

Contract-based Cache Partitioning and Pricing Mechanism in Wireless Network Slicing

Tra Huong Thi Le¹, Nguyen H. Tran¹, Phuong Luu Vo², Zhu Han^{1,3}, Mehdi Bennis⁴, Choong Seon Hong¹

¹ Dept. of Computer Science and Engineering, Kyung Hee University, Korea.

² School of Computer Science and Engineering, International University - VNUHCM, Vietnam.

³ Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004-4005 USA.

⁴ Centre for Wireless Communications-CWC, University of Oulu, Finland.

E-mail: {huong_tra25,nguyenth,cshong}@khu.ac.kr, vtlphuong@hcmui.edu.vn, zhan2@uh.edu, bennis@ee.oulu.fi.

Abstract—In the commercial caching system, both Infrastructure Provider (InP), who owns the infrastructure and wireless network resource and Service Providers (SPs), who provide service to its users based on the virtual resource provided by the InP, are beneficial in leasing and renting the cache space. By partitioning the cache space at the BS into slices and leasing each partition to the SPs, the InP can receive a payment. Meanwhile, the SPs can serve their users with faster download service with local caching. However, both SPs and InP are selfish and want to maximize their own benefits. In addition, in practice, there is asymmetric information between SPs and InP. Thus, some SPs may declare inaccurate private information to get more cache space or less payment. To deal with these problems, in this paper, we propose an incentive mechanism based on contract theory, in which the InP, the employer, who designs and offers contracts to SPs, the employees. In particular, SPs are specified into types based on their valuation parameters and request rates. Different from the traditional contract model with two feasible contract conditions, we impose cache capacity constraints, which induces the interaction among SPs and makes the contract design more complicated. We propose an algorithm that achieves the optimal contract so that the InP can motivate SPs to participate into renting caching space while maximize its utility. Simulation results show that the proposed approach not only ensures no SP has incentive to select another contract but also outperforms the baseline allocation algorithm.

I. INTRODUCTION

The demand for mobile data is increasing rapidly, which results from the growth of mobile users and their bandwidth-hungry mobile applications such as video downloading. The global cellular data traffic is expected to increase nearly tenfold in the next five years and mobile video streaming will account for 72 percent of the total mobile data traffic by 2019. One promising approach to meet the demand of data traffic is deployment of small cell base stations (BSs) [1]. Such deployment brings the BS closer to the users, hence increases the spectral efficiency and network capacity. However, in a small cell BS, a new challenge for backhaul is raised for supporting the aggregate data rate with the reliable connection. In addition, video downloading involves repeated wireless transmission multiple times, leading to reduce video

transmission. Therefore, caching at BSs has been an essential solution with threefold advantages: alleviating the backhaul burden, avoiding congestion and reducing the delay [2].

Compared with strategy design in cooperative transmission with caching, caching placement and coding for caching [3]–[5], a fewer work has done in the cache size allocation. The work [6] considered the storage allocation problem with limited backhaul transmission. The authors investigated the cache size allocation to maximize the user success probability, considering wireless channel statistic, backhaul capacities and file popularity distribution. The authors in [7] proposed a utility-driven cache partitioning approach for multiple content providers. A cache is partitioned into slices with each partition being dedicated to a particular content provider. A formal proof partitioning the cache yields better performance compared to sharing it was also given. A similar work to [7] was explored in [8] from a game-theoretic cache allocation standpoint.

From a commercial perspective, pricing the rental of local storage is a promising topic. A commercialized caching system may consist of Service Providers (SPs) and Infrastructure Provider (InP). In such setting, the SPs revenue is acquired from providing video services to the users. By locally storing these videos, nearer users, SPs can gain profits from providing faster transmission. However, the SPs do not own cache infrastructure and need to rent local cache of the InP. This offers the InP some profitable opportunities from leasing its cache space at the BS. In this sense, both the SPs and InPs are beneficial of the local caching system. However, each entity is selfish and wants to get maximal benefit, raising a competition problem among these entities, which can be solved effectively by applying game theory. Game theory, a well-developed mathematical tool that studies the interactions among rational users, has been used to analyze the behaviors of caching system. Recently, the works in [9] and [10] used the Stackelberg game to investigate economics caching. However, these works typically assume that the players are truthful and reveal all private information, which is not always a valid assumption in practice. For example, some SPs can declare inaccurate information about their request rates so as to defraud the InP into charging them much lower prices or give them more cache space. Such possibilities can therefore incentivize some of the SPs not to share their truthful private information, which makes the results in [9] and [10] no longer hold. This problem is information asymmetry between InP and SPs. To this end, to handle information asymmetry issue, we turn to contract

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (R0126-16-1009, Development of Smart Mediator for Mashup Service and Information Sharing among ICBMS Platform) *Dr. CS Hong is the corresponding author.

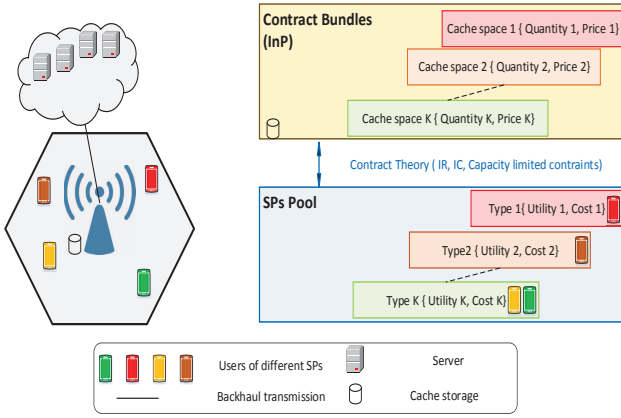


Fig. 1: System model

theory [11]–[13]. Using contract theory, one can analyze the interactions between an employer who is trying to offer proper contracts to employees whose skills are not known a priori. In a caching system, this contractual situation can be used to study the interactions between the InP, acting as employers and, the SPs, whose valuation and request rate are unknown to the BSs [14]. The contribution of the paper is to propose a partitioning and pricing mechanism for the InP so that it can optimally lease cache space to the SPs without knowing the SPs’ exact types. The problem is formulated using a contract model in which the InP calculates the fraction of cache space lease to each type of SPs and corresponding payment. Specially, we classify the SPs into types depending on their valuation parameters and request rates. For the formulated problem, we analyse the condition for derive the optimal contract. We also propose the algorithm to get the optimal contract which maximizes the InP’s utility and motivate the SPs to leverage their private information and choose a proper contract intended for their types.

The rest of this paper is organized as follows. The system model is provided in Section II, followed by the utility model in Section III. The contract formulation is presented in Section IV. The simulation results are shown in Section V. Finally, conclusions are drawn in Section VI.

II. SYSTEM MODEL

A. Network Model

We consider a caching system which includes one InP and S SPs as shown in Fig. 1. The InP owns a single cell network in which the BS transmits data to users of SPs. The cache space at the BS has capacity of Q . The InP partitions the cache space into virtual slices and then leases them to SPs. Traditionally, in the virtualization wireless network, the resources of the InP are subchannel, power and antennas [15]. To the best of our knowledge, this is the first work that deals with caching resources of the InP. When user of SP k requests a file, it can be served directly from the BS if the requested file is cached at the BS. Otherwise, the BS has to fetch the file from the SP’s server through limited capacity backhaul before transmitting this file to user.

B. Caching Model

The caching policy in each slices is assumed time to live caches (TTL caches), in which the eviction of file $f \in \mathcal{F}$ occurs upon the expiration of timer T_f . Based on the previous work on the analysis of TTL caches [16], the hit probability of file f can be expressed as $h_f = 1 - e^{-\lambda_f T_f} \approx \lambda_f T_f$. The average hit rate for the set of N files through TTL cache can be given by $H = \sum_{f=1}^N \lambda_f h_f \approx \sum_{f=1}^N (\lambda_f)^2 T_f$. Assume files in \mathcal{F} have equal TTL values, i.e., $T_f = T, \forall f$, which is consistent with Che’s approximation [16]. We can rewrite the hit rate as follows $H \approx \sum_{f=1}^N (\lambda_f)^2 T$.

III. UTILITY MODEL

A. Utility of SP k

Each SP k has file library $\mathcal{F}_k = \{f_{1k}, f_{2k}, \dots, f_{N_k k}\}$ of N_k equal size files. We denote the size of each file as l . The request for file $f_{ik} \in \mathcal{F}_k$ is modeled as the inhomogeneous Poisson point process with rate λ_{ik} . Suppose SP k rents a fraction β_k of the BS cache space. Thus, we have $\sum_{i=1}^{N_k} h_{ik} = \beta_k Q$, where h_{ik} is the hit probability of file f_{ik} with the timer T_{ik} of SP k , i.e. $h_{ik} = 1 - e^{-\lambda_{ik} T_{ik}} \approx \lambda_{ik} T_{ik}$. Assume $T_{ik} = T_k$, we have $T_k = \frac{\beta_k Q}{(\sum_{i=1}^{N_k} \lambda_{ik})}$. Thus, the hit rate of SP k can be expressed as:

$$H_k = \sum_{i=1}^{N_k} \lambda_{ik} h_{ik} = \frac{\sum_{i=1}^{N_k} \lambda_{ik}^2 \beta_k Q}{\sum_{i=1}^{N_k} \lambda_{ik}}. \quad (1)$$

Correspondingly, the miss rate of SP k is $M_k = \sum \lambda_{ik} - H_k = \sum_{i=1}^{N_k} \lambda_{ik} - \frac{\sum_{i=1}^{N_k} \lambda_{ik}^2 \beta_k Q}{\sum_{i=1}^{N_k} \lambda_{ik}}$. We can see that hit rate H_k and miss rate M_k depend on the fraction β_k of cache space that SP k rent.

When the file requested from user of SP k is cached at the BS then request is served directly. Therefore, SP k can make profit by providing the faster service. Let ω_k denote SP k ’s content hit rate, we have the valuation function of SP k as

$$V_k(\beta_k) = \omega_k H_k = \omega_k \frac{\sum_{i=1}^{N_k} \lambda_{ik}^2}{\sum_{i=1}^{N_k} \lambda_{ik}} \beta_k Q. \quad (2)$$

We note that ω_k and $\frac{\sum_{i=1}^{N_k} \lambda_{ik}^2}{\sum_{i=1}^{N_k} \lambda_{ik}}$ are personal parameters for SP k . We further define $\theta_k = \omega_k \frac{\sum_{i=1}^{N_k} \lambda_{ik}^2}{\sum_{i=1}^{N_k} \lambda_{ik}}$ as the type of SP k , which depends on the valuation parameter and traffic load of SP k . The physical meaning of θ_k is the hit rate H_k of SP k can turn into revenue. We also define the utility of the SP k as the difference between valuation function of renting resource of SP k and the price charged by the InP as follows:

$$U_k(\beta_k, P_k) = V_k(\beta_k) - P_k = \theta_k \beta_k Q - P_k. \quad (3)$$

We assume there are total K different types and we denote the types by the set $\mathcal{K} = \{\theta_1, \theta_2, \dots, \theta_K\}$. Without loss of generality, we assume $\theta_1 < \dots < \theta_k < \dots < \theta_K$. The InP does not have exact information on the individual θ_k of every SP k . Instead, the InP can know with which probability a SP can belong to a certain type. Therefore, we let q_k be the probability with which SP belongs to a certain type k .

B. Utility of InP

The utility of the InP is determined by the difference between the payment from leasing cache to all SPs and the backhaul delay cost. When file requested from user of SP k is cached at the BS and then request is served directly. Otherwise, the file has to be retrieved from server through the backhaul link with transmission rate of D_0 , and then is transmitted to users, which induces backhaul delay cost. If we model the backhaul transmission missed file of SP k as an M/M/1 queue with mean of service time of $\frac{l}{D_0}$. Then backhaul delay of missed files of SP k is

$$\tau_k(\beta_k) = \frac{M_k}{\frac{D_0}{l} - M_k}. \quad (4)$$

Thus, the utility of the InP is expressed as

$$R(\beta, P) = \sum_{k=1}^K q_k(P_k - \tau_k). \quad (5)$$

The selfish and rational InP, who wants to maximize its utility, has to apply different strategies to different types of the SPs, since the SPs of different types have different valuation of cache renting or different willing to renting cache. In the caching trading market, the private information of SPs is unknown by the InP. Some SPs may claim wrong information valuation parameter and/or request rate so that the cache partition capacity allocate for it is higher or the price charged is lower. Thus, it is difficult for the InP to know exactly the type of a certain SP. It is well known that contract theory is an useful technique in modeling the incentive mechanism under asymmetry information. In the following, the interactions of the InP and SPs are formulated as a game and based on the framework of contract theory, we design a feasible contract as solution.

IV. CONTRACT FORMULATION

A. Contract Formulation

We denote the InP's contract designed for type k - SP as (β_k, P_k) . To incentivize SP type k to rent cache space, it must be individual rational (IR), i.e.,

$$\theta_k \beta_k Q - P_k \geq 0, \forall k \in \mathcal{K}. \quad (6)$$

The contracts should bring a non-negative utility to the SP, which motivate the SP to actively participate in the trade. In addition, SP type k would be prefer to choose the contact item (β_k, P_k) rather than contract item design for other types. So it should be incentive compatible (IC), i.e.,

$$\theta_k \beta_k Q - P_k \geq \theta_k \beta_j Q - P_j, \forall k, j \in \mathcal{K}, k \neq j. \quad (7)$$

The IC condition ensures that the SP will accept the contract designed for it rather than choosing the other contracts. These two constraints guarantee that the optimal contract can provide participation incentive for users.

Besides the IC and IR constraints, the InP will need to design the contract such that the expected cache size leased meets the cache size limit of the BS, i.e.,

$$S \left(\sum_{k=1}^K q_k \beta_k \right) = 1. \quad (8)$$

The probability q_k can be interpreted as the proportion of SPs of type θ_k . Thus, Sq_k can be considered as the number of SPs of type θ_k . The left-hand side of the constraint in (12) is the total fraction of cache that can be leased to K types of SPs, which is equal 1.

We assume that the InP does not know exactly the type of a particular SPs but knows the percentage of each type. We can obtain these percentage of each type of SPs by applying the online learning. The InP offers contract items designed not only to make profit (exploit) but also learn the SP's types (explore) by record if the contract items is accepted or rejected. During learning, the InP has a tradeoff between choosing the best contract item to maximize its utility or trying different item to improve the estimation. To find the solutions to such trade off we apply multi-armed bandit solution and to obtain the percentage of SPs of each type [17]. The optimal contract design can be formulated as the InP's utility maximization problem, i.e.,

$$\begin{aligned} \max_{\beta, P} R(\beta, P) &= \sum_{k=1}^K q_k (P_k - \tau_k), \\ \text{s.t.} &: (6), (7), (8), 1 \geq \beta_k \geq 0. \end{aligned} \quad (9)$$

where β and P are denoted as vector of cache fraction and payment, respectively.

B. Simplifying Contracts

The main difficulty in solving the problem in (9) is that there are a large number of IR and IC constraints (K IR constraints and $K \times (K - 1)$ IC constraints). Therefore, to make (9) more tractable, we need to simplify the IR and IC constraints before solving it.

Lemma 1. *For any feasible contract (β, P) , $\beta_i \geq \beta_j$ if and only if $\theta_i \geq \theta_j$.*

Proof: We can refer [14] for the proof. This Lemma simply proves that the InP must provide more cache capacity to the SPs with higher types, which is the monotonicity property of contract design. ■

Lemma 2. *For any feasible contract (β, P) , the utility of each type of SPs must satisfy*

$$0 < U_1 < \dots < U_k < \dots < U_K. \quad (10)$$

Proof: We can refer [14] for the proof. ■

Thus, the higher type SPs receive more utility than the lower type SPs. From the IC constraint and the two lemmas above, we can easily infer the following. If a SP selects the contract designed for a high type, even though it receive more cache space, the profit of the hit rate can not compensate the payment to the InP. Moreover, if a SP selects a contract intended for a low type, although less payment to the InP, the less cache space it gets. The SP can receive the maximum utility if and only if it selects the contract that best fit into its preference. Thus, we can guarantee that the contract is truthfully self reveal.

Lemma 3. (Reduce IR Constraints): *Given that the IC constraint holds, for the optimal contract under incomplete*

information, the IR constraint can be reduced by

$$\theta_1 \beta_1 Q - P_1 = 0. \quad (11)$$

Proof: With definition of types : $\theta_1 < \dots < \theta_k < \dots < \theta_K$, we have $\theta_k \beta_k Q - P_k \geq \theta_k \beta_1 Q - P_1 \geq \theta_1 \beta_1 Q - P_1 \geq 0$. So, if the IR constraint of type-1 user is satisfied, the other IR constraints will automatically hold. Moreover, to maximize its utility, the InP will adapt its schedule by raising price P_k as much as possible, leading to the decrease of utility of SP type 1. Accordingly, the condition in (11) will be satisfied when contract is optimal. ■

Lemma 4. (Reduce LDIC Constraints): *If Local Downward Incentive Constraints (LDIC) are satisfied for all user type $\theta_k, k \in \{1, 2, \dots, K\}$, i.e.,*

$$\theta_k \beta_k Q - P_k \geq \theta_k \beta_{k-1} Q - P_{k-1} \quad (12)$$

then IC constraints will hold for any $h \leq k$, i.e.,

$$\theta_k \beta_k Q - P_k \geq \theta_k \beta_h Q - P_h. \quad (13)$$

Proof: We have two LDIC as follows:

$$\theta_k \beta_k Q - P_k \geq \theta_k \beta_{k-1} Q - P_{k-1}. \quad (14)$$

$$\theta_{k-1} \beta_{k-1} Q - P_{k-1} \geq \theta_{k-1} \beta_{k-2} Q - P_{k-2}. \quad (15)$$

We have $\theta_k > \theta_{k-1}$, so the inequality in (15) becomes

$$\begin{aligned} \theta_k (\beta_{k-1} - \beta_{k-2}) Q &\geq \theta_{k-1} (\beta_{k-1} - \beta_{k-2}) Q \\ &\geq P_{k-1} - P_{k-2}. \end{aligned} \quad (16)$$

Additionally, (12) is equivalent to

$$\theta_k (\beta_k - \beta_{k-1}) Q \geq P_k - P_{k-1}. \quad (17)$$

Summing (16) and (17), we have:

$$\theta_k \beta_k Q - P_k \geq \theta_k \beta_{k-2} Q - P_{k-2}. \quad (18)$$

Therefore, if for type- $k-1$ SP the LDIC holds, the incentive constraint with respect to type $k-2$ hold. This process can be extended downward from type $k-2$ to 1 SPs, which prove that all the downward incentive constraints hold. In view of the random selection of θ_k , we have completed the proof. ■

Lemma 5. (Reduce LUIC Constraints): *If Local Upward Incentive Constraints (LUIC) are satisfied for all user type $\theta_k, k \in \{1, 2, \dots, K\}$, i.e.,*

$$\begin{aligned} \theta_k \beta_k Q - P_k &\geq \theta_k \beta_{k+1} Q - P_{k+1}, \\ &\text{for all type } \theta_k, k \in \{1, 2, \dots, K\} \end{aligned} \quad (19)$$

then IC constraints will be satisfied for any $l \geq k$,

$$\theta_k \beta_k Q - P_k \geq \theta_k \beta_l Q - P_l. \quad (20)$$

Proof: Proof is similar to Lemma 4. ■

Lemma 6. *Let $\{\beta_k, P_k\}$ be a feasible contract then the unique optimal price satisfy:*

$$\begin{aligned} P_1^* &= \theta_1 \beta_1 Q \\ P_k^* &= P_{k-1}^* + \theta_k (\beta_k Q - \beta_{k-1} Q). \end{aligned} \quad (21)$$

Proof: Let us proceed contradiction. Given the fixed cache allocation, the utility of the InP is decided by $\sum_{k=1}^K q_k P_k$.

Suppose that there exists another feasible payment $\{P'_k, \forall k\}$ which better solution than $\{P_k^*, \forall k\}$ in (21). Thus, there is at least one price $P'_k > P_k^*$ for one type θ_k . If $k = 1$, then $P'_1 > P_1^*$. Since $P_1^* = \theta_1 \beta_1 Q$, then $P'_1 > \theta_1 \beta_1 Q$, which violates the IR constraints for type θ_1 . If $k > 1$. Since $\{P'_k, \forall k\}$ must satisfy the LDIC: $\theta_k \beta_k Q - P'_k \geq \theta_k \beta_{k-1} Q - P'_{k-1}$ or $P'_k \leq P'_{k-1} + \theta_k (\beta_k Q - \beta_{k-1} Q)$. By substituting $P_k^* = P_{k-1}^* + \theta_k (\beta_k Q - \beta_{k-1} Q)$ into this equality, we have $P'_{k-1} \geq P_{k-1}^*$. By the induction method, we have $P'_1 > P_1^*$, which violates the IR constraint for type θ_1 again. So we have ??.

By iterating (21) we can conclude that

$$P_k^* = \theta_1 \beta_1 Q + \sum_{i=1}^k Z_i \quad (22)$$

for any $k \in \{1, 2, \dots, K\}$, where

$$Z_i = \begin{cases} 0, & \text{if } i = 1, \\ \beta_i Q - \beta_{i-1} Q, & \text{if } i = 2, 3, \dots, K. \end{cases}$$

C. Optimality of Contract

Based on the above lemmas, we can present the problem (9) as followed:

$$\begin{aligned} \max_{\{\beta\}} R &= \sum_k q_k \{ \theta_1 \beta_1 Q + \sum_{i=1}^k Z_i - \tau_k \} \\ \text{s.t. : } &\beta_1 \leq \beta_2 \leq \dots \leq \beta_K, (8) \text{ and } 1 \geq \beta_k \geq 0. \end{aligned} \quad (23)$$

Solving problem (23) is summarized in phase 1 of Algorithm 1. Detail analysis is explained as follows. In order to solve this problem, one standard approach is to ignore the monotonicity condition at first and then to check whether the obtained solution satisfies this condition.

After removing the monotonicity condition, we can represent the objective function of problem (23) as follows:

$$\begin{aligned} \max_{\{\beta\}} R &= \sum_{k=1}^K R_k \\ \text{s.t. : } &(8) \text{ and } 1 \geq \beta_k \geq 0. \end{aligned} \quad (24)$$

where $R_k = q_k \theta_k \beta_k Q + \Delta_k \sum_{i=k+1}^K q_i - q_k \tau_k$ and $\Delta_k = (\theta_k - \theta_{k+1}) \beta_k Q, \forall k < K$ and $\Delta_k = 0$ when $k = K$. Obviously, all $R_k, \forall k \in K$ are the convex function of β_k . We have the Langrangian function of (24) as:

$$L = \sum_{k=1}^K R_k + \mu (S \left(\sum_{k=1}^K q_k \beta_k \right) - 1) + \sum_{k=1}^K \nu_k \beta_k \quad (25)$$

where μ, ν_k are the Langrange multipliers. Applying KKT condition, we have $\beta_k^*, k \in K$ are solutions of

$$\begin{cases} R'_k(\beta_k) + \mu q_k + \nu_k = 0, & \forall k \in \mathcal{K}, \\ \nu_k \beta_k = 0, & \forall k \in \mathcal{K}, \\ M \left(\sum_{k=1}^K q_k \beta_k \right) = 1, \end{cases}$$

where R'_k is the first order derivative of R_k . Furthermore, we need to check if these solutions satisfy the monotonicity condition. If β_k^* satisfies the monotonicity condition, it can

Algorithm 1: Optimal Contract Algorithm

Phase 1: Partitioning
for $k = 1, \dots, K$ **do**

 set $R_k = q_k \theta_k \beta_k Q + \Delta_k \sum_{i=k+1}^K q_i - q_k \tau_k$

 set $\beta_k^* = \begin{cases} R'_k(\beta_k) + \mu q_k + \nu_k = 0 \\ \nu_k \beta_k = 0 \\ M(\sum_{k=1}^K q_k \beta_k) = 1. \end{cases}$
end
while $\{\beta_k^*\}$ *is not feasible* **do**

 find an infeasible subsequence $\{\overline{\beta}_i^*, \overline{\beta}_{i+1}^*, \dots, \overline{\beta}_j^*\}$

 set $\beta_i^* = \beta_{i+1}^* = \dots = \beta_j^* = \frac{q_k \overline{\beta}_i^* + q_{i+1} \overline{\beta}_{i+1}^* + q_j \overline{\beta}_j^*}{q_i + q_{i+1} + \dots + q_j}$
end
Phase 2: Pricing
for $k = 1, \dots, K$ **do**

 set $Z_i^* = \begin{cases} 0, & \text{if } i = 1, \\ \beta_i^* Q - \beta_{i-1}^* Q, & \text{if } i = 2, 3, \dots, K. \end{cases}$

 set $P_k^* = \theta_1 \beta_1^* Q + \sum_{i=1}^k Z_i^*$
end

be regarded as our desired optimal quality β_k^* . Otherwise, we need to make some adjustments, which based on following proposition:

Proposition 1. *Let $X_1(x)$ and $X_2(x)$ be concave functions on x . If $\widehat{x}_1 \geq \widehat{x}_2$ where $\widehat{x}_1 = \operatorname{argmax}_{x_1} X_1(x_1)$ and $\widehat{x}_2 = \operatorname{argmax}_{x_2} X_2(x_2)$, then $\widehat{x}_1 = \widehat{x}_2$ where*

$$\{\widehat{x}_1, \widehat{x}_2\} = \operatorname{argmax}_{x_1, x_2} \sum_{i=1}^2 X_i(x_i) \text{ s.t. } x_1 \leq x_2.$$

Proof: We can refer to [18] for the detailed proof of this Proposition 1. ■

Proposition 1 can be extended to a more general form: if $\widehat{x}_1 \widehat{x}_2 \dots \widehat{x}_K$ where $\widehat{x}_i = \operatorname{argmax}_{x_i} X_i(x_i)$, then $\widehat{x}_1 = \widehat{x}_2 = \dots =$

\widehat{x}_K where $\{\widehat{x}_k\} = \operatorname{argmax}_{x_i} \sum_{i=1}^K X_i(x_i) \text{ s.t. } x_1 \leq x_2 \leq \dots \leq$

x_K . We denote a subsequence of $\{\beta_k^*\}$, say $\{\overline{\beta}_i^*, \overline{\beta}_{i+1}^*, \dots, \overline{\beta}_j^*\}$ as an infeasible subsequence, if $\overline{\beta}_i^* \geq \overline{\beta}_{i+1}^* \geq \dots \geq \overline{\beta}_j^*$. For example, in a quality allocation $\{\beta_k^*\} = \{1, 4, 4, 3, 7, 5\}$, there are two feasible subsequence, i.e., $\{\beta_2^*, \beta_3^*, \beta_4^*\}$ and $\{\beta_5^*, \beta_6^*\}$. According to Proposition 1, adjusted values satisfy $\beta_i^* = \beta_{i+1}^* = \dots = \beta_j^*$. Moreover, $\beta_i^*, \beta_{i+1}^*, \dots, \beta_j^*$ should keep the capacity constraints, means that

$$q_i \beta_i^* + q_{i+1} \beta_{i+1}^* + \dots + q_j \beta_j^* = q_i \overline{\beta}_i^* + q_{i+1} \overline{\beta}_{i+1}^* + \dots + q_j \overline{\beta}_j^*.$$

Therefore, $\beta_i^* = \beta_{i+1}^* = \dots = \beta_j^* = \frac{q_k \overline{\beta}_i^* + q_{i+1} \overline{\beta}_{i+1}^* + q_j \overline{\beta}_j^*}{q_i + q_{i+1} + \dots + q_j}$. Substituting the feasible allocation $\{\beta_k^*\}$ into (22), we obtain the corresponding optimal price P_k^* as follows:

$$P_k^* = \theta_1 \beta_1^* Q + \sum_{i=1}^k Z_i^* \quad (26)$$

for any $k \in \{1, 2, \dots, K\}$, where

$$Z_i^* = \begin{cases} 0, & \text{if } i = 1, \\ \beta_i^* Q - \beta_{i-1}^* Q, & \text{if } i = 2, 3, \dots, K. \end{cases}$$

This is pricing phase of Algorithm 1. Therefore, we have the optimal contract algorithm as shown in Algorithm 1.

V. SIMULATION RESULTS

For our simulations, we consider $M = 7$ SPs with $K = 7$ types of SPs. We consider a uniform distribution of SP types, i.e., $q_k = 1/K$. The cache capacity of the BS is $Q = 2 \times 10^5$. seven SPs serve $2 \times 10^5, 2.3 \times 10^5, 3.4 \times 10^5, 4.5 \times 10^5, 4.6 \times 10^5, 4.7 \times 10^5, 4.8 \times 10^5$ contents, respectively.

Firstly, we check the feasibility of our contract. Fig.2a shows the utility of SPs when SP selects the contract designed by the InP for each type. From Fig. 2a, one SP gets the maximum utility when this SP chooses the contract designed for its type, which is suitable for the IC condition of the contract design. Additionally, when each SP chooses the optimal contract designed for its type, the utility of SP increases according its type. For example, at the optimal contract, SP type 1, 4, 7 receive utility of 0, 4, 15.8, respectively. This result is an evidence of Lemma 2, in which the higher type SP will get the higher utility. Fig. 2b illustrates the fraction of cache divided among types of SPs and the corresponding payment. The higher type SPs get the higher fraction of cache and pay more for it. Fig. 2c compares the case when cache allocation is deployed based on contract and the case when cache space is equally allocated among SPs. While the latter scheme outperforms the former one for lower type, the former scheme outperforms for the higher types. This is as a result of the cache capacity limit constrain. Moreover, it demonstrates for interdependence among SPs. When the cache fraction for one certain type SPs increases, their utilities increases. However, this means cache for other types decreases, or their utilities decreases.

Secondly, we evaluate the performance of the caching system by seeing the impact of different parameters on the utility of the SPs and InP. We compare the proposed mechanism with two baseline mechanisms : equal allocation and DC allocation [11]. Fig. 2d shows the utility of the InP in the our scheme is higher than in the baseline scheme, which is suitable for the target of design contract in order to maximize the utility of the InP. In addition, when Q increases, more cache space for lease, more payment the InP gets, therefore, the InP's utility increases. Fig. 2e presents the backhaul delay of three schemes during changing the cache capacity of Q . When Q increases, more files are cached, so the backhaul delay decreases. However, in our scheme, the backhaul delay decreasing is just insignificant. The reason is that in our scheme, the large portions of cache space is taken the SPs with the higher types while just small portions of cache space is offered for small type SP. Therefore, while the backhaul delay of higher type SPs decrease, the backhaul delay of smaller SP almost remains. Fig. 2f shows the utility of the InP when the number of types of SPs increases. Although the number of types increases, the backhaul capacity, and cache capacity are fixed. Therefore, the backhaul delay increases, which results in the InP's utility decreases. In addition, the utility of the proposed scheme is the best among the three observed schemes.

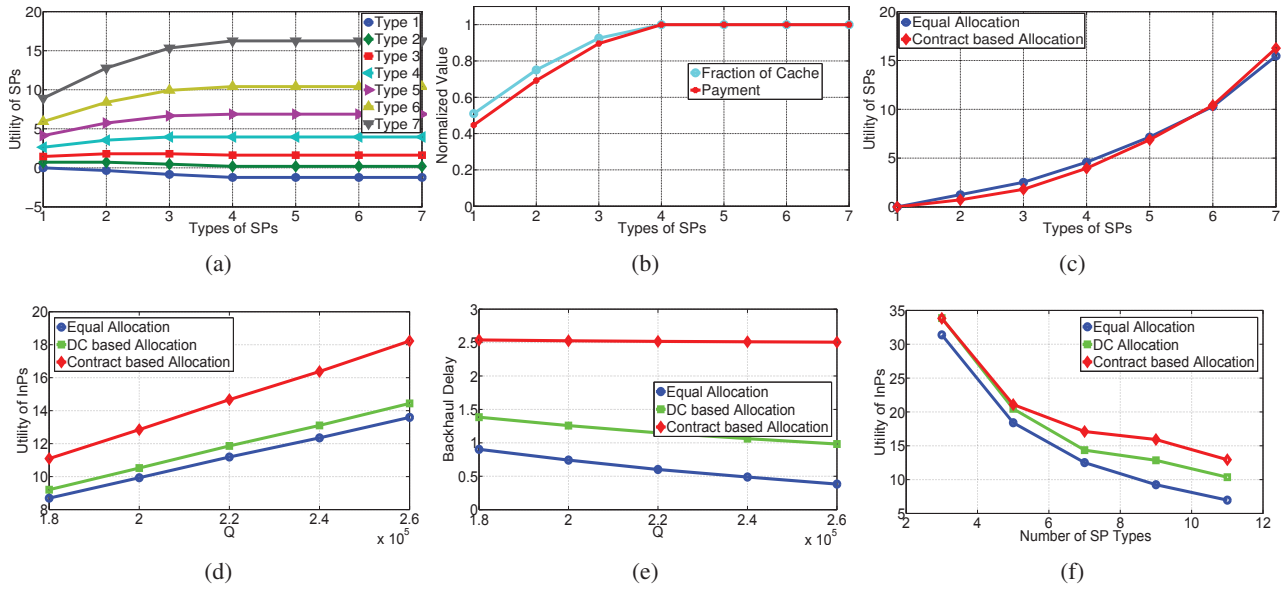


Fig. 2: Numerical results for a) The Utility of SP with contract designed for each type of SP, b) The fraction of cache and payment for each type of SPs, c) The Utility of SPs with respect to each type of SPs, d) The InP's Utility with respect to cache capacity, e) The backhaul delay of all SPs with respect to cache capacity, f) The InP's Utility with respect to the number of types of SPs.

VI. CONCLUSION

In this paper, we have proposed a new incentive framework to motivate the SPs to rent cache space at the BS of the InP. Based on contract theory, we have designed an incentive mechanism that allows the InPs to offer a contract for each SP type in which it sets the allocated storage for the SP and corresponding payment of caching service. This model accounts for asymmetry information. We have then derived the optimal pricing mechanisms and contracts that motivate the CPs to cache their content and reveal their private information while considering the cache capacity limit at the BS. Simulations have shown the advantage of proposed mechanism in motivating the participation of SPs in caching.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2016R1D1A1B01015320)

REFERENCES

- [1] A. Anpalagan, M. Bennis, and R. Vannithamby, *Design and Deployment of Small Cell Networks*. Cambridge University Press, 2015.
- [2] 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, 2014.
- [3] Z. Chen, J. Lee, T. Q. Quek, and M. Kountouris, "Cooperative caching and transmission design in cluster-centric small cell networks," *arXiv preprint arXiv:1601.00321*, 2016.
- [4] J. Sung, M. Kim, K. Lim, and J. K. Rhee, "Efficient cache placement strategy in two-tier wireless content delivery network," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1163–1174, 2016.
- [5] K. Poularakis and L. Tassiulas, "Code, cache and deliver on the move: A novel caching paradigm in hyper-dense small-cell networks," *IEEE Transactions on Mobile Computing*, 2016.
- [6] X. Peng, J. Zhang, S. Song, and K. B. Letaief, "Cache size allocation in backhaul limited wireless networks," in *Proceedings of IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.
- [7] W. Chu, M. Dehghan, D. Towsley, and Z.-L. Zhang, "On allocating cache resources to content providers," in *Proceedings of the 2016 conference on 3rd ACM Conference on Information-Centric Networking*, 2016, pp. 154–159.
- [8] S. Hoteit, M. El Chamie, D. Saucez, and S. Secci, "On fair network cache allocation to content providers," *Computer Networks*, vol. 103, pp. 129–142, 2016.
- [9] 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 Journal on Selected Areas in Communications*, vol. 34, no. 8, pp. 2115–2129, 2016.
- [10] F. Shen, K. Hamidouche, E. Baştuğ, and M. Debbah, "A stackelberg game for incentive proactive caching mechanisms in wireless networks," *arXiv preprint arXiv:1609.02596*, 2016.
- [11] L. Duan, L. Gao, and J. Huang, "Cooperative spectrum sharing: A contract-based approach," *IEEE Transactions on Mobile Computing*, vol. 13, no. 1, pp. 174–187, 2014.
- [12] Y. Zhang, L. Song, W. Saad, Z. Dawy, and Z. Han, "Contract-based incentive mechanisms for device-to-device communications in cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2144–2155, 2015.
- [13] B. Zhang, C. Jiang, J.-L. Yu, and Z. Han, "A contract game for direct energy trading in smart grid," *IEEE Transactions on Smart Grid*, 2016.
- [14] P. Bolton and M. Dewatripont, *Contract theory*. Cambridge, MA, USA: MIT press, 2005.
- [15] K. Zhu and E. Hossain, "Virtualization of 5g cellular networks as a hierarchical combinatorial auction," *IEEE Transactions on Mobile Computing*, vol. 15, no. 10, pp. 2640–2654, 2016.
- [16] N. C. Fofack, M. Dehghan, D. Towsley, M. Badov, and D. L. Goeckel, "On the performance of general cache networks," in *Proceedings of the 8th International Conference on Performance Evaluation Methodologies and Tools*, 2014, pp. 106–113.
- [17] S. Maghsudi and E. Hossain, "Multi-armed bandits with application to 5g small cells," *IEEE Wireless Communications*, vol. 23, no. 3, pp. 64–73, 2016.
- [18] L. Gao, X. Wang, Y. Xu, and Q. Zhang, "Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 843–855, 2011.