5G Network Communication, Caching, and Computing Algorithms Based on the Two-Tier Game Model

Sungwook Kim

In this study, we developed hybrid control algorithms in smart base stations (SBSs) along with devised communication, caching, and computing techniques. In the proposed scheme, SBSs are equipped with computing power and data storage to collectively offload the computation from mobile user equipment and to cache the data from clouds. To combine in a refined manner the communication, caching, and computing algorithms, game theory is adopted to characterize competitive and cooperative interactions. The main contribution of our proposed scheme is to illuminate the ultimate synergy behind a fully integrated approach, while providing excellent adaptability and flexibility to satisfy the different performance requirements. Simulation results demonstrate that the proposed approach can outperform existing schemes by approximately 5% to 15% in terms of bandwidth utilization, access delay, and system throughput.

Keywords: 5G network, Bargaining models, Cache splitting, Computation offloading, Two-tier game approach, Stackelberg game.

Manuscript received July 31, 2017; revised Nov. 23, 2017; accepted Nov. 30, 2017.

Sungwook Kim (corresponding author, swkim01@sogang.ac.kr) is with the Department of Computer Science and Engineering, Sogang University, Seoul, Rep. of Korea.

I. Introduction

As the next-generation of mobile networks, 5G networks are intended to increase the network's capacity to support the immense amount of global mobile traffic services. Along with recent advances in mobile communication technologies, 5G networks are expected to be well suited to addressing the increasing number of wireless devices as well as the need for omnipresent network access. Among the main characteristics of the 5G network, its data rates are tens of megabits per second for tens of thousands of users; it provides several hundreds of thousands of simultaneous wireless connections; it offers enhanced spectral and signaling efficiency; it improves coverage; and it reduces latency [1]. However, the capacities of wireless links, front-haul radio access networks, and back-haul core networks are unable to match the extraordinarily rapid growth of traffic services owing to the limitations of traditional cellular network architectures [2]. In 4G network architecture, the main concern is to efficiently deploy the network resource for the system performance maximization. This classic approach cannot adaptively meet end-user demands that are repeatedly presented by multiple users because each request should be processed with different quality of service (QoS) requirements [2], [3].

With the development of mobile user equipment (UE), new types of computation-intensive and delay-sensitive mobile applications are drawing increasing attention, such as augmented reality and recognition assistance [4]. However, UE is usually resource-constrained; it has a limited computation capacity. As a result, it is difficult to smoothly implement these computation-intensive applications at the UE. To solve this problem, cloud computing can be used to expand mobile device resources. Cloud computing may be viewed as a decentralized proxy cloud server that offers cloud services to mobile UE while enriching a limited available infrastructure. Fueled by this cloud-based offloading technique, the communicating and computing functionalities are beginning to converge in the so-called 5G ecosystem, thereby evoking the notion of "computing for communications" [5], [6].

The inter-communications between applications running UE and remote cloud servers act as bridges in the cooperation of UE devices and their corresponding clouds. In this inter-communication mechanism, computation execution and data access are always tightly coupled. Usually, applications running UE access their required data from the cloud server. In this process, a transmission delay will inevitably occur To reduce the delay latency, data caching technology was introduced Data caching can greatly reduce the number of duplicate data transmissions while preventing the front-haul capacity bottleneck [7], [8]. With the data caching technique for computation services, content caching for communication services is considered for popular distribution. By caching the popular content, UE effectively obtains its requested content instead of straining the front-haul connections [5], [9].

To effectively support 5G network communication, caching, and computing services, smart base stations (SBSs) have been widely viewed as a key infrastructure. In SBSs, computation offloading processing, data caching and mobile communication technologies are collectively employed. Therefore, the computation can be offloaded to the SBS, and data and contents can be cached at the SBS to effectively decrease the delays generated during the inter-communication process between the UE and its cloud server [7]. To maximize the 5G network performance, all relevant control factors must be comprehensively considered to leverage the full synergy of the converged communication, caching, and computing operations in SBSs [5].

To design a novel 5G network SBS control scheme, a new control paradigm is needed. The interaction between the rational agents—who have conflicting objectives—is often characterized using game theory. Game theory is the study of strategic interactions between multiple intelligent rational decision makers, who consistently pursue their own objectives in striving to maximize the expected value of their own payoffs.

Game theory has been successfully applied to wireless communications for solving competition problems of network resources [10]. Being the control theory of multiple goal-driven agents, game theory provides many effective solutions for dealing with 5G network situations and questions. Motivated by the above background, we adopted a theoretical gaming approach to develop practical SBS control algorithms. Accordingly, we can ease the heavy computational burden of theoretically optimal centralized solutions [10].

In this study, we designed a two-tier hierarchical game to model the interaction among SBSs and UE. At the first-tier, SBSs are game players, and the total bandwidth is distributed to each SBS based on a dynamic bargaining model. At the second-tier, each individual SBS and its corresponding UE are game players; their interactions are modeled as a Stackelberg game model. As a leader, the SBS splits its cache capacity for data and content caching, as well as decides the price for communication and computation services. As followers, UE devices monitor their leader's decision and select their best strategies. Using the proposed two-tier game approach, we can capture various mobile communication, caching, and computing characteristics, and a balanced solution can be obtained under diversified 5G network situations.

In addition, we address the challenges of effective computation offloading, cache splitting and bandwidth allocation for communications. These algorithms are combined in an integrated SBS control scheme that can harness the synergies between SBS computations and communications with caching capabilities. The basic concept of our integrated scheme is to design a two-tier game-model-based interactive mechanism. As game players, SBSs and UE make control decisions according to the step-by-step timed learning approach. Under dynamically changing 5G network environments, our game model achieves greater and reciprocal advantages for all players, who can strengthen their competitive advantages by cooperation. Therefore, the main contributions of this paper are to ensure adaptability and flexibility to handle the wide range of 5G network control issues, and to obtain the finest solution to effectively adapt to current system conditions.

Based on the game-based-learning approach, our proposed scheme mainly considers three operational decision issues: i) the decision of computation offloading in each individual UE, ii) the decision of splitting radio for data and content caching capacities, and iii) the decision of bandwidth allocation for each SBS to ensure mobile communications. To effectively address these issues, we focus on design principles, such as feasibility, self-interactivity, and the holistic combination of different control algorithms, which are mutually dependent, to resolve conflicting performance criteria under multiple highly diversified 5G network situations. Although several 5G communication, caching, and computing algorithms have been proposed, no systematic study based on the unified and combined approach has been conducted. To the best of our knowledge, the present study is the first to design a game-based 5G network control scheme by integrating offloading, caching and bandwidth allocation algorithms.

1. Related Work

Considerable research has been conducted on the design of communication, caching, and computing algorithms for 5G networks. Reference [15] presents a system architecture and potential caching scheme for 5G networks that can optimize the average latency and energy cost in content transmission. Unlike the traditional caching schemes, the caching strategy in [15] is designed to exploit the advantage of multicasting and cooperating.

Reference [16] presents a comprehensive survey of the state-of-the-art mobile edge computing (MEC) research with a focus on joint radio and computational resource management. In addition, that study summarizes the modeling methodologies on key components of MEC systems, such as the computation tasks and communications, as well as the computation of mobile devices and MEC servers [18]. Reference [18] introduces a unified computing, caching, and communication solution for the upcoming 5G network environment that enables service, content, and function providers to deploy their services/content/functions near the end users, and to allow network providers to virtually deploy their connectivity services over commodity hardware. In addition, a designed architecture is presented that is intended to reshape the network landscape to fulfill forthcoming 5G requirements [18].

In [19], a collaborative cache allocation and computation offloading scheme is proposed, whereby the MEC servers collaborate for executing computation tasks and data caching. In their proposal, caching and computational resources are allocated to many service requesters based on their demands and payments. The mobile network operator allocates resources based on weighted proportional allocation [19]. In [7], Fan and others propose a joint computation offloading and caching (JCOC) scheme. In the scheme, base stations are equipped with computing power and data storage to jointly offload the computation from terminals and cache the data from clouds. In this scheme, joint computation offloading and data caching capabilities are utilized by base stations to reduce the delay experienced in the process of cloud computations and communications. Using a genetic algorithm, those authors developed a resource management algorithm, which jointly schedules computation offloading and allocates data caching for computation or data requests sent by UE. Finally, the efficiency of the JCOC scheme is shown and its superior performance is demonstrated via the comparisons among different schemes [7].

Reference [20] presents a collaborative computation offloading and caching (CCOC) scheme to improve the performance of cloud intercommunications. Based on the queuing theory, the CCOC scheme formulates the total delays, which consist of transmission and computation delays in the cloud intercommunications of all UE under the base station coverage. To minimize the total delays, data allocation, offload scheduling and resource management algorithms are presented to jointly schedule computation offloading from UE to base stations while allocating data caching from clouds to base stations. To reduce the iterations, which are needed by the offloading scheduling algorithm, constraint relaxation and revision technologies are also proposed [20].

Hajimirsadeghi and others proposed a joint caching and pricing control (JCPC) scheme for information-centric networks [21]. To determine caching and pricing strategies, a theoretical game approach is presented to study Nash strategies for a non-cooperative game using a probabilistic model. To this end, it is assumed that access requests generally follow the generalized Zipf distribution. The authors show that the Nash equilibrium is unique and that the caching policy is determined by a content popularity threshold. In addition, they provide a monetary incentive model to collaborate in caching and distributing content. In this model, the caching costs vary with respect to content popularity, while the content provider cost per unit data is fixed for all content types. When content with different popularities are available in the network, this price-based control approach is an effective solution [21].

Some earlier studies [5], [7], [20], [21] have attracted considerable attention while introducing unique challenges in handling the 5G computation, caching, and communication control problems. In the present paper, we demonstrate that our proposed scheme significantly outperforms these existing JCOC [7], CCOC [20], and JCPC [21] schemes.

II. Proposed 5G Network Integrated Control Scheme

1. Two-Tier Hierarchical Game Model

During the 5G network operation, SBSs and UE make control decisions individually while considering their mutual relationship. In this paper, we consider a scenario formed by n SBSs $\mathfrak{B} = \{\mathcal{B}_1 \cdots \mathcal{B}_n\}$ and m UE $\mathcal{E} = \{\mathcal{E}_1 \cdots \mathcal{E}_m\}$. Each SBS $\mathcal{B}_i \in \mathfrak{B}$ is connected to the

Internet via a limited front-haul link with capacity $C_{\mathcal{B}_i}$, whereas each UE $\mathcal{E}_j \in \mathcal{E}$ is connected to its serving SBS via a wireless link. The caching at each SBS is an important means of offloading the traffic and tackling the front-haul bottleneck to reduce the latency and cost of services.

There are two kinds of caching for 5G network services: data caching for computation offloading and content caching for mobile communications. Therefore, the capacity of $\mathcal{B}_i$'s cache storage $\mathcal{Q}_{\mathcal{B}_i}$ is divided. With the computation offloading and caching services, wireless bandwidth is assigned to each SBS to support the communications between the UE and its corresponding SBS. Usually, the bandwidth needs for each SBS would vary temporally and spatially. To dynamically adapt the traffic fluctuations, our principle of allocating bandwidth to SBSs is a fair and efficient bargaining approach based on self-monitoring.

To combine computation offloading, caching, and bandwidth allocation algorithms, we present a two-tier hierarchical game model ($\mathbb{G}$) assuming dynamic 5G network situations. Formally, we define $\mathbb{G} = \{\mathbb{G}^{\text{first}}, \mathbb{G}^{\text{second}}_{1\le i\le n}\}$, where $\mathbb{G}^{\text{first}}$ is the first-tier bargaining game to formulate interactions among SBSs, and $\mathbb{G}^{\text{second}}_{1\le i\le n}$ comprises multiple second-tier games to formulate interactions between the $\mathcal{B}_{1\le i\le n}$ and its corresponding UE. They work in parallel and independently. Firstly, the $\mathbb{G}^{\text{first}}$ can be defined as $\mathbb{G}^{\text{first}} = \{\mathfrak{B}, \mathbb{C}^{\mathcal{B}}, \langle \mathbb{S}_{\mathcal{B}_1} \cdots \mathbb{S}_{\mathcal{B}_n}\rangle, \ \mathbb{U}_{\mathcal{B}_i \in \mathfrak{B}}, T\}$ at each time period t of game play.

- $\mathfrak{B} = \{\mathcal{B}_1 \cdots \mathcal{B}_n\}$ represents a set of SBSs; they are the first-tier game players.
- The bandwidth capacity of the 5G network system is $\mathbb{C}$; it is divided into SBSs.
- $\langle \mathbb{S}_{\mathcal{B}_1} \cdots \mathbb{S}_{\mathcal{B}_n}\rangle$ is a strategy vector, which corresponds to the assigned bandwidth amount for each SBS, where $\mathbb{S}_{\mathcal{B}_i}$ is the allocation strategy set of $\mathcal{B}_i$.
- $\mathbb{U}_{\mathcal{B}_i}$ is the payoff received by $\mathcal{B}_i$ during the bandwidth allocation operation.
- $T = \{\mathcal{H}_1, \ldots, \mathcal{H}_t, \mathcal{H}_{t+1}, \ldots\}$ denotes time, which is represented by a sequence of time steps with imperfect information for the $\mathbb{G}^{\text{first}}$ game process.

Secondly, $\mathbb{G}^{\text{second}}_i$ is the ith second-tier game, and it can be defined as

$$\mathbb{G}^{\text{second}}_i = \left\{ \{\mathcal{B}_i \in \mathfrak{B}, \mathcal{E}_{\mathcal{B}_i}\}, \left\{ \{\mathcal{S}^{\mathcal{P}}_{\mathcal{B}_i}, \mathcal{S}^{\mathcal{Q}}_{\mathcal{B}_i}\}, \mathcal{S}_{\mathcal{E}^j_{\mathcal{B}_i} \in \mathcal{E}_{\mathcal{B}_i}} \right\}, \left\{ U_{\mathcal{B}_i}, \mathcal{U}^{\mathcal{B}_i}_{\mathcal{E}^j_{\mathcal{B}_i} \in \mathcal{E}_{\mathcal{B}_i}} \right\}, T \right\},$$

at each time period t of gameplay. There are a total of n second-tier games ($\mathbb{G}^{\text{second}}_{1\le i\le n}$) in our two-tier game structure.

- $\mathcal{E}_{\mathcal{B}_i} = \{\mathcal{E}^1_{\mathcal{B}_i} \cdots \mathcal{E}^r_{\mathcal{B}_i}\}$ is the set of UE devices that are in the $\mathcal{B}_i$ coverage area. $\mathcal{B}_i$ and $\mathcal{E}_{\mathcal{B}_i}$ are game players for the second-tier $\mathbb{G}^{\text{second}}_i$ game, which is formulated as a Stackelberg model, where $\mathcal{B}_i$ is a leader and $\mathcal{E}_{\mathcal{B}_i}$ are followers.
- $\{\mathcal{S}^{\mathcal{P}}_{\mathcal{B}_i}, \mathcal{S}^{\mathcal{Q}}_{\mathcal{B}_i}\}$ is the strategy sets of $\mathcal{B}_i$, where $\mathcal{S}^{\mathcal{P}}_{\mathcal{B}_i} = \{\prod_{\min\le k\le\max} \mathfrak{P}^{\mathcal{B}_i}_k | \mathfrak{P}^{\mathcal{B}_i}_k \in [\mathfrak{P}^{\mathcal{B}_i}_{\min} \cdots \mathfrak{P}^{\mathcal{B}_i}_k \cdots \mathfrak{P}^{\mathcal{B}_i}_{\max}]\}$ and $\mathfrak{P}^{\mathcal{B}_i}_k$ means the kth service price strategy. $\mathcal{S}^{\mathcal{Q}}_{\mathcal{B}_i}$ is the $\mathcal{B}_i$ strategy for cache splitting, where $0 \le \mathcal{S}^{\mathcal{Q}}_{\mathcal{B}_i} \le 1$ is the ratio for data caching. Therefore, the $[\mathcal{S}^{\mathcal{Q}}_{\mathcal{B}_i} \times \mathcal{Q}_{\mathcal{B}_i}]$ cache capacity is assigned to data caching ($\mathcal{Q}^{\mathbb{D}}_{\mathcal{B}_i}$) for computation offloading services. The remaining cache capacity $[(1 - \mathcal{S}^{\mathcal{Q}}_{\mathcal{B}_i}) \times \mathcal{Q}_{\mathcal{B}_i}]$ is assigned to content caching ($\mathcal{Q}^{\mathbb{M}}_{\mathcal{B}_i}$) for mobile communication services.
- $\mathcal{S}_{\mathcal{E}^j_{\mathcal{B}_i} \in \mathcal{E}_{\mathcal{B}_i}}$ is the strategy set of $\mathcal{E}^j_{\mathcal{B}_i}$, where $\mathcal{S}_{\mathcal{E}^j_{\mathcal{B}_i}} = \{(0,0), (0,1), (1,0), (1,1)\}$ is the decision combination for computation offloading and communication services, respectively; 0 represents no service and 1 denotes the service activation.
- In $\{U_{\mathcal{B}_i}, \mathcal{U}^{\mathcal{B}_i}_{\mathcal{E}^j_{\mathcal{B}_i} \in \mathcal{E}_{\mathcal{B}_i}}\}$, $U_{\mathcal{B}_i}$ is the $\mathcal{B}_i$ payoff and $\mathcal{U}^{\mathcal{B}_i}_{\mathcal{E}^j_{\mathcal{B}_i}}$ is the $\mathcal{E}^j_{\mathcal{B}_i}$ payoff received from the $\mathbb{G}^{\text{second}}_i$ game process, respectively.
- T is a time period. $\mathbb{G}^{\text{second}}_i$ is repeated $t \in T < \infty$ time periods with imperfect information.

During our hierarchical two-tier game ($\mathbb{G}$) operations, $\mathbb{G}^{\text{first}}$ and $\mathbb{G}^{\text{second}}_{1\le i\le n}$ games work together in a coordinated manner. In the $\mathbb{G}^{\text{first}}$ game process, SBSs bargain with each other to fairly distribute $\mathbb{C}^{\mathcal{B}}$. According to their contributions, the SBSs make joint agreements that give mutual advantage. In the multiple $\mathbb{G}^{\text{second}}_{1\le i\le n}$ game processes, all individual game players select their strategies selfishly to maximize their payoffs. At the end of each game iteration, the players examine their payoffs periodically and they dynamically adapt their decisions in an entirely distributed fashion. During the step-by-step iteration, this feedback process is repeated until the best solution has been found.

2. First-Tier Game Model for Proposed Bandwidth Allocation Algorithm

One of the most significant challenges of the 5G network is how to accommodate the extraordinary increases in data volume and performance expectations. The SBS has been deemed a promising approach to meet the increasing demand of cellular network capacity. For providing better service to users and improving resource utilization, coordination protocols for bandwidth allocation problems are needed to effectively handle spectrum-sharing negotiations. Decisions in bargaining game models are exactly coincident with those in the bandwidth allocation process. Accordingly, we develop our first-tier bargaining game model to distribute bandwidth for SBSs.

In our game model for the bandwidth allocation, SBSs are game players and strategies are assigned bandwidth amounts. To represent the amount of satisfaction of a player toward the game outcome, we construct a utility function for each game player. By relying on the reciprocal interaction between the resource utilization and user's QoS, the utility function for $\mathcal{B}_i$ $(\mathbb{U}_{\mathcal{B}_i})$ can be derived from [22]:

$$\mathbb{U}_{\mathcal{B}_i} = \frac{1}{1-\omega_{\mathcal{B}_i}} \times \max\left(0, \left[-\left(\omega_{\mathcal{B}_i} \times \exp(\mathfrak{R}_{\mathcal{B}_i})\right)\right]\right), \quad (1)$$

$$\text{s.t.,} \quad \mathfrak{R}_{\mathcal{B}_i} = \frac{E^m}{m!} \bigg/ \sum_{c=0}^{m} \frac{E^c}{c!} \text{ and } 0 \le \omega_{\mathcal{B}_i} \le 1,$$

where $\mathfrak{R}_{\mathcal{B}_i}$ is the service blocking probability, and $\omega_{\mathcal{B}_i}$ is a control parameter of the $\mathcal{B}_i$ utility function. Based on the traffic information, such as bandwidth capacity, channel size, and call duration time, $\mathfrak{R}_{\mathcal{B}_i}$ can be estimated according to Erlang's formula [17]. E is the total amount of offered traffic in $\mathcal{B}_i$ and m is the number of wireless channels for $\mathcal{B}_i$. Owing to the current traffic characteristics, the $\omega_{\mathcal{B}_i}$ value is defined as the cache hitting ratio of $\mathcal{B}_i$; the larger is the value of $\omega_{\mathcal{B}_i}$, the more sensitive is the utility to the $\mathfrak{R}_{\mathcal{B}_i}$

By employing a bargaining approach, SBSs in our first-tier game can be team players, can cooperate with each other, and can make a collective decision Based on the concept of relative utilitarianism [11], [12], we adopt the relative utilitarian bargaining solution (RUBS), and develop an adaptive feedback bargaining model to iteratively adjust the bandwidth allocation. RUBS fulfills the axioms of i) Pareto efficiency, ii) independence of equivalent utility representations, and iii) symmetry [12]. Under dynamically changing 5G network environments, this approach is adaptable to approximate a fair-efficient system performance.

In the bargaining solution, the solution set $(\mathbb{W})$ is normally interpreted as the set of feasible utility payoffs to the game players. A point $\mathbf{P} = \{\mathbb{U}_{\mathcal{B}_1} \cdots \mathbb{U}_{\mathcal{B}_m}\} \in \mathbb{W}$ can be achieved if all SBSs agree to it. In the case of agreement, game player $\mathcal{B}_i$ receives $\mathbb{U}_{\mathcal{B}_i}$. According to the relative utilitarianism, the desirable best bargaining solution for a utility vector $\{\mathbb{U}^*_{\mathcal{B}_1} \cdots \mathbb{U}^*_{\mathcal{B}_m}\}$ is obtained as:

$$\left\{\mathbb{U}^* = \{\mathbb{U}^*_{\mathcal{B}_1} \cdots \mathbb{U}^*_{\mathcal{B}_m}\} \middle| \mathbb{U}^* \text{ solves} \right.$$
$$\left. \max_{\langle \mathbb{S}_{\mathcal{B}_1} \cdots \mathbb{S}_{\mathcal{B}_n} \rangle} \sum_{\mathcal{B}_i \in \mathfrak{B}} \left(\frac{\mathbb{U}_{\mathcal{B}_i}}{\mathbb{U}^{\max}_{\mathcal{B}_i} - \mathbb{U}^{\min}_{\mathcal{B}_i}} \right)^{1-\rho_{\mathcal{B}_i}} \right\}, \quad (2)$$

$$\text{s.t.,} \quad 0 \le \frac{\mathbb{U}_{\mathcal{B}}}{\mathbb{U}^{\max}_{\mathcal{B}} - \mathbb{U}^{\min}_{\mathcal{B}}} \le 1 \text{ and } \rho_{\mathcal{B}_i} = \frac{\mathfrak{P}^{\mathcal{B}_i}}{\sum_{\mathcal{B}_j \in \mathfrak{B}} \mathfrak{P}^{\mathcal{B}_j}},$$

where $\rho_{\mathcal{B}_i}$ is the $\mathcal{B}_i$ service price strategy. It is the relative ability to exert influence over other SBSs. By considering the system efficiency of SBSs, we can give appropriate incentives or punishments to each individual SBS; the SBS with a higher service price obtains more bandwidth resources than other SBSs.

3. Second-Tier Game Model for Caching Algorithms

UE is currently expected to support complex applications, such as multimedia processing, online gaming, virtual reality, and sensing. However, owing to the limited computation resources and power supplies, cloud computing has been introduced as state-of-the-art technology. In the 5G network infrastructure, UE can offload its computationally intensive tasks to resource-rich SBSs, which enable their corresponding UE to elastically utilize resources in an on-demand fashion [5]. In most scenarios of cloud computing services, computation execution and required data access are always tightly coupled [20]. Therefore, inter-communication among the UE, SBS, and remote cloud server is a key factor of delay latency. To effectively address this control issue, further integration of cloud computing in the wireless communication environment promotes many practical challenges related to system performance.

To harness the synergy effect, we develop a second-tier game by using the Stackelberg game model. Traditionally, in a Stackelberg game, one player acts as a leader and the others act as followers. The main goal is to find an optimal strategy for the leader, assuming that the followers react in such a rational way while optimizing their objective functions with consideration of the leader's actions [10]. In our second-tier game model, each SBS is a leader, and its corresponding UE are followers. The main goal of a leader is to translate the selfish motives of followers into socially desirable actions. As followers, UE devices attempt to maximize their utility functions in a distributed online fashion. The UE decisions are applied as the input to the SBS control adjustment procedure. Therefore, control decisions are coupled with each other in a hierarchical interaction relationship. This feedback-based iterative process continues until a satisfactory solution is obtained. As a leader, each SBS make decisions for caching management and the service price. As followers, UE devices select their strategies for offloading computations and communications.

SBS caching technology was recently proposed to cache popular data and content in SBSs. It is becoming a promising solution to transcend the service delays of 5G networks. Usually, SBS caching has advantages. First, the distances between the required computation data or communication contents and UE are further decreased. Hence, the end-to-end access delay can potentially be further reduced. Second, SBS caching mitigates the front-haul traffic congestion problem by replacing the front-haul capacity with a cache capacity at the local SBSs. Without a caching technique, the UE should download these files via the front-haul. Third, the request load to origin servers in a remote area can be reduced. Caching at the SBSs can directly provide required data and content instead of straining the server connections [5].

In this study, we assume that a holistic caching structure on each SBS can be split, thereby leveraging a split-cache. One part caches popular content for communication services; the other part caches popular data for computation offloading services. This method can improve the system performance while balancing communication and computation services. In practice, each SBS's cache size is limited; thus it is imperative to make two important decisions: one regarding cache splitting, which is used to adaptively split the cache capacity for data and content, and the other regarding the cache placement strategy, which is used to decide caching data and content with consideration of the frequency of data and content requests.

In our second-tier game, each individual SBS adaptively splits the total cache capacity $(\mathcal{Q}_{\mathcal{B}})$ for data file caching $(\mathcal{Q}_{\mathcal{B}}^{\mathbb{D}})$ and content file caching $(\mathcal{Q}_{\mathcal{B}}^{\mathbb{M}})$, where $\mathcal{Q}_{\mathcal{B}} = \mathcal{Q}_{\mathcal{B}}^{\mathbb{D}} + \mathcal{Q}_{\mathcal{B}}^{\mathbb{M}}$. To find the best solution to the cache splitting problem, we adopt the concepts of Kalai–Smorodinsky bargaining solution (KSBS) and egalitarian bargaining solutions (EBS). The main feature of KSBS is that the increase of the bargaining set size in a direction favorable to a specific player always benefits that player. In short, it can be considered that the KSBS is the maximal point that maintains the ratios of gains.

Unlike KSBS, the EBS has even stronger monotonicity requirements, while satisfying independence conditions [10]. Therefore, KSBS equalizes the ratios between the players' payoffs and their ideal payoffs, while the EBS equalizes the players' payoffs. To design our cache splitting algorithm, data caching $(\mathcal{Q}_{\mathcal{B}}^{\mathbb{D}})$ and content caching $(\mathcal{Q}_{\mathcal{B}}^{\mathbb{M}})$ are assumed as game players, and we split $\mathcal{Q}_{\mathcal{B}_i}$ of $\mathcal{B}_i$ as follows:

$$\frac{\delta_{\mathcal{H}_t}(\mathcal{Q}_{\mathcal{B}_i}^{\mathbb{D}})}{\Omega_{\mathcal{H}_t}(\mathcal{Q}_{\mathcal{B}_i}^{\mathbb{M}})} = \left(\frac{\delta_{\mathcal{H}_t}(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{D}}) = \{\mathbb{O}_{\mathcal{B}_i}^{\mathbb{D}} = \mathcal{X} | \text{optimal}\left(\delta_{\mathcal{H}_t}(\mathcal{X})\right)\}}{\Omega_{\mathcal{H}_t}(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{M}}) = \{\mathbb{O}_{\mathcal{B}_i}^{\mathbb{M}} = \mathcal{Y} | \text{optimal}\left(\Omega_{\mathcal{H}_t}(\mathcal{Y})\right)\}}\right)^{\gamma}, \tag{3}$$

where $\delta_{\mathcal{H}_t}(\cdot)$ and $\Omega_{\mathcal{H}_t}(\cdot)$ are the cache hit ratio per cache size of $\mathcal{Q}_{\mathcal{B}_i}^{\mathbb{D}}$ and $\mathcal{Q}_{\mathcal{B}_i}^{\mathbb{M}}$, respectively, at time $\mathcal{H}_t$. $\mathbb{O}_{\mathcal{B}_i}^{\mathbb{D}}$ and $\mathbb{O}_{\mathcal{B}_i}^{\mathbb{M}}$ guarantee the ideal gains of $\delta_{\mathcal{H}_t}(\cdot)$ and $\Omega_{\mathcal{H}_t}(\cdot)$; they are the maximal possible ratios. The KSBS and EBS correspond to $\gamma = 1$ and $\gamma = 0$, respectively, where $0 \le \gamma \le 1$.

Under the dynamic 5G network environment, a fixed value of γ cannot effectively adapt to the changing conditions. In this scheme, it is treated as an online decision problem and adaptively modifies the γ value. To fine-tune the system performance, it is a suitable approach. When the difference of $\delta(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{D}})$ and $\Omega(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{M}})$ is high, we can place more emphasis on the axiom of individual monotonicity. In this case, KSBS is more suitable. When the difference of $\delta(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{D}})$ and $\Omega(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{M}})$ is low, we strongly depend on the axiom of strong equality. In this case, EBS is more suitable. In the proposed algorithm, the value of γ is dynamically adjusted based on the current values of $\delta(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{D}})$ and $\Omega(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{M}})$. Therefore, the system can be more responsive to current 5G network conditions by the real-time network monitoring. The value of γ at time $\mathcal{H}_t$ is given by

$$\gamma = \frac{|\delta_{\mathcal{H}_t}(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{D}}) - \Omega_{\mathcal{H}_t}(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{M}})|}{\max\left(\delta_{\mathcal{H}_t}(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{D}}), \Omega_{\mathcal{H}_t}(\mathbb{O}_{\mathcal{B}_i}^{\mathbb{M}})\right).} \tag{4}$$

Owing to the limited cache capacity, it is impossible to cache all requested data and content. Therefore, popular data and content are properly cached in each SBS to reduce delays and system overhead. To develop a cache placement algorithm, we assume there is a computation data file set $\mathbb{D}$ and a communication content file set $\mathbb{M}$, and files in $\mathbb{D}$ and $\mathbb{M}$ can be possibly cached in each SBS. The popularity distributions among $\mathbb{D}$ and $\mathbb{M}$ are represented by vectors, which are frequently requested by UE devices. Generally, the vectors can be modeled by a Zipf distribution, which is a discrete probability distribution commonly used in the modeling of rare events [13]. As most of the cache-placing work assumes perfect content popularity, we also adopt the Zipf distribution to model the file popularity, and we implement our cache placement algorithm according to this distribution.

4. Second-Tier Game Model for 5G Services

In our second-tier game model, SBSs and UE are assumed to be self-reflective game players who make their decisions for the goal of maximizing their perceived payoffs. By using a repeated interactive approach, each player's behavior might affect the behavior of other players. Therefore, control decisions are joined, which causes cascading interactions of other players. Based on their utility functions, all game players make their decisions to find the most profitable strategy.

We begin to describe a follower's utility function by using two terms that capture the tradeoff between cooperative and non-cooperative propensities. Cooperative propensity is modeled using a Taguchi loss function, which interprets a follower's dissatisfaction as increasing as the variation increases from their desired communication bandwidth amount [14]. This approach guides selfish UE toward a socially desirable outcome. Non-cooperative propensity is modeled according to the entity's own payoff, which corresponds to the received benefit minus the incurred cost. Based on this assumption, the utility function of $\mathcal{E}_{\mathcal{B}_i}^k$ ($\mathcal{U}_{\mathcal{E}_{\mathcal{B}_i}^k}^{\mathcal{B}_i}$) is defined as:

$$
\begin{aligned}
&\mathcal{U}_{\mathcal{E}_{\mathcal{B}_i}^k}^{\mathcal{B}_i}\left(\mathcal{P}_{\mathcal{B}_i}, \Gamma(\mathcal{E}_{\mathcal{B}_i}^k), \Theta(\mathcal{E}_{\mathcal{B}_i}^k)\right)\\
&= \left(\mu_{\mathcal{E}_{\mathcal{B}_i}^k} \times \log\left(\Gamma(\mathcal{E}_{\mathcal{B}_i}^k)\right)\right) + \left(\psi_{\mathcal{E}_{\mathcal{B}_i}^k} \times \log\left(\Theta(\mathcal{E}_{\mathcal{B}_i}^k)\right)\right)\\
&- \frac{\tau_{\mathcal{E}_{\mathcal{B}_i}^k}}{|\mathcal{E}_{\mathcal{B}_i}| - 1}\left(\sum_{\substack{\mathcal{E}_{\mathcal{B}_i}^l \in \mathcal{E}_{\mathcal{B}_i}\\ \mathcal{E}_{\mathcal{B}_i}^l \neq \mathcal{E}_{\mathcal{B}_i}^k}} \max\left\{\left(\Gamma(\mathcal{E}_{\mathcal{B}_i}^l) - \Gamma(\mathcal{E}_{\mathcal{B}_i}^k), 0\right)\right\}\right)\\
&- \mathcal{C}\left(\mathcal{P}_{\mathcal{B}_i}, \Gamma(\mathcal{E}_{\mathcal{B}_i}^k)\right)\\
&- \frac{\phi_{\mathcal{E}_{\mathcal{B}_i}^k}}{|\mathcal{E}_{\mathcal{B}_i}| - 1}\left(\sum_{\substack{\mathcal{E}_{\mathcal{B}_i}^l \in \mathcal{E}_{\mathcal{B}_i}\\ \mathcal{E}_{\mathcal{B}_i}^l \neq \mathcal{E}_{\mathcal{B}_i}^k}} \max\left\{\left(\Gamma(\mathcal{E}_{\mathcal{B}_i}^k) - \Gamma(\mathcal{E}_{\mathcal{B}_i}^l), 0\right)\right\}\right)\\
&- \mathcal{D}\left(\mathcal{P}_{\mathcal{B}_i}, \Theta(\mathcal{E}_{\mathcal{B}_i}^k)\right),
\end{aligned}
\tag{5}
$$

$$
\text{s.t.,}\begin{cases}
\mathcal{D}\left(\mathcal{P}_{\mathcal{B}_i},\Theta(\mathcal{E}_{\mathcal{B}_i}^k)\right)\\
=\begin{cases}\frac{1}{\zeta}\times\left[\mathfrak{X}\times\log\left(\Theta(\mathcal{E}_{\mathcal{B}_i}^k)\right)\right]\times\mathcal{P}_{\mathcal{B}_i}, & \text{if data is cacheed}\\ \left[\mathfrak{X}\times\log\left(\Theta(\mathcal{E}_{\mathcal{B}_i}^k)\right)\right]\times\mathcal{P}_{\mathcal{B}_i}, & \text{otherwise}\end{cases}\\
\mathcal{C}\left(\mathcal{P}_{\mathcal{B}_i},\Gamma(\mathcal{E}_{\mathcal{B}_i}^k)\right)\\
=\begin{cases}\frac{1}{\delta}\times\left[\mathfrak{Y}\times\log\left(\Gamma(\mathcal{E}_{\mathcal{B}_i}^k)\right)\right]\times\mathcal{P}_{\mathcal{B}_i}, & \text{if content is cacheed}\\ \left[\mathfrak{Y}\times\log\left(\Gamma(\mathcal{E}_{\mathcal{B}_i}^k)\right)\right]\times\mathcal{P}_{\mathcal{B}_i}, & \text{otherwise}\end{cases}
\end{cases},
$$

where $\Gamma(\mathcal{E}_{\mathcal{B}_i}^k), \Theta(\mathcal{E}_{\mathcal{B}_i}^k)$ are the $\mathcal{E}_{\mathcal{B}_i}^k$ requested communication bandwidth and offloading computation amounts, respectively. $\mu_{\mathcal{E}_{\mathcal{B}_i}^k}$ and $\psi_{\mathcal{E}_{\mathcal{B}_i}^k}$ denote the $\mathcal{E}_{\mathcal{B}_i}^k$ satisfaction factors for communication and computation services. $\mathcal{D}(\cdot)$, $\mathcal{C}(\cdot)$ are offloading computation and communication cost functions of $\mathfrak{I}_k^h$, respectively. Let $\tau_{\mathcal{E}_{\mathcal{B}_i}^k}$ and $\phi_{\mathcal{E}_{\mathcal{B}_i}^k}$ be the $\mathcal{E}_{\mathcal{B}_i}^k$ degree of envy and degree of guilt. Based on the concept of inequality aversion, social welfare preference is implemented in the follower's utility function.

As a leader, each individual SBS considers its own payoff and social welfare. To address this multi-objective control problem, it is necessary to find the best compromise solution while maintaining a good balance. Based on the reciprocal relationship of two objectives, the utility function of $\mathcal{B}_i$ at time $\mathcal{H}_t$ ($U_{\mathcal{B}_i}^{\mathcal{H}_t}(\cdot)$) is defined as:

$$
\begin{aligned}
&U_{\mathcal{B}_i}^{\mathcal{H}_t}(\mathcal{P}_{\mathcal{B}_i},\mathcal{E}_{\mathcal{B}_i},\mathcal{S}_{\mathcal{B}_i}^{\mathcal{Q}})\\
&=\left((1-\alpha)\times\sum_{\mathcal{E}_{\mathcal{B}_i}^l\in\mathcal{E}_{\mathcal{B}_i}}\left(\mathcal{D}\left(\mathcal{P}_{\mathcal{B}_i},\Theta(\mathcal{E}_{\mathcal{B}_i}^l)\right)+\mathcal{C}\left(\mathcal{P}_{\mathcal{B}_i},\Gamma(\mathcal{E}_{\mathcal{B}_i}^l)\right)\right)\right)\\
&+\left(\alpha\times\sum_{\mathcal{E}_{\mathcal{B}_i}^l\in\mathcal{E}_{\mathcal{B}_i}}\mathcal{U}_{\mathcal{E}_{\mathcal{B}_i}^l}^{\mathcal{B}_i}\left(\mathcal{P}_{\mathcal{B}_i},\Gamma(\mathcal{E}_{\mathcal{B}_i}^l),\Theta(\mathcal{E}_{\mathcal{B}_i}^l)\right)\right),
\end{aligned}
\tag{6}
$$

$$
\text{s.t.,}\,\alpha=\frac{\left(\sum_{\mathcal{E}_{\mathcal{B}_i}^l\in\mathcal{E}_{\mathcal{B}_i}}\mathcal{U}_{\mathcal{E}_{\mathcal{B}_i}^l}^{\mathcal{B}_i}\left(\mathcal{P}_{\mathcal{B}_i},\Gamma(\mathcal{E}_{\mathcal{B}_i}^l),\Theta(\mathcal{E}_{\mathcal{B}_i}^l)\right)\right)^2}{|\mathcal{E}_{\mathcal{B}_i}|\times\sum_{\mathcal{E}_{\mathcal{B}_i}^l\in\mathcal{E}_{\mathcal{B}_i}}\left(\mathcal{U}_{\mathcal{E}_{\mathcal{B}_i}^l}^{\mathcal{B}_i}\left(\mathcal{P}_{\mathcal{B}_i},\Gamma(\mathcal{E}_{\mathcal{B}_i}^l),\Theta(\mathcal{E}_{\mathcal{B}_i}^l)\right)\right)^2}.
$$

To learn the traffic situation in 5G networks, $\mathcal{B}_i$ periodically monitors the current local and global information; local information is its ($\mathcal{B}_i$'s) own payoff, and global information represents the situation of neighboring SBSs. If the strategy $\mathfrak{P}_k^{\mathcal{B}_i}$ is selected at time

$\mathcal{H}_{t-1}$ by $\mathcal{B}_i$, the latter updates the strategy of the $\mathfrak{P}_k^{\mathcal{B}_i}$ learning value $(\mathbb{L}_{\mathcal{H}_t}(\mathfrak{P}_k^{\mathcal{B}_i}, \mathcal{B}_i))$ for the next time $\mathcal{H}_t$, as follows:

$$\mathbb{L}_{\mathcal{H}_t}(\mathfrak{P}_k^{\mathcal{B}_i},\mathcal{B}_i)=\left((1-\chi)\times\mathbb{L}_{\mathcal{H}_{t-1}}(\mathfrak{P}_k^{\mathcal{B}_i},\mathcal{B}_i)\right)+\left(\chi\times[\mathbb{H}+\mathbb{Y}]\right), \tag{7}$$

$$\text{s.t.,}\,\mathbb{H}=\xi\times U_{\mathcal{B}_i}^{\mathcal{H}_{t-1}}(\mathcal{P}_{\mathcal{B}_i},\mathcal{E}_{\mathcal{B}_i},\mathcal{S}_{\mathcal{B}_i}^{\mathcal{Q}}),$$

$$\mathbb{Y}=\left\{\frac{(1-\xi)}{|\mathfrak{B}|-1}\times\sum_{\mathcal{B}_j\in\mathfrak{B},\mathscr{B}_j\neq\mathcal{B}_i}\mathbb{L}_{\mathcal{H}_{t-1}}(\mathfrak{P}_k^{\mathcal{B}_j},\mathcal{B}_j)\right\},$$

and ξ

$$=\min\left(\frac{1}{2},\left(\frac{1}{|\mathfrak{B}|-1}\times\sum_{\mathcal{B}_j\in\mathfrak{B},\mathscr{B}_j\neq\mathcal{B}_i}\frac{|\mathfrak{M}_{\mathcal{H}_{t-1}}(\mathcal{B}_j)-\mathfrak{M}_{\mathcal{H}_{t-1}}(\mathcal{B}_i)|}{\mathfrak{M}_{\mathcal{H}_{t-1}}(\mathcal{B}_i)}\right)\right),$$

where χ is the learning rate that models how the $\mathbb{L}$-values are updated. In (7), $\mathbb{H}$ and $\mathbb{Y}$ represent local and global learning values, and ξ is a control factor for the weighted average between different learning approaches. $\mathfrak{M}_{\mathcal{H}_{t-1}}(\mathcal{B}_j)$ is the total traffic amount of $\mathcal{B}_j$ at time $\mathcal{H}_{t-1}$.

Based on the $\mathbb{L}(\cdot)$ values, a strategy selection distribution $(\mathbb{P})$ for each SBS is defined. To respond to the current 5G network situation at time $\mathcal{H}_t$, we determine $\mathbb{P}_{\mathcal{H}_t}^{\mathcal{B}_i}=\{\mathcal{P}_{\mathcal{H}_t}(\mathfrak{P}_{\min}^{\mathcal{B}_i})\cdots\mathcal{P}_{\mathcal{H}_t}(\mathfrak{P}_k^{\mathcal{B}_i})\cdots\mathcal{P}_{\mathcal{H}_t}(\mathfrak{P}_{\max}^{\mathcal{B}_i})\}$ as the probability distribution of the $\mathfrak{P}^{\mathcal{B}_i}$ strategy selection; it is sequentially modified over time. From $\mathbb{P}_{\mathcal{H}_t}^{\mathcal{B}_i}$, the $\mathfrak{P}_k^{\mathcal{B}_i}$ strategy selection probability by $\mathcal{B}_i$ at time $\mathcal{H}_t$ $(\mathcal{P}_{\mathcal{H}_t}(\mathfrak{P}_k^{\mathcal{B}_i}))$ is defined as:

$$\mathcal{P}_{\mathcal{H}_t}(\mathfrak{P}_k^{\mathcal{B}_i})=\frac{\mathrm{EXP}\left(\mathbb{L}_{\mathcal{H}_t}(\mathfrak{P}_k^{\mathcal{B}_i},\mathcal{B}_i)\right)}{\sum_{e=\min}^{\max}\mathrm{EXP}\left(\mathbb{L}_{\mathcal{H}_t}(\mathfrak{P}_e^{\mathcal{B}_i},\mathcal{B}_i)\right)}. \tag{8}$$

In each game round, the learning process sequentially proceeds according to (5) to (8), and the $\mathcal{B}_i$ stochastically selects the $\mathfrak{P}^{\mathcal{B}_i}$ strategy using its strategy selection distribution $(\mathbb{P}^{\mathcal{B}_i})$. In this study, we effectively implement the second-tier game model by adopting the dynamic-learning-based Stackelberg model. In each game period, the leaders and followers attempt to maximize their payoffs by modifying their respective strategies.

$$\max_{\mathcal{S}_{\mathcal{B}_i}^{\varepsilon^j}}\left(\mathcal{U}_{\mathcal{E}_{\mathcal{B}_i^k}}^{\mathcal{B}_i}\left(\mathcal{P}_{\mathcal{B}_i},\Gamma(\mathcal{E}_{\mathcal{B}_i}^k),\Theta(\mathcal{E}_{\mathcal{B}_i}^k)\right)\right)$$
$$\text{and}\ \max_{\mathcal{P}_{\mathcal{B}_i}\in\mathcal{S}_{\mathcal{B}_i}^{\mathcal{P}},\mathcal{S}_{\mathcal{B}_i}^{\mathcal{Q}}}\left(U_{\mathcal{B}_i}^{\mathcal{H}_t}(\mathcal{P}_{\mathcal{B}_i},\mathcal{E}_{\mathcal{B}_i},\mathcal{S}_{\mathcal{B}_i}^{\mathcal{Q}})\right). \tag{9}$$

III. Performance Evaluation

To ensure a fair comparison, the following assumptions and system scenario were used.

- The simulated system consisted of 50 SBSs, and the number of UE devices was 1,000. The front-haul link capacity $(\mathcal{C}_B)$ of each SBS was 4 Gbps.
- The total wireless bandwidth $(\mathbb{C})$ was 100 Gbps, which was distributed to each SBS, and $\mathcal{Q}_B$ was 2 Gbyte.
- The service price strategies in $\mathcal{S}_{\mathcal{B}_i}^{\mathcal{P}}$ are defined as $\mathfrak{P}_{\min=1}^{\mathcal{B}}=1$, $\mathfrak{P}_2^{\mathcal{B}}=1.2$, $\mathfrak{P}_3^{\mathcal{B}}=1.4$, $\mathfrak{P}_4^{\mathcal{B}}=1.6$, $\mathfrak{P}_5^{\mathcal{B}}=1.8$ and $\mathfrak{P}_{\max=6}^{\mathcal{B}}=2$.
- The wireless channel size was 128 Mbps, and the bandwidth assignment for application services was specified in terms of channels.
- According to the UE characteristics, service requests were generated based on the Poisson process with rate λ (services/s), and the range varied from zero to three.
- There were eight different service requests, which were randomly generated from the UE.
- To represent various application services, eight different traffic types were assumed based on the connection duration and bandwidth requirement. They were generated with equal probability.
- The durations of service applications were exponentially distributed.
- The system performance measures obtained on the basis of 100 simulation runs were plotted as functions of the service request generation rate.
- For simplicity, we assumed the absence of physical obstacles in the experiments.

To demonstrate the validity of our proposed method, we measured the bandwidth utilization, access delay, and system throughput. Table 1 shows the system parameters used in the simulation. The major system control parameters of the simulation, as presented in the table, facilitated the development and implementation of our simulator.

Figure 1 gives the performance comparison of each scheme in terms of the bandwidth utilization. From this figure, we can observe that all schemes exhibit a similar trend. However, our first-tier game efficiently allocates available bandwidth to each SBS via RUBS, and it iteratively adjusts the bandwidth allocation based on the adaptive feedback bargaining model. It leads to higher bandwidth utilization and provides an ideal solution characterized by 5G network environments.

Figure 2 presents the normalized access delay for each scheme. The access delay specifies the time required for a data bit to traverse the network from the source to

Table 1. System parameters used in the simulation experiments.

Type	Comp. offloading	Min. require.	Bandwidth require.	Duration average
I	100 MHz/s	128 Mbps	256 Mbps	1,800 s (30 min)
II	120 MHz/s	256 Mbps	512 Mbps	1,800 s (30 min)
III	150 MHz/s	384 Mbps	768 Mbps	300 s (5 min)
IV	180 MHz/s	512 Mbps	1.28 Gbps	300 s (5 min)
V	200 MHz/s	640 Mbps	1.28 Gbps	1,800 s (30 min)
VI	240 MHz/s	768 Mbps	1.54 Gbps	1,800 s (30 min)
VII	270 MHz/s	896 Mbps	1.82 Gbps	3,000 s (50 min)
VIII	300 MHz/s	1.28 Gbps	2.56 Gbps	1,200 s (20 min)

Parameters	Value	Description
n	50	Number of SBSs in the system
m	1,000	Number of UEs in the system
μ	$1.5 \leq \mu \leq 2.5$	Satisfaction factor of communication service; randomly selected
ψ	$1.5 \leq \psi \leq 2.5$	Satisfaction factor of computation service; randomly selected
τ	0.2	Degree of envy in UE
ϕ	0.1	Degree of guilty in UE
ζ	2	Data cache profit factor
δ	2	Content cache profit factor
$\mathfrak{X}$	0.9	Cost factor for computation service
$\mathfrak{Y}$	0.8	Cost factor for communication service
χ	0.3	Learning rate to update the $\mathbb{L}$-values

destination. The access delay quickly increases under the heavy traffic situation. We also observe that all the schemes have almost identical performances when the traffic load is light. During the 5G network operation, there is no doubt that cache-enabled SBSs can obviously reduce the access delay. Based on our cache splitting and placing methods, each SBS intelligently monitors every service request from the UE and effectively manages the cache operation.

The curves in Fig. 3 show normalized system throughput in the 5G cellular network system. For the system operator, throughput maximization is the main

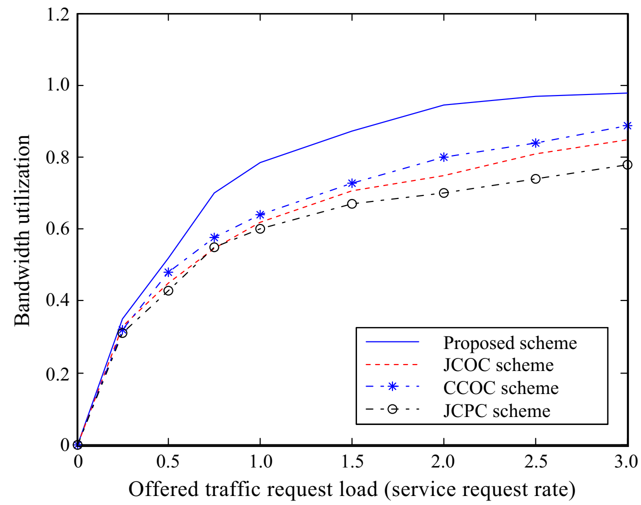


Fig. 1. Bandwidth utilization of the network system.

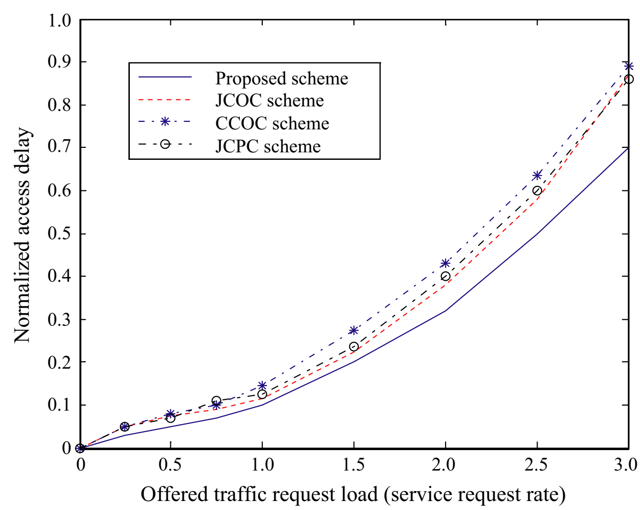


Fig. 2. Normalized access delay.

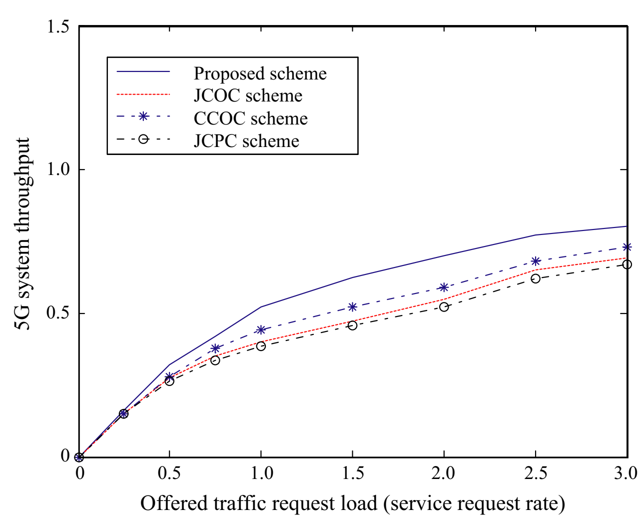


Fig. 3. System throughput.

concern. As shown in Fig. 1, all schemes exhibit a similar trend. This is intuitively correct; usually the network throughput is obtained by using the bandwidth resource. In our second-tier game, SBSs and UE devices are game players and adaptively select strategies to maximize their payoffs. According to an interactive feedback mechanism, game players effectively learn the current system environments and attempt to improve their payoffs, which correlates with the throughput maximization.

IV. Conclusions

To address the extraordinarily rapid growth of future traffic services, the synergistic combination of communication, caching, and computing algorithms is a promising technique for 5G networks. In this paper, we proposed an integrated holistic control scheme by integrating bandwidth allocation, caching splitting, and computation offloading mechanisms. Based on our two-tier game model, we explore effective solutions to the fundamental problems of determining a means to decide control decisions for enabling adequate network performance. From the viewpoint of SBSs, their own payoffs and social welfare are important factors. From the viewpoint of UE, it is necessary to capture the tradeoff between cooperative and non-cooperative propensities.

Using the step-by-step interactive feedback process, SBSs and UE sequentially interact and select their respective strategies to maximize their expected benefits. Under incomplete information situations, it is a practical and suitable operational approach. According to the dynamic bargaining and repeated Stackelberg game procedures, synergistic and complementary features can be provided to adapt to dynamic 5G network situations.

We verified the effectiveness of our proposed scheme using extensive simulations. Numerical results showed the superiority of our integrated scheme compared to existing schemes. Furthermore, our game-based 5G network control approach is expected to be an interesting topic for future work. New interesting research topics can include socially aware concepts to further improve the quality of experience of 5G networks. To date, security issues require examination. Therefore, another interesting direction is to address the network security issues in the 5G network system from the operator's perspective.

Acknowledgments

This research was supported by the Ministry of Science and ICT (MSIT) of the government of the Rep. of Korea under the Information Technology Research Center (ITRC) support program (IITP-2017-2014-0-00636) and supervised by the Institute for Information and Communications Technology Promotion (IITP). It was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01060835).

References

[1] Q. Wang, D. Chen, N. Zhang, Z. Win, and Z. Qin, "LACS: A Lightweight Label-Based Access Control Scheme in IoT-Based 5G Caching Context," *IEEE Access*, vol. 5, Mar. 2017, pp. 4018–4027.

[2] X. Li, X. Wang, K. Li, and V.C.M. Leung, "CaaS: Caching as a Service for 5G Networks," *IEEE Access*, vol. 5, Mar. 2017, pp. 5982–5993.

[3] K. Kim, S. Uno, and M.W. Kim, "Adaptive QoS Mechanism for Wireless Mobile Network," *J. Comput. Sci. Eng.*, vol. 4, no. 2, 2010, pp. 153–172.

[4] X. Ma, S. Zhang, W. Li, P. Zhang, C. Lin, and X. Shen, "Cost-Efficient Workload Scheduling in Cloud Assisted Mobile Edge Computing," *IEEE/ACM Int. Symp. Quality Service*, Vilanova i la Geltru, Spain, June 2017, pp. 1–10.

[5] S. Andreev et al., "Exploring Synergy between Communications, Caching, and Computing in 5G-Grade Deployments," *IEEE Commun. Mag.*, vol. 54, no. 8, Aug. 2016, pp. 60–69.

[6] K. Lee and I. Shin, "User Mobility Model Based Computation Offloading Decision for Mobile Cloud," *J. Comput. Sci. Eng.*, vol. 9, no. 3, 2015, pp.155–162.

[7] W. Fan, Y. Liu, B. Tang, F. Wu, and H. Zhang, "Exploiting Joint Computation Offloading and Data Caching to Enhance Mobile Terminal Performance," *IEEE Globecom.*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[8] D. Aga Bulti and K. Raimond, "Optimizing Caching in a Patch Streaming Multimedia-on-Demand System," *J. Comput. Sci. Eng.*, vol. 9, no. 3, 2015, pp. 134–141.

[9] K.C. Chen, "Machine-to-Machine Communications for Healthcare," *J. Comput. Sci. Eng.*, vol. 6, no. 2, 2012, pp. 119–126.

[10] S. Kim, "*Game Theory Applications in Network Design*," Hershey, PA, USA: IGI Global, 2014.

[11] X. Huang, Z. Zhao, and H. Zhang, "Cooperate Caching with Multicast for Mobile Edge Computing in 5G Networks," *IEEE Veh. Technol. Conf.*, Sydney, Australia, June 2017, pp. 1–6.

[12] Y. Mao, C. You, J. Zhang, K. Huang, and K.B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Commun. Surveys Tutor.*, vol. 19, no. 4, 2017, pp. 2322–2358.

[13] E.K. Markakis, K. Karras, A. Sideris, G. Alexiou, and E. Pallis, “Computing, Caching, and Communication at the Edge: The Cornerstone for Building a Versatile 5G Ecosystem,” *IEEE Commun. Mag.*, vol. 55, no. 11, 2017, pp. 152–157.

[14] A. Ndikumana, S. Ullah, T. LeAnh, N.H. Tran, and C.S. Hong, “Collaborative Cache Allocation and Computation Offloading in Mobile Edge Computing,” *IEEE Asia-Pacific Netw. Operation Manage. Symp.*, Seoul, Rep. of Korea, Sept. 2017, pp. 366–369.

[15] W. Fan, Y. Liu, B. Tang, F. Wu, and H. Zhang, “TerminalBooster: Collaborative Computation Offloading and Data Caching via Smart Basestations,” *IEEE Wirel. Commun. Lett.*, vol. 5, no. 6, Dec. 2016, pp. 612–615.

[16] M. Hajimirsadeghi, N.B. Mandayam, and A. Reznik, “Joint Caching and Pricing Strategies for Popular Content in Information Centric Networks,” *IEEE J. Selected Areas Commun.*, vol. 35, no. 3, Mar. 2017, pp. 654–667.

[17] D. Niyato and E. Hossain, “A Noncooperative Game-Theoretic Framework for Radio Resource Management in 4G Heterogeneous Wireless Access Networks,” *IEEE Trans. Mobile Comput.*, vol. 7, no. 3, 2008, pp. 332–345.

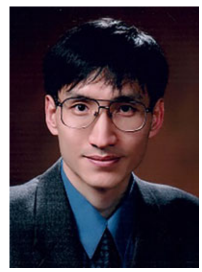
[18] D. Niyato and E. Hossain, “A Noncooperative Game-Theoretic Framework for Radio Resource Management in 4G Heterogeneous Wireless Access Networks,” *IEEE Trans. Mobile Comput.*, vol. 7, no. 3, 2008, pp. 332–345.

[19] A. Dhillon and J.F. Mertens, “Relative Utilitarianism,” *Econometrica*, vol. 67, no. 3, 1999, pp. 471–498.

[20] J. Sobel, “Manipulation of Preferences and Relative Utilitarianism,” *Games Econ. Behav.*, vol. 37, no. 1, 2001, pp. 196–215.

[21] J. Li, J. Sun, Y. Qian, F. Shu, M. Xiao, and W. Xiang, “A Commercial Video-Caching System for Small-Cell Cellular Networks Using Game Theory,” *IEEE Access*, vol. 4, June 2016, pp. 7519–7531.

[22] I.C. Konstantakopoulos et al., “Smart Building Energy Efficiency via Social Game: A Robust Utility Learning Framework for Closing the Loop,” *IEEE SCOPE-GCTC*, Vienna, Austria, Apr. 2016, pp. 1–6.



Sungwook Kim received BS and MS degrees in computer science from Sogang University, Seoul, Rep. of Korea, in 1993 and 1995, respectively. In 2003, he received a PhD degree in computer science from Syracuse University, NY, USA, supervised by Prof. Pramod K. Varshney. He has held faculty positions in the Department of Computer Science of ChoongAng University, Seoul, Rep. of Korea. In 2006, he returned to Sogang University, where he is currently an associate professor in the Department of Computer Science and Engineering, and he is a research director of the Internet Communication Control Research Laboratory. His research interests include resource management, online algorithms, multimedia network management, bandwidth allocation, adaptive quality-of-service control, and game theory for wireless networks.