Harvesting Short-lived White Spaces via Opportunistic Traffic Offloading between Mobile Service Providers

Diep N. Nguyen[1], Eryk Dutkiewicz[1], and Marwan Krunz[12]
[1]Faculty of Engineering and Information Technology, University of Technology Sydney,
E-mail:{diep.nguyen,eryk.dutkiewicz}@uts.edu.au
[2]Department of Electrical and Computer Engineering, University of Arizona, E-mail:krunz@email.arizona.edu

***Abstract*—Currently, Wi-Fi (IEEE 802.11) is the most widely adopted wireless technology for mobile traffic offloading at hot spots. Despite its great success, Wi-Fi is constrained by the over-crowded unlicensed spectrum, which leads to poor user experience, especially in urban areas. This work introduces an opportunistic cooperation framework that allows mobile service providers (MSPs) to offload traffic onto each other's network by harvesting short-lived spectrum/resources of cellular systems. Specifically, through traffic offloading, MSPs aim to maximize their profit while maintaining their quality of service (QoS) commitments. For that purpose, we model the strategic cooperation between MSPs as a stochastic Markov game in which the dynamics of resource availability and user behaviors are captured via a Markov decision process (MDP). We prove that the game is irreducible and admits a Nash Equilibrium at which all MSPs benefit from traffic offloading. A practical algorithm that uses only local information to govern traffic offloading at MSPs is then developed. Numerical simulations show that by designing appropriate profit sharing contracts, our proposed algorithm can achieve almost the same performance as that of a socially optimal solution. The derived traffic offloading strategies not only improve QoS and revenue for MSPs, but also can be used to guide MSPs on when to turn off their base stations while the traffic volume is light (e.g., during nighttime).**

***Index Terms*—Traffic offloading, short-lived whitespaces, cooperation, incentives, cellular systems, noncooperative game, stochastic Markov game, Markov decision process, queuing theory.**

I. Introduction

Mobile data traffic has grown dramatically in recent years, and is estimated to increase by more than a thousand-fold in the next 10 years [2]. To accommodate such ever-increasing traffic demand, mobile service provider/operators (MSPs) will have to either acquire additional radio spectrum or improve spectrum utilization by deploying additional small cells and reuse the spectrum more intensively. However, both approaches can be costly and time-consuming. Another alternative is to offload mobile traffic to Wi-Fi networks. According to Cisco [3], by 2021, 63% of the total mobile data traffic will be offloaded. Currently, Wi-Fi (IEEE 802.11) is the most widely adopted solution for offloading mobile traffic at hot spots (through dual-mode devices). In 2016, more traffic was offloaded from cellular networks (onto Wi-Fi) than the traffic that remained in cellular networks. Despite its great success, Wi-Fi is constrained by the over-crowded unlicensed spectrum, which leads to poor user experience, especially in urban areas [4].

In this work, we take a further step to enable traffic offloading between MSPs. Specifically, when receiving service requests, an MSP can decide whether to serve its customers (we use the terminology "customers" or "service requests" in queuing theory to mean communications sessions in cellular systems) or redirect them to other MSPs. An MSP should also decide to either serve or reject customers offloaded from other MSPs. The decision depends not only on the MSP's resource availability, but also on the reward/payment from the customer, and the quality of service (QoS) commitment. This *commitment* requires that an MSP will only share its resources if its QoS target is guaranteed. In this article, QoS is measured in terms of the probability that a given customer is not served.

Traffic offloading between MSPs is in line with the well-known spectrum sharing/trading concept, in which temporarily unused spectrum can be traded for a profit. Despite the many spectrum economics/sharing and auctioning mechanisms proposed in the literature, a dynamic spectrum market is still unlikely to be deployed in the near future. There are various hurdles to overcome, including both policy and economic aspects, e.g., MSPs unwillingness to help improve the performance of a competitor. Additionally, temporarily unused spectrum chunks are highly dynamic in both temporal and spatial dimensions. Consequently, MSPs are not willing to share/trade their spectrum but rather maintain their exclusive ownership so that they can access the spectrum whenever and wherever needed.

Existing works on spectrum and base stations sharing (e.g., [5] [6] [7]) capture the user traffic using a random variable with known pdf. In [5], the authors use the newsvendor model in operation research to study the revenue-sharing between MSPs for their traffic roaming/offloading service. They also investigate the impact of traffic roaming on MSPs' infrastructure investment strategies. A set of revenue-sharing contracts that provide incentives for both MSPs is introduced and characterized. In [7] and [6], the energy efficiency through base station sharing is proposed using game theory. The competition of MSPs for customers under spectrum sharing is explored in [8]. The authors of [8] find that competition in the unlicensed spectrum can potentially decrease social welfare. However, competition under shared spectrum leads to a always non-decreasing social welfare function.

This work has been supported in part by the Australian Research Council (Discovery Early Career Researcher Award DE150101092), and in part by US NSF (grants IIP-1265960, CNS-1513649, CNS-1563655, and CNS- 1731164. Preliminary results in this paper were presented at the IEEE ICC Conference, London, 2015 [1].

The traffic model in the above works neglects time variability, e.g., the arising of a service request and its service time (e.g., file size). In other words, these works did not take into account the dynamics of cellular user and hence fail to harvest short-lived whitespaces. In fact, short-lived whitespaces in the cellular band are very significant. Real-life cellular system traces show that short-lived whitespaces account for more than one third of the entire frequency-time resources of the cellular bands, even in urban areas [9], [10]. Recent spectrum sharing architectures by FCC (e.g., Spectrum Access Systems or SAS) and ETSI (namely Licensed Shared Access or LSA) [11] [12] tend to ignore the short-lived whitespaces. On the other hand, the long-lived whitespaces like the ones found in the TV bands (Figure 1) are very limited (0, 1 or 2 channels at a time) and becoming less and less available, especially in the populated urban areas where most traffic demand arises.

Traffic offloading is also in line with the infrastructure sharing, e.g., [13], [14], [15], [16]. The authors in [13] propose a model to study the network planning while accounting for both possible cooperative resource sharing and competition regulation amongs MSPs. [14] investigates the infrastructure sharing using game theory and Poisson process to model the traffic arrival. The outcomes of [14] are probabilities for a base station to be switched off/on. The base station switching problem is also considered in [16] under the joint consideration of both uplink and downlink traffic. The network infrastructure sharing has been recently addressed under the concept of network slicing, e.g., [15] where the authors consider the slicing of radio access network resources by multiple service providers. All these works focus on the action of MSPs at a macro and long-term level (e.g. at the base station). By contrast, the focus of our work is the strategic admission of *every single* service request/user to harvest short-lived spectrum whitespace.

To that end, we formulate the traffic offloading problem between MSPs as a constrained Markov game [17] [18] in which MSPs or players aim to maximize their *revenue rate* (i.e., average revenue over the time horizon) while maintaining the QoS commitment. We model the traffic load using a random process and more importantly we consider the decision (admit, offload, or reject) for every single request/user using a Markovian game of queues. The dynamics of user behaviors and resource availability are captured by an underlying Markov Decision Process (MDP). Doing so allows us to utilize spectrum opportunities/whitespace that are as small as those used to serve a single user/request. The time-scale of spectrum/resource sharing in our case is also as fine as that of a user/service request. Our proposed model and results can be applicable to any user mobility model.

For simplicity, we assume there are two MSPs in the game (the case of more than two MSPs follows similarly). We show that the game admits at least one Nash Equilibrium (NE) at which both MSPs gain higher average revenue by offloading traffic, especially when one experiences heavy traffic. When both MSPs have very light traffic (e.g., during nighttime), the reward rates at the NE can be used as a benchmark so that MSPs can make decisions on switching base stations on/off to save operating cost (e.g., energy). The theoretical results herein are not only applicable to cellular systems but also to extend to more general settings involving competitive and cooperative admission control in queuing systems. It is worth noting that although there is a rich literature in queuing theory, the study on strategic admission control between two queues has yet been reported.

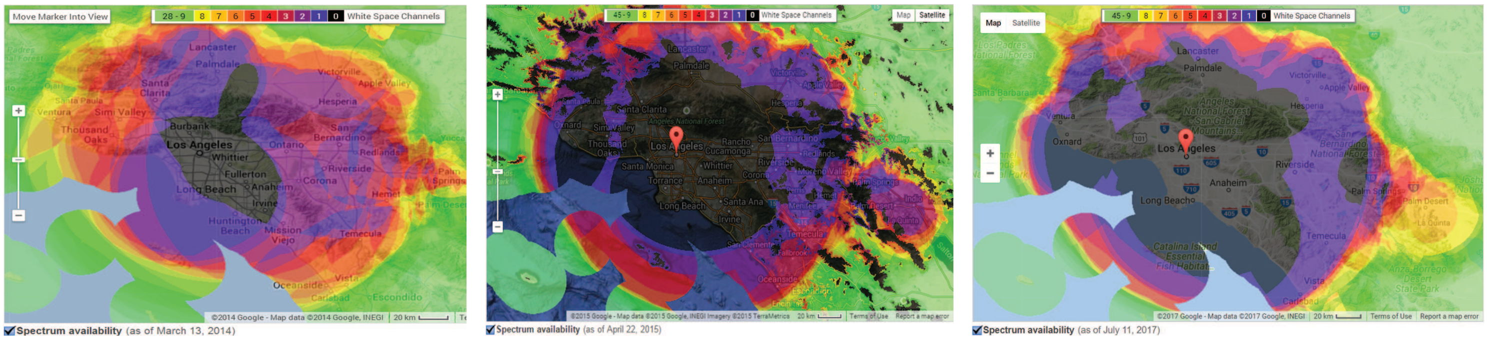
To facilitate practical implementation, we design an efficient algorithm that achieves a tight lower-bound on the reward rate of the above Markov game using *trunk reservation* policy [19] [20]. The algorithm requires only local information to guide MSPs on making their decision of rejecting/admiting customers. Note that our proposed framework that enables traffic offloading *between MSPs* to harvest short-lived whitespaces in cellular bands differs from existing opportunistic traffic offloading, e.g. [21] [22], that facilitates offloading *between mobile devices* or Wi-Fi offloading.

The problem statement is presented in Section II, where the stochastic Markov game is formulated (Section II-A) and its NE existence and characterization are presented (Section II-B). The practical algorithm that uses only local information is derived in Section III. The application of the stochastic Markov game in switching on/off base stations (or base station sharing) is discussed in Section V. Numerical results are presented in Section VI, followed by conclusions in Section VII.

II. Problem Statement

Consider two MSPs that provide coverage to the same residential area, each with its own base stations. Customers arrive at MSP k, $k = 1, 2$, according to a Poisson process with rate λ_i. Note that the Poisson process has been widely used in cellular systems to model call arrivals, e.g., [14]. Although this model is understandably idealized, it captures the aggregate effect from a large population of users that generate ON/OFF calls at a call center. Even when the per-user call arrival/departure process follows an arbitrary random ON/OFF process, the Palm theorem guarantees that the superposition of many such processes converges to a Poisson process. Each customer arrival can be interpreted as a service request (either voice or data connection) in practical cellular systems. Network resource (e.g., spectrum) or capacity of MSP 1 and MSP 2 allows them to serve N_1 and N_2 customers simultaneously. The service time for each customer is exponentially distributed with average μ (for brevity, we normalize $\mu = 1$). We assume that when a customer arrives, it either gets served or rejected, i.e., the two network operators do not queue unserved customers. MSP 1 (MSP 2) gets p_1 (p_2) monetary units after admitting/serving one of its own customers for service.

We consider three scenarios. In scenario 1 (S_1), the two MSPs operate independently; in scenario 2, (S_2), they partially cooperate; and finally in the third scenario (S_3), the two MSPs fully cooperate. In S_1, each MSP does not share its resource with the other, i.e., an MSP only serves its customers (as in conventional cellular systems). In S_2 a MSP shares its resource with the other by serving customers from the other MSP. However, for the purpose of maximizing its own revenue, a MSP reserves the right to reject or serve the other MSP's customers. It also decides whether to serve its own customers or redirect its customers to the other MSP for service. By serving a customer of MSP 1 (or 2), MSP 2 (or 1) is paid $\beta_1 p_1$ (or $\beta_2 p_2$) and the



(a) (b) (c)

Fig. 1. Snapshots from the Google spectrum database that we recorded in March 2014(a), April 2015(b), and July 2017(c) for the Los Angeles region: TV whitespaces are very limited in populated urban area (0, 1 or at most 2). Compare these snapshots in 2014, 2015, and 2017 we find that TV whitespaces are even getting scarcer and scarcer.

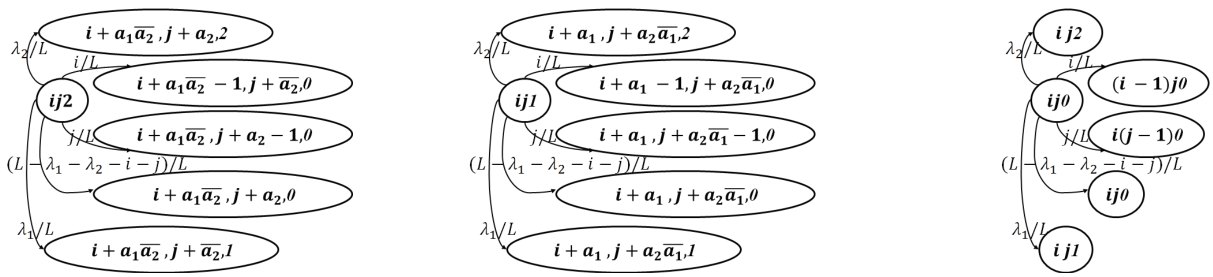


Fig. 2. Transition probabilities

other fraction $(1-\beta_1)p_1$ (or $(1-\beta_2)p_2$) is paid for MSP 1 (or MSP 2), where $\beta_1, \beta_2 \in [0,1]$. The "partial cooperation" in S_2 refers to the fact that the two MSPs can offload traffic/customers to each other (i.e., sharing resource is possible) but each has its own interest in maximizing its individual revenue (by reserving rights to make strategic decision on accepting or rejecting a customer).

To facilitate theoretical analysis, in S_2, we also assume that MSPs share their current resource availability information and admitting/rejecting policies with each other. We later design practical algorithms that guide MSPs' decisions with only local information. Our work aims to come up with a dynamic resource sharing decision on a short-term basis (per service request) while taking the longer-term revenue-sharing contract β_1 and β_2 as input parameters. The impact of the admission and offloading decision on the longer-term revenue-sharing contract that would involve long-term issues (e.g., competition, infrastructure investment plan) is an interesting issue, yet out of scope of our work. The inter-play between revenue-sharing contract, investment, and strategic decision on admission/offloading remains an open problem.

Unlike S_2, in S_3 the two MSPs fully cooperate by chipping in their resources so that they can serve at most (N_1+N_2) customers of both type 1 (i.e., with reward p_1) and type 2 (i.e., with reward p_2). The two MSPs under S_3 have a common interest in maximizing their total revenue.

We aim to answer the following questions:

- $\boldsymbol{Q_1}$: Will the two MSPs under S_2 benefit from sharing their resources, i.e., they both get higher revenue, compared with S_1?
- $\boldsymbol{Q_2}$: If so, then how to design such a cooperation policy and what are the best strategies for each MSP in accepting/rejecting/redirecting customers?
- $\boldsymbol{Q_3}$: What is the best strategies in accepting/rejecting/redirecting customers if MSPs do not share information on their resource/capacity, strategies as well as traffic load (i.e., the customer arrival rate)?
- $\boldsymbol{Q_4}$: How is the total revenue of the two MSPs under S_1 and S_2 compared with that under S_3? How to design a revenue-sharing (β_1, β_2) mechanism between the two so that both have incentives to fully cooperate in S_3?

A. Stochastic Markov Game Formulation

For S_1, each MSP can be modeled by a classical $M/M/N_1/N_1$ (or $M/M/N_2/N_2$) queue. For S_3, this is a $M/M/(N_1+N_2)$ queue with two classes of customers. If one removes the QoS commitment, the optimal strategy in maximizing the revenue in S_3 is the trunk reservation policy [19][1]. In this article, to enforce the QoS commitment, we will rely on constrained Markov decision process (MDP) [23] to find the optimal strategy for S_3, as discussed in Section III.

For S_2, we model the strategic traffic offloading between MSPs as a stochastic Markov game [17]. First, we define the underlying Markov decision process. Let i and j denote, respectively, the number of customers which are being served concurrently at MSP 1 and MSP 2 at a given point in time. Let x

[1]Trunk reservation policy states that customers with higher payment/reward are always admitted while customers with lower payment/reward will be admitted into the system only if the system's size is less than a given threshold.

denote the customer types that arrive at either MSP: $x = 0$ if no customer arrives, $x = 1$ if a customer of MSP 1 arrives, $x = 2$ if a customer of MSP 2 arrives. $\mathbb{S} \overset{\text{def}}{=} \{s\}$, with $s \overset{\text{def}}{=} (ijx)$, denotes the system state space. At any time instance, the system can be in one of these states: $(ij1)$ (a customer of MSP 1 arrives), $(ij2)$ (a customer of MSP 2 arrives), or $(ij0)$ (no one arrives). The cardinality of $\mathbb{S}$, $|\mathbb{S}| = 3(N_1 + 1)(N_2 + 1)$.

Let $\mathbb{A} \overset{\text{def}}{=} \{a_k\} = \{0, 1\}$ denote the pure action/strategy space of each MSP. A *pure action/strategy* $a_k = 1$ of MSP k means the MSP admits the newly arrived customer, and $a_k = 0$ if it refuses to admit the customer. If a customer of MSP k (also referred to as a customer of type k) is rejected by its MSP k, it is then directed to the other MSP for service. At the second MSP, this customer can be admitted or rejected, depending on this MSP's strategy. Note that a customer who has been refused by both MSPs will be discarded.

The state space $\mathbb{S}$ is countable and the transition rate is bounded. Thus, there exists an equivalence between the continuous- and discrete-time domains for the MDP [24]. Hence, we can study this continuous MDP in its equivalent discrete-time domain. Let $P(s'|s, a_1, a_2)$ denote the transition probability (corresponding to the transition rate in the continuous time domain) to state $s' \overset{\text{def}}{=} (i'j'x')$ when actions (a_1, a_2) are implemented (by the two MSPs) at state s. For $i, j > 0$ and $i < N_1; j < N_2$, the transition probabilities are illustrated in Figure 2 and computed as follows [2] (the operator $(\bar{.})$ refers to the binary bit flip):

$$P(i'j'x'|ij2, a_1, a_2) = \frac{1}{L}\begin{cases} \lambda_2 \text{ if } x' = 2; i' = i + a_1\overline{a_2}; j' = j + a_2 \\ \lambda_1 \text{ if } x' = 1; i' = i + a_1\overline{a_2}; j' = j + a_2 \\ i \text{ if } x' = 0; i' = i + a_1\overline{a_2} - 1; j' = j + a_2 \\ j \text{ if } x' = 0; i' = i + a_1\overline{a_2}; j' = j + a_2 - 1 \\ (L - \lambda_1 - \lambda_2 - i - j) \text{ if } x' = 0; i' = i + a_1\overline{a_2}; j' = j + a_2 \\ 0 \text{ for all other states} \end{cases} \tag{1}$$

$$P(i'j'x'|ij1, a_1, a_2) = \frac{1}{L}\begin{cases} \lambda_2 \text{ if } x' = 2; i' = i + a_1; j' = j + a_2\overline{a_1} \\ \lambda_1 \text{ if } x' = 1; i' = i + a_1; j' = j + a_2\overline{a_1} \\ i \text{ if } x' = 0; i' = i + a_1 - 1; j' = j + a_2\overline{a_1} \\ j \text{ if } x' = 0; i' = i + a_1; j' = j + a_2\overline{a_1} - 1 \\ (L - \lambda_1 - \lambda_2 - i - j) \text{ if } x' = 0; i' = i + a_1; j' = j + a_2\overline{a_1} \\ 0 \text{ for all other states} \end{cases} \tag{2}$$

$$P(i'j'x'|ij0, a_1, a_2) = \frac{1}{L}\begin{cases} \lambda_2 \text{ if } x' = 2; i' = i; j' = j \\ \lambda_1 \text{ if } x' = 1; i' = i; j' = j \\ i \text{ if } x' = 0; i' = i - 1; j' = j \\ j \text{ if } x' = 0; i' = i; j' = j - 1 \\ (L - \lambda_1 - \lambda_2 - i - j) \text{ if } x' = 0; i' = i; j' = j \\ 0 \text{ for all other states} \end{cases} \tag{3}$$

where $L = \lambda_1 + \lambda_2 + N_1 + N_2$.

Let $\mathbf{F}_1$ and $\mathbf{F}_2$ be the $|\mathbb{S}|$-by-2 matrices that denote the *mixed/stationary strategies* of MSPs 1 and 2, respectively. $\mathbf{F_k}(s, :)$ denote a distribution vector whose element $\mathbf{F_k}(s, 0)$ is the probability that MSP k rejects (i.e., action $a_k = 0$ is taken) and $\mathbf{F_k}(s, 1)$ is the probability that MSP k accepts (i.e., action $a_k = 1$ is taken) the arriving customer when the system is in state s. We have a stochastic probability transition matrix $|\mathbb{S}|$-by-$|\mathbb{S}|$ $\mathbf{P}(\mathbf{F}_1, \mathbf{F}_2)$, where element (s', s) is denoted as $P(s'|s, \mathbf{F}_1, \mathbf{F}_2)$ with:

$$P(s'|s, \mathbf{F}_1, \mathbf{F}_2) = \sum_{a_1 \in \mathbb{A}} \sum_{a_2 \in \mathbb{A}} P(s'|s, a_1, a_2)\mathbf{F}_1(s, a_1)\mathbf{F}_2(s, a_2) \tag{4}$$

The reward of operator k at state s when actions a_1, a_2 are executed by the two MSPs is denoted by $r_k(s, a_1, a_2)$, where:

$$r_1(ij1, a_1, a_2) = \begin{cases} p_1 \text{ if } a_1 = 1 \text{ and } i < N_1, \\ p_1(1 - \beta_1) \text{ if } (a_1 = 0 \text{ or } i = N_1) \text{ and } a_2 = 1 \text{ and } j < N_2, \\ 0 \text{ for all other cases} \end{cases} \tag{5}$$

and

$$r_2(ij1, a_1, a_2) = \begin{cases} p_1\beta_1 \text{ if } (a_1 = 0 \text{ or } i = N_1) \text{ and } a_2 = 1 \text{ and } j < N_2, \\ 0 \text{ for all other cases} \end{cases} \tag{6}$$

and

$$r_1(ij2, a_1, a_2) = \begin{cases} p_2\beta_2 \text{ if } (a_2 = 0 \text{ or } j = N_2) \text{ and } a_1 = 1 \text{ and } i < N_1, \\ 0 \text{ for all other cases} \end{cases} \tag{7}$$

and

$$r_2(ij2, a_1, a_2) = \begin{cases} p_2 \text{ if } a_2 = 1 \text{ and } j < N_2, \\ p_2(1 - \beta_2) \text{ if } (a_2 = 0 \text{ or } j = N_2) \text{ and } a_1 = 1 \text{ and } i < N_1, \\ 0 \text{ for all other cases} \end{cases} \tag{8}$$

and $r_1(ij0, a_1, a_2) = r_2(ij0, a_1, a_2) = 0$.

Let $V_k^{(a)}(s, \mathbf{F}_1, \mathbf{F}_2)$ denote the *reward rate* (or the *average reward* over time) that MSP k earns when the two MSPs start at state s[3]. By definition:

$$V_k^{(a)}(s, \mathbf{F}_1, \mathbf{F}_2) = \lim_{T \to \infty} \frac{1}{1 + T} \sum_{t=0}^{T} r_k^{(t)}(s, \mathbf{F}_1, \mathbf{F}_2) \tag{9}$$

where $r_k^{(t)}(s, \mathbf{F}_1, \mathbf{F}_2)$ is the *expected reward at time t* (w.r.t. $\mathbf{F}_1$ and $\mathbf{F}_2$) of MSP k when the system starts in state s.

Define an $|\mathbb{S}| \times 1$ vector $\mathbf{r}_k^{(t)}(\mathbf{F}_1, \mathbf{F}_2) \overset{\text{def}}{=} [r_k^{(t)}(1, \mathbf{F}_1, \mathbf{F}_2), \ldots, r_k^{(t)}(|\mathbb{S}|, \mathbf{F}_1, \mathbf{F}_2)]$. We have:

$$\mathbf{r}_k^{(t)}(\mathbf{F}_1, \mathbf{F}_2) = \mathbf{P}(\mathbf{F}_1, \mathbf{F}_2)^t \mathbf{r}_k(\mathbf{F}_1, \mathbf{F}_2) \tag{10}$$

where the $|\mathbb{S}| \times 1$ vector $\mathbf{r}_k(\mathbf{F}_1, \mathbf{F}_2) \overset{\text{def}}{=}$

[2]When either one of the MSPs is full (e.g., $j = N_2$) or empty (e.g., $i = 0$), the transition probabilities need to be revised accordingly. We omit these simple cases here due to space limitation.

[3]Note that the discounted reward criterion which is easier to analyze due to its guaranteed convergence/existence of the reward function [17] can also be studied in a similar manner.

$[r_k(1,\mathbf{F}_1,\mathbf{F}_2),\ldots,r_k(s,\mathbf{F}_1,\mathbf{F}_2),\ldots,r_k(|\mathbb{S}|,\mathbf{F}_1,\mathbf{F}_2)]$ and $r_k(s,\mathbf{F}_1,\mathbf{F}_2)$ is the *initial/immdediate expected reward* when the system starts in state s:

$$r_k(s,\mathbf{F}_1,\mathbf{F}_2) = \sum_{a_1\in\mathbb{A}}\sum_{a_2\in\mathbb{A}} r_k(s,a_1,a_2)\mathbf{F}_1(s,a_1)\mathbf{F}_2(s,a_2). \tag{11}$$

The following proposition states the existence of the reward rate in (9):

Proposition 1: If MSPs aim to maximize the reward rate $V_k^{(a)}(s,\mathbf{F}_1,\mathbf{F}_2)$, i.e., the *average criterion* is used, the underlying Markov decision process with state space $\mathbb{S}$ and the transition probability matrix $\mathbf{P}(\mathbf{F}_1,\mathbf{F}_2)$ is irreducible, and $V_k^{(a)}(s,\mathbf{F}_1,\mathbf{F}_2)$ in (9) is well-defined.

Proof: Our idea is to prove that starting from any state, we can go to state (000) to empty the system (due to the service completion). Then, from the empty state, the system can go to any other state. We now can cite the Theorem 5.1.5 [17] that states $V_k^{(a)}(s,\mathbf{F}_1,\mathbf{F}_2)$ in (9) is well-defined and identical for all initial states. The detailed proof is in the Appendix A. □

While sharing its resource, an MSP needs to maintain its *QoS commitment*. Specifically, the QoS of MSP k is measured by the probability that a customer of that MSP does not get served (the lower this probability the higher QoS), denoted by $R_k(\mathbf{F}_1,\mathbf{F}_2)$. The QoS commitment ensures that while sharing its resource, an operator k *either* meets its QoS target, QoS_k, (i.e., $QoS_k \geq R_k(\mathbf{F}_1,\mathbf{F}_2)$) *or* at least achieves the same level of QoS as if it did not share its resources with the other MSP, defined as $P_b(\lambda_k,N_k)$ (i.e., $P_b(\lambda_k,N_k) \geq R_k(\mathbf{F}_1,\mathbf{F}_2)$). Without sharing its resource, $P_b(\lambda_k,N_k)$ is exactly the Erlang-B blocking probability, computed as:

$$P_b(\lambda_k,N_k) = \frac{\lambda_k^{N_k}}{N_k!\sum_{i=0}^{N_k}\frac{\lambda_k^i}{i!}}.$$

$R_k(\mathbf{F}_1,\mathbf{F}_2)$ is given by $\boldsymbol{\pi}\boldsymbol{Re}_k$ where $\boldsymbol{\pi}$ is the stationary distribution vector of the MDP under the stationary strategies $(\mathbf{F}_1,\mathbf{F}_2)$; $\boldsymbol{Re}_k$ is an $|\mathbb{S}|\times 1$ vector whose element $\boldsymbol{Re}_k(s)$ is the probability that a customer of MSP k does not get served *given the system is in state* s. For MSP 1, $\boldsymbol{Re}_1(s)=0$ if $s\in\{(i0j0),(i0j2)\}$ and $\boldsymbol{Re}_1(s)=\mathbf{F}_1(s,0)\mathbf{F}_2(s,0)$ if $s=(i1j0)$ (as a customer does not get served if and only if it is rejected by both operators). Similarly, $\boldsymbol{Re}_2(s)=0$ if $s\in\{(i0j0),(i1j0)\}$ and $\boldsymbol{Re}_2(s)=\mathbf{F}_1(s,0)\mathbf{F}_2(s,0)$ if $s=(i0j2)$.

The objective of each MSP is to optimize its own stationary strategy given the other MSP's strategy so as to maximize its reward rate while maintaining its QoS commitment. From the above, $V_k^{(a)}(s,\mathbf{F}_1,\mathbf{F}_2)$ does not depend on which state the system starts from. Thus, for brevity, let $V_k^{(a)}(\mathbf{F}_1,\mathbf{F}_2)$ denote the reward rate of operator k. Formally, each operator k needs to solve the following problem:

$$\begin{aligned} &\underset{\mathbf{F_k}}{\text{maximize}} \;\; V_k^{(a)}(\mathbf{F}_1,\mathbf{F}_2)\\ \text{s.t.}\quad &\text{C1: } \sum_{a_k\in\mathbf{A}}\mathbf{F}_k(s,a_k)=1,\;\;\forall s\\ &\text{C2: } 1\geq \mathbf{F}_k(s,a_k)\geq 0,\;\;\forall s,\forall a_k\in\mathbf{A}\\ &\text{C3: } \max(QoS_k,P_b(\lambda_k,N_k))\geq R_k(\mathbf{F}_1,\mathbf{F}_2)\end{aligned} \tag{12}$$

where constraints C1 and C2 are to ensure that each row of $\mathbf{F}_k$ is a probability distribution vector. C3 enforces the QoS commitment.

B. NE Existence and Characterization

Theorem 1: There exists a NE for the game (12) in which MSPs aim to maximize their reward rates.

Proof: Game (12) belongs to the class of constrained Markov game [18] in which players' strategies/rewards depend on the system state. The system state transitions, in return, depend on players' strategies and make an MDP. For the NE to exist, we rely on the results in [18]. Theorem 2.1 in [18] states that a constrained Markov game admits at least one NE if the two following conditions hold:

- *(Ergodicity)* If the average criterion is used, then the state process is an irreducible Markov chain.
- *(Strong Slater)* For any stationary strategy from the other player, a player can still find its stationary strategy to ensure that the constraint is met.

As a consequence of Proposition 1, it is easy to see that the ergodicity condition holds. Game (12) is then an irreducible stochastic game, and without considering C3, the game admits at least one NE (according to Theorem 5.4.5 in [17]). In our case, the slater condition also holds. Specifically, for MSP 1 the LHS of C3 is greater than or equal to the blocking/rejecting probability $P_b(\lambda_1,N_1)$. Hence, C3 can always be met if MSP 1 refuses to serve customers from MSP 2. This is realized by executing the stationary strategy with $\mathbf{F}_1^+(s,0)=1,\forall s\in\{ijx|x=2,0\}$, regardless of strategies from the other player. In other words, the Strong Slater condition holds. Thus, there exists at least one NE to the constrained Markov game (12). □

Let $(\mathbf{F}_1^*,\mathbf{F}_2^*)$ are the stationary strategies at an NE. We have the following corollary (answering the question $\boldsymbol{Q_1}$ in Section II).

Corollary 1: Cooperation in game (12) is rational, i.e., both MSPs have incentives to share their resources.

Proof: If MSPs do not share their resources, the reward rates are $(1-P_b(\lambda_1,N_1))\lambda_1 p_1$ and $(1-P_b(\lambda_2,N_2))\lambda_2 p_2$ for MSP 1 and MSP 2, respectively. Let $\mathbf{F}_1^+$ be the strategy of MSP 1 when it does not accept customer type 2, i.e., $\mathbf{F}_1^+(s,0)=1,\forall s\in\{ijx|x=2,0\}$. By definition of NE strategies $(\mathbf{F}_1^*,\mathbf{F}_2^*)$, we have:

$$\begin{aligned} V_1^{(a)}(\mathbf{F}_1^*,\mathbf{F}_2^*) &\geq V_1^{(a)}(\mathbf{F}_1^+,\mathbf{F}_2^*)\geq(1-P_b(\lambda_1,N_1))\lambda_1p_1\\ V_2^{(a)}(\mathbf{F}_1^*,\mathbf{F}_2^*) &\geq V_2^{(a)}(\mathbf{F}_1^*,\mathbf{F}_2^+)\geq(1-P_b(\lambda_2,N_2))\lambda_2p_2\end{aligned} \tag{13}$$

In the above, $V_1^{(a)}(\mathbf{F}_1^+,\mathbf{F}_2^*)\geq(1-P_b(\lambda_1,N_1))\lambda_1p_1$ because although MSP 1 rejects customers from MSP 2, MSP 1's customers are still offloaded and can be accepted by MSP 2 under MSP 2's policy $\mathbf{F}_2^*$. Corollary 1 is proved. □

Intuitively, Corollary 1 also guarantees that each MSP does at least as good as he would if he does not participate in the traffic offloading game. The following theorem states necessary and sufficient conditions of a NE of the game (12) that can be used to find the game's NE(s).

Theorem 2: Any pair $(\mathbf{F}_1,\mathbf{F}_2)$ is a NE of the constrained Markov game (12) *if and only if* $\mathbf{z}\overset{\text{def}}{=}(\mathbf{v}_1,\mathbf{v}_2,\mathbf{F}_1,\mathbf{F}_2,\mathbf{u}_1,\mathbf{w}_1,\mathbf{u}_2,\mathbf{w}_2)$ is the globally optimal solution of the following problem, and its optimal value is 0:

$$\begin{aligned}
&\underset{\boldsymbol{z}}{\text{minimize}} \quad \sum_{k=1}^{2} \mathbf{1}^T[\mathbf{v}_k - \mathbf{P}(\mathbf{F}_1, \mathbf{F}_2)\mathbf{v}_k] \\
\text{s.t.} \quad & \text{C1: } \mathbf{T}(s, \mathbf{v}_1)\mathbf{F}_2 \le v_1(s)\mathbf{1}, \forall s \\
& \text{C2: } \mathbf{r}_1(s)\mathbf{F}_2 + \mathbf{T}(s, \mathbf{u}_1)\mathbf{F}_2 \le (v_1(s) + u_1(s))\mathbf{1}, \forall s \\
& \text{C3: } \mathbf{F}_1\mathbf{T}(s, \mathbf{v}_2) \le v_2(s)\mathbf{1}, \forall s \\
& \text{C4: } \mathbf{F}_1\mathbf{r}_2(s) + \mathbf{F}_1\mathbf{T}(s, \mathbf{u}_2) \le (v_2(s) + u_2(s))\mathbf{1}, \forall s \\
& \text{C5: } r_k(\mathbf{F}_1, \mathbf{F}_2) + \mathbf{P}(\mathbf{F}_1, \mathbf{F}_2)\mathbf{w}_k = \mathbf{v}_k + \mathbf{w}_k, k = 1, 2 \\
& \text{C6: } \sum_{a_k \in \mathbf{A}} \mathbf{F_1}(\mathbf{s}, \mathbf{a_k}) = 1; \sum_{a_k \in \mathbf{A}} \mathbf{F_2}(\mathbf{s}, \mathbf{a_k}) = 1 \ \ \forall s \\
& \text{C7: } 1 \ge \mathbf{F_k}(\mathbf{s}, \mathbf{a_k}) \ge 0, \ \ \forall s, \forall a_k \in \mathbf{A} \\
& \text{C8: } \max(QoS_k, P_b(\lambda_k, N_k)) \ge R_k(\mathbf{F}_1, \mathbf{F}_2), k = 1, 2.
\end{aligned} \tag{14}$$

where $\mathbf{1}$ is a column vector with all ones; $\mathbf{u}_k$ and $\mathbf{w}_k$ are 1 by $|\mathbb{S}|$ vectors of auxiliary variables $u_k(s)$, $w_k(s)$, respectively; $\mathbf{T}(s, \mathbf{v}_k)$, $\mathbf{T}(s, \mathbf{u}_k)$, and $\mathbf{r}_k(s)$ are 2 by 2 matrices whose elements are $\sum_{s' \in \mathbb{S}} P(s'|s, a_1, a_2)\mathbf{v}_k(s')$, $\sum_{s' \in \mathbb{S}} P(s'|s, a_1, a_2)\mathbf{u}_k(s')$, and $r_k(s, a_1, a_2)$ for $a_1, a_2 \in \{0, 1\}$, respectively.

Proof: The proof is similar to that of Theorem 3.8.4 in [17], except the constraint C8. The detailed proof is in Appendix B. □

The NE strategies are also the optimal strategies for each MSP in accepting/rejecting/redirecting customers (answering the question $\boldsymbol{Q_2}$ in Section II). Problem (14) may have multiple solutions, i.e., multiple NEs. It is possible to characterize the case in which the NE is unique by converting (14) to its nonlinear complementarity problem then relying on variational inequality theory [25] (details are omitted due to space limitation). Using a gradient-based algorithm, we can numerically obtain a solution very close (within 10^{-7}) to the optimal value of (14) that is lower-bounded by 0.

Remark 1: First, (14) is a nonlinear problem (with nonlinear constraints and its objective function) that involves $10 \times 3 \times (N_1 + 1) \times (N_2 + 1)$ variables. For a reasonable cell size (e.g., 10 customers per pico-cell), the number of variables $30 \times 11 \times 11 = 3630$ is very large. Hence, the computational complexity involved in solving (14) is significant. In fact, we attempted to solve (14) using a gradient-based algorithm but the running time is quite long. Second, computing the NE via problem (14) requires a MSP to reveal its capacity and resource availability status to the other MSP. Besides the MSPs' willingness to share their business privacy with other MSPs, this approach requires additional communication overhead for them to exchange this information. In the following, we derive a practical algorithm that only relies on local information and achieves a lower-bound (it is shown to be tight via simulations) for NE utilities of game (12).

III. Practical Implementation to Achieve a Tight Lower Bound

Even ignoring the complexity of solving (14), if one operator in game (12) does not want to share its state (i.e., its on-going number of customers or the resource availability), the the other operator is unable to derive its NE strategies. In such a case, it is vital to derive an alternative strategy by looking into only the local information at the operator (addressing the question $\boldsymbol{Q_3}$ in Section II). In this section, we will derive a simple policy that yet achieves a low-bound on the reward rate of (14) which is shown to be tight via simulations.

Note that the offloaded traffic from a MSP is not an *overflow process*[4]. In fact, its statistical characteristics depend on the MSP's strategy. That makes the approach in [26] not readily applicable. However, we observe that the actual offloaded traffic process from MSP k (in game (14)) is comprised of not only overflow customers (rejected because of not having enough resources, with rate $P_b(\lambda_k, N_k)\lambda_k$) but also customers rejected even when having enough resources. Thus, we can find a lower bound for the reward rate under the optimal strategy (derived from (14)) by replacing the offloaded traffic process with an overflow process with rate $P_b(\lambda_k, N_k)\lambda_k$. The resulting reward rate is a lower bound because its derived accept/reject policy is suboptimal for (14) (by always rejecting customers offloaded by the MSP who still has enough resources).

We now limit our interest to MSP 1, and the following results/analysis also apply to MSP 2. MSP 1 serves two types of customers, one arriving according to a Poisson process of rate λ_1 and the other following an overflow process of rate $P_b(\lambda_2, N_2)\lambda_2$. The authors in [26] pointed out that the average reward rate of MSP 1 can be well approximated by assuming the overflow process also follows a Poison process. Now, MSP 1 serves two types of customers, arriving according to Poisson distribution with rates λ_1 and $\lambda_2' \overset{\text{def}}{=} P_b(\lambda_2, N_2)\lambda_2$.

Remark 2: Note that the following approach is also applicable to scenario S_3 in which two MSPs fully cooperate by contributing their resource and maximize their total reward rate (i.e., a $M/M/(N_1 + N_2)$ queue with two classes of customers arriving at rates λ_1 and λ_2, and rewards p_1 and p_2, respectively).

To solve for the optimal admission policy, we denote the system state at MSP 1 as $s = (ix)$, where i is the number of customers being served and x is the type of the incoming customer at the MSP ($x = 0$ if no one arrives). The system space $\mathbb{S}' \overset{\text{def}}{=} \{s\}$. The transition probability of the state process and corresponding rewards are as follows:

$$P(i'x'|i0, a_1) = \tfrac{1}{N_1+\lambda_1+\lambda_2'} \begin{cases} \lambda_2' \text{ if } \mathrm{x}' = 2; i' = i \\ \lambda_1 \text{ if } \mathrm{x}' = 1; i' = i \\ i \text{ if } \mathrm{x}' = 0; i' = i - 1 \\ (N_1 - i) \text{ if } \mathrm{x}' = 0; i' = i \\ 0 \text{ for all other states} \end{cases}$$

$$\begin{aligned}
P(i'x'|i1, a_1) &= \frac{1}{N_1 + \lambda_1 + \lambda_2'} \begin{cases} \lambda_2' \text{ if } \mathrm{x}' = 2; i' = i + a_1 \\ \lambda_1 \text{ if } \mathrm{x}' = 1; i' = i + a_1 \\ i \text{ if } \mathrm{x}' = 0; i' = i + a_1 - 1 \\ (N_1 - i) \text{ if } \mathrm{x}' = 0; i' = i + a_1 \\ 0 \text{ for all other states} \end{cases} \\
&= P(i'x'|i2, a_1) \\
r(i0, a_1) &= 0 \\
r(i1, a_1) &= a_1 p_1 \\
r(i2, a_1) &= a_1 p_2 \beta_2
\end{aligned}$$

Without the QoS commitment constraint, the optimal admission strategy that maximizes the average reward is the (deterministic) trunk reservation policy [20]. However, with the

[4]For a finite queue/capacity system, an overflow process captures customers who are not admitted due to overflow.

QoS commitment, one faces a constrained MDP whose optimal admission policy is generally not deterministic. As MSPs do not reveal their strategies, MSP 1 has to maintain its QoS commitment regardless of MSP 2's strategy. The probability that a customer of MSP 1 does not get served $R'_1(\mathbf{F}_1)$ is computed as $R'_1(\mathbf{F}_1) = \boldsymbol{\pi'}\boldsymbol{Re'}_1$, where $\boldsymbol{\pi'}$ is the stationary distribution of the MDP with state space $\mathbb{S}'$. $\boldsymbol{Re'}_1$, similar to $\boldsymbol{Re}_1$, is a vector whose element $\boldsymbol{Re'}_1(s) = \boldsymbol{F}_1(s,0)$ if $s = (i1)$ and $\boldsymbol{Re'}_1(s) = 0$ if $s = (i0)$.

The optimal stationary policy $\boldsymbol{F}_1(s,a_1)$ that maximizes the average reward $V'^{(a)}_1$ for the constrained MDP with the states and transition probabilities above is obtained by solving the following problem [27]:

$$
\begin{aligned}
& \underset{\boldsymbol{f}}{\text{maximize}} \;\; V'^{(a)}_1 = \sum_{s=1}^{\mathbb{S}'} \sum_{a \in \boldsymbol{A}} r(s,a_1) f(s,a_1) \\
\text{s.t.} \quad & \text{C1: } \boldsymbol{W}\boldsymbol{f} = 0 \\
& \text{C2: } \mathbf{1}\boldsymbol{f} = 1 \\
& \text{C3: } \max(QoS_1, P_b(\lambda_1, N_1)) \geq R'_1(\mathbf{F}_1) \\
& \text{C4: } \boldsymbol{f} \geq 0.
\end{aligned}
\tag{15}
$$

where $\boldsymbol{W}$ is an $|\mathbb{S}'| \times 2|\mathbb{S}'|$ matrix with $w_{s',(s,a_1)} = \delta(s,s') - P(s'|s,a_1)$ ($\delta(s,s')$ is a Kronecker delta function). C3 is needed to enforce the QoS commitment. $\boldsymbol{f}$ is a vector of nonnegative auxiliary variables that are used to compute the optimal stationary policy as $\boldsymbol{F}_1(s,a_1) = \frac{f(s,a_1)}{\sum\limits_{a_1 \in \boldsymbol{A}} f(s,a_1)}$.

Remark 3: The two operators do not need to share information regarding their resources, capacity, and traffic load (i.e., λ_1, λ_2). The only external input for each operator to make its decision is the rate of offloaded traffic from the other. This rate can be estimated/learnt locally and accurately with initial training time, so are customer/traffic arrival rates (i.e., λ_1, λ'_2).

The above resource sharing framework is feasible for implementation in existing and future cellular systems. MSPs just need to negotiate the revenue-sharing contracts, i.e., parameters β_1 and β_2. The values of β_1 and β_2 that achieve the social optimality (i.e., maximizing the total utilities of the two MSPs) can be found numerically, e.g., in Figures 4 to 7. After setting β_1 and β_2, the two MSPs do not need to exchange any control messages. The length of the contract is up to MSPs but can be as short as days or as long as years before renegotiating. The proposed methods allow MSPs to harvest resource/spectrum opportunity as fine as to serving a single service request (as it guides MSPs to make decision for every single service arrival).

IV. K MSPs Case ($K > 2$)

As aforementioned, the above results are still valid when more than two MSPs are considered. Theoretically, for more than two MSPs, one can follow similar analysis in [28] of a Markov stochastic game with K users ($K > 2$) under both discounted and limiting average reward rate. Specifically, the state of the underlying MDP is then a $(K+1)$-tuple of which K elements present the number of customers being served by K MSPs and the last element presents the type of the coming request. The state space cardinality is $(K+1)(N_1+1)...(N_K+1)$ (hence finite). The action space for each MSP is the same as the 2-MSP case. The transition probabilities can be calculated similarly as in equations (1-3). NE existence and a NE of the game can be proved and found similarly by solving a nonlinear programming in Theorems 2.1 and 3.2 of [28]. For K-user case, the cooperation is also rational to all MSPs.

Practically, computing the NE of the game will require MSPs to reveal their capacity, resource availability status to each other. Additionally, the number of variables involved in NE computation is $10 \times 3 \times (N_1+1)...(N_K+1)$. Even for a reasonable capacity N_k, this is a large number of variables. As such, one can rely a practical solution in Section III. For the case with K MSPs, the transition probabilities are similar to the case with 2 MSPs and as follows:

$$
P(i'x'|i0,a_1) = \frac{1}{N_1+\lambda_1+\lambda'_2} \begin{cases} \lambda'_k \text{ if } \mathrm{x}' = k; i' = i, k \neq i \\ \lambda_1 \text{ if } \mathrm{x}' = 1; i' = i \\ i \text{ if } \mathrm{x}' = 0; i' = i-1 \\ (N_1 - i) \text{ if } \mathrm{x}' = 0; i' = i \\ 0 \text{ for all other states} \end{cases}
$$

$$
\begin{aligned}
P(i'x'|i1,a_1) &= \frac{1}{N_1+\lambda_1+\lambda'_2} \begin{cases} \lambda'_k \text{ if } \mathrm{x}' = k; i' = i+a_1, k \neq i \\ \lambda_1 \text{ if } \mathrm{x}' = 1; i' = i+a_1 \\ i \text{ if } \mathrm{x}' = 0; i' = i+a_1-1 \\ (N_1 - i) \text{ if } \mathrm{x}' = 0; i' = i+a_1 \\ 0 \text{ for all other states} \end{cases} \\
&= P(i'x'|ik,a_1), k = 2,...,K
\end{aligned}
$$

Given the fact that the offloading rate λ_k with $k = 2,...,K$ can be estimated, the optimal admission strategy of the MSP k can be obtained by solving the problem (15).

V. Application in Sharing Base Stations

We now consider an extreme, but interesting, case of the above, in which an MSP wishes to turn off its base stations and relies on the other MSP to carry its traffic. This case is of great interest when both MSPs experience light traffic (e.g., at nighttime) and base stations can be switched off to save operational costs (e.g., energy consumption). In the sequel, we determine when a MSP should turn its base station on or off based on traffic conditions.

For that purpose, we formulate the interactions between MSPs as a noncooperative game in which MSPs are players, aiming to maximize their own profit rates. The profit rate of MSP 1 (similarly for MSP 2) under different pure strategies of the two MSPs (turning on/off their base stations) is defined as:

$$
\begin{aligned}
&U_1(\textit{on,on}) = V^{(a)}_1(\textit{on,on}) - E_1 \\
&U_1(\textit{on,off}) = V^{(a)}_1(\textit{on,off}) - E_1 \\
&U_1(\textit{off,on}) = \begin{cases} V^{(a)}_1(\textit{off,on}) \text{ if MSP 1's QoS commit' is honored,} \\ -\infty \text{ otherwise} \end{cases} \\
&U_1(\textit{off,off}) = -\infty
\end{aligned}
$$

where E_1 is the expense to operate a BS of MSP 1 per time unit, (*on, on*) refers to the case when both operators turn on their BSs. The other cases (*on, off*), (*off, on*), and (*off, off*) are defined in an analogous manner.

Note that the above utility equations implicitly enforce the QoS commitment. Specifically, the *only* possibility that QoS commitment of a MSP is violated is when its BS is turned off (however, turning off a BS does not *always* lead to QoS commitment violation, thanks to traffic offloading). Such a case is eliminated by setting the MSP's utility to $-\infty$ (to discourage the MSP from turning its BSs off).

Let (o_1, o_2) denote the strategies of the two MSPs; $o_k = on$ (or $o_k = off$) if MSP k keeps its BS on (or turns its BS off). By definition of a NE, (o_1^*, o_2^*) is an optimal NE strategy if and only if $U_1(o_1^*, o_2^*) \geq U_1(o_1, o_2^*)$ and $U_2(o_1^*, o_2^*) \geq U_1(o_1^*, o_2)$. We have the following theorem:

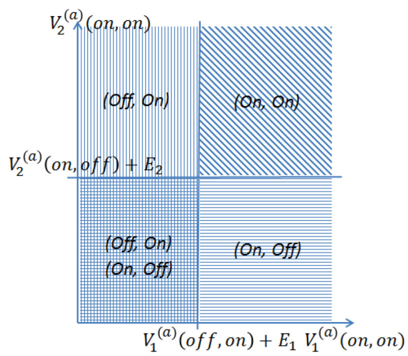


Fig. 3. Pure NE of BSs switching on/off game.

Theorem 3: For MSP k, the optimal BS switching on/off strategies are as follows: turn off if $V_k^{(a)}(\textit{on, on}) < V_k^{(a)}(\textit{off, on}) + E_k$ and its QoS commitment is honored (case A_k); turn on otherwise (i.e., either turning off leading to QoS commitment violation (case B_k) or $V_k^{(a)}(\textit{on, on}) > V_k^{(a)}(\textit{off, on}) + E_k$ (case C_k)).

The proof is straightforward by recalling the utility definition above. □

The NEs for the BS switching on/off game are then stated in the following corollary (depicted in Figure 2):

Corollary 2: For cases (B_1C_2), (B_1B_2), (C_1B_2), and (C_1C_2), the NE is (*on, on*); for cases (B_1A_2) and (C_1A_2), the NE is (*on, off*); for cases (A_1B_2) and (A_1C_2), the NE is (*off, on*).

Proof: The corollary can be easily verified from the above definition of NE (o_1^*, o_2^*). □

Note that there are two pure NEs when $V_2^{(a)}(\textit{on, on}) < V_2^{(a)}(\textit{on, off}) + E_2$ and $V_1^{(a)}(\textit{on, on}) < V_1^{(a)}(\textit{off, on}) + E_1$. In this region, if one MSP turns its BSs off, then the other has to turn its BS on. If the utilities of both MSPs at a NE are higher than those under the other NE, then they both should implement that NE.

Nonetheless, if only one MSP is better off by moving from one NE to the other while the other MSP's utility reduces, the MSP with utility reduction should be provided with incentives so that both can move to the NE with higher welfare (i.e., at which the total utility of both operators is higher). Let $(U_1^{(1)}, U_2^{(1)})$ and $(U_1^{(2)}, U_2^{(2)})$ be the utilities of the two MSPs at the NE 1 and NE 2, respectively. Without loss of generality, we assume the social welfare at NE 2 is higher (i.e., $U_{11} + U_{21} \leq U_{12}$ and $U_{11} + U_{21} \leq U_{22}$) and MSP 2 is better off by moving NE 1 to NE 2. To incentivise MSP 1 to move to NE 2, MSP 2 proposes a payment of Δ to MSP 1. Applying a Nash bargaining mechanism [29], one can find that setting $\Delta = \frac{U_1^{(2)} + U_2^{(1)} - U_2^{(2)} - U_1^{(1)}}{2}$ ensures that both MSPs get the same amount of utility improvement $(\frac{U_1^{(2)} - U_1^{(1)} + U_2^{(2)} - U_2^{(1)}}{2})$ when moving from NE 1 to NE 2.

VI. Numerical Results

In this section, we use Matlab simulations to evaluate the average reward under scenarios S_1, S_3, and the lower-bound of game (12) in S_2. The admission policy for the lower-bound of game (12), numerically obtained by solving (15), is then used to govern the admission policy of two real MSPs in simulations.

A. Moderate Traffic at both MSPs

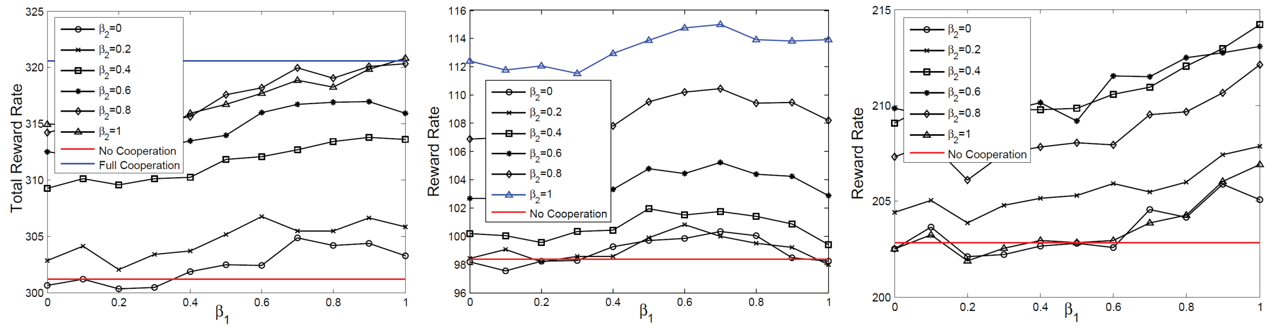
Let $N_1 = 10$, $N_2 = 15$, $\lambda_1 = 8$, $\lambda_2 = 14$, $p_1 = 14$, and $p_2 = 17$. The corresponding blocking rates at two MSPs are 0.1217 and 0.1478. The total reward rate when the two MSPs do not cooperate is $(1-P_b(\lambda_1, N_1))\lambda_1 p_1 + (1-P_b(\lambda_2, N_2))\lambda_2 p_2 = 302$ (S_1). The total reward rate when the two MSPs fully cooperate (S_3) is obtained by solving (15). It is 321 in our case. The cooperation gain is about 6.3%. Figure 4 and Figure 5 depict the total reward rate, reward rate of MSP 1, and reward rate of MSP 2 vs. β_1 and β_2, respectively (lower bounds in S_2).

B. Unbalanced Traffic

$N_1 = 10$; $N_2 = 15$; $\lambda_1 = 5$; $\lambda_2 = 25$; $p_1 = 14$; $p_2 = 17$. The corresponding blocking rates at two MSPs are 0.0184 and 0.443. The total reward rate when two MSPs do not cooperate is $(1 - P_b(\lambda_1, N_1))\lambda_1 p_1 + (1 - P_b(\lambda_2, N_2))\lambda_2 p_2 = 277$ (S_1). The total reward rate when two MSPs fully cooperate (S_3) is obtained by solving (15). It is 329.5 in our case. The cooperation gain is about 20%. This is very significant when accumulating over the time horizon. Figure 6 and Figure 7 depict the total reward rate, reward rate of MSP 1, and reward rate of MSP 2 vs. β_1 and β_2, respectively (lower bounds in S_2).

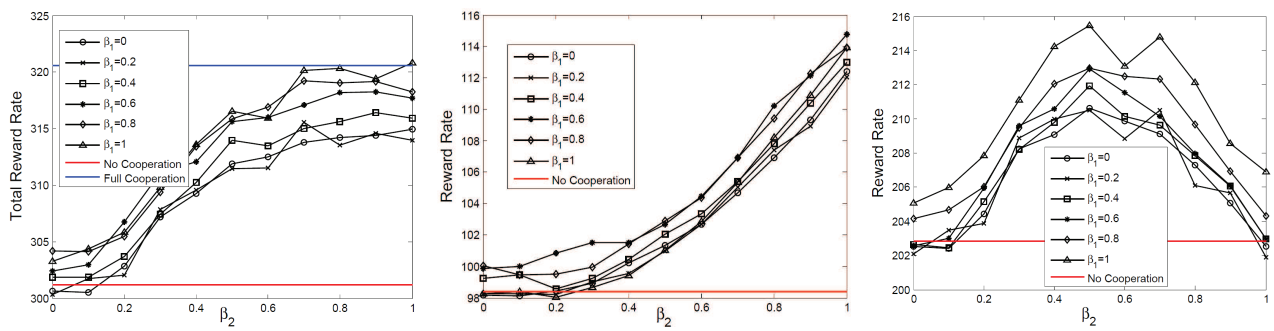
As can be seen, by selecting appropriate β_1 and β_2, the lower bound on total reward rate under S_2 (obtained by solving (15)) is almost that under S_3 when both MSPs fully cooperate (Figure 4(a), 5(a), 6(a), 7(a)). This means that the lower-bound for the reward rate in S_2 can be made very tight by tuning β_1 and β_2 (addressing the question $\boldsymbol{Q_4}$ in Section II). Additionally, the reward rate of a MSP (e.g., MSP 1) monotonically increases w.r.t. the fraction of reward (β_2) it gets from serving the other MSP's customers (Figure 4(c), 5(b), 6(c), and 7(b)). However, the reward rate of a MSP (e.g., MSP 1) has a almost concave shape w.r.t. the fraction of reward (β_1) it pays for the other MSP to carry its traffic (Figure 4(b), 5(c), 6(b), and 7(c)). This is because if the reward from serving traffic offloaded from a MSP is too small, then the other MSP will reserve less resource for offloaded customers. This leads to the loss of revenue for the traffic owner. On the other hand, the traffic owner also earns less if it pays the other MSP too much for carrying its traffic. The critical values β_1 and β_2 that shape the revenue sharing contract can be found numerically.

The average reward of both MSPs vs. traffic loads λ_1 and λ_2 are shown in Figure 8(a) and 8(b). As can be seen, the higher the traffic load the higher the gain can be harvested via traffic offloading and the gain can be up to 60%, compared with the case without traffic offloading.



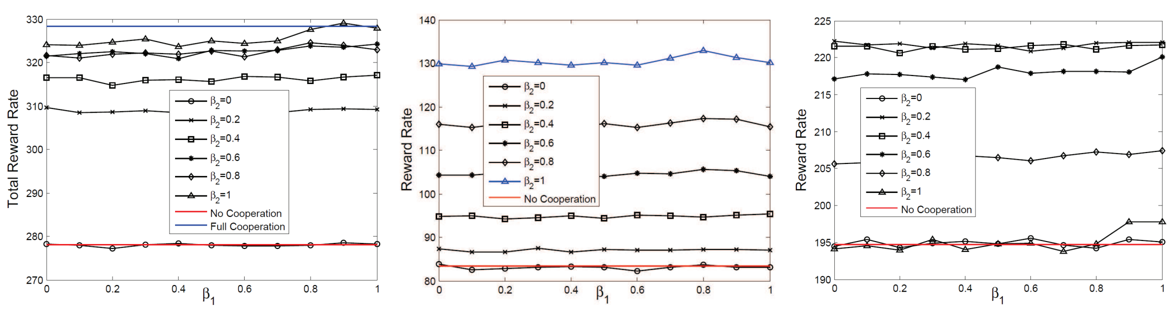
(a) (b) (c)

Fig. 4. (a) Total reward rate, (b) reward rate of MSP 1, (c) reward rate of MSP 2 vs. β_1 (moderate traffic at both MPSs).



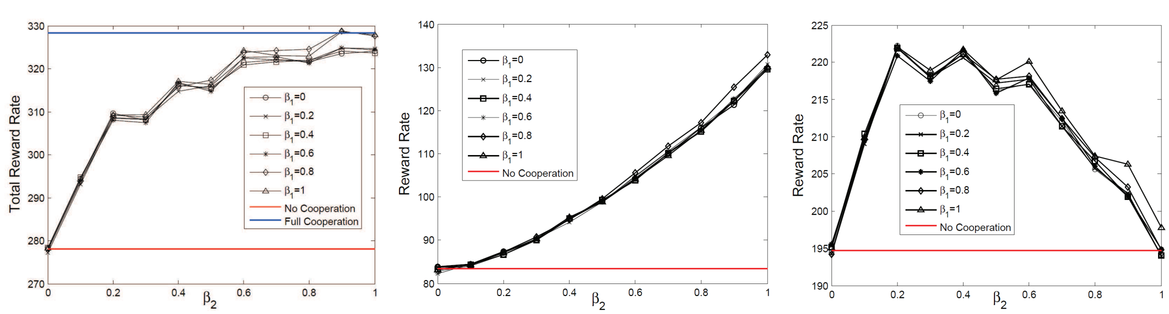
(a) (b) (c)

Fig. 5. (a) Total reward rate, (b) reward rate of MSP 1, (c) reward rate of MSP 2 vs. β_2 (moderate traffic at both MPSs).



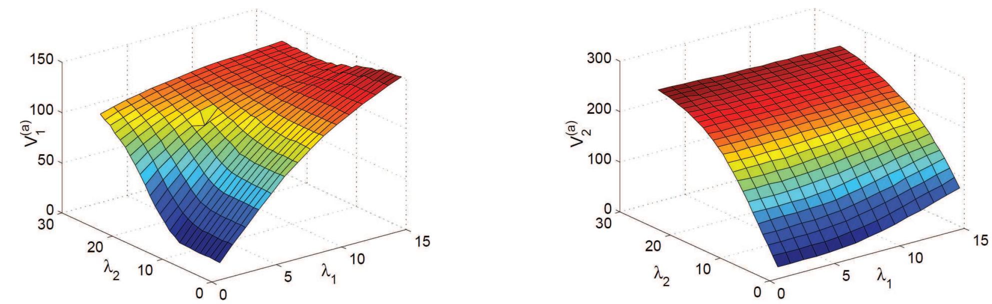
(a) (b) (c)

Fig. 6. (a) Total reward rate, (b) reward rate of MSP 1, (c) reward rate of MSP 2 vs. β_1 (unbalanced traffic).



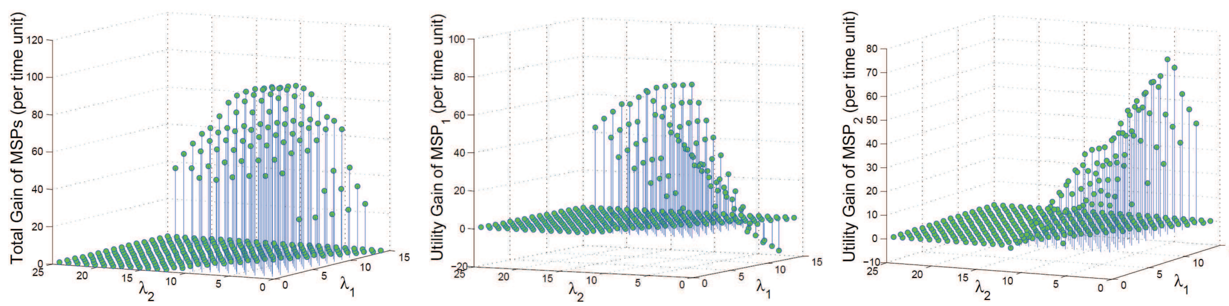
(a) (b) (c)

Fig. 7. (a) Total reward rate, (b) reward rate of MSP 1, (c) reward rate of MSP 2 vs. β_2 (unbalanced traffic).



(a) (b)

Fig. 8. (a) Reward rate of MSP 1 and (b) MSP 2 vs. λ_1 and λ_2 ($\beta_1 = 0.7$, $\beta_2 = 0.8$).



(a) (b) (c)

Fig. 9. (a) Total utility gain, (b) utility gain of MSP 1, and (c) utility gain of MSP 2 vs. λ_1 and λ_2 by switching on/off BSs ($\beta_1 = 0.7$, $\beta_2 = 0.8$).

C. Light Traffic at both Cells: Base Station Sharing

Let $N_1 = 10$, $N_2 = 15$, $\lambda_1 = 5$, $\lambda_2 = 8$, $p_1 = 14$, and $p_2 = 17$. The corresponding blocking rates at two MSPs are 0.0184 and 0.0091. The total reward rate when two MSPs do not cooperate is $(1 - P_b(\lambda_1, N_1))\lambda_1 p_1 + (1 - P_b(\lambda_2, N_2))\lambda_2 p_2 = 203.5$ (S_1). The total reward rate when two MSPs fully cooperate (S_3) is obtained by using the trunk reservation policy by solving (15). It is 205.8 in our case. As we can see, the cooperation gain is marginal due to the fact that both MSPs have light traffic demands.

However, in such a case, it is very likely that traffic from both MSPs can be fulfilled by only one of them while the other is turned off for the sake of saving operational cost. Lets study the case that $E_1 = 80$, $E_2 = 100$, $\beta_1 = 0.7$, $\beta_2 = 0.8$. Figure 10 shows the NEs of the BSs switching on/off game vs. traffic loads in which NEs with crossly-filled patterns are in the area having two NEs (Figure 3). As can be seen, if *both* MSPs have light traffic (crossly-filled patterns), BSs of MSP 2 with higher capacity (and also higher operating cost) are turned on only if the total traffic is greater than a threshold (5 in this example). Otherwise, BSs of MSP 1 with lower operating cost are turned on (crossly- and green-filled patterns). The total and individual profit gains of MSPs by sharing their BSs vs. traffic load are plotted in Figure 9. It shows that the gain from sharing BSs is very significant (about 43%) when MSPs have light traffic. From Figure 8, we observe that both MSPs benefit from applying the proposed sharing framework if their traffic are both heavy (i.e., high arrival rates). The "worst" case of the proposed framework

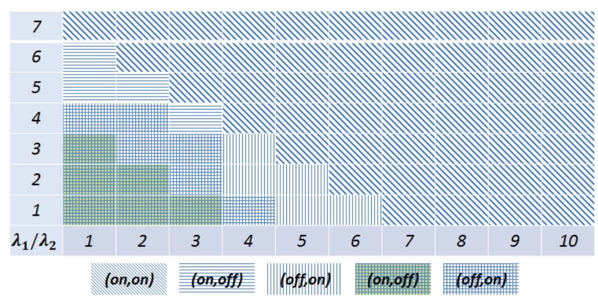


Fig. 10. Pure NEs of BSs switching on/off game vs. λ_1 and λ_2.

is when both the two MSPs have very light traffic (i.e., blue area in Figure 8). However, under such a case, the base station sharing method can be applied. The gain from sharing base station is highest when the traffic at both MSPs is light. We summarize key numerical results and their insights in the table in Figure 11.

VII. Conclusions

We proposed a cooperation framework that allows mobile/cellular service providers (MSPs) to opportunistically offload traffic onto each other while maintaining their own QoS commitments. Using Markov decision processes and a constrained stochastic Markov game, we proved that there exists NEs at which all MSPs improve their performance by strategically sharing their available resources (i.e., carrying traffic from other MSPs as well as offloading traffic to other MSPs).

Traffic condition	Cooperation scheme	Benefit
Heavy	Spectrum sharing (S2)	As much as full cooperation (~60% higher than no cooperation)
Moderate	Spectrum sharing (S2)	As much as full cooperation (e.g., ~6% higher than no cooperation)
Unbalanced	Spectrum sharing (S2)	As much as full cooperation (e.g., ~20% higher than no cooperation)
Light	Base station sharing (S2) via bargaining	As much as full cooperation (e.g., ~40% higher than no cooperation)

Fig. 11. Summary of the gain and resource sharing methods for different traffic scenarios.

The optimal offloading strategy for each MSP was derived by solving constrained Markov decision processes. Numerical results showed that the profit gain is very significant when one MSP is in need of resources while the other experiences light traffic. For the case that both MSPs experience light traffic, the traffic offloading framework was used to guide them on turning on/off their base stations to save operating costs. The theoretical results herein are not only applicable to cellular systems but also to a more general area of competitive and cooperative admission control in queuing systems.

ACKNOWLEDGEMENT

Part of this work has been done while Dr. Diep N. Nguyen was at Macquarie University. Dr. Diep N. Nguyen would like to thank Prof. Stephen Hanly, Dr. Phil Whiting, Dr. Iain Collings for insightful discussion, especially in the conference version of this article [1].

REFERENCES

[1] D. N. Nguyen, I. B. Collings, S. V. Hanly, and P. Whiting, "A cooperation framework for traffic offloading among cellular systems," in *Proceedings of the IEEE International Conference on Communications (ICC)*, 2015, pp. 1637–1642.

[2] A. Osseiran, "Challenges and scenarios of the fifth generation (5G) wireless communications system," Ericsson Research, Tech. Rep., Nov. 2013.

[3] "Cisco visual networking index: Global mobile data traffic forecast update, 2016-2021," Cisco Systems," Cisco White Paper, Feb. 2017.

[4] N. Golmie, "Interference in the 2.4 ghz ism band: Challenges and solutions," Networking for Pervasive Computing: Research from the USA National Institute of Standards and Technology," Tech. Rep., 2013.

[5] R. Berry, M. Honig, T. Nguyen, V. Subramanian, H. Zhou, and R. Vohra, "On the nature of revenue-sharing contracts to incentivize spectrum-sharing," in *Proceedings of the IEEE INFOCOM Conference*, 2013, pp. 845–853.

[6] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2126–2136, May 2013.

[7] B. Leng, P. Mansourifard, and B. Krishnamachari, "Microeconomic analysis of base-station sharing in green cellular networks," in *Proceedings of the IEEE INFOCOM Conference*, April 2014, pp. 1132–1140.

[8] C. Liu and R. A. Berry, "Competition with shared spectrum," in *IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, April 2014, pp. 498–509.

[9] P. Marques, H. Marques, J. Ribeiro, and A. Gameiro, "Coexistence analysis and cognitive opportunities selection in GSM bands," in *Proceedings of the IEEE 69th Vehicular Technology Conference (VTC)*, April 2009, pp. 1–5.

[10] S. Kandeepan, A. Sierra, J. Campos, and I. Chlamtac, "Periodic sensing in cognitive radios for detecting UMTS/HSDPA based on experimental spectral occupancy statistics," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, April 2010, pp. 1–6.

[11] Electronic Communications Committee, "ETSI ts 103 379, Licensed Shared Access (LSA)," April 2017.

[12] Federal Communications Commission, "Order on reconsideration and second report and order," vol. FCC-16-55, April 2016.

[13] P. D. Francesco, F. Malandrino, T. K. Forde, and L. A. DaSilva, "A sharing- and competition-aware framework for cellular network evolution planning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 2, pp. 230–243, June 2015.

[14] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Game-theoretic infrastructure sharing in multioperator cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3326–3341, May 2016.

[15] P. Caballero, A. Banchs, G. de Veciana, and X. Costa-Prez, "Multi-tenant radio access network slicing: Statistical multiplexing of spatial loads," *IEEE/ACM Transactions on Networking*, vol. 25, no. 5, pp. 3044–3058, Oct 2017.

[16] E. Oh and K. Son, "A unified base station switching framework considering both uplink and downlink traffic," *IEEE Wireless Communications Letters*, vol. 6, no. 1, pp. 30–33, Feb 2017.

[17] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. Springer Press, 1997.

[18] E. Altman and A. Shwartz, "Constrained markov games: Nash equilibria," *Advances in Dynamic Games and Applications*, vol. 5, pp. 213–221, 2000.

[19] G. Foschini and B. Gopinath, "Optimum alocation of servers to two types of competing customers," *IEEE Transactions on Communication*, vol. 29, pp. 1051–1055, 1981.

[20] S. Stidham, "Optimal control of admission to a queueing system," *IEEE Transactions on Automated Control*, vol. 30, pp. 705–713, 1985.

[21] B. Han, P. Hui, V. S. A. Kumar, M. V. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Transactions on Mobile Computing*, vol. 11, no. 5, pp. 821–834, 2012.

[22] Z. Lu, X. Sun, and T. L. Porta, "Cooperative data offloading in opportunistic mobile networks," in *Proceedings of the 35th Annual IEEE International Conference on Computer Communications (INFOCOM)*, 2016, pp. 1–9.

[23] A. Hordijk and F. Spieksma, "Constrained admission control to a queueing system," *Advances in Applied Probability*, vol. 21, no. 2, pp. 409–431, 1989.

[24] R. F. Serfozo, "An equivalence between continuous and discrete time markov decision processes," *Operations Research*, vol. 27, no. 3, pp. 616–620, 1979.

[25] G. Scutari, D. Palomar, F. Facchinei, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory," *IEEE Signal Processing Magazine*, May 2010.

[26] V. Nguyen, "On the optimality of trunk reservation in overflow process," *Probability in the Engineering and Informational Sciences*, vol. 5, pp. 369–390, 1991.

[27] A. Hordijk and F. Spieksma, "Constrained admission control to a queueing system," *Advances in Applied Probability*, vol. 21, no. 2, pp. 409–431, 1989.

[28] F. T. J. A. Filar, T. A. Schultz and O. J. Vrieze, "Nonlinear programming and stationary equilibria in stochastic games," *Mathematical Programming*, vol. 50, pp. 227–237, 1991.

[29] J. Nash, "The bargaining problem," *Econometrica*, vol. 18, pp. 155–162, Apr. 1950.

APPENDIX A
PROOF OF PROPOSITION 1

The key idea is that starting from any state, we can go to state (000) to empty the system (due to the service completion). From the empty state, we can go to any other state. Specifically, first, note that the MDP can always move from any state $(ij0)$ to either state $(ij1)$ or $(ij2)$ due to the arrival of customers. Similarly, due to the departures of customers, from any state (ijx) with $i > 1$ or $j > 1$, the process can move to state $((i-1)jx')$ or $(i(j-1)x')$.

Consider MSP 1. If this MSP aims to maximize its average reward, for any state $s \in \{(ij1), (ij2)\}$ in which it still has available resource (i.e., $i < N_1$) while a customer arrives, it should not *always* reject service to *both* types of customers. If not (i.e., $\mathbf{F}_1(ij1, 1) = 0$ *and* $\mathbf{F}_1(ij2, 1) = 0$), from (11), the immediate expected reward $r_k(s, \mathbf{F}_1, \mathbf{F}_2)$ can always be improved by having either $\mathbf{F}_1(ij1,1) > 0$ or $\mathbf{F}_1(ij2,1) > 0$, so is the average reward in (9). Hence, from any state $s \in \{(ij1),(ij2)\}$, the process can move to state $((i+1)jx)$ for $i < N_1$. In a similar manner for MSP 2, starting from any state $s \in \{(ij1),(ij2)\}$, the process can move to state $(i(j+1)x)$ for $j < N_2$.

Thus, the state space $\mathbb{S}$ contains only one communicating class or $P(s'|s,\mathbf{F}_1,\mathbf{F}_2) > 0 \ \forall s,s'$. In other words, the MDP with states in $\mathbb{S}$ is irreducible.

As the underlying Markov process is irreducible for any pair of stationary strategies $(\mathbf{F}_1,\mathbf{F}_2)$ that maximize the average reward, according to Theorem 5.1.5 [17], $V_k^{(a)}(s,\mathbf{F}_1,\mathbf{F}_2)$ in (9) is well-defined and identical for all initial states s. □

APPENDIX B
PROOF OF THEOREM 2

Sufficiency: We assume $\mathbf{z}^* = (\mathbf{v}_1,\mathbf{v}_2,\mathbf{F}_1^*,\mathbf{F}_2^*,\mathbf{u}_1,\mathbf{w}_1,\mathbf{u}_2,\mathbf{w}_2)$ is the globally optimal solution of (14) and its optimal value is 0. We will prove that $(\mathbf{F}_1^*,\mathbf{F}_2^*)$ is the NE of the constrained Markov game (12).

From (9) and (10), we can rewrite $V_k^{(a)}(\mathbf{F}_1,\mathbf{F}_2)$ as:

$$\begin{aligned} V_k^{(a)}(\mathbf{F}_1,\mathbf{F}_2)\mathbf{1} &= \lim_{T\to\infty}\frac{1}{1+T}\sum_{t=0}^{T}\mathbf{P}(\mathbf{F}_1,\mathbf{F}_2)^t\mathbf{r}_k(\mathbf{F}_1,\mathbf{F}_2) \\ &= \mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2)\mathbf{r}_k(\mathbf{F}_1,\mathbf{F}_2) \end{aligned} \tag{16}$$

where $\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2)$ is the Cesaro-limit matrix [17], defined as:

$$\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2) \overset{\text{def}}{=} \lim_{T\to\infty}\frac{1}{1+T}\sum_{t=0}^{T}\mathbf{P}^t(\mathbf{F}_1,\mathbf{F}_2) \tag{17}$$

As the underlying Markov process is irreducible, the above $\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2)$ exists (Theorem 5.1.3 in [17]). Additionally:

$$\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2) = \mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2)\mathbf{P}(\mathbf{F}_1,\mathbf{F}_2) \tag{18}$$

C1 in (14) implies that:

$$\mathbf{v}_1 \geq \mathbf{P}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{v}_1, \forall \mathbf{F}_1. \tag{19}$$

Hence, together with the definition of $\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2)$, we also have:

$$\mathbf{v}_1 \geq \mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{v}_1, \forall \mathbf{F}_1 \tag{20}$$

Since the objective function in (14) is zero under $\mathbf{z}^*$, from the above, we must have: $\mathbf{v}_1 = \mathbf{P}(\mathbf{F}_1^*,\mathbf{F}_2^*)\mathbf{v}_1$. Recall the definition of $\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2)$. We then have

$$\mathbf{v}_1 = \mathbf{Q}(\mathbf{F}_1^*,\mathbf{F}_2^*)\mathbf{v}_1 \tag{21}$$

Multiply both sides of C5 with $\mathbf{Q}(\mathbf{F}_1^*,\mathbf{F}_2^*)$ on the left and recall (18) and (21):

$$V_1^{(a)}(\mathbf{F}_1^*,\mathbf{F}_2^*)\mathbf{1} = \mathbf{v}_1 \tag{22}$$

On the other hand, C2 in (14) implies that

$$\mathbf{v}_1 + \mathbf{u}_1 \geq \mathbf{r}_1(\mathbf{F}_1,\mathbf{F}_2^*) + \mathbf{P}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{u}_1, \forall \mathbf{F}_1. \tag{23}$$

Multiply both sides of (23) with $\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2^*)$ on the left and recall (18):

$$\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{v}_1+\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{u}_1\geq\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{r}_1(\mathbf{F}_1,\mathbf{F}_2^*)+\mathbf{Q}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{u}_1 \tag{24a}$$

$$\mathbf{v}_1 \geq V_1^{(a)}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{1} \tag{24b}$$

where the last inequality follows by appealing (20).

From (22) and (24b): $V_1^{(a)}(\mathbf{F}_1^*,\mathbf{F}_2^*)\mathbf{1} \geq V_1^{(a)}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{1}$. In a similar way, from C3, C4, and C5, we can also show that: $V_2^{(a)}(\mathbf{F}_1^*,\mathbf{F}_2^*)\mathbf{1} \geq V_2^{(a)}(\mathbf{F}_1^*,\mathbf{F}_2)\mathbf{1}$. In other words, $(\mathbf{F}_1^*,\mathbf{F}_2^*)$ is the NE of the constrained Markov game (12).

Necessity: We need to prove that if $(\mathbf{F}_1^*,\mathbf{F}_2^*)$ is the NE of the constrained Markov game (12), then there exists $(\mathbf{v}_1,\mathbf{v}_2,\mathbf{u}_1,\mathbf{w}_1,\mathbf{u}_2,\mathbf{w}_2)$ so that we can construct $\mathbf{z}^* = (\mathbf{v}_1,\mathbf{v}_2,\mathbf{F}_1^*,\mathbf{F}_2^*,\mathbf{u}_1,\mathbf{w}_1,\mathbf{u}_2,\mathbf{w}_2)$ to be the globally optimal solution of (14) and its optimal value is 0.

For that purpose, we need to construct a feasible solution $\mathbf{z}^*$ (i.e., all constraints in (14) hold) and show that 0 is the optimal value of (14) which can be then attained by $\mathbf{z}^*$. First set $\mathbf{v}_k = V_k^{(a)}(\mathbf{F}_1^*,\mathbf{F}_2^*)\mathbf{1}$. Note that for a given stationary strategy from an MSP, e.g., $\mathbf{F}_2^*$, MSP 1 finds its optimal stationary strategy $\mathbf{F}_1^*$ by solving for the optimal stationary policy of an MDP with transition probability $\mathbf{P}(\mathbf{F}_1,\mathbf{F}_2^*)$. Applying Proposition 2.8.4 and 2.8.5 in [17] to the MDP with transition probability $\mathbf{P}(\mathbf{F}_1,\mathbf{F}_2^*)$, we have:

$$V_1^{(a)}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{1} \geq \mathbf{P}(\mathbf{F}_1,\mathbf{F}_2^*)V_1^{(a)}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{1}, \forall \mathbf{F}_1 \tag{25}$$

and there exists $\mathbf{u}_1$ such that:

$$V_1^{(a)}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{1}+\mathbf{u}_1\mathbf{1} \geq \mathbf{r}_1(\mathbf{F}_1,\mathbf{F}_2^*)+\mathbf{P}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{u}_1, \forall \mathbf{F}_1 \tag{26}$$

As $(\mathbf{F}_1^*,\mathbf{F}_2^*)$ is the NE and recalling that $\mathbf{v}_k = V_k^{(a)}(\mathbf{F}_1^*,\mathbf{F}_2^*)\mathbf{1}$, the inequalities below must hold:

$$\begin{aligned} \mathbf{v}_1 &\geq \mathbf{P}(\mathbf{F}_1,\mathbf{F}_2^*)V_1^{(a)}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{1}, \forall \mathbf{F}_1 \\ \mathbf{v}_1 + \mathbf{u}_1 &\geq \mathbf{r}_1(\mathbf{F}_1,\mathbf{F}_2^*) + \mathbf{P}(\mathbf{F}_1,\mathbf{F}_2^*)\mathbf{u}_1, \forall \mathbf{F}_1 \end{aligned} \tag{27}$$

Since the above inequalities hold $\forall \mathbf{F}_1$, constraints C1 and C2 must hold. Similarly, we can also show that there exists $\mathbf{u}_2$ such that C3 and C4 also hold.

According to Theorem 5.1.3 in [17], for the MDP with transition probability $\mathbf{P}(\mathbf{F}_1^*,\mathbf{F}_2^*)$ (at the NE), constraint C5 holds for both MSPs by setting:

$$\mathbf{w}_k = (\mathbf{I}-\mathbf{P}(\mathbf{F}_1^*,\mathbf{F}_2^*)+\mathbf{Q}(\mathbf{F}_1^*,\mathbf{F}_2^*))^{-1}(\mathbf{r}_k(\mathbf{F}_1^*,\mathbf{F}_2^*)-\mathbf{v}_k). \tag{28}$$

The strategy pair $(\mathbf{F}_1^*,\mathbf{F}_2^*)$ is the NE of the constrained Markov game (12). Thus, $(\mathbf{F}_1^*,\mathbf{F}_2^*)$ has to be within the strategy space of both MSPs, defined by constraints C1, C2, and C3 in (12). Hence, constraints C6, C7, and C8 of (14) must hold. We have just constructed a feasible solution $\mathbf{z}^*$ of (14).

Note that for any feasible solutions of (14) constraints C1 and C3 imply that $\mathbf{v}_k \geq \mathbf{P}(\mathbf{F}_1,\mathbf{F}_2)\mathbf{v}_k$. In other words, the objective function of (14) is lower-bounded by 0. By recalling (16) and (18), the NE stationary strategy pair $(\mathbf{F}_1^*,\mathbf{F}_2^*)$ from $\mathbf{z}^*$ can attain this lower bound. The proof is completed. □

Diep N. Nguyen is a faculty member of the Faculty of Engineering and Information Technology, University of Technology Sydney (UTS). He received M.E. and Ph.D. in Electrical and Computer Engineering from the University of California San Diego (UCSD) and The University of Arizona (UA), respectively. Before joining UTS, he was a DECRA Research Fellow at Macquarie University, a member of technical staff at Broadcom (California), ARCON Corporation (Boston), consulting the Federal Administration of Aviation on turning detection of UAVs and aircraft, US Air Force Research Lab on anti-jamming. He has received several awards from LG Electronics, University of California, San Diego, The University of Arizona, US National Science Foundation, Australian Research Council. His recent research interests are in the areas of computer networking, wireless communications, and machine learning application, with emphasis on systems' performance and security/privacy.

Eryk Dutkiewicz received his B.E. degree in Electrical and Electronic Engineering from the University of Adelaide in 1988, his M.Sc. degree in Applied Mathematics from the University of Adelaide in 1992 and his PhD in Telecommunications from the University of Wollongong in 1996. His industry experience includes management of the Wireless Research Laboratory at Motorola in early 2000s. He has co-authored 17 US patents filings and published over 280 journal and conference papers. He is currently the Head of School of Electrical and Data Engineering at the University of Technology Sydney, Australia. He has held visiting professorial appointments at several institutions including the Chinese Academy of Sciences, Shanghai JiaoTong University and Macquarie University. His current research interests cover 5G networks and medical body area networks.

Marwan Krunz is the Kenneth VonBehren Endowed Professor in the ECE Department at the University of Arizona. He is also an affiliated faculty member of the University of Technology Sydney (UTS). He co-directs the Broadband Wireless Access and Applications Center, a multi-university industry-focused NSF center that includes affiliates from industry and government labs. He previously served as the UA site director for Connection One, an NSF IUCRC that focuses on wireless communication circuits and systems. In 2010, Dr. Krunz was a Visiting Chair of Excellence at the University of Carlos III de Madrid. He previously held visiting research positions at UTS, INRIA-Sophia Antipolis, HP Labs, University of Paris VI, University of Paris V, University of Jordan, and US West Advanced Technologies. Dr. Krunzs research interests lie in the areas of wireless communications and networking, with emphasis on resource management, adaptive protocols, and security issues. He has published more than 250 journal articles and peer-reviewed conference papers, and is a co-inventor on several US patents. He is an IEEE Fellow, an Arizona Engineering Faculty Fellow (2011-2014), and an IEEE Communications Society Distinguished Lecturer (2013 and 2014). He was the recipient of the 2012 IEEE TCCC Outstanding Service Award. He received the NSF CAREER award in 1998. He currently serves as the Editor-in-Chief for the IEEE Transactions on Mobile Computing (TMC). He previously served on the editorial boards for the IEEE Transactions on Cognitive Communications and Networks, IEEE/ACM Transactions on Networking, IEEE TMC, IEEE Transactions on Network and Service Management, Computer Communications Journal, and IEEE Communications Interactive Magazine. He was the general vice-chair for WiOpt 2016 and general co-chair for WiSec12. He was the TPC chair for WCNC 2016 (Networking Track), INFOCOM04, SECON05, WoWMoM06, and Hot Interconnects 9. He has served and continues to serve on the steering and advisory committees of numerous conferences and on the panels of several funding agencies. He was a keynote speaker, an invited panelist, and a tutorial presenter at numerous international conferences. See http://www2.engr.arizona.edu/ krunz/ for more details.