A Cooperative Hybrid Caching Strategy for P2P Mobile Network

Mo Zhou, Bo Ji, Kunpeng Han, Hongsheng Xi

Department of Automation

University of Science and Technology of China

Hefei, Anhui, 230027, China

{zhoumo,jibo,icymaple}@mail.ustc.edu.cn, xihs@ustc.edu.cn

Abstract—**Recently mobile network technologies develop quickly. To meet the increasing demand of wireless users, many multimedia proxies have been deployed over wireless networks. The caching nodes constitute a wireless caching system with an architecture of P2P and provide better service to mobile users. In this paper, we formulate the caching system to optimize the consumption of network bandwidth and guarantee the response time of mobile users. Two strategies: single greedy caching strategy and cooperative hybrid caching strategy are proposed to achieve this goal. Single greedy caching aims to reduce bandwidth consumption from the standpoint of each caching node, while cooperative hybrid caching allows sharing and coordination of multiple nodes, taking both bandwidth consumption and popularity into account. Simulation results show that cooperative hybrid caching outperforms single greedy caching in both bandwidth consumption and delay time.**

Keywords-mobile network; multimedia; caching strategy

I. INTRODUCTION

The recent development in wireless access technologies and the increasing capacity of mobile devices have enabled users to access information more easily. To meet the increasing demand of wireless users, many multimedia applications such as VOD have been deployed over wireless network. However, considering the high bandwidth requirements of multimedia applications and mobility of mobile users, it is more challenging to satisfy the quality of service (QoS) requirements of multi-media applications. An efficient solution is to cache frequently accessed media data at locations close to mobile users to improve the performance of mobile applications as caching popular video data can alleviate network congestion, reduce latency on the client side and reduce load on the multi-media servers.

In our research, the infrastructure-based wireless network which is still dominant right now is studied. The mobile system is P2P organized, and our goal is to design a strategy to optimize the consumption of bandwidth network while guaranteeing the response time of mobile users. Based on the theory of pure P2P systems where individual caching nodes have to make decisions independently, we propose two distributed strategies : single greedy caching strategy and cooperative hybrid caching strategy. The evaluation results show that cooperative hybrid caching performs better under the metrics of both bandwidth consumption and latency.

The rest of this paper is organized as follows: In Section II, we present the architecture of multimedia cache proxies across wireless network. Section III formulates the problem we
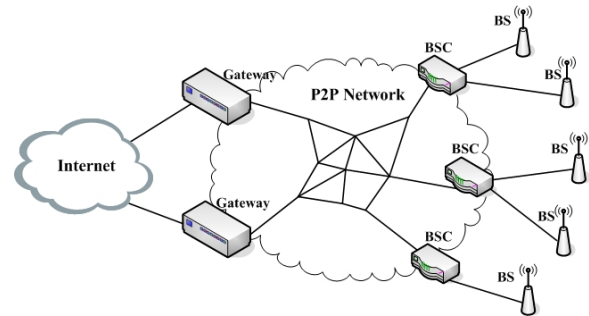

Figure 1.   Network Scenario

study. In Section IV we propose the caching strategy. Section V gives simulation results to demonstrate the performance of the caching systems. Finally, in Section VI, we conclude our study.

A. Related Work

As for wired caching, there has been sophisticated caching and searching schemes [1,2], yet these existing schemes mainly aim to increase bit hit ratio. J. Kangasharju et al. [3] discuss heuristics of all servers cooperating in making decisions and provide insight into peer-to-peer content distribution. Y. Chen et al. [4] present a replica placement protocol that minimizes the number of replicas in the dissemination tree while meeting the QoS constraints. W.M. Aio et al. [5] propose a wireless dynamic content distribution network framework and X. Fu et al. [6] address the QoS-aware replica placement problem in a tree based mobile grid environments. J. Dai et al. [7] propose a collaborative caching mechanism that aims at maximizing the social welfare in the context of Vickrey-Clarke-Groves auctions. The existing schemes mainly assume that the cache servers are configured into hierarchical or tree topology. In contrast, we will study the more practical scenario based on P2P caching systems.

II. NETWORK SCENARIO

Figure 1 illustrates the scenario of multimedia caching system in our research. The system is composed of base stations(BS), base station controllers(BSC), multimedia servers, and gateways. The multimedia servers store all the video objects provided by the system to support streaming service to mobile users. The caching proxies installed in base stations [8] are utilized to serve mobile the requests from users with their limited bandwidth and storage capacity. The base stations

attached to the multimedia servers are interconnected with each other, forming a highly connected network. Considering the fact that each base station can only serve the clients within its communication range, video objects have to be transmitted between base stations.In general, our systems allow caching proxies to fetch video contents either from remote servers or from peer nodes.

III. PROBLEM FORMULATION

As analyzed above, our goal is to minimize the traffic among the network as well as reduce response time for mobile users of the caching system. In order to achieve this goal we establish the mathematical model for the system. In the problem formulation, the mobile caching system consists of a original streaming server S which provides K video objects and N interconnected caching nodes, each with capacity s-pace S_i, where $i \in \{1, 2..., N\}$. The bandwidth and storage requirement of each video object is denoted by b_i and s_i, where $i \in \{1, 2..., K\}$.

Several other notations are: $p_{i,j}$ indicates the probability that a client at server i requests video object j; $d(i, k)$ indicates the shortest distance for caching node i to locate the copy of video object j, if j is locally cached, the value equals to zero; $x_{i,j}$ is a selection parameter that equals to 1 if caching node i holds a copy of video object j, otherwise equals to 0; the problem is then formulated as:

$$min \sum_{i=1}^{N} \sum_{k=1}^{K} p_{i,k} d(i, k) b_k \qquad (1)$$

Subject to

$$\sum_{j=1}^{K} x_{i,j} s_j \leq S_i, i = 1, 2...N \qquad (2)$$

The objective function minimizes the traffic over the network. In our model, the variable $d(i, k)$ guarantees the QoS of response time to some degree by locating the nearest copy of video objects. The consumption of bandwidth that locates video information is negligible compared with that of transferring video objects, thus not be included in the cost of bandwidth in our research. The value of $x_{i,j}$ will describe the placement of video objects in each caching node.

However, the problem above is NP hard, thus it is computationally complicated to get the optimal solution. Moreover the system we study is a pure P2P system without controlling nodes that collect information and assign video objects to the whole system according to the value of $x_{i,j}$. In P2P systems where any centralized algorithm is not suitable, each node has to make decisions based on its own information. Considering all these factors, a distributed strategy based on the individual nodes is proposed.

IV. CACHING STRATEGIES

To minimize the objective function (1), there are several factors needed to be taken into account: popularity, distance and video size. All these factors can make impact on the performance of the caching system, therefore these factors will be considered in our strategy.

A. Single Greedy Caching

As for each individual caching node, statistics about the popularity of the video objects requested by clients can be collected, and we assume the value of $p_{i,j}$ is known in advance in our research. Due to the limited storage capacity of each caching node, only parts of the objects can be stored. To minimize (1), it is intuitive that each caching node should cache video objects with relatively higher values of $p_{i,k} d(i, k) b_k$. Thus, when video objects with smaller value of $p_{i,k} d(i, k) b_k$ are requested, they will be transferred from other caching nodes generating bandwidth consumption.

This solution will involve cache replacement procedure. Each caching node should maintain the $p_{i,k} d(i, k) b_k$ list of its cached video objects. When a request for an un-cached video object comes, if its weight is higher than the smallest one among all cached video objects, cache replacement process will happen by evicting the smallest weight video object. Through the above procedure, each caching node will cache video objects with higher value of $p_{i,k} d(i, k) b_k$ at each moment.

The solution we discuss above may not be optimal because the accumulation of all single optimal solutions of each caching node may not result in wholly optimal solution. Moreover, if the replacement procedure happens frequently, the value of $p_{i,k} d(i, k) b_k$ cached in each caching node is easily invalid and the updating cost in real time is too large.

B. Cooperative Hybrid Caching

Among all the factors affecting the performance of the caching system, popularity plays an important role. An intuitive strategy related to popularity is that each caching node caches its most popular video objects locally. While single greedy caching may not lead to optimal results as analyzed above, our goal is to study if caching nodes can save bandwidth by cooperating with nearby caching nodes. To explore the impact of cooperative hybrid caching based on popularity, we analyze the following scenario:

A and B are caching nodes with similar popularity in the system and the distance between them is $d(A, B)$ as shown in Figure 2. They both can cache N video objects, where $N \ll K$ and K denotes the total number of objects in the caching system. And we assume that small video number has high popularity, for example video 1 has the highest popularity.

The figure describes two strategies based on popularity. The upper part shows that each caching node acts independently. They cache the most locally popular video objects and the objects cached are similar due to similar interests of the two caching nodes. To optimize the system performance, the following strategy as the below part showed is proposed. The two caching nodes crossly store the most popular video objects, node A caches N video objects including $\{1, 3..., 2N - 1\}$, node B caches $\{2, 4..., 2N\}$. We propose this strategy from the perspective that node A fetches video
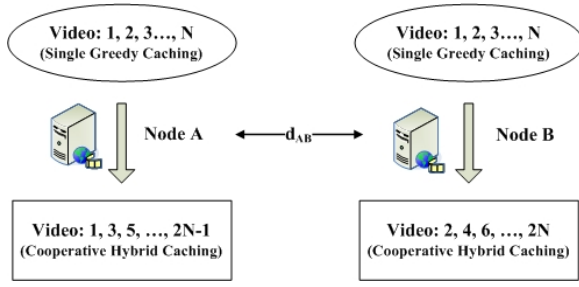
Figure 2.   Video Storage Strategies

Figure 3.   Caching Nodes Network

objects $\{N+1, N+3..., 2N-1\}$ from nearby node B rather than other distant nodes as described in strategy above will save bandwidth, and the incremental bandwidth cost caused by fetching video objects $\{2, 4..., N\}$ from node B will be small if the distance between A and B is close enough. We will compare the two strategies in the following discussion. The notation $d_{A,j}$ denotes the distance for A to fetch j, and the other notations have the same definitions as above. The bandwidth difference between the two strategies is:

$$diff = \sum_{j \epsilon \{N+1,...2N\}} (p_{A,j}d_{A,j}b_j + p_{B,j}d_{B,j}b_j)$$
$$- \sum_{j \epsilon \{1,2,...2N\}} (p_{A,j} + p_{B,j})d(A,B)b_j \qquad (3)$$

So from the above expression we can see that if $diff$ is greater than 0, the bandwidth can be saved, and $diff > 0$ equals to the following expression:

$$d(A,B) < \frac{\sum_{j \epsilon \{N+1,...2N\}} p_{A,j}d_{A,j}b_j + p_{B,j}d_{B,j}b_j}{\sum_{j \epsilon \{(p_{A,j}+p_{B,j})b_j\}}} \qquad (4)$$

We can conclude that similar popularity of two nodes is the preconditions of cooperating and on this basis if the the distance satisfies (4), the bandwidth consumption can be saved.

C. How to Judge The Similarity of Two Nodes

In our system, the video sets in each caching node can serve as an indication of the user's interest, as the composition of the caching node is not decided in advance by the popularity of video contents, but rather triggered by user's requests. To describe the similarity between different video sets, the following definition is given:

$$s_{i,j} = \sum_k (p_{i,k} - p_{j,k})^2 \qquad (5)$$

If node i does not cache video object k, then $p_{i,k} = 0$, then if $s_{i,j}$ is less than a certain threshold, node i and j can be considered as similar nodes. As for the real mobile system, nearby neighbor nodes always have the similar interests considering the fact that mobile users always move between neighbor nodes.
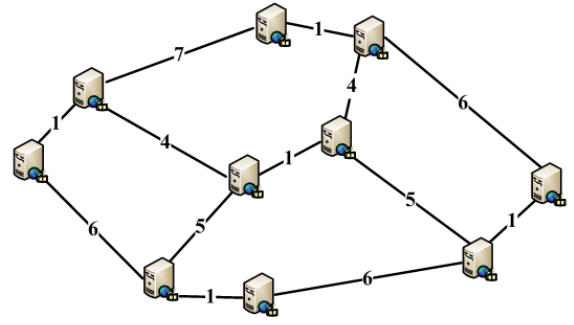
D. Cooperative Hybrid Caching

According to the above analysis, the cooperative strategy combining weight and popularity is proposed as follows: 1. Partition the system into several sets. Each set consists of the nodes that have similar interests and their distance satisfies (4). 2. Crossly store popular video objects on the node of each set. Then similar replacement procedure like strategy A is implemented during client request. 3. Periodically refresh the system. Replace the non-original cached video object of higher weight with the original popular video objects of smaller weight until no replacement will happen.

V. EVALUATION

In this section, we evaluate the performance of the cooperative hybrid caching scheme and compare it with single greedy caching scheme. The virtual network topology and the distance between caching nodes are shown in figure 3. We assume that the two nodes with the shortest distance have similar video popularity that also conforms to the common situation in wireless network. The client query generate time follows exponential distribution with mean value 4s. The client request distribution is shown as figure 4. The size of video objects is randomly assigned and the popularity of video objects follow a Zipf-like distribution. The access probability of the ith ($1 \leq i \leq K$) video object is showed as follows:

$$P_i = \frac{1}{i^\alpha \sum_{n=1}^{K} \frac{1}{k^\alpha}} \qquad (6)$$

where $0 \leq \alpha \leq 1$ and larger α represents more skewed access distribution.

We first plot the overall bandwidth consumption under the two strategies in figure 5. The x-axis shows the storage capacity of the caching node as the percentage of the total size of the video objects. From the picture we can see that, the performance of cooperative hybrid caching outperforms single greedy caching in bandwidth consumption. As the storage capacity increases, the bandwidth consumption of both schemes decreases. The bandwidth consumption of cooperative hybrid caching decreases relatively slowly after the storage exceeds 35%, as the cooperative nodes have already cached most parts of the popular video objects, redundant storage capacity utilized to cache less popular video will contribute
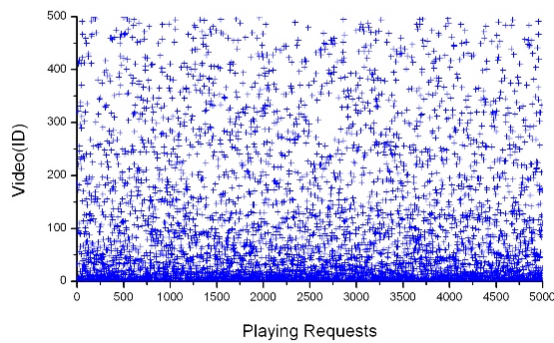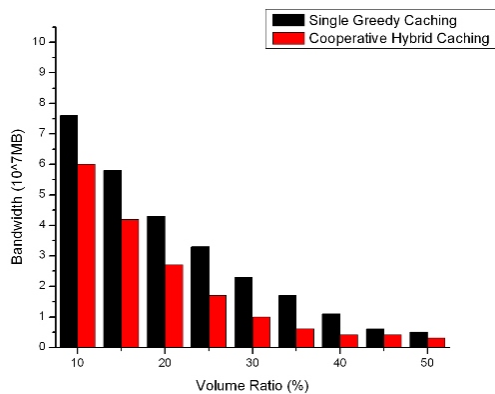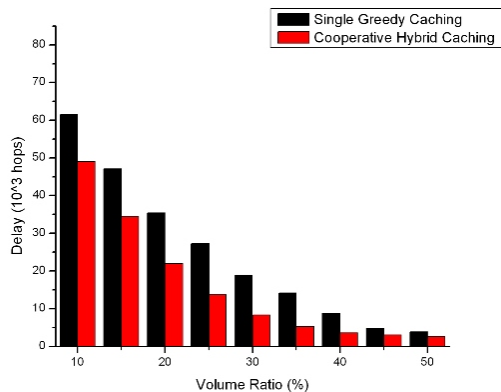
Figure 4.   Request Distribution

Figure 7.   Single Greedy Caching

Figure 5.   Total Bandwidth When α=0.8
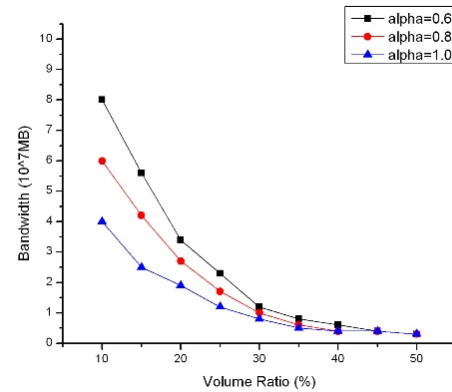
Figure 8.   Cooperative Hybrid Caching

Figure 6.   Total Hops When α=0.8

as plotted by figure 6. Under cooperative hybrid caching, crossly caching ensures caching nodes to fetch most un-cached video objects from cooperative nodes, thus delay can be saved. While the single caching node will fetch un-locally cached video objects randomly.

From figure 7 and figure 8, we can see the influence of the parameter α in Zipf's law to our proposed caching scheme. For certain storage capacity, if α is larger or more requests are aggregated to most popular videos, the more bandwidth consumption can be saved for both two schemes. But the influence to cooperative hybrid caching is more significant and the strategy can deal with dynamic situation better.

VI. CONCLUSION

In this paper we propose two caching strategies, single greedy caching strategy and cooperative hybrid caching strategy to optimize the caching system, and cooperative hybrid caching strategy outperforms single greedy caching in both bandwidth consumption and delay. Because we just assume the popularity distribution of video objects in advance, we will focus on introducing predicting algorithm for video object popularity in future work.

less to decrease in bandwidth consumption. While bandwidth consumption under single greedy caching decreases almost linearly, since the storage capacity is a mainly important factor affecting the performance of single greedy caching.

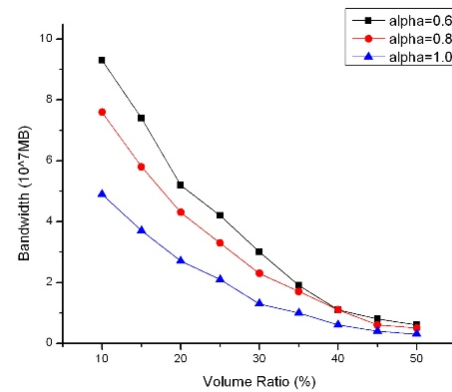Similar performance can be observed in the achieved delay

REFERENCES

[1] A. Rousskov and D. Wessels, "Cache digests", Computer Networks and ISDN Systems, June 1998.

[2] L. Fan, P. Cao, J. Almeida and A. Broder, "Summary cache: a scalable wide-area web cache sharing protocol", IEEE/ACM Transactions on Networking, 2000.

[3] J. Kangasharju, J. Roberts, and K. Ross, "Object replication strategies in content distribution networks", Web Caching and Content Distribution Workshop, Boston, MA, June 2001.

[4] Y. Chen, R. Katz, and J. Kubiatowicz, "Dynamic replica placement for scalable content delivery", International Workshop on Peer-to-Peer Systems, Cambridge, MA, March 2002.

[5] W. M. Aioffi, G. R. Mateus, J. M. Almeida, and R. C. Melo, "Dynamic content placement for mobile content distribution networks", WCW 2004, Lecture Notes in Computer Science(LNCS), 2004.

[6] X. Fu, R. Wang, Y. Wang, and S. Deng, "A replica placement algorithm in mobile grid environments", International Conference on Embedded Software and Systems, IEEE Computer Society, 2009.

[7] J. Dai, F. M. Liu, B. Li, B. C. Li and J. C. Liu, "Collaborative caching in wireless video streaming through resource auctions", IEEE Journal on Selected Areas in Communications, February 2012.

[8] Z. Jiang, L. F. Chang, B. J. Kim and K. K Leung, "Incorporating proxy services into wide area cellular IP networks", Proceedings of the IEEE WCNC, 2000.