This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings.

# Collaborative Caching for Video Streaming among Selfish Wireless Service Providers

Jie Dai†, Bo Li†, Fangming Liu‡, Baochun Li*, Jiangchuan Liu§
†*Hong Kong University of Science and Technology*, **University of Toronto*
‡*Huazhong University of Science and Technology*, §*Simon Fraser University*

***Abstract*—Video streaming is now at the fingertips of mobile users with recent advances in wireless communications and mobile networking. Caching has been widely deployed by wireless service providers (WSPs) to facilitate video content dissemination. Yet, capacity provisioning of cache servers is challenging given dynamic user demands and limited wireless bandwidth resources available. With increased densities of wireless service deployment, it is common that mobile users are now covered by more than one WSP within a geographical region. This brings both challenges and opportunities towards a collaborative caching paradigm among cache servers that are deployed by different WSPs. This paper explores the benefits of collaborative caching for wireless video streaming services, addressing challenges related to both incentives and truthfulness of selfish WSPs. We propose a collaborative mechanism that aims to maximize the social welfare in the context of Vickrey-Clarke-Groves (VCG) auctions, which encourages cache servers to spontaneously cooperate for trading their resources in a self-enforcing manner. Results from simulations demonstrate significant performance improvements with respect to video streaming quality.**

## I. Introduction

The popularity of video streaming services has substantially changed the landscape of wireless multimedia applications, with an increasing number of content providers (*e.g.,* YouTube and Netflix) offering streaming content to mobile users. Meanwhile, a diverse range of wireless access technologies, including Wi-Fi, HSPA+ cellular access, and recent advances in 4G deployment such as LTE [1], have made broadband wireless connections (*e.g.,* at downlink peak rates of 100 Mbps in LTE) a near-term reality. Proxy caching has been commonly utilized by wireless service providers (WSPs) to improve the streaming quality [2]: cache servers are typically deployed at Mobile Switching Centers (MSCs), so as to locate video content closer to end users, thus making more efficient use of bandwidth. Nevertheless, the inherent dynamics of mobile users have made resource provisioning at cache servers a significant challenge. Considering the fact that mobile users are often covered by multiple WSPs in the same geographical region, all of which deliver video content to their own clients, a more efficient way to provision bandwidth is to incentivize these autonomous and selfish WSPs to collaborate with one another. Such a collaboration improves the degree of multiplexing available resources in WSPs, as well as the general availability of video content in proxy servers.

†The research was support in part by a grant from Huawei Technologies Co. Ltd. under the contract HUAW18-15L0181011/PN, by a grant from HKUST under the contract RPC11EG29. Fangming Liu is with the Services Computing Technology and System Lab, Cluster and Grid Computing Lab in the School of Computer Science and Technology, Huazhong University of Science and Technology.

Prior works in wireless video caching have primarily focused on *independent* caching, where no collaboration among WSPs takes place. Zhang *et al.* [3] presented a cost-based cache replacement algorithm for a single cache server and a server selection algorithm for multiple cache servers in wireless multimedia proxy caching. Xie *et al.* [4] proposed a cache-assisted wireless Video-on-Demand (VoD) system in which relaying stations stored a portion of ongoing video streams. Tan *et al.* [5] introduced a smart caching design that duplicated video contents in access points, to effectively reduce the redundant traffic in the WLAN while improving the response delay of video streaming. Collaborative caching has been explored in various video streaming systems. Borst *et al.* [6] developed distributed caching algorithms that maximized the traffic volume served from a cache, while minimizing the bandwidth costs in a VoD system.

Our work addresses collaborative caching that encourages cache servers to actively deliver content across different domains, while both fairness and truthfulness are guaranteed. We argue such a collaboration can maximize the social welfare by improving the quality of video streaming to all end users. The main challenge towards implementing such a collaboration, however, is the inherent *selfish* nature of the autonomous WSPs, often belonging to different for-profit corporations. Specifically, we focus on engineering the *incentives* to promote the cache servers owned by different WSPs to *truthfully* cooperate with one another.

Inspired by the design of Vickrey-Clarke-Groves (VCG) auctions in game theory [7], we design a new collaborative caching mechanism that can be employed by co-locating WSPs. VCG auctions are known as non-cooperative games, in which any decision of cooperation is “self-enforcing.” We treat server bandwidth as commodities in these auctions, with different valuations based on dynamics of streaming systems. The virtual payments associated with the valuation of bandwidth ensure that the contribution of cache servers is acknowledged by other participants. Truthfulness is also guaranteed, which encourages the cache servers to faithfully reveal their true valuation when “bids” are submitted. Our simulation results show that the performance of streaming systems can be substantially improved by maximizing the social welfare in these auctions, in which bandwidth units are used to serve more valuable demands.

The remainder of the paper is organized as follows. In

Sec. II, we formulate the model of resource auctions in collaborative caching, and present the design of VCG-based bandwidth trading mechanisms. In Sec. III, we present our simulation studies to evaluate the effectiveness of collaborative caching. We conclude the paper in Sec. IV.

## II. Resource Auctions in Collaborative Caching

In wireless video streaming, cache servers deployed by WSPs at their own MSCs utilize their constrained bandwidth and storage resources to serve video content requests from mobile users, which results in better streaming quality for a group of video *channels*. In *independent* caching used by each individual WSP, such resources are always allocated to local users within its own domain. Our proposed collaborative caching mechanism further explores the cooperation among cache servers across different administrative domains. Intuitively, such a collaboration helps alleviate the negative impact from a sudden surge of requests to popular channels. As such, the dynamics of the overall video streaming system can be smoothly dispersed across multiple WSPs.

### A. Model Formulation

We consider a WSP set $\mathcal{I} = \{1, 2, ..., N_i\}$ with a total of $N_i$ WSPs that provide streaming services in the same geographical region. Cache servers controlled by a single WSP are integrated into one cache server for the sake of simplicity. The notation $i \in \mathcal{I}$ is interchangeably used to represent WSP $i$ or the cache server in WSP $i$. The relationship among cache servers is represented by $G = (\mathcal{I}, E)$, where $E = \{e_{ij}\}_{|\mathcal{I}|\times|\mathcal{I}|}$. The notation $e_{ij}$ reflects the degradation of benefits from bandwidth provisioning among different domains, and $0 < e_{ij} \leq 1$. Let the set of channels in the video streaming system be denoted by $\mathcal{K} = \{1, 2, ..., N_k\}$, where $N_k$ denotes the total number of channels. The bandwidth and storage capacity at server $i$ are denoted by $W_i$ and $S_i$.

The efficient utilization of bandwidth resources across cache servers is critical when dealing with dynamics of the system. The VCG auction from game theory serves as a candidate that fits the needs of designing such a mechanism. First, buyers have to *pay* after each successful trade, which serves as *incentives* to contribute. In addition, the payment method in the VCG auction ensures *truthfulness*, in which buyers are willing to truthfully reveal their bidding information in the auction. We now present a basic model inspired by the VCG auction theory, in order to achieve efficient allocations that benefit users in different WSPs.

When an *auction* is held in the context of collaborative caching, participants in the auction are cache servers, with bandwidth resources being commodities. Each cache server can act as a seller or a bidder that trades resources in multiple rounds. When acting as a seller, a cache server $i$ provides bandwidth $W_i^r$ to remote bidders in a certain round. The bandwidth unit $w$ is defined as the basic trading unit. Therefore, $W_i^r$ has been divided into $W_i^r/w$ homogeneous bandwidth units while $W_i^r/w$ is assumed to be integral. The assigned bandwidth can be implemented as an external source residing in a remote MSC, which adaptively adjusts its uplink capacity. We then have $W_i^r = W_i - \sum_{k\in\mathcal{K}} W_i^k$, in which $W_i^k$ represents the local bandwidth assignment of WSP $i$ to channel $k \in \mathcal{K}$.

For the buyer cache server $i$, the number of bandwidth units requested for channel $k$ from server $j$ is denoted as $n_{ij}^k$. The buyer $i$ submits bids that include the value of $n_{ij}^k$ and the corresponding valuation to seller $j$. After collecting all bids, the winner determination process decides the bandwidth allocation in the next round based on the valuation of channels. Consequently, buyers have to pay by *virtual payments* if the streaming contents are successfully delivered. Such payments can later be used for bandwidth resource demands to other domains, which serve as an incentive for contributions. The truthfulness will also be guaranteed by the payment method with which no bids deviating from the true valuation will be benefited in the VCG auction.

### B. VCG-based Bandwidth Trading

Channels supported by server $i$ have different requirements on the streaming quality. We thus introduce the concept of a *virtual bidder*, so that the demand on different channels can be treated differently based on the channel status. Each virtual bidder $b_i^k$ submits bids to request $n_{ij}^k$ units on behalf of users on channel $k$. $b_i^k$ has a privately known valuation function: $v_i^k : \{0, 1, ..., \min(n_{ij}^k, \frac{W_j^r}{w})\} \rightarrow \Re, \forall i \in \mathcal{I}, k \in \mathcal{K}$, where $v_i^k(n)$ denotes the valuation on the benefit of receiving $n$ units. If we restrict $n_{ij}^k \leq \frac{W_j^r}{w}$, virtual bidder $b_i^k$ then submits bids $\mathbf{b}_i^k = \{(0, 0), (1, v_i^k(1)), (2, v_i^k(2)), ..., (n_i^k, v_i^k(n_{ij}^k))\}$ to seller $j$, since submitting one's true valuation is the dominant strategy in VCG auctions. The request for channel $k$ will only be delivered to a single server within one round.

When we consider a video streaming system, the streaming quality experienced by online users is of great importance to evaluate the system efficiency. Therefore, it can indeed be used to reveal the benefit of receiving a certain amount of bandwidth units. We also consider important observations inside real-world peer-assisted streaming systems through extensive measurements [8] [9]. According to statistical results in [8], there exists a positive correlation between the per-user server bandwidth provisioned and the proportion of users that experience a smooth playback. In the context of independent caching, the streaming quality of channel $k$ in domain $i$ can therefore be defined as:

$$q_i^k = \gamma(\frac{W_i^k}{r_k \cdot x_i^k})^\alpha \tag{1}$$

In this definition, the channel with streaming rate $r_k$ has the number of concurrent users $x_i^k$. $\gamma$ is an adjustable scaling parameter. The value of $\alpha$ satisfies the constraint $0 < \alpha < 1$ which is observed in [9]. The constraint indicates that the streaming quality could be improved when the provisioned bandwidth increases, but with a decreasing marginal gain.

*Definition 1: Valuation function*. Based on formulation (1), the valuation function that reflects the streaming quality improvement of channel $k$ by receiving $n_{ij}^k$ units is defined as:

$$v_i^k(n_{ij}^k) = \gamma(x_i^k)^{1-\alpha}((\frac{W_i^k + e_{ij} \cdot w \cdot n_{ij}^k}{r_k})^\alpha - (\frac{W_i^k}{r_k})^\alpha) \quad (2)$$

*Definition 2: Winner determination*. Cache server $j$ needs to determine the allocation of $W_j^r$ after receiving all bids. Winner determination is considered to be an *efficient allocation* in a VCG auction if it maximizes the social welfare as:

$$\begin{aligned} \text{Maximize} \quad & \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} v_i^k(\tilde{n}_{ij}^k) \\ \text{Subject to:} \quad & \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} \tilde{n}_{ij}^k \le \frac{W_j^r}{w} \\ & 0 \le \tilde{n}_{ij}^k \le n_{ij}^k \quad \forall i \in \mathcal{I}, k \in \mathcal{K} \end{aligned} \quad (3)$$

In this formulation, $\tilde{n}_{ij}^k$ is the optimization variable that determines the bandwidth allocation. Since received bids only include discrete values for each possible number of acknowledged units, this problem becomes an integer programming problem, which is in general NP-hard. Fortunately, we will show in Sec. II-C that a polynomial-time optimal solution can be achieved through exploring the unique structure of the valuation function.

*Definition 3: VCG payments*. The VCG-based auction mechanism results in a payment for bidder $b_i^k$ as:

$$p_i^k = v_i^k(\tilde{n}_{ij}^{k*}) + \max_{\substack{\mathbf{b}_i^k=\\ \{(0,0)\}}} \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} v_i^k(\tilde{n}_{ij}^k) - \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} v_i^k(\tilde{n}_{ij}^{k*}) \quad (4)$$

The notation $\tilde{n}_{ij}^{k*}$ denotes the optimal solution to (3). The implications of the virtual payment are two-fold. First, since the payment can be used for future trades, it provides an incentive to resource contributors. Second, the truthfulness of collaborative caching is also guaranteed by the payment method, which will be proven in Sec. II-C.

*Definition 4:* The utility gained by virtual bidders in a VCG auction is $u_i^k = v_i^k(\tilde{n}_{ij}^{k*}) - p_i^k$.

### *C. Optimal Bandwidth Allocation Strategies*

The winner determination problem in the form of integer programming is of high computational complexity [7]. However, the valuation function in our proposed mechanisms satisfies a downward sloping property [10], which results in a polynomial-time optimal allocation.

*Theorem 1:* The valuation function given in (2) satisfies the downward sloping property.

*Proof:* The first order derivative of the valuation is:

$$\frac{\partial v_i^k}{\partial n_{ij}^k} = \frac{\gamma \cdot \alpha \cdot e_{ij} \cdot w (x_i^k)^{1-\alpha}}{{r_k}^\alpha (W_i^k + e_{ij} \cdot w \cdot n_{ij}^k)^{1-\alpha}} > 0 \quad (5)$$

This ensures that the valuation monotonically increases with a growing number of provisioned bandwidth units. The second order derivative of the valuation is:

$$\frac{\partial^2 v_i^k}{\partial {n_{ij}^k}^2} = \frac{\gamma(\alpha^2 - \alpha)(e_{ij} \cdot w)^2 (x_i^k)^{1-\alpha}}{{r_k}^\alpha (W_i^k + e_{ij} \cdot w \cdot n_{ij}^k)^{2-\alpha}} < 0 \quad (6)$$

The second order derivative is constantly smaller than 0, which indicates that the valuation function (2) satisfies the downward sloping property. ■

Thus, the seller can achieve the optimal solution to problem (3) in polynomial time without obtaining any details of the valuation function. The procedure of the winner determination is given in Algorithm 1.

**Algorithm 1** A Collaborative Caching Framework through VCG Resource Auctions

```
n_j^r ← ⌊W_j^r / w⌋
if n_j^r ≥ Σ_{k∈K} Σ_{i∈I} n_ij^k then
    ñ_ij^k* ← n_ij^k, ∀i ∈ I, k ∈ K
else
    ñ_ij^k* ← 0, ∀i ∈ I, k ∈ K
    for n_j^r > 0 do
        {i, k} = argmax_{i∈I,k∈K} v_i^k(ñ_ij^k* + 1) − v_i^k(ñ_ij^k*)
        ñ_ij^k* ← ñ_ij^k* + 1
        n_j^r ← n_j^r − 1
    end for
end if
```

*Theorem 2:* The utility $u_i^k$ gained by virtual bidder $b_i^k$ is a non-negative value.

*Proof:* Suppose $\tilde{n}_{ij}^{k*}$ is determined in the winner determination process. The resulting payment is $p_i^k$ given by Eq. (4). In the optimal solution to the problem of $\max \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} v_i^k(\tilde{n}_{ij}^k)$ in case of $\mathbf{b}_i^k = \{(0,0)\}$, $\tilde{n}_{ij}^k = 0 \le n_{ij}^k$. Therefore, this solution also satisfies the constraints in Problem (3) if the remaining bids are not changed. Therefore, its social welfare is smaller than or equal to $\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} v_i^k(\tilde{n}_{ij}^{k*})$. Then $p_i^k - v_i^k(\tilde{n}_{ij}^{k*}) \le 0$, $u_i^k \ge 0$. ■

*Theorem 3:* Bidding on one's true valuation $v_i^k(\tilde{n}_{ij}^{k*})$ maximizes the utility obtained by virtual bidders.

*Proof:* The utility gained from the true valuation is $u_i^k = \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} v_i^k(\tilde{n}_{ij}^{k*}) - \max_{\mathbf{b}_i^k=\{(0,0)\}} \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} v_i^k(\tilde{n}_{ij}^k)$. Consider that $b_i^k$ submits a false valuation such that ${v'}_i^k(\tilde{n}_{ij}^k) \ne v_i^k(\tilde{n}_{ij}^k)$. In this case, the utility gained from the false valuation is ${u'}_i^k = \max \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} {v'}_i^k(\tilde{n}_{ij}^k) - {v'}_i^k(\tilde{n'}_{ij}^{k*}) - \max_{\mathbf{b}_i^k=\{(0,0)\}} \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} v_i^k(\tilde{n}_{ij}^k) + v_i^k(\tilde{n'}_{ij}^{k*})$, where $\tilde{n'}_{ij}^{k*}$ is the optimal allocation to $\max \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} {v'}_i^k(\tilde{n}_{ij}^k)$. The utility differentiation between two scenarios is denoted as $u_i^k - {u'}_i^k = (\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} v_i^k(\tilde{n}_{ij}^{k*}) - v_i^k(\tilde{n'}_{ij}^{k*})) - (\max \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} {v'}_i^k(\tilde{n}_{ij}^k) - {v'}_i^k(\tilde{n'}_{ij}^{k*}))$. Since $\tilde{n}_{ij}^{k*}$ is an efficient allocation to problem (3) in the context of true valuations, the authenticity of remaining bids ensures that $u_i^k - {u'}_i^k \ge 0$. Therefore, we have $u_i^k \ge {u'}_i^k$. ■

All bids submitted by virtual bidders are therefore *truthful* in collaborative caching, according to Theorem 2 and 3. Meanwhile, the demands that contain higher valuations are satisfied with higher priorities, which optimize the social

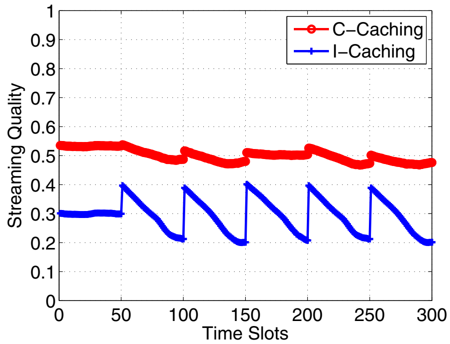


Fig. 1. The streaming quality of collaborative caching versus independent caching.

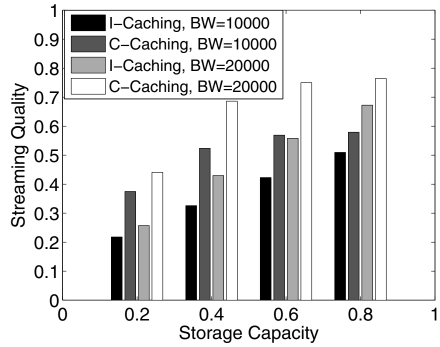


Fig. 2. The streaming quality of collaborative caching versus independent caching, under different settings of bandwidth and storage capacity.

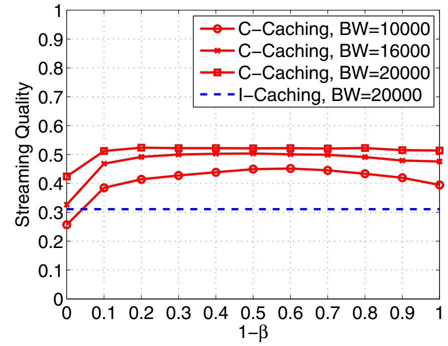


Fig. 3. The streaming quality versus the percentage of bandwidth provisioning to other WSPs, under different settings of bandwidth capacities.

welfare. The bidding strategy is also simplified to promote practical deployment of the proposed mechanisms.

### D. Implementation Issues

*First, local resource allocation*. In each trading round, a fraction of $\beta$ ($0 \leq \beta \leq 1$) of the bandwidth capacity is used for local demands such that $\sum_{k \in \mathcal{K}} W_i^k = \beta W_i$. When the value of $\beta$ is set to 1, the collaborative caching mechanism degenerates to independent caching in which no bandwidth is reserved for users of other WSPs.

*Second, the storage update*. Besides the bandwidth allocation scheme derived from the above resource auctions, the storage resource of cache servers also needs periodic updates to keep valuable contents that benefit their own payoffs and the social welfare. It can be observed from the valuation function, Eq. (2), that it is more beneficial to assign bandwidth resources to popular channels. Therefore, caching mechanisms prefer to cache highly popular videos. In practice, statistical results of the video popularity over the most recent storage update interval will be used for making storage update decisions.

## III. Performance Evaluation

In this section, we examine the performance of the proposed collaborative caching in comparison with the conventional independent caching mechanism. Our evaluation is based on a time-slotted simulator implemented in Python. The performance of different caching mechanisms is analyzed in terms of the overall streaming quality, *i.e.*, the percentage of online users that experience a smooth playback. We also take a close look at the potential overhead of collaborative caching, which is critical in practical system design.

### A. Performance Improvement through Collaborative Caching

We simulate $4$ WSPs that provide video streaming services in a geographical region. Each WSP deploys cache servers at their own MSC that connect to access points. There are $500$ video channels and the peak popularity is distributed over $10$ to $500$. The size of the bandwidth unit in auctions is set as the same as the streaming rate of $500$ kbps. In a general setting, each cache server is able to store $25\%$ of all the existing video channels, and their bandwidth capacity is set to $16,000$ bandwidth units which approximately accounts for $20\%$ of the total bandwidth requirement in the streaming system. A fraction of $60\%$ of the bandwidth capacity is dedicated to local demands. The allocation update interval is set to $1$ time slot and the storage update interval is set to $50$ time slots.

We first plot the overall streaming quality improvement brought by the proposed collaborative caching mechanism (denoted as C-Caching) in Fig. 1. We observe that the overall streaming quality of C-Caching consistently outperforms that of independent caching (denoted as I-Caching). In the beginning, the streaming qualities of C-Caching and I-Caching stay around $50\%$ and $30\%$, respectively. After the first storage update, most recently popular videos are then fetched by cache servers. The streaming quality of I-Caching increases to $40\%$ as more bandwidth units are now utilized for supporting channels with high valuation. Meanwhile, the streaming quality of C-Caching does not explicitly increase since the resource auctions already fulfill the demands to those unavailable popular videos before the storage update. In subsequent time slots, the streaming quality of I-Caching fluctuates between around $20\%$ and $40\%$, while C-Caching can always maintain a satisfactory level of performance. This also demonstrates the stability of the proposed collaborative caching mechanism.

Fig. 2 compares different scenarios of bandwidth and storage provisioning. We observe that C-Caching still outperforms I-Caching under different capacity settings. Specifically, the streaming quality improvement under C-Caching is significant initially when the storage capacity increases from $20\%$ to $40\%$ and it becomes less significant if the storage capacity exceeds $40\%$. This clearly demonstrates that collaborative caching through multiplexing substantially reduces the overall storage requirement to achieve a desirable streaming quality.

Fig. 3 depicts the impact from the portion of bandwidth resources that are allowed to be utilized by other WSPs. All three C-Caching with different bandwidth settings exhibit a similar pattern, in which (1) the overall streaming quality of all C-Caching is similar to that of I-Caching initially when all bandwidth are allocated for local use ($1-\beta=0$), (2) the streaming quality is improved when more portion of bandwidth is allocated for other WSPs (*i.e.*, for auctions) as the value of $1-\beta$ increases. The maximal streaming quality is achieved

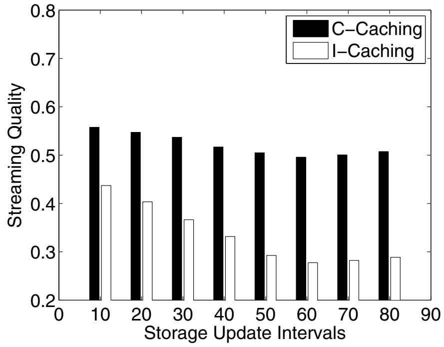


Fig. 4. The streaming quality versus the cache storage update interval.

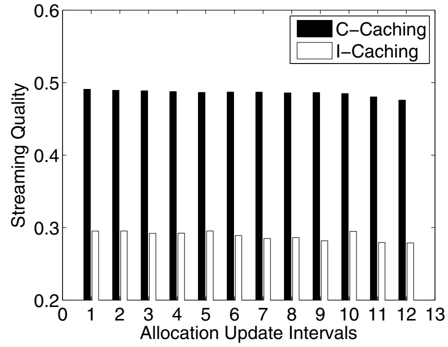


Fig. 5. The streaming quality versus the bandwidth allocation update interval.

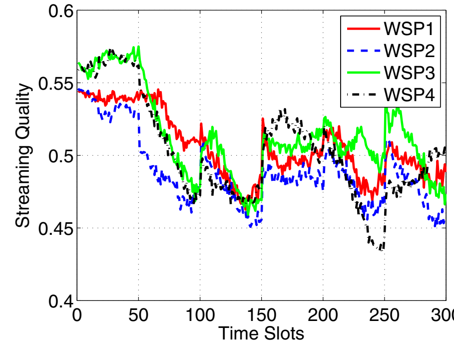


Fig. 6. The streaming quality of different WSPs in collaborative caching over time.

when the bandwidth is adequately allocated for both local and other WSPs, *i.e.*, when the value of $\beta$ varies between $0.4$ and $0.6$. (3) The streaming quality begins to deteriorate as we further increase the portion of bandwidth for other WSPs. Fig. 3 suggests that we can properly design $\beta$ to achieve better streaming quality in practice.

### B. Overhead with Collaborative Caching

We next examine the overhead of C-Caching by analyzing the impact of the update interval of cache storage and that of bandwidth allocation. The former causes overhead on content dissemination between content servers and cache servers while the latter incurs computational overhead. Fig. 4 illustrates the tradeoff between the streaming quality and the frequency of cache storage updates. When the storage is updated each $10$ slots, the streaming quality of C-Caching achieves $55\%$ while that of I-Caching is $45\%$. However, the performance gap between them becomes more significant when the update interval gradually increases. The increasing gap is caused by the fluctuation of channel popularity, which can be easily resolved with inter-WSP resource auctions. I-Caching has to frequently update its storage to reach a comparable level of performance as C-Caching. In sharp contrast, the performance of C-Caching remains stable regardless of specific cache update strategies.

Further, we analyze the tradeoff between the streaming quality and the update interval of bandwidth allocation in Fig. 5. Although we have obtained a polynomial-time optimal solution to the VCG auction in Sec. II-C, the computational overhead is still not negligible if we require the trade to be conducted in each time slot. Inspired by empirical experiences that system dynamics are usually limited over a short period of time, we evaluate the scenarios where a bandwidth allocation decision from resource auctions is used over multiple rounds. Fig. 5 clearly shows that the overall streaming quality only decreases slightly along with an increase of the allocation update interval. This implies that in practical system design, it is feasible to reuse the bandwidth allocation result that further reduces the overhead of auction mechanisms.

### C. Fairness among WSPs

Although incentives and truthfulness have been incorporated into our auction settings, it is still indispensable to evaluate fairness in the system when auctions are practically conducted among multiple WSPs. Fig. 6 plots the respective streaming quality of different WSPs in our simulation. It shows that variations of the streaming quality in each WSP exhibit a similar pattern over time. This shows that the proposed resource auction can fairly improve social welfare by assisting one another.

## IV. Conclusion

In this paper, we have proposed a collaborative caching mechanism for wireless video streaming that coordinates cache resource provisioning among multiple selfish WSPs. In particular, we focus on engineering the incentives to promote and encourage cache servers from different WSPs to cooperate with one another in the context of VCG auctions while both fairness and truthfulness are guaranteed. We treat server bandwidth as commodities in these auctions, with different valuations based on dynamics of streaming systems. We then derive solutions to allocate server bandwidth. Simulation results have demonstrated that the video streaming quality can be significantly improved with the maximization of social welfare in auctions conducted by the proposed collaborative caching mechanism.

## References

[1] F. Khan, *LTE for 4G Mobile Broadband: Air Interface Technologies and Performance*. Cambridge University Press, 2009.

[2] "Explanation of Optimization Deployment," http://www.verizonwireless.com/vzwoptimization/.

[3] Q. Zhang, Z. Xiang, W. Zhu, and L. Gao, "Cost-Based Cache Replacement and Server Selection for Multimedia Proxy across Wireless Internet," *IEEE Transactions on Multimedia*, vol. 6, no. 4, 2004.

[4] F. Xie and K. A. Hua, "A Caching-based Video-on-Demand Service in Wireless Relay Networks," in *Proc. IEEE Wireless Communications and Signal Processing*, Nov. 2009.

[5] E. Tan, L. Guo, S. Chen, and X. Zhang, "SCAP: Smart Caching in Wireless Access Points to Improve P2P Streaming," in *Proc. IEEE ICDCS*, Jun. 2007.

[6] S. Borst, V. Gupta, and A. Walid, "Distributed Caching Algorithms for Content Distribution Networks," in *Proc. IEEE INFOCOM*, Mar. 2010.

[7] N. Nisan and A. Ronen, "Computationally Feasible VCG Mechanisms," *Journal of Artificial Intelligence Research*, vol. 29, no. 1, 2007.

[8] C. Wu, B. Li, and S. Zhao, "Diagnosing Network-wide P2P Live Streaming Inefficiencies," in *Proc. IEEE INFOCOM*, Apr. 2009.

[9] C. Wu and B. Li, "Multi-channel Live P2P Streaming: Refocusing on Servers," in *Proc. IEEE INFOCOM*, Apr. 2008.

[10] S. Dobzinski and N. Nisan, "Mechanisms for Multi-Unit Auctions," *Journal of Artificial Intelligence Research*, vol. 37, no. 1, 2010.