DYNAMIC SERVER REDIRECT FOR MULTIMEDIA SERVICE IN DISTRIBUTED PEER-TO-PEER NETWORK

Ming-Ho Hsiao and Suh-Yin Lee

Department of Computer Science and Information Engineering,
National Chiao Tung University, Hsinchu, Taiwan, R.O.C
{mhhsiao, sylee}@csie.nctu.edu.tw

ABSTRACT

Peer-to-Peer (P2P) multimedia application is expected to be one of the most important services supported by the next generation networks. However, how to balance server load in reducing the peak network traffic for a given P2P network is still a challenge. This paper performs dynamic server redirect (DSR) techniques that blocked requests from high load peer can be redirect to the peer which has high available bandwidth and very low request arrival rate. A novel and approximate model for evaluating the relationship between the request blocking probability and increased cost is proposed. Based on the performance analysis of request blocking, the P2P system can dynamically find efficient ways of using available bandwidth within the acceptable increased cost. Numerical results show that the dynamic server redirect scheme significantly reduces request blocking rate.

1. INTRODUCTION

Peer-to-Peer (P2P) multimedia service is expected to be one of the most important applications supported by the next generation networks [1]. P2P multimedia distributed techniques are design to provide efficient and scalable multimedia service. To date, the P2P systems, including Napster [2], and Gnutella [3] are not scalable [4]. The scalable multimedia service has the ability that the required bandwidths of network and the load of server can be evaluated in a limited range even when the number of clients or movies is extensively increased.

Although recent research in have been significant effort P2P network, a number of challenges intrinsic in P2P multimedia service. Resource management is one of the most important issue in the P2P based multimedia service, especially the bandwidth management The main challenges are: how many bandwidth we need to increase in reducing the peak network traffic and balancing server load for a given P2P network, what is the key issue to effect the bandwidth management, and how to evaluate the increase cost, so that high scalability can be maintained?

The contribution of this paper is to analyze the probabilities of peers request blocking in a model for a given a distributed P2P network. Based on the performance analysis of request blocking, the system can dynamically find efficient ways of using available bandwidth within the acceptable increased cost. Through our numerical results, we show that request service policy based on limited redirect of blocked requests to other peer can achieve better scalability.

The rest of this paper is organized as follows. The overview of the proposed server redirect scheme is described in Section 2. Section 3 presents the model for the requesting blocking probability and increased cost evaluation. Section 4 evaluates the system performance using numerical results. Some conclusions are drawn in Section 5.

2. DYNAMIC SERVER REDIRECT SCHEME

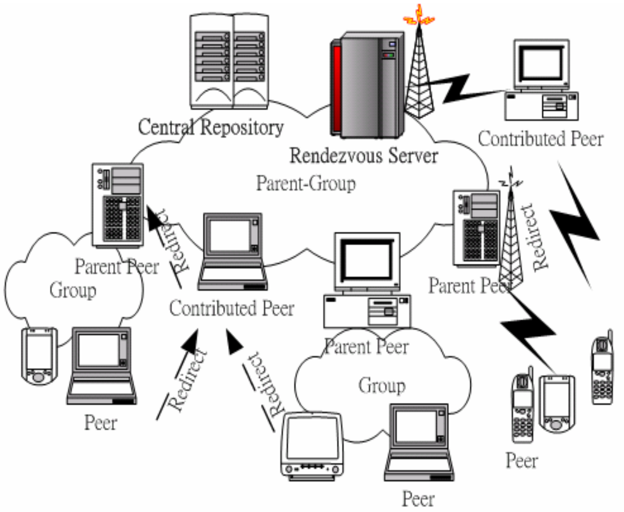

Figure 1. Dynamic peer redirect scheme

Figure 1 shows the basic topology of the proposed server redirect scheme, peers are self-organized in a multi-layer hierarchy of groups. The head of a group is called parent peer which usually has high available bandwidth. The parent peer is responsible for streaming the multimedia content within the peer group and is called contribution peer if it has high available bandwidth and very low request arrival rate. The rendezvous server, parent peers, and contributed peers are organized as a parent group. One of the tasks of the rendezvous server is to collect the network conditions such as peer request arrival rate and capability. In this paper, if the bandwidth of a parent peer is available to service a request, it means the parent peer has a *channel*. The rendezvous server concentrates the bandwidth information of contribution peers and then forms them as the *contributed server*. The contributed server, CServer, logically is overlaid the parent peers in the same parent group.

Channels can be utilized more efficiently by multi-tier peer-to-peer networks. Requests blocking probabilities of this proposed architecture can be proved by server redirect techniques described as follows. Suppose a new request arrival *Rn* to the *i*th parent peer. Assume the *i*th parent peer is blocked and request *Rn* will be redirect and served by the CServer. If channel are available in the *i*th parent peer later, request *Rn* can be transferred to that parent peer again. The request exchange from the CServer to the parent peer is called "*server redirect*". This scheme increases the number of idle channels in the CServer, and more CServer channels can be shared by blocked parent peers. Based on collecting the network condition, the Rendezvous Server can determine how many channels we need to assign for satisfying the system blocking rate requirements. Based on this scheme, multiple sender distributed streaming technique can also be applied [5].

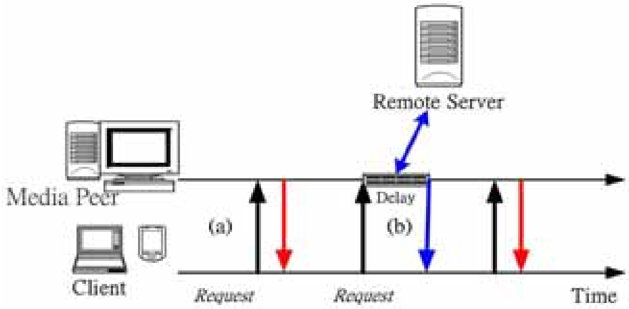

Figure 2. Flow chart of multimedia content delivery

To redirect the request from blocked peer can service more users but it means that we need to increase the total cost. We are interested in what is the relationship between the reduced blocking rate and the total increased cost. We assume a sufficiently stable network, and fault recovery and service interruptions in the process of video transmission are not the key factor [6]. We take consider that the cost is associated with the storage and channel requirements. For the CServer, if the content of the new arrival request from the *i*th parent peer is stored in the CServer, the cost for this content is storage cost (as shown in figure 2(a)). Figure 2(b) shows that if requested content is not stored in the CServer, the CServer needs an allocation of a network channel of duration content length to stream the content from remote server to the client, which causes a delay, and the cost is the transmission cost. The rendezvous server can dynamically assign how many channels form the CServer overlaid the parent peers to reducing blocking probability cccording the limitation of the increased cost.

3. SCHEME ANALYSIS

This section proposes an analytic model to evaluate the request blocking performance of the proposed DSR techniques. Based on the result of performance analysis, the rendezvous server can evaluate the total cost of the storage and channel requirement, and then dynamically redirect the requests overlaid the parent peers to low load server for reducing the blocking probability according the minimum cost requirement.

3.1. Analytic Model of Request Blocking Probability

We assume that the CServer has *B* channels and the *i*th parent peer has b_i channels, where $1 \le i \le N$. We assume that the request arrival (for both outgoing and incoming requests) to the CServer and the *i*th parent peer are a Poisson distribution with rate λ_0 and λ_i req/min, respectively. The arrival rate of CServer is usually is far smaller than it of any parent peer. The request service time of the CServer and the *i*th parent peer with rate $1/u_0$ and $1/u_i$, respectively. Thus the traffic intensity is $\rho_j \equiv \lambda_j / u_j$, where $0 \le j \le N$.

Assume *Pb* be the request blocking probability. Let r_0 and r_i represent the number of outstanding requests generated from the CServer and the *i*th parent peer, respectively. The state of the DSR system is defined by a vector $r = [r_0, r_1, \ldots r_N]$. Suppose the unrestricted DSR system where the CServer has unlimited number of channels. Let $R = [R_0, R_1, \ldots R_N]$ be the random vector representing the state of the unrestricted DSR system, where R_0 and R_i represent the number of outstanding requests generated from the CServer and the *i*th parent peer, respectively. According to the $M/M/\infty$ model [7], the equilibrium probability for $p(R_i = r_j)$ is defined as

$$p(R_j = r_j) = \frac{(\lambda_j / u_j)^{r_j}}{r_j!} e^{-\lambda_j / u_j}, 0 \le j \le N \qquad (1)$$

Let Rcs_0 and Rcs_i represent the number of outstand requests from the CServer itself and the *i*th parent peer such that they are served by the CServer, where $Rcs_i = \max(R_i - r_i, 0)$. Thus we have the probability for Rcs_j is determined by

$$p(Rcs_j = k) = \begin{cases} p(R_j = k), if\ j = 0 \\ \sum_{g=0}^{b_i} p(R_j = g), if\ j \ne 0, k = 0 \\ p(R_j = b_i + k), if\ j \ne 0, k > 0. \end{cases} \qquad (2)$$

Assume the channel capability $C_j(R) = \sum_{g=0}^{j} Rcs_g$, and from (2), we have

$$p(C_j(R) = k) = \begin{cases} p(Rcs_j = k), if\ j = 0 \\ \sum_{g=0}^{k} [p(C_{j-1}(R) = g) p(Rcs_j = k - g)], \\ \qquad if\ j > 0 \end{cases} . \qquad (3)$$

If $j = N$, $C_N(R) = \sum_{j=0}^{N} Rcs_j$ is the total number of outstanding requests in the CServer. For $0 \le j \le x \le N$ and x>0, let $C_{j,x}(R) = \sum_{g=0, g \ne x}^{j} Rcs_g$. $C_{N,x}(R)$ is the number of outstanding request in the CServer excluding the request generated from *x*-th parent peer. Suppose event F_i = {r| $\sum_{g=0, g \ne x}^{j} Rcs_g = B$ } that we obtain

$$p(F_i) = p(C_{N,x}(R) = B) \qquad (4)$$

To compute *Pb,* let event L = {r| $0 \le \sum_{j=0}^{N} Rcs_j \le B$ }. From (2), the probability *p(L)* is defined as

$$p(L) = \sum_{k=0}^{B} p(C_N(R) = k) \qquad (5)$$

Let event W_i represents the number R_i of outstanding requests in the *i*th parent peer is less than b_i , which means the *i*th server peer is not blocked. Then from (1), we have

$$p(W_i) = \sum_{k=0}^{b_i - 1} \frac{(\lambda_i / u_i)^k}{k!} e^{-\lambda_j / u_j} \qquad (6)$$

Suppose event Q = {r| $\sum_{j=0}^{N} Rcs_j = B$ } From (2), the probability *p(Q)* is defined as

$$p(Q) = p(C_N(R) = B) \qquad (7)$$

In [8], Hung *et al* proved that the probability measure in restricted can be expressed by the probability measure in the unrestricted system. We have

$$p_L(R = r) = p(R = r) / p(L) \qquad (8)$$

In proposed scheme, a request attempt from the *i*th server peer is blocked if the *i*th parent peer is blocked and the CServer is blocked. Thus the probability of the request attempt from the *i*th parnet peer is defined as

$$Pb_i = p_L(Q \cap \overline{Wi}) = (p(Q) - p(W_i) p(F_i)) / p(L) \qquad (9)$$

From (9), we can compute the total request blocking probability *Pb* which is derived as

$$Pb = (\sum_{i=1}^{N} \lambda_i Pb_i + \lambda_0 p(Q)) / \sum_{j=0}^{N} \lambda_j \qquad (10)$$

3.2. Cost Evaluation

We assume all multimedia contents can be found in the P2P system. After choosing the *B* value, we can compute the blocking probability of each parent peer. Then the successful increase of the request arrival rate for the CServer is derived as

$$\lambda_t = (\sum_{i=1}^{N} \lambda_i (1 - p(W_i)))(1 - p(Q)) \qquad (11)$$

Assume that there are *M* contents, the size and the popularity of multimedia content cn_y is TL_y and z_y , respectively where y=1,…*M.* According the storage capacity, *SC*, the CServer stores the most popularity content in local side. Requests for the content cn_y arrive at the server with rate $(z_y)\lambda_t$. If content cn_y is stored in the CServer, the cost for this content is $\delta \times TL_y$, where δ is the storage cost of per unit. If requested content is not stored in the CServer, the CServer needs an allocation of a network channel of duration minutes to stream the content from remote server to the client, which causes a delay Dx minutes. By Little's formula, the average number of streams required given the content request rate $(z_y)\lambda_t$, and hence the normalized cost (because there is no storage cost) is $S = \beta \times (TL_y / (1 / z_y \lambda_t + D_x))$, where β is the stream transmission cost per unit. Thus we can compute the total increased cost $\hat{C}$ is derived as

$$\hat{C} = \sum_{y=1}^{M} [(\delta TL_y A_y) + (\beta (\frac{TL_y}{\frac{1}{z_y \lambda_t} + D_x})(1 - A_y))] \qquad (12)$$

where $A_y = \begin{cases} 1, & if\ \sum_{g=0}^{y} TL_g \le SC \\ 0, & else \end{cases}$, (13)

4. ILLUSTRATIV EXAMPLE AND COMPARISION

To evaluate the performance of the proposed serve redirect scheme for multimedia service, in this section we illustrate the result of our proposed architecture. Based on the analytic model developed in the previous section, we compared the traffic intensity effect in terms of the blocking probability. We suppose there are total 500 videos in our system and the size of each content is 300Mb. Assume the content popularity follows a Zipf distribution $z_y = y^{-(1-\alpha)} / V$, where $V = \sum_{u=1}^{M} y^{-(1-\alpha)}$ .To simply discuss, we consider the homogeneous DSR system where $bi = bx$ and $\lambda_i = \lambda_x$ for $1 \le i, x \le N$.

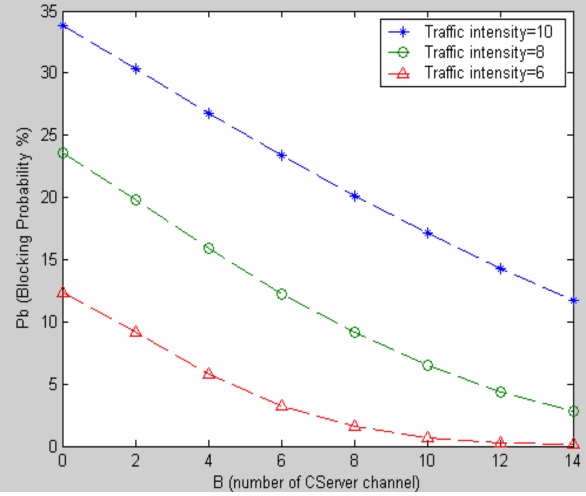

Figure 3. Request blocking probability

Figure 3 shows the effect of the CServer channel number *B*. We assume the traffic intensity of CServer is 2. In Figure 3, the request service times are exponentially distributed with mean=30 Min., N=8, the channel of parent peers is 10 .Usually the high traffic intensity will cause the high request blocking rate of the P2P system. To increase *B* significantly improves *Pb.* If the high traffic intensity is low, when *B* is large, the CServer channels are no longer bottleneck for different request arrival rate. Through these results, we show that even with only limit redirection (channel) it is possible to improve the performance of the P2P system.

The traffic intensity of the CServer and parent peers is 2 and 8, respectively. .Figure 4 shows the increased cost of the CServer, when the rendezvous server dynamically assigns channel number *B* overlaid the parent peers. Cost increase as *B* increase, that is because more request arrival to the CSever, it spent more transmission cost from remote server to the requested peer.

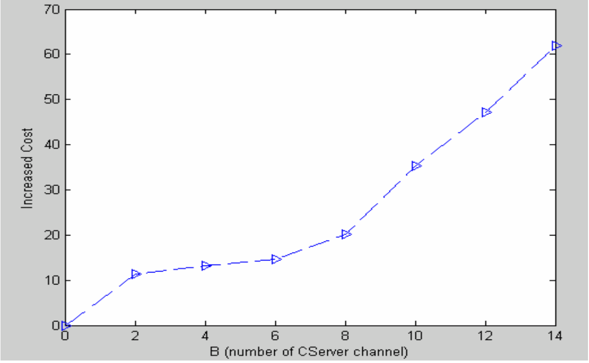

Figure 4. Relationship between number of CServer channels and increased cost

5. CONCLUSION

In this paper, the blocking probability of P2P is further improved by taking the server redirect techniques. Requests form blocked peer are redirect to CServer which has high available channel and low request arrival rate, as a result, reducing the peak network traffic and server load. Numerical results show that the dynamic server redirect scheme significantly reduces request blocking rate. In future work, we will integrate the redirect technique and the replication strategies to increase the scalability of P2P systems.

6. REFERENCES

[1] Mundur, P., Simon, R., and Sood, A.K., "End-to-end analysis of distributed video-on-demand system," *IEEE Trans. on Multimedia,* 129-141, Feb., 2004.

[2] [online]. Available:http://www.napster.com/

[3] [online]. Available:http://www.gnutella.com/

[4] Zhe Xiang, Qian Zhang, Wenwu Zhu, Zhensheng Zhang, and Ya-Qin Zhang, "Peer-to-Peer Based multimedia Distributed Service," *IEEE Trans. on Multimedia,* 343-354, Apr., 2004.

[5] Thinh Nguyen and Acideh Zakhor, "Multiple Sender Distributed Video Streaming," *IEEE Trans. on Multimedia,* 315-326, Apr., 2004.

[6] Chan,S.H.G. and Tobagi, F., "Distributed servers architecture for networked video services," *IEEE Trans. on Networking,* 125-136, Apr., 2004.

[7] Kleinrock, L. *Queueing Systems; Volume I: Theory.* Wiley, 1975.

[8] Hung, H.N., Lin, Y.-B., Peng, N.F., and Tsai, H.M, "Repacking on Demand for Two-tier Wireless Local Loop," *IEEE Trans. on Wireless Communications,* 3(3):745-757, 2004.