Auction-based Cache Trading for Scalable Videos in Multi-Provider Heterogeneous Networks

Behrouz Jedari and Mario Di Francesco

Department of Computer Science, Aalto University, Finland

Email:{behrouz.jedari,mario.di.francesco}@aalto.fi

Abstract—Content providers (CPs) are keen to cache their popular contents in small-cell base stations (SBSs) provided by mobile network operators (MNOs). In fact, they can serve the requests of their subscribers with low latency, thereby increasing user satisfaction. Employing advanced video encoding techniques, such as scalable video coding (SVC), improves the utilization of wireless resources and the network infrastructure. However, the cache trading policies for SVC videos in multi-provider networks have not been studied yet. In this article, we design a commercial trading system in which multiple CPs, each owning SVC videos, compete over renting the cache in multiple SBSs provided by an MNO. We model cache trading between the MNO and CPs as a social welfare maximization problem, whose objective is to maximize the trading profit while achieving the economic properties of rationality, balanced budget, and truthfulness. Since optimal allocation of random-size caches to multiple CPs is NP-hard, we devise an iterative trading mechanism based on double auction called DOCAT, wherein the cache of SBSs is segmented and traded in multiple rounds. In each round of the auction, the MNO and CPs price the cache segments based on their profit, then submit their asking and buying bids, respectively. Next, a many-to-one matching algorithm is run to efficiently find perfect matches between the cache segments and winning CPs. Numerical results based on a real video dataset show that DOCAT increases the social welfare of the system while satisfying the desired economic properties.

Index Terms—Heterogeneous networks, edge caching, scalable video coding, network economics, game theory, auction.

I. INTRODUCTION

The global mobile traffic is expected to grow about eight times by 2023 and the share of video data will reach 73% [1]. Such booming data traffic can greatly increase the revenue of mobile network operators (MNOs) and content providers (CPs). However, the capacity of data storage and delivery over cellular networks should be expanded to cope with the massive demands of end-users. Generally, MNOs (e.g., AT&T) acquire new radio spectrum or apply multi-antenna techniques, and CPs (e.g., YouTube) establish dedicated data centers, both of which are often costly and time consuming. With the advent of heterogeneous networks (HetNets), caching contents at the network edge through small-cell base stations (SBSs) is considered as a low-cost solution to increase the network capacity, while reducing the network backhaul traffic [2]. In this context, SBSs (e.g., femto and pico cells) with different caching and data transmission capabilities are deployed in strategic places to make the contents closer to end-users.

A key challenge in content caching is to determine which contents should be placed in each cache so that the requests of mobile users (MUs) are served with minimum delay. Existing caching algorithms mainly employ the concept of *content popularity* to increase the cache hit ratio [3–6]. FemtoCaching [4], for instance, demonstrated that caching popular videos in SBSs increases the system throughput up to five times. Indeed, CPs store a portion of their videos in SBSs owned by MNOs to serve the requests of subscribers with minimum delay, thus improving their quality of experience (QoE). However, data storage and transmission resources in SBSs are limited, especially considering CPs' huge volume of data and user requests [7]. Thus, CPs have to compete with each other over renting caches in SBSs.

Recent studies have employed economic models (e.g., game theory) for infrastructure sharing and cache trading in Het-Nets [8–12]. However, they did not consider the structure of cached content on the pricing policies of MNOs and CPs. This is extremely important because the properties of video data significantly affect the operation of caching and delivery algorithms, thus, also the revenues and costs of CPs and MNOs. This issue becomes challenging when CPs employ advanced encoding techniques, such as scalable video coding (SVC), to stream their videos. SVC is the most prominent HTTP-based video encoding scheme that allows for multiple spatiotemporal resolutions and frame rates [13]. An SVC (or layered) video includes a basic layer for low-quality demands, combined with enhancement layers to provide high-quality videos. Layered encoding is particularly suited for wireless communications as video data can be adapted to fast varying wireless links without re-encoding [14]. However, the dependency between different layers of an SVC video has to be considered during the caching and delivery process. Despite some works on caching SVC videos [15–17], the economic aspects of cache trading have not been studied in such a context.

We model a cache trading system in which an MNO (seller) aims at leasing its caches in SBSs to multiple CPs, where each CP owns a set of layered videos. We identify the profit of the MNO and CPs from cache trading based on their revenues and costs. We assume that the MNO values an SBS located at the border of the network higher than an SBS at the center (e.g., due to its higher power consumption). The profit of the MNO from cache leasing is identified as the monetary income it receives from CPs plus the traffic reduction it experiences in the backhaul network due to edge caching. Furthermore,

the profit of each CP from renting a cache is quantified in terms of the time saved by its subscribers in downloading their videos thanks to the edge caching. We identify the optimal set of cached videos given the cache size in such a way that the video download time is minimized. However, due to the different popularity of CPs, the number of video requests CPs receive within a certain period can be considerably different [18], which impacts on the profit of both of the MNO and CPs. Thus, the following problem arises: *how to allocate the caches in SBSs to multiple CPs so that the profit of both the MNO and CPs is maximized?*

To address the problem above, we apply the concept of *social welfare maximization* to model cache trading between the MNO and CPs. Our objective is to maximize the profit of the MNO and CPs from cache trading while achieving the economic properties of rationality, balanced budget, and truthfulness. However, optimal allocation of random-size caches to multiple CPs is NP-hard [19]. A double auction is a suitable approach to deal with this problem. Accordingly, market commodities (the caches) in a double auction should be traded in single units to achieve certain economic criteria, such as individual rationality and truthfulness [20].

To make double auction feasible, we first divide the cache at each SBS into equal-sized cache segments. Next, we devise an iterative double auction-based cache trading (DOCAT) mechanism, in which multiple cache segments (one cache segment from each SBS) are traded in each round of the auction. In DOCAT, a third-party broker (auctioneer) who is unaware of the actual profits of the MNO and CPs runs a sequential double auction. In each round of the auction, the MNO and CPs calculate their profit from trading cache segments, then submit their sealed asking prices and buying bids (respectively). Next, the broker applies a many-to-one matching to find the perfect matches between the trading cache segments and winning CPs, so that the social welfare is maximized while satisfying the desired economic properties.

In summary, the contributions of this article are as follows.

- To the best of our knowledge, this is the first trading approach that determines the profit of the MNO and CPs in cache trading based on the structure of caching contents, particularly, for scalable videos (Section VI).
- We model cache trading between the MNO and CPs as a social welfare maximization problem, where the profit of the MNO and CPs is identified based on their revenues and costs (Sections II and III).
- We propose a double auction-based cache trading (DOCAT) mechanism with polynomial time complexity that maximizes the profit of the MNO and CPs (Section IV). DOCAT is rational (the MNO and CPs willingly participate in trading), maintains a weakly balanced budget (the broker does not lose money), and is incentive-compatible (the MNO and CPs bid truthfully).
- A numerical evaluation based on a real video dataset demonstrates that DOCAT is efficient and achieves the desired economic properties in trading caches (Section V). In particular, popular CPs achieve the highest
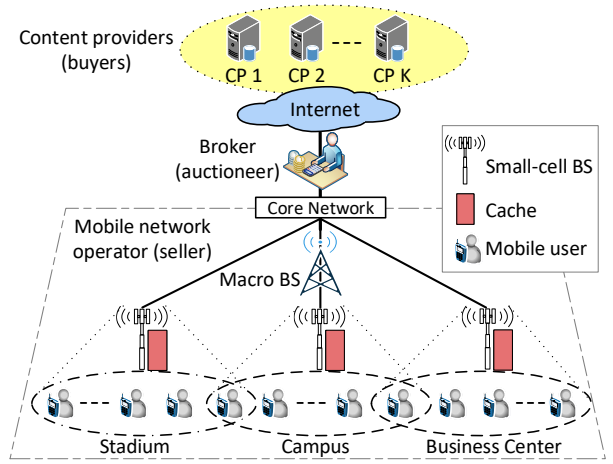

Fig. 1: A commercial cache trading system in HetNets.

social welfare when they bid truthfully, as the broker divides as much as possible segments between different CPs.

II. System Model

This section models a cache trading system consisting of an MNO with multiple SBSs, CPs, and MUs (Fig. 1). First, we introduce the main characteristics of the MNO and CPs. Next, we present the properties of layered videos, as well as their caching and delivery models. The key parameters of the system are summarized in Table I.

A. Network Model

The MNO: There is an MNO consisting of a single macro BS (MBS) M that serves a geographical area. M is connected to the core network through a high-capacity backhaul link (e.g., fiber optics), whose average transmission capacity is denoted as L_M Mbit/s. In addition, the MNO deploys a set $\mathcal{S}=\{1, 2, \ldots, S\}$ of S cache-enabled SBSs in strategic locations, where the physical distance between M and each SBS $s \in \mathcal{S}$ is $d_{M,s}$ meters. The SBSs are connected to the core network through wired or wireless backhaul links. Each SBS s has F_s downlink channels whose average download capacity is L_s Mbit/s. Each channel serves multiple users simultaneously through proportional fair scheduling. The storage size and transmission range of SBS s are indicated as C_s MB and R_s meters, respectively. MBS M and SBSs transmit in orthogonal channels, thus, they do not interfere with each other.

CPs: There exist a set $\mathcal{K}=\{1, 2, \ldots, K\}$ of K CPs, where the popularity of CPs can be different [18]. Similar to [21], we characterize the popularity of CPs by using a Zipf-like distribution [22] in a descending order as $Z=[z_1, z_2, ..., z_K]$, where $\sum_{i=1}^{K} z_i=1$. The popularity of the $i-$th ranked CP is:

$$z_i = \frac{i^{-\gamma}}{\sum_{j=1}^{K} j^{-\gamma}} \quad (1)$$

where γ is the popularity skewness of CPs. Here, $\gamma=0$ implies that the popularities of CPs are the same, while γ near to 1 implies that few CPs are more popular than others.
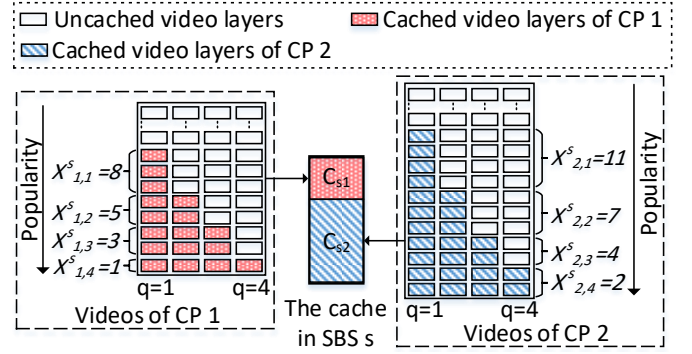
TABLE I: Key System Parameters

Symbol(s)	Definition
M, L_M	MBS owned by MNO and its transmission capacity
$d_{M,s}$	Physical distance between MBS M and SBS s
$\mathcal{S}, s$	Set of SBSs and a single SBS
L_s, C_s	Transmission capacity and cache size of SBS s
C_{sk}	Size of cache rented by CP k in SBS s
$\mathcal{U}, i$	Set of MUs and a single MU
$\mathcal{U}_k^t$	Number of CP k's subscribed MUs at time t
$\mathcal{U}_{ks}^t$	Number of CP k's subscribers in SBS s at time t
$\mathcal{K}, \mathcal{K}_i, k$	Set of CPs, CPs subscribed by MU i, and a single CP
z_k, γ	Popularity of CP k and the popularity skewness of CPs
ζ_{ik}^t	Demand rate of MU i from CP k per time slot
$\mathcal{V}_k, v_k$	Set of videos owned by CP k and a single video in $\mathcal{V}_k$
Q	Number of layers or quality levels of videos
o_{kvq}, O_{kvq}	Size of $q-$th layer and quality level of video v_k in $\mathcal{V}_k$
p_{ki}	Popularity of $i-$th ranked video in $\mathcal{V}_k$
η_k	Popularity skewness of videos in $\mathcal{V}_k$
X_k^s	Placement parameters of CP k in SBS s
P_{ks}^q	Probability that q-th quality of a video in $\mathcal{V}_k$ is hit in s

Fig. 2: A sample video placement in SBS s by CPs 1 and 2.

MUs: The network consists of a set $\mathcal{U}=\{1, 2, \ldots, U\}$ of U mobile users (MUs), where each MU (covered by one or more SBSs) only communicates with its nearest SBS. MUs download their requested data only from either SBSs or MBS M. We evaluate the system in T time slots, each with a duration of $\triangle t$. The location of MUs and their transmission rates are assumed to be fixed within each time slot, but they may change in different time slots. Each MU can subscribe to multiple CPs, where $\mathcal{K}_i \in \mathcal{K}$ denotes the set of subscribed CPs of MU i. We identify the number of subscribers of a CP and the video demand ratio of each MU based on the popularity of CPs. Moreover, $\mathcal{U}_{ks}^t \subseteq \mathcal{U}$ denotes the set of subscribers of CP k associated with SBS s and $\mathcal{U}_k^t = \sum_{i=1}^{S} \mathcal{U}_{ki}^t$ the set of subscribers of CP k in the network at time t. Finally, ζ_i^t is the average demand of MU i at time t, thus, the number of i's video requests from CP $k \in \mathcal{K}_i$ per $\triangle t$ is $\zeta_{ik}^t = \zeta_i^t z_k$.

B. Video Coding and Popularity Models

Video Encoding: Each CP $k \in \mathcal{K}$ publishes a set $\mathcal{V}_k = \{1_k, 2_k, \ldots, V_k\}$ of V_k videos, where each video is encoded with the SVC scheme [13]. An SVC video consists of Q layers including one base layer and $Q - 1$ enhancement layers. Layer 1 realizes the first quality level of the video, the combination of layers 1 and 2 realizes the second quality level of the video, and so on. Following real SVC videos (e.g., [23]), the size of $q-$th layer of video v_k owned by CP k, denoted as o_{kvq}, decreases with q ($o_{kv1} > o_{kv2} > \ldots > o_{kvQ}$). When a MU requests the $q-$th quality level of video v_k, q layers of v_k from o_{kv1} up to o_{kvq} are delivered, where the size of $q-$th quality level of video v_k is $O_{kvq} = \sum_{l=1}^{q} o_{kvl}$.

Video Popularity: We describe the request probability of videos of a CP by using the Zipf distribution. We assume that the popularity of videos can be predicted by using learning methods (e.g., [24]) and is known in advance. In addition, the content popularity is almost constant within a certain

period (e.g., few hours or days) [25]. Thus, we sort the videos of CP $k \in \mathcal{K}$ in the descending order of their popularity as $P_k = [p_{k1}, p_{k2}, \ldots, p_{kK}]$, $\sum_{i=1}^{V_k} p_{ki} = 1$. Again, the popularity of the i-th ranked video in V_k is:

$$p_{ki} = \frac{i^{-\eta_k}}{\sum_{j=1}^{V_k} j^{-\eta_k}} \tag{2}$$

where η_k is the popularity skewness of videos of CP k (V_k).

C. Video Caching and Delivery Models

Caching Policy: In general, CPs place their video layers in their rented caches during off-peak hours (e.g., midnight). Without loss of generality, we study a case multiple CPs aim at caching the desired videos in a single SBS $s \in \mathcal{S}$. Since the cache of SBS s can be rented by one or multiple CPs, we denote the size of the cache rented by CP k in SBS s as C_{sk}, where $C_{sk} \in \mathbb{Z}^+$ and $0 \leq C_{sk} \leq C_s$. Next, we explain how to identify the placement strategy of CP k with respect to cache size C_{sk}.

As a general principle, CP k caches the $q-$th layer of video v_k only if its previous layers (i.e., layer 1 up to layer $q-1$) have already been cached [15]. We define the placement policy of CP k in SBS s with $X_k^s = [x_{k,Q}^s, x_{k,Q-1}^s, \ldots, x_{k,1}^s]$, where $x_{k,Q}^s \leq x_{k,Q-1}^s \leq \ldots \leq x_{k,1}^s$, $x_{k,Q}^s \geq 0$, $x_{k,1}^s \leq V_k$, and $x_q^s \in \mathbb{Z}^+ (\forall q, 1 \leq q \leq Q)$. Accordingly, CP k caches the following video layers of V_k in SBS s: Q layers of $x_{k,Q}^s$ most popular videos, $Q-1$ layers of videos that their popularity is ranked between $x_{k,Q}^s + 1$ and $x_{k,Q-1}^s$, and so on. Since the total size of video layers cached by CP k cannot exceed C_{sk}, the elements of X_k^s should be identified in a way that satisfies:

$$\sum_{i=1}^{Q} \sum_{j=1}^{n} \sum_{q=1}^{i} o_{kzq} \leq C_{sk} \tag{3}$$

where n is the number of videos in V_k that their first i layers are cached in SBS s. If $i=Q$ then $n=x_{k,Q}^s$, and if $1 \leq i < Q$ then $n=x_{k,i}^s - x_{k,i+1}^s$. In addition, z is the index of the first video that its first i layers are cached. If $i=Q$ then $z=1$, and if $1 \leq i < Q$ then $z=x_{k,i+1}^s + j$. We explain how to identify optimal video layers from each CP for caching (i.e., X_k^s) in Sec. III.

Fig. 2 illustrates a sample placement of video layers of CPs 1 and 2 in the cache of SBS s. CPs 1 and 2 own sets of V_1 and V_2 videos, respectively, where each video is encoded

in 4 quality levels. CPs 1 and 2 rented C_{s1} and C_{s2} MB cache space in SBS s, respectively, where $C_{s1}+C_{s2}=C_s$. For instance, CP 1 caches the 4-th quality level of video 1, the 3-rd quality level of videos 2 and 3, the 2-nd quality level of videos 4 and 5, and the 1-st quality level of videos 6-8.

Delivery Procedure: Once CPs place the desired video layers in their rented caches, video delivery operates as follows. When MU i requests $q-$th quality level of video $v_k\in\mathcal{V}_k$, layers 1 up to q of v_k are served by i's associated SBS (SBS-response). If any of the requested layers are not cached in that SBS, they are served by MBS M (MBS-response).

III. PROFIT MODEL AND PROBLEM FORMULATION

In this section, we first identify the revenues and costs of the MNO and CPs from cache trading. Next, we formulate cache trading as a profit maximization problem. We quantify the profit of the MNO and CPs in terms of price of trading caches with certain size for a certain time period (PCT), e.g., unit price (\$) of trading 1MB cache for period $\triangle t$.

Profit of the MNO: We identify the profit of the MNO from leasing the caches in SBSs in two parts: (i) the income gained from leasing the cache in SBSs to CPs, and (ii) the cost reduction due to the traffic reduction in the backhaul network.

To derive (i), we suppose that the MNO prices SBSs based on their distance from MBS M, where SBSs at the border of the network are valued higher than SBSs closer to MBS M. The reason is that serving MUs at the border can be very costly to the MNO (e.g., due to higher power consumption and poor link quality) [26]. Thus, the MNO income from leasing a cache with size C_{sk} in SBS s to CP $k\in\mathcal{K}$ is given by:

$$s_{LP}^t(C_{sk}) = d_\psi g_{LP}^t C_{sk} \tag{4}$$

where $d_\psi=\frac{d_{M,s}}{d_{M,S}^{Avg}}$ is a *distance factor* in which $d_{M,S}^{Avg}$ is the average distance between MBS M and SBSs, and g^t is the cache price in PCT units.

To derive (ii), the MNO has to estimate the average number of video requests it receives from each CP in each SBS because it is unaware of CPs' properties (e.g., their popularity and the number of their subscribers). Assuming that MNO receives on average $E[\zeta_{sk}^t]$ video demands from MUs$\in\mathcal{U}_{ks}^t$ per $\triangle t$, the cost reduction for the MNO (in PCTs) from leasing cache with size C_{sk} in SBS s to CP k per $\triangle t$ is given by:

$$s_{CR}^t(C_{sk}) = E[\zeta_{sk}^t]g_{CR}^t p(C_{sk}) \tag{5}$$

where $p(C_{sk})$ is the probability of cache hit for requests of MUs$\in\mathcal{U}_{ks}^t$ in SBS s given cache size C_{sk}, and g_{CR}^t is the cost reduction for the MNO from serving the video request of a MU by its associated SBS (rather than MBS M) in PCT units. Thus, the total profit of MNO from leasing cache with size C_{sk} in SBS s to CP k is calculated as:

$$Prof_{MNO}^t(C_{sk}) = s_{LP}^t(C_{sk})+s_{CR}^t(C_{sk}) - g_{cost}^t C_{sk} \tag{6}$$

where g_{cost}^t is the cost of maintenance and operation of caches for the MNO in PCT units.

Profit of CPs: We identify the profit of a CP from cache renting in form of quality of experience (QoE) improvement of its subscribed MUs due to content caching, in terms of shorter time to download videos. In particular, we consider the *startup delay*, namely, the time from the moment a video is requested by a MU to the time the last packet of the video is delivered. We take the popularity of caching videos and the demand ratio of MUs into account to measure the delay saving. Thus, when k rents the cache with size C_{sk} in SBS $s\in\mathcal{S}$, the download delay saving for the subscribers of CP k associated with SBS s ($\mathcal{U}_{ks}^t$) is calculated as follows. If CP k does not cache any content in SBS s (i.e., the requests of MUs in $\mathcal{U}_{ks}^t$ are served by MBS M), the aggregated download delay for the requests of MUs in $\mathcal{U}_{ks}^t$ per $\triangle t$ is given by:

$$D_{M,s,k}^{NoCache} = \sum_{i\in\mathcal{U}_{ks}^t} \sum_{v\in\mathcal{V}_k} \sum_{q\in Q} \zeta_{ik}^t p_{kv} \frac{O_{kvq}}{QL_M} \tag{7}$$

where $\frac{1}{Q}$ is the probability of requesting different quality levels of a video. In case CP k rents cache with size C_{sk} in SBS s, the aggregated download delay for the requests of MUs in $\mathcal{U}_{ks}^t$ per $\triangle t$ that are served by MBS M is calculated as:

$$D_{M,s,k}^{Cache}(C_{sk})= \sum_{i\in\mathcal{U}_{ks}^t} \sum_{v\in\mathcal{V}_k} \sum_{q\in Q} \left(1-P_q(C_{sk})\right)\zeta_{ik}^t p_{kv} \frac{O_{kvq}}{QL_M} \tag{8}$$

In turn, the aggregated download delay of requests of $\mathcal{U}_{ks}^t$ that are served by SBS s per $\triangle t$ is:

$$D_{s,k}^{Cache}(C_{sk})= \sum_{i\in\mathcal{U}_{ks}^t} \sum_{v\in\mathcal{V}_k} \sum_{q\in Q} P_q(C_{sk})\zeta_{ik}^t p_{kv} \frac{O_{kvq}}{QL_s} \tag{9}$$

where $P_q(C_{sk})$ denote the hit rate probability of $q-$th quality level of a video in $\mathcal{V}_k$ if CP k rents cache with size C_{sk} in SBS s, which is given by:

$$P_q(C_{sk}) = \frac{\sum_{i=1}^{x_{k,q}^s} i^{-\eta_k}}{\sum_{j=1}^{V_k} j^{-\eta_k}} \approx \frac{{x_{k,q}^s}^{1-\eta_k}}{(1-\eta_k)\sum_{j=1}^{V_k} j^{-\eta_k}} \tag{10}$$

where $x_{k,q}^s$ indicates the number of videos in $\mathcal{V}_k$ that at least their first q layers are cached in SBS s. For a smoother flow of discussion, we postpone the derivation of the optimal video layers in $\mathcal{V}_k$ for caching in SBS s (i.e., X_k^s) given a cache size C_{sk} to Appendix A.

According to Eqs. (7)-(9), the cost reduction for CP k due to renting the cache with size C_{sk} in SBS s is given by:

$$k_{QoE}^t(C_{sk})= \left(D_{M,s,k}^{NoCache}-\left(D_{M,s,k}^{Cache}(C_{sk})+D_{s,k}^{Cache}(C_{sk})\right)\right)g_{QoE}^t \tag{11}$$

where g_{QoE}^t is a unit profit of CP k from download delay saving per $\triangle t$. Thus, the total profit of CP k from renting C_{sk} MB cache space in SBSs $\mathcal{S}$ per period t is given by:

$$Prof_{CP_k}^t(C_{sk})=k_{QoE}^t(C_{sk})-g_{rent}^t C_{sk} \tag{12}$$

where g_{rent}^t is paid by CP k to MNO as the rent in PCT units.

Problem Formulation: The MNO attempts to lease its caching resources to those CPs that maximize its profit. Meanwhile, each CP selfishly tries to rent caches with a certain size in each SBS so that its profit is maximized. We employ the concept of *social welfare* to maximize the joint profit of

both the MNO and CPs [27]. In particular, we state the social welfare maximization (SWM) problem as follows:

$$\max_{C_{SK}} \sum_{s\in\mathcal{S}} \sum_{k\in\mathcal{K}} Prof^t_{MNO}(C_{sk}) + \sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} Prof^t_{CP_k}(C_{sk})$$

(13)

$$\text{subject to } \sum_{k\in\mathcal{K}} C_{sk} \le C_s, \ \forall s\in\mathcal{S} \tag{14}$$

$$C_{sk} \ge 0, \ C_{sk}\in\mathbb{Z}, \forall s\in\mathcal{S}, \forall k\in\mathcal{K} \tag{15}$$

where C_{SK} is a $S\times K$ matrix, whose element C_{sk} denotes the amount of cache in SBS s traded between MNO and CP k. The SWM problem is an instance of a random-size knapsack problem, which is NP-hard [19]. In the next section, we employ an auction mechanism to find an efficient solution for a simplified version of the problem above.

IV. DOUBLE AUCTION-BASED CACHE TRADING

In this section, we present our proposed double auction-based cache trading (DOCAT) mechanism, followed by the proof of the economic properties it achieves.

A. Overview of DOCAT

In DOCAT, we first divide the cache of each SBS into equal-sized cache segments, where each cache segment is considered as a single trading commodity. We then design a sequential double auction mechanism in which multiple cache segments (i.e., one cache segment from each SBS) are traded in each round of the auction. In each round, the MNO (*seller*) and CPs (*buyers*) calculate their profit from trading the cache segments based on the profit they can gain (Sec. III). Next, the MNO and CPs respectively submit their sealed asking prices and buying bids regarding currently trading cache segments (*trading commodities*) to a broker (*auctioneer*). Finally, an efficient and fair matching algorithm is designed to find optimal matchings between the selling cache segments and buying CPs. Each CP can win renting zero, one or multiple cache segments in each round of the auction.

It is impossible for double auction mechanisms to simultaneously accomplish both the maximum system efficiency (here social welfare) and any of the economic properties (e.g., truthfulness) at the same time [20]. Nevertheless, DOCAT aims at achieving the maximum possible social welfare while satisfying the following properties.

- *Computational efficiency (CE)*: the trading algorithm should be computed in polynomial time.
- *Individual rationality (IR)*: both the MNO and CPs obtain non-negative profit by participating in the auction.
- *Weakly balanced budget (WBB)*: the total payment of CPs should not be less than the total charging of the MNO.
- *Incentive compatibility (IC) or Truthfulness*: the dominant strategy for the MNO and each CP is to bid their true valuation, otherwise their profit will be reduced.

B. The DOCAT Mechanism

DOCAT works as follows: first, the broker divides the cache in each SBS into l equal-sized segments and announces holding the auction in multiple rounds. Let C_s be the cache size in SBS $s\in\mathcal{S}$; the size of cache segment f ($1\le f\le l$) is denoted as $c_s^f = \frac{C_s}{l}$, $c_s^f\in\mathbb{Z}^+$. In this step, the broker initializes the allocation matrix $\Gamma_{l\times S}$ to zero, where entry Γ_{fs} indicates the identity of the CP that wins renting the f−th cache segment in SBS s. Next, each round of the auction runs as follows: the MNO and CPs valuate and submit their proposed bids regarding the currently trading cache segments based on their preferences. Next, the broker applies an equilibrium matching to identify winning CPs. Finally, actual charging and paying prices are determined. Specifically, in each round of the auction, the broker aims at solving the following problem:

$$\max_{C_{SK}^f} \sum_{s\in\mathcal{S}} \sum_{k\in\mathcal{K}} Prof^t_{MNO}(c_{sk}^f) + \sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{S}} Prof^t_{CP_k}(c_{sk}^f) \tag{16}$$

$$\text{s.t. } c_{sk}^f\in\{0,c_s^f\}, \ \forall s\in\mathcal{S}, \forall k\in\mathcal{K} \tag{17}$$

$$0\le \sum_{k\in\mathcal{K}} c_{sk}^f \le c_s^f \ \forall s\in\mathcal{S} \tag{18}$$

where C_{SK}^f is a matrix wherein entry c_{sk}^f indicates whether cache CP k wins renting f−th cache segment in SBS s or not. The constraint in Eq. (17) implies that the amount of cache rented by each CP in each SBS $s\in\mathcal{S}$ can be either 0 or c_s^f. Furthermore, the constraint in Eq. (18) implies that at most one CP can rent cache segment f in each SBS $s\in\mathcal{S}$.

Cache Valuation of the MNO: Regarding the profit of the MNO in Sec. III, its minimum asking price for leasing cache segment f with size c_s^f in SBS s (denoted as ϕ_s^f) should satisfy $\phi_s^f = s_{LP}^t(c_s^f) + s_{CR}^t(c_s^f) \ge g_{cost}^t c_s^f$ [Eq. (6)]. Again, we set the asking price of MNO for leasing cache segment f in SBS s to be $\phi_s^f = g_{cost}^t c_s^f$. While the MNO announces the asking prices for leasing the cache segments of SBSs based on the distance factor, it assigns the same asking price for leasing the cache segments of a SBS in different rounds of the auction.

Cache Valuation of CPs: Each CP values a cache segment in each SBS based on the QoE improvement of its subscribers associated with the SBS. Since a CP does not intend to lose money in the cache trading, its minimum buying bid for renting cache segment f with size c_s^f MB in SBS s, denoted as $\theta_{s,k}^f$, should satisfy $\theta_{s,k}^f = k_{QoE}^t(C_{sk}) \ge g_{rent}^t C_{sk}$ [Eq. (12)]. We set the asking price of CP k for renting cache segment f in SBS s to be $\theta_{s,k}^f = g_{rent}^t C_{sk}$. As discussed in Sec. III, the bidding value of a CP is calculated based on the demand ratio of its subscribers and the popularity of its caching videos. Thus, the most popular CP owning the most popular uncached videos will assign the highest bidding value to a particular cache segment. As a result, the biding value of a CP for renting cache segments in each SBS decreases in consecutive rounds of the auction, as the CP wins renting cache segments in the SBS, i.e., $\theta_{s,k}^1 \ge \theta_{s,k}^2 \ge \ldots \ge \theta_{s,k}^l$. In this way, less popular CPs
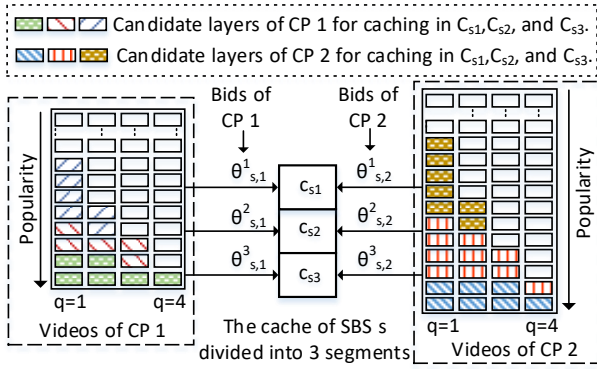
Fig. 3: Valuation of cache segments in SBS s by CPs 1 and 2.

owning less popular videos will also have a chance to win renting cache segments in this SBS.

Fig. 3 demonstrates the valuation of cache segments in SBS s by CPs 1 and 2, where the cache space is divided into 3 equal-sized segments. Since the videos of each CP have skewed popularity, their valuation for the cache segments is decreasing, i.e., $\theta_{s,1}^1 \geq \theta_{s,1}^2 \geq \theta_{s,1}^3$ and $\theta_{s,2}^1 \geq \theta_{s,2}^2 \geq \theta_{s,2}^3$. This valuation policy is rational because the profit of CPs from renting extra cache segments decreases, as the popularity of their candidate videos for caching decreases. Assuming that the popularity of CPs and the number of their subscribed MUs associated to SBS s are equal (i.e., $z_1 = z_2$ and $\mathcal{U}_{1s}^t = \mathcal{U}_{2s}^t$), the CP with the higher popularity skewness will value a particular cache segment higher than another one. For instance, if the popularity skewness of videos of CP 1 is higher than CP 2 ($\eta_1 \geq \eta_2$), it will assign higher price to a cache segment than CP 2, i.e., $\theta_{s,1}^f \geq \theta_{s,2}^f (\forall\ 1 \leq f \leq 3)$.

Matching Algorithm: We design an efficient matching in the sense that it matches perfect asking prices and buying bids so that the social welfare of the system (i.e., the joint profit of the MNO and CPs) is maximized. Moreover, the matching considers fair trading in such a way that a CP with fewer cache segments rented in a particular SBS in previous rounds of the auction has a higher priority to win the auction. The matching is many-to-one, implying that each CP can win one or multiple cache segments in each round of the auction.

Let us denote the set of trading cache segments in each round of the auction as $c_{\mathcal{S}i}$ ($1 \leq i \leq l$). The MNO's asking prices and CP k's buying bids for the cache segments in $c_{\mathcal{S}i}$ are denoted as $\phi_1^{(.)}, \phi_2^{(.)}, \ldots, \phi_S^{(.)}$ and $\theta_{1,k}^{(.)}, \theta_{2,k}^{(.)}, \ldots, \theta_{S,k}^{(.)}$, respectively. The matching procedure for trading a particular cache segment, say segment $c_{si} \in c_{\mathcal{S}i}$, operates as follows. Given the MNO's asking price ϕ_s^i, the buying bids of CPs for c_{si} are sorted in a descending order as $\theta_{s,(1)}^{(i)} \geq \theta_{s,(.)}^{(i)} \geq \ldots \geq \theta_{s,(.)}^{(i)}$. Next, equilibrium buying bids are identified, where a bid submitted by CP $k \in \mathcal{K}$ is an equilibrium if $\theta_{s,k}^{(i)} \geq \phi_s^{(i)}$. If $\theta_{s,k}^{(i)} < \phi_s^{(i)}, \forall k \in \mathcal{K}$ (i.e., there is no equilibrium price), then no trade occurs and c_{si} remains with the MNO. If there are one or more equilibrium prices, a CP with the fewest number of cache segments rented in SBS s in previous rounds is selected as the winner. This policy satisfies the fairness in trading by providing a winning chance to CPs that rented fewer cache segments in an SBS.

Algorithm 1 The DOCAT Mechanism

Data: size of caches $C_{s \in \mathcal{S}}$, number of segments l
Result: allocation matrix $\Gamma_{l \times S}$

1   divide each $C_{s \in \mathcal{S}}$ into l equally-sized segments
2   set the entries of $\Gamma_{l \times S}$ to zero
3   **foreach** auction round l **do**
4   set c_{Si} with one cache segment from each SBS
5   broker announces trading the cache segments in c_{Si}
6   MNO submits asking prices $\phi_1^i, \phi_2^i, \ldots, \phi_S^i$ ▷ Eq. (6)
7   each CP k bids $\theta_{1,k}^i, \theta_{2,k}^i, \ldots, \theta_{S,k}^i$ ▷ Eq. (12)
8   **foreach** cache segment $c_{si} \in c_{Si}$ **do**
9   sort $\theta_{s,1}^i, \theta_{s,2}^i, \ldots, \theta_{s,K}^i$ in a descending order
10   identify equilibrium prices $\theta_{s,k}^i \geq \phi_s^i, \forall k \in \mathcal{K}$
11   **if** at least one equilibrium price exists **then**
12   identify the winning CP and allocate c_{si} to it
13   charge the winning CP and pay the MNO
14   set entry Γ_{is} to the identity of the winning CP

Once the winner of cache segment c_{si} is identified, entry Γ_{is} in matrix $\Gamma_{l \times S}$ is updated with the identity of the winning CP.

Payment Strategy: In identifying the winning CPs in each round of the auction, we apply a payment method to satisfy the truthful bidding in DOCAT in a way that cheating CPs should lose their profit in trading. However, determining the payment in a double auction in a way that all the economic properties are satisfied in non-trivial [28]. In DOCAT, we employ the Vickrey–Clarke–Groves (VCG) payment model to maintain the WBB property while ensuring the IC property. Accordingly, winning CP k pays the damage done to other bidders as $\overline{\theta}_{s,k}^{(i)}$ if $\overline{\theta}_{s,k}^{(i)} \geq \phi_s^{(i)}$. This strategy gives an incentive to CPs to bid with their true bidding valuation. Algorithm 1 shows the pseudo-code of the DOCAT mechanism.

C. Analysis of DOCAT

Here, we prove that DOCAT satisfies the economic properties mentioned in Sec. IV-A. Since the pricing policy in each round of the auction is independent from others, we prove the satisfaction of each economic property in only one round.

Theorem 1: DOCAT has a polynomial complexity (CE).

Proof: As shown in Algorithm 1, $|S|$ cache segments are traded in each round of the auction in the best case, thus l loops contribute a factor of $O(|l|)$ to the algorithm complexity. In the worst case, one cache segment is traded in each round of the auction, thus the complexity of the algorithm is $O(|l| \cdot |S|)$.

Theorem 2: DOCAT maintains a weakly balanced budget (WBB).

Proof: Given the MNO's asking price $\phi_s^{(i)}$ for leasing cache segment c_{si}, the bid of winning CP $k \in \mathcal{K}$ is $\theta_{s,k}^{(i)}$, where $\theta_{s,k}^{(i)} \geq \phi_s^{(i)}$. In addition, the actual payment of CP k based on the VCG payment model is $\overline{\theta}_{s,k}^{(i)} \leq \theta_{s,k}^{(i)}$, where $\overline{\theta}_{s,k}^{(i)} \geq \phi_s^{(i)}$. Thus, the overall difference between the paying amount of CPs and charging prices by the MNO is:

$$\sum_{s \in \mathcal{S}} \sum_{i \in s} (\overline{\theta}_{s,k}^{(i)} - \phi_s^{(i)}) \geq 0 \qquad (19)$$

which is non-negative. In other words, the broker does not pay extra surplus in the auction, thus DOCAT maintains a WBB.

Theorem 3: DOCAT satisfies individual rationality (IR).

Proof: the profit of the MNO is zero if it cannot lease any caching resources (i.e., $Prof_{MNO}^t(0)=0$) [Eq. (6)]. Similarly, the profit of CP $k \in \mathcal{K}$ is zero if it cannot rent any cache resources (i.e., $Prof_k^t(0)=0$) [Eq. (12)]. If the MNO leases its cache in a SBS, it is paid the same as its asking price. In addition, the paying price of a winning CP is not more than its bidding price, according to the VCG payment model. Thus, the MNO and each CP obtain non-negative profit by participating in the auction. Thus, DOCAT satisfies IR.

Theorem 4: DOCAT achieves incentive compatibility (IC).

Proof: A truthful CP $k \in \mathcal{K}$ wins renting cache segment c_{si} by bidding value $\theta_{s,k}^{(i)}$, where it pays the second highest bidding value, which is $\overline{\theta}_{s,k}^{(i)} \leq \theta_{s,k}^{(i)}$. We consider the two possible untruthful bidding behaviors next. If CP k bids $\theta'^{(i)}_{s,k}$ where $\theta'^{(i)}_{s,k} > \theta_{s,k}^{(i)}$, CP k still can win the auction but its paying price will not change because it pays the second highest bidding value. In the second untruthful behavior, if CP k bids with bid $\theta'^{(i)}_{s,k}$ where $\theta'^{(i)}_{s,k} < \theta_{s,k}^{(i)}$, two conditions can occur; (i) CP k may win the auction but the amount of its payment does not change, (ii) CP k loses the auction, thus its profit will be zero. As a result, there is no incentive for a CP to bid untruthfully.

V. TRACE-DRIVEN EVALUATION

In this section, we present numerical results to demonstrate the effectiveness of DOCAT based on a real video dataset.

Setup: We consider the scenario illustrated in Fig. 1 with 5 CPs, with CP 1 being the most popular and CP 5 the least popular one. The MBS is located at the center of the network. Similar to [29], we assume that each MU is able to achieve the nominal download rate with the MBS and its associated SBS. Without loss of generality, we set $g_{CR}^t = g_{cost}^t = g_{QoE}^t = g_{rent}^t = 1$ in deriving the revenues and costs of the MNO and CPs (Sec. III). We consider $T=100$ time slots, each with a duration of $\triangle t=1$ hour. The values of the other parameters are shown in Table II. We use the real SVC dataset in [23] to characterize the videos of the CPs. The dataset includes 19 SVC videos, where each video is encoded in 5 qualities. The average size of layers 1 to 5 is 483, 247, 130, 72, and 46 MB, respectively. We generate the videos for each CP by selecting them at random from those in the dataset, according to a uniform distribution. The total size of the videos of each CP is about 1 TB.

Methodology: We compare the performance of DOCAT with two other schemes: an *upper bound*, namely, a modified version of DOCAT (namely, DOCAT-UB) that maximizes the social welfare without satisfying the economic properties introduced in Sec. IV-A; and a *non-truthful* version of DOCAT (namely, DOCAT-NT) that does not attempt to enforce the IC property, implying that CPs can increase their profit by submitting less than their actual valuations as bids. For all

TABLE II: Default Simulation Settings

Parameter	Value
Network radius	300 meters
Maximum download rate of MBS M: L_M	50 Mbit/s [29]
Number of SBSs: $|\mathcal{S}|$, (uniformly placed)	9
Transmission range of each SBS $s \in \mathcal{S}$: R_s	100 meters
Number of channels in each SBS $s \in \mathcal{S}$: F_s	7 [29]
Max. download rate of each channel: L_s (4G LTE)	12 Mbit/s [29]
Storage capacity of each SBS $s \in \mathcal{S}$: C_s	500 GB
Number of cache segments in each SBS $s \in \mathcal{S}$: l	20
Number of MUs: $|U|$ (randomly placed)	500
Demand rate of each MU $i \in \mathcal{U}$ per time slot: ζ_i^t	10
Number of CPs: $|\mathcal{K}|$	5
Skewness of the popularity of CPs: γ	0.8
Number of videos of each CP $k \in \mathcal{K}$: $|\mathcal{V}_k|$	1,000 [23]
Number of layers of each video: Q	5
Skewness of video popularity of each CP $k \in \mathcal{K}$: η_k	0.8

schemes, 3 of the CPs (namely, CP 1, 2, and 5) are partially malicious, i.e., they submit 50% of their bids with half of the price of their actual valuation. We run 10 independent replications for each considered scenario. The reported results show the average value obtained, together with the corresponding standard deviations as error bars. In the following, we first focus on the efficiency of the considered schemes, namely, on the social welfare of the system. We then study the profits obtained by the CPs and the MNO during the auction.

Efficiency: Fig. 4 shows the social welfare obtained by the three considered schemes. As shown in Fig. 4a, the social welfare always increases as the skewness of the CP popularity γ grows. This is because the valuation of caches by the most popular CPs significantly increases with γ. DOCAT-UB and DOCAT-NT reach the highest and lowest amount of social welfare, respectively. The social welfare achieved by DOCAT is located in between, not much lower than DOCAT-UB but considerably higher than DOCAT-NT. For instance, when γ is 0.8, the social welfare with DOCAT becomes \$1,725, which is 35% higher than DOCAT-NT and only 15% less than DOCAT-UB. This is because CPs in DOCAT have no incentive to bid less than their actual cache valuations.

Fig. 4b illustrates that the social welfare obtained by the schemes raises as the demand ratio $\zeta_i^t (\forall i \in \mathcal{U})$ increases. This is because the profit of CPs rises as an increasing number of requests from their subscribers are served by the caches. Furthermore, the social welfare achieved by DOCAT-UB and DOCAT increases faster than DOCAT-NT as ζ_i^t grows. For instance, when ζ_i^t is 30, the social welfare achieved by DOCAT is \$2,814, which is 37% higher than DOCAT-NT. The reason for is that malicious CPs value the caches less than their actual prices in DOCAT-NT; as a consequence, they fail to rent caches, thus, their profit reduces.

Last, Fig. 4c shows that the social welfare obtained by all the schemes constantly increases as the number l of segments in a cache increases too. In particular, DOCAT-UB achieves the highest social welfare and DOCAT-NT obtains the lowest
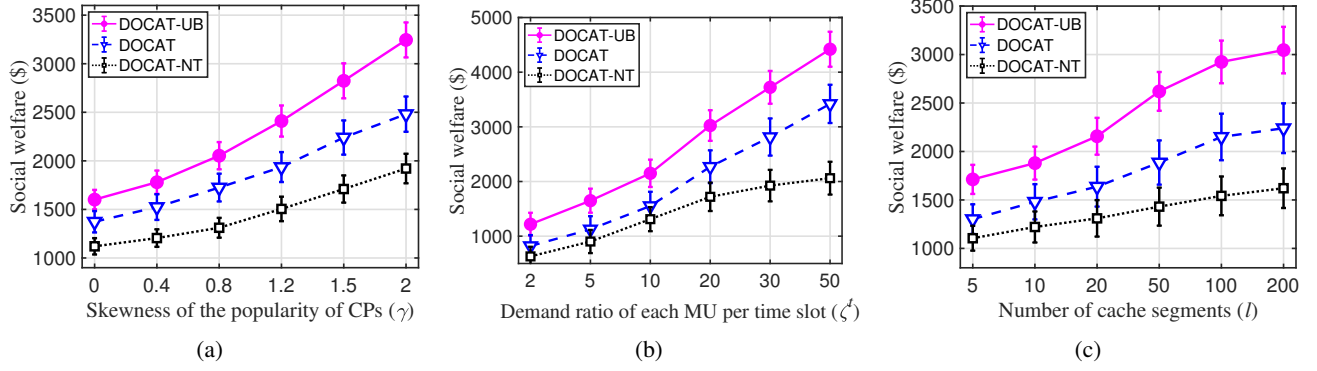
Fig. 4: The social welfare achieved by the different schemes as a function of (a) the skewness (γ) in CP popularities, (b) the demand ratio (ζ_i^t) of MUs, and (c) the number of cache segments in each SBS.
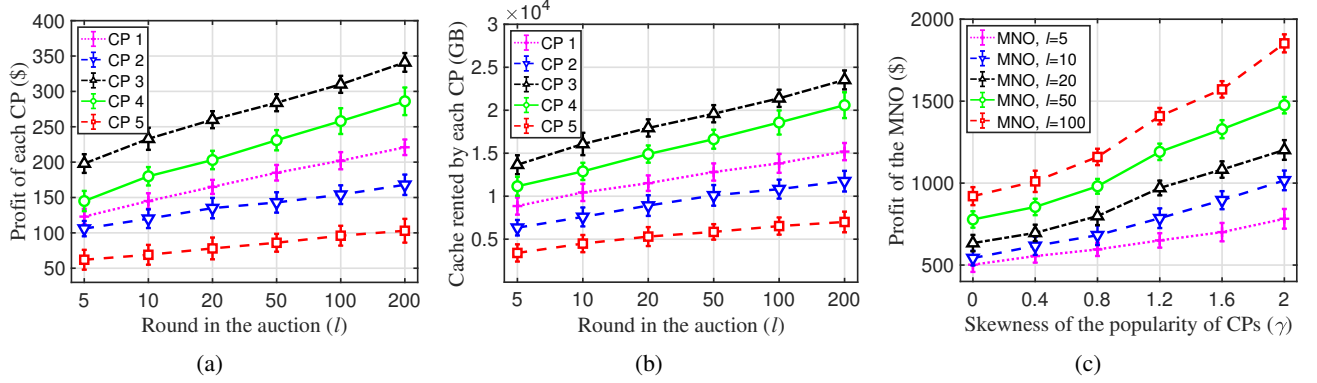

Fig. 5: (a) The profit and (b) the size of caches rented by CPs as a function of the auction rounds. (c) The profit of the MNO as a function of the skewness (γ) in CP popularities.

social welfare, both irrespective of l. Furthermore, the increase in the social welfare of DOCAT has a similar trend to that of DOCAT-UB as l grows; in contrast, DOCAT-NT exhibits a linear growth. For instance, when l increases from 20 to 100, the social welfare in DOCAT grows by about \$500, while it only increases by \$300 for DOCAT-NT. This happens as malicious CPs win less number of cache segments in DOCAT-NT when the number of rounds in the auction increases, thus, the corresponding social welfare overall reduces.

Individual profit: Fig. 5a shows the cumulative profit of CPs over the auction rounds. The figure clearly shows how honest CPs (i.e., CPs 3 and 4) obtain higher profits than malicious CPs (i.e., CPs 1, 2 and 5), even if their popularity is lower than others (e.g., CP 3 is less popular than CP 1). Within each group (i.e., honest and malicious), more popular CPs still obtain higher profits (e.g., CP 1 obtains a higher profit than CPs 2 and 5). This behavior is confirmed by Fig. 5b, which shows the cumulative size of caches rented by CPs over the auction rounds. Honest CPs rent larger caches, while malicious CPs are disadvantaged by their valuation scheme.

Fig. 5c shows the profit of the MNO as a function of the skewness (γ) in CP popularities for different numbers of the segments in the cache. It is apparent that the profit increases not only with γ, but also with the number of segments in the cache. This happens as the number of auction rounds

corresponds to the number of segments in the cache. As a consequence, the MNO has a higher chance to sell segments at higher prices because CPs that did not have a chance to get many of them (if at all) are more likely place higher bids.

VI. RELATED WORK

Some studies have addressed the implications of resource sharing between MNOs and CPs. Zhao et al. [9] applied a Stackelberg game to analyze the competition among CPs over renting the cache of an SBS. The works in [10, 11] designed incentive contracts to stimulate CPs to cache their contents in SBSs. You et al. [12] proposed an auction approach to model cache trading between an MNO and CPs. However, existing works did not study how the structure of caching contents affects the profit of MNOs and CPs in cache trading.

Caching layered videos is non-trivial because the dependency between different layers of a video has to be considered during the caching and delivery process. Zhan and Wen [15] proposed a heuristic solution to identify video layers for caching so that the video delivery delay is minimized. Zhang et al. [17] designed fractional and random caching algorithms to minimize the energy consumption in content delivery of SVC videos. Poularakis et al. [16] studied cooperative layered video caching and delivery in multi-operator cellular networks. The analysis therein showed that the cooperation of co-located SBSs can reduce the delivery delay of video downloads up to

25%. However, all the above-mentioned works did not study the economic aspects of layered video caching.

VII. Conclusion

This article studied the implications of cache trading in a commercial HetNet with multiple CPs, where each CP aims at caching a portion of their layered videos in SBSs provided by an MNO. Our objective was to achieve a trading solution that jointly maximizes the profit of the MNO and CPs. Therefore, we formulated the caching trading as a social welfare maximization problem and proposed a sequential double auction mechanism to efficiently find the perfect matches of selling caches and buying CPs in a fair manner. The main observations in our experiments are the following: CPs with high popularity achieve the highest social welfare if they bid truthfully, as they find it more valuable; the social welfare of the system increases when the cache contains many segments, which enables the broker to divide as much as possible the cache between CPs; malicious (i.e., non-truthful) CPs lose their profit even if their popularity is high. Future work may address protecting the cache trading system against collusion of CPs. Considering multiple MNOs is also a promising research direction.

References

[1] Ericsson, "Ericsson mobility report," June 2018, [Online]. Available: https://www.ericsson.com/en/mobility-report/reports/june-2018.

[2] N. Bhushan *et al.*, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Comm. Mag.*, vol. 52, no. 2, pp. 82–89, 2014.

[3] L. Li, G. Zhao, and R. S. Blum, "A survey of caching techniques in cellular networks: Research issues and challenges in content placement and delivery strategies," *IEEE Comm. Surveys & Tutorials*, 2018.

[4] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless video content delivery through distributed caching helpers," in *IEEE INFOCOM*, 2012, pp. 1107–1115.

[5] D. Liu, B. Chen, C. Yang, and A. F. Molisch, "Caching at the wireless edge: design aspects, challenges, and future directions," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 22–28, 2016.

[6] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 131–145, 2016.

[7] T. Sanguanpuak, S. Guruacharya, E. Hossain, N. Rajatheva, and M. Latva-aho, "Infrastructure sharing for mobile network operators: Analysis of trade-offs and market," *IEEE Transactions on Mobile Computing*, 2018.

[8] J. Dai *et al.*, "Collaborative caching in wireless video streaming through resource auctions," *IEEE JSAC*, vol. 30, no. 2, pp. 458–466, Feb. 2012.

[9] K. Zhao, S. Zhang, N. Zhang, Y. Zhou, Y. Zhang, and X. Shen, "Incentive mechanism for cached-enabled small cell sharing: A Stackelberg game approach," in *IEEE GLOBECOM*, 2017, pp. 1–6.

[10] K. Hamidouche, W. Saad, and M. Debbah, "Breaking the economic barrier of caching in cellular networks: Incentives and contracts," in *IEEE Global Communications Conference*, 2016, pp. 1–6.

[11] T. Liu, J. Li, F. Shu, M. Tao, W. Chen, and Z. Han, "Design of contract-based trading mechanism for a small-cell caching system," *IEEE Trans. on Wireless Comm.*, vol. 16, no. 10, pp. 6602–6617, 2017.

[12] F. You, J. Li, J. Lu, and F. Shu, "On the auction-based resource trading for a small-cell caching system," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1473–1476, 2017.

[13] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 17, no. 9, pp. 1103–1120, Sept 2007.

[14] Cisco, "Emerging video technologies: H.265, SVC, and WebRTC," 2014, [Online]. Available: https://www.ciscolive.com.

[15] C. Zhan and Z. Wen, "Content cache placement for scalable video in heterogeneous wireless network," *IEEE Comm. Lett.*, vol. 21, no. 12, pp. 2714–2717, 2017.

[16] K. Poularakis, G. Iosifidis, A. Argyriou, I. Koutsopoulos, and L. Tassiulas, "Caching and operator cooperation policies for layered video content delivery," in *IEEE INFOCOM*, 2016, pp. 1–9.

[17] X. Zhang *et al.*, "Energy-efficient caching for scalable videos in heterogeneous networks," *IEEE J. on Selected Areas in Comm.*, 2018.

[18] H. Yan *et al.*, "On the understanding of video streaming viewing behaviors across different content providers," *IEEE Trans. on Netw. and Service Manag.*, vol. 15, no. 1, pp. 444–457, 2018.

[19] S. Martello and P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*. New York, NY, USA: John Wiley & Sons, Inc., 1990.

[20] R. B. Myerson and M. A. Satterthwaite, "Efficient mechanisms for bilateral trading," *J. of Econ. Theory*, vol. 29, no. 2, pp. 265–281, 1983.

[21] J. Li, H. Chen, Y. Chen, Z. Lin, B. Vucetic, and L. Hanzo, "Pricing and resource allocation via game theory for a small-cell video caching system," *IEEE JSAC*, vol. 34, no. 8, pp. 2115–2129, 2016.

[22] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system," in *ACM Conf. on Internet Meas.*, 2007, pp. 1–14.

[23] "Video trace library," http://trace.eas.asu.edu/.

[24] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, Aug 2014.

[25] G. Ma, Z. Wang, M. Zhang, J. Ye, M. Chen, and W. Zhu, "Understanding performance of edge content caching for mobile video streaming," *IEEE J. on Selected Areas in Comm.*, vol. 35, no. 5, pp. 1076–1089, 2017.

[26] G. Iosifidis, L. Gao, J. Huang, and L. Tassiulas, "An iterative double auction for mobile data offloading," in *WiOpt*, 2013, pp. 154–161.

[27] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, 2006.

[28] T. Sandholm and S. Suri, "Market clearability," in *International Joint Conference on Artificial Intelligence*, 2001, pp. 1145–1151.

[29] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4G LTE networks," in *ACM MobiSys*, 2012, pp. 225–238.

[30] B. Jedari and M. Di Francesco, "Delay analysis of layered video caching in crowdsourced heterogeneous wireless networks," in *IEEE Global Communications Conference (GLOBECOM)*, 2018.

Appendix A

Let C_{sk} be the cache size rented by CP k in SBS s. The optimal video layers in $\mathcal{V}_k$ for caching in SBS s maximizes the aggregated delay saving of video downloads requested by the subscribers of CP k. By considering Eqs. (7)–(9) in Sec. III, such a problem can be formulated as follows:

$$\max_{X_k^s} D_{M,s,k}^{NoCache} - \left(D_{M,s,k}^{Cache}(C_{sk}) + D_{s,k}^{Cache}(C_{sk}) \right) =$$

$$\sum_{i \in \mathcal{U}_{ks}^t} \sum_{v \in \mathcal{V}_k} \sum_{q \in \mathcal{Q}} \left(\zeta_{ik}^t p_{kv} \frac{O_{kvq}}{QL_M} - \right.$$

$$\left(\left(\left(1 - \frac{x_{k,q}^{s\,1-\eta_k}}{(1-\eta_k)\sum_{j=1}^{V_k} j^{-\eta_k}} \right) \right) \zeta_{ik}^t p_{kv} \frac{O_{kvq}}{QL_M} + \right.$$

$$\left. \left. \frac{x_{k,q}^{s\,1-\eta_k}}{(1-\eta_k)\sum_{j=1}^{V_k} j^{-\eta_k}} \zeta_{ik}^t p_{kv} \frac{O_{kvq}}{QL_s} \right) \right) \quad (20)$$

subject to $C1 : \sum_{i=1}^{Q} \sum_{j=1}^{n} \sum_{q=1}^{i} o_{kzq} \leq C_{sk}$

$C2 : 0 \leq x_{k,Q}^s \leq x_{k,Q-1}^s \leq \ldots \leq x_{k,1}^s \leq V_k, (x_q^s \in \mathbb{Z}^+, \forall q (1 \leq q \leq Q))$

The details on the problem in Eq. (20) can be found in [30].