, x axis represents chunk ID, y axis means nodes number. Coordinate of $(50,100)$ means that number of watching progress with no more than 50 is 100. From figure 6, we can see that watching speed mainly concentrates on location of small chunk ID. Between the two curves is a size gap. There are just 10% users have watching progress more than 50% while 20% have downloading progress more than 50%.

Figure 7 gives the deducted results and collected results. We can see there is not much difference between the two results. That is to say the model can precisely describe the actual condition. The model can also precisely transform from watching progress into download progress.

Figure 8 shows the download progress between simulation and experimental results. We can see that watching progress is more concentrate on the start of video compared with download progress. At the end of download progress curve, the peak means there are many users who completely download the video. But, among of them, only little part of the users are watching the video. These users play important role in normally running of channel. Peak point of download progress is locate on $l = L$, and peak point of watching progress would be locate on $x = x^*$ due to the model deducted.

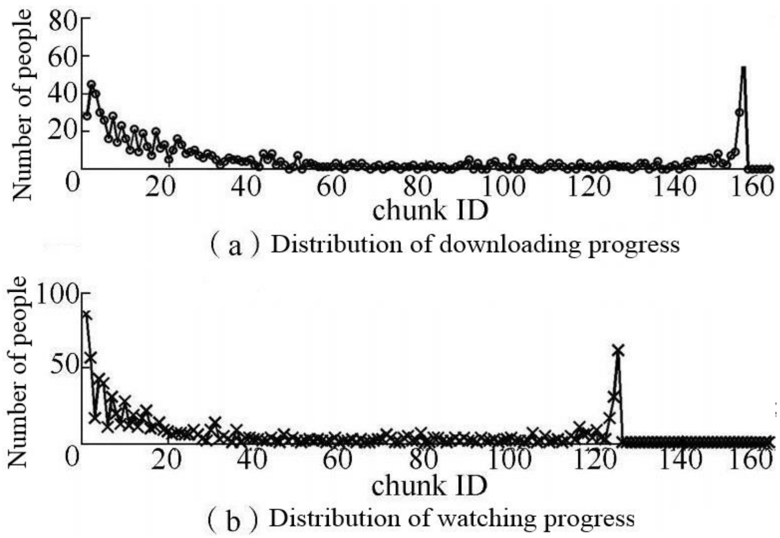

Figure 8. Distribution of Downloading and Watching Speed in a Channel

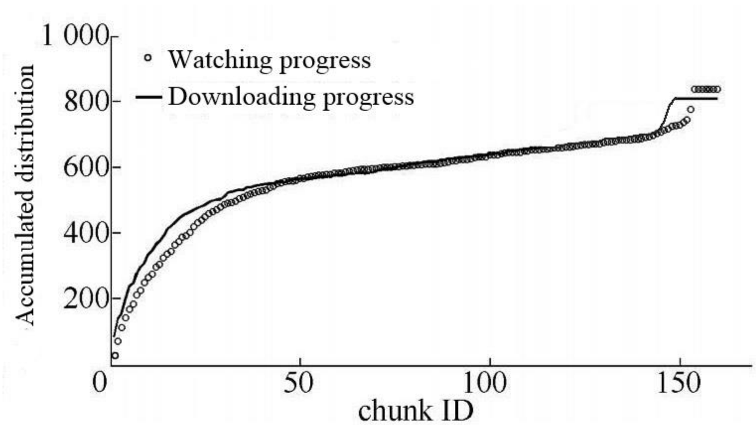

Figure 9. Accumulated Distribution of Watching and Downloading Progress

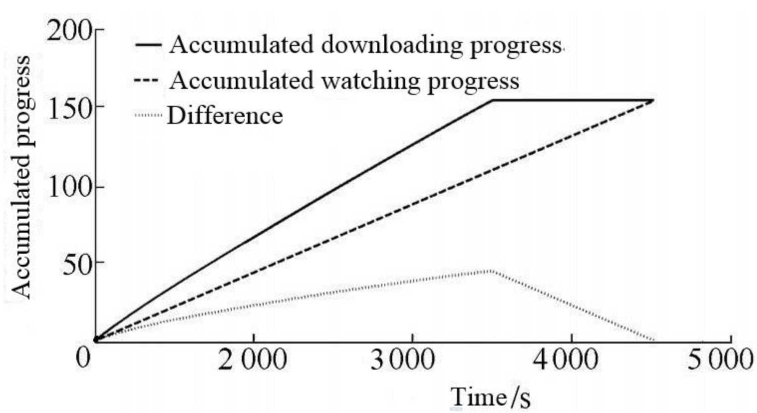

Figure 10. Difference between Watching and Downloading Progress

Figure 9 shows the accumulative distribution of download and watching progress. In the figure, watching speed is almost the same as download speed. This can't be deducted by the simulation model. Main reasons may include the follows: (1) user number is less than that in simulation; (2) not each bandwidth can get the upper limit of demand. However, we can still see the big difference between download speed and watching speed with chunk ID less than 50.

Figure 10 shows that we can't neglect the difference between download speed and watching speed. The most difference can get almost 40 chunks. It is in normal in the system we studied. Because of the node distribution follows power-law principle, so, in large size channel, the model can give approximate result. It also can cover most of the users' condition. With more and more nodes join in the network, results of model would be more and more accuracy.

5. Conclusion

Watching fluency has a close relation with downloading speed. The paper is based on watching fluency to calculate downloading speed and downloading progress. Also, downloading speed and progress is also calculated with watching fluency. The model is established to forecast the downloading progress and watching progress. The simulation results show that the model has good precision. The model can greatly help to improve and keep watching fluency.

References

[1] Y. Zhao and H. Shen, "A Simple Analysis on P2P Streaming with Peer Playback Lags", 2011 IEEE 3rd International Conference on Communication Software and Networks, Xi'an, China, **(2011)** May 27-29.

[2] M. Zhang, Q. Zhang, L. Sun and S. Yang, "Understanding the power of pull-based streaming protocol: Can we do better?", IEEE Journal on Selected Areas in Communications focuses on all telecommunications, including telephone, telegraphy, facsimile, and point-to-point television, by electromagnetic propagation, vol. 25, no. 9, **(2007)**, pp. 1678-1694.

[3] A. Dan, D. Sitaram and P. Shahabuddin, "Scheduling policies for an on-demand video server with batching", Proceedings of the 2nd ACM Multimedia Conference, **(1994)**, pp. 15-23.

[4] C. C. Aggarwal, J. L. Wolf and P. S. Yu, "On optimal batching policies f or video-on-demand storage servers", Proceedings of the 3rd IEEE International Conference on Multimedia Systems, **(1996)**, vol. 253-258.

[5] K. A. Hua, Y. Cai and S. Sheu, "Patching: A multi cast technique for true video-on-demand services", Proceedings of the 6th ACM International Conference on Multimedia. Bristol, **(1998)**, pp. 191-200.

[6] X. Hei, C. Liang, Y. Liu and K. W. Ross, "A measurement study of a large-scale P2P iptv system", IEEE Transactions on Multimedia, vol. 9, no. 8, **(2007)**, pp. 1672-1687.

[7] G. Bianchi, N. Blefari Melazzi, L. Bracciale, F. Lo Piccolo and S. Salsano, "Fundamental delay bounds in peer-to-peer chunk-based real-time streaming systems", 21st International Teletraffic Congress, **(2009)**, pp. 1-8.

[8] Y. Liu, "Delay Bounds of Chunk-Based Peer-to-Peer Video Streaming", IEEE/ACM Transactions on Networking, vol. 18, no. 4, **(2010)**, pp. 1195-1206.

[9] Y. Liu, "On the minimum delay peer-to-peer video streaming: how Real-time can it be?", Proceedings of the 15th international conference on Multimedia, **(2007)**, pp. 127-136.

[10] R. Kumar, Y. Liu and K. W. Ross, "Stochastic Fluid Theory for P2P Streaming Systems", Proc. of IEEE INFOCOM, **(2007)**, pp. 919-927.

[11] D. Wu, Y. Liu and K. W. Ross, "Queuing Network Models for Multi-Channel P2P Live Streaming Systems", In Proc. of IEEE INFOCOM, **(2009)**, pp. 73-81.

[12] C. Wu, B. Li and S. Zhao, "Multi-Channel Live P2P Streaming: Refocusing on Servers", Proc. of IEEE INFOCOM, **(2008)**.

[13] Y. Zhou, D. Ming Chui and J. C. S. Lui, "A Simple Model for Analyzing P2P Streaming Protocols", IEEE ICNP 2007, **(2007)**, pp. 226-235.

[14] L. Ying, R. Srikant and S. Shakkottai, "The Asymptotic Behavior of Minimum Buffer Size Requirements in large P2P Streaming Networks", Information Theory and Applications Workshop (ITA), **(2010)**, pp. 1-6.

[15] A.-Y. Zhou, H.-Q. Wang and P.-Y. Song, "Graph partitioning method to determine servers placement in CDN", Journal of Harbin Institute of Technology (New Series), vol. 20, no. 2, **(2013)**, pp. 61-68.
[16] S. Yao, L. Zhu and M. Zhu, "The popularity of movies predict system based on data mining technology for CDN", Proceedings - 2010 3rd IEEE International Conference on Computer Science and Information Technology", ICCSIT 2010, vol. 7, **(2010)**, pp. 604-607.
[17] B. Zhu, R. Pi and K. Xu, "A local content routing method for cache group in P2P-CDN", Journal of Information and Computational Science, vol. 9, no. 1, **(2012)**, pp. 103-111.
[18] C. X. Zheng, G. B. Shen and S. P. Li, "Distributed Prefetching Scheme for Random Seek Support in Peer- t o- Peer Streaming Applications", Proceedings of the ACM Workshop on Advances in Peer- to- Peer Multimedia Streaming Hilton, ACM Press, **(2005)**.
[19] H. L. Yu, D. D. Zh Eng, B. Y. Zhao and W. M. Zh Eng, "Understanding User Behavior in Large Scale Video-on-Demand Systems", Proceedings of Eurosys' 06 Belgium, ACM Press, **(2006)**, pp. 333- 3441.
[20] L Chunxi and Chen Changjia, "A Measurement Based Study on the Relations between Users", Watching Behavior and Network Sharing in P2P VoD Systems. Computer Networks, vol. 1, no. 5, **(2010)**, pp. 13-271.
[21] B. Cohen, "Incentives Build Robustness in Bittorrent", Proceeding of 1st Workshop on Economics of Peer-to- Peer Systems (P2PECON. 03), Berkeley, **(2003)**, pp. 1-5.
[22] H. Yan, Z. J. Fu Tom and C. Dah-Ming, "Design and Analysis of a Large- scale P2P-VoD System", SIGCOMM. 08: Proceeding s of the ACM SIGCOMM 2008 Conference on Data Communication, Berkeley: ACM Press, **(2008)**, pp. 375- 388.
[23] R. Kumar, Y. Liu and K. Ross, "Stochastic Fluid Theory for P2P Streaming Systems", INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE, Anchorage, AK, **(2007)** May 6-12.
[24] N. Parvez, C. Williamson and A. Mahanti, "Analysis of Bittorrent- like Protocols for on-Demand Stored Media Streaming", Proceedings of the 2008 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, Annapolis: ACM Press, **(2008)**, pp. 2-61.
[25] L. Abeni, G. Lipari and G. Buttazzo, "Constant Bandwidth vs Proportional Share Resource Allocation", Multimedia Computing and Systems, 1999. IEEE International Conference, vol. 2, **(1999)**, pp. 107-111.

Authors

Yubin Wang, he received his bachelor's degree of engineering in Yanshan University, Qinhuangdao, Hebei. (2000) and master's degree of engineering in Yanshan University (2009), now he is a lecturer in Hebei Normal University of Science & Technology, Qinhuangdao, Hebei. His major fields of study are computer network technology.

Jingyi Bo, her received her bachelor's degree of engineering in Hebei Normal University, Shijiazhuang, Hebei. (2002) and master's degree of engineering in Yanshan University (2008), now she is a lecturer in Hebei Normal University of Science & Technology, Qinhuangdao, Hebei. His major fields of study are computer network technology.

Weili Xu, her received her bachelor's degree of engineering in Hebei University, Tianjin. (2002) and master's degree of engineering in Yanshan University (2005), now she is a lecturer in Hebei Normal University of Science & Technology, Qinhuangdao, Hebei. His major fields of study are Information Retrieval.

