Wireless Access Network Selection Game with Negative Network Externality

Yu-Han Yang, *Student Member, IEEE,* Yan Chen, *Member, IEEE,* Chunxiao Jiang, *Member, IEEE,* Chih-Yu Wang, *Member, IEEE,* and K. J. Ray Liu, *Fellow, IEEE*

***Abstract*—Network service acquisition in a wireless environment requires the selection of a wireless access network. A key problem in wireless access network selection is to study the rational strategy considering the negative network externality, i.e, the influence of subsequent users' decisions on an individual's throughput due to the limited available resources. In this work, we formulate the wireless network selection problem as a stochastic game with negative network externality and show that finding the optimal decision rule can be modelled as a multi-dimensional Markov Decision Process (MDP). A modified value iteration algorithm is proposed to efficiently obtain the optimal decision rule with a simple threshold structure, which enables us to reduce the storage space of the strategy profile. We further investigate the mechanism design problem with incentive compatibility constraints, which enforce the networks to reveal the truthful state information. The formulated problem is a mixed integer programming problem which in general lacks an efficient solution. Exploiting the optimality of substructures, we propose a dynamic programming algorithm that can optimally solve the problem in the two-network scenario. For the multi-network scenario, the proposed algorithm can outperform the heuristic greedy approach in a polynomial-time complexity. Finally, simulation results are shown to validate the analysis and demonstrate the effectiveness of the proposed algorithms.**

***Index Terms*—Game theory, stochastic game, negative network externality, Markov decision process, network selection, mechanism design, dynamic programming.**

I. Introduction

NOWADAYS, wireless network services such as Femtocells [1] and Wi-Fi access points are widely deployed to provide Internet access in areas such as homes, offices, airports, hotels, etc. While there may be multiple available wireless networks, a user can only choose one to join. Figure 1 shows an example of the Wi-Fi network selection from a smart phone. Since the networks can be owned by different operators, the network selection problem, which used to be resolved in a centralized manner by admission control [2], [3], should be investigated in a distributed perspective by considering users' own interests. In the wireless access network selection problem, a myopic strategy can usually be adopted by choosing the one with the strongest signal. A consequence of this strategy is the congestion of users to communicate with certain network controllers such as access points (APs), switches, or routers. The concentration of users creates an unbalanced load in the network, which leads to an inefficient resource utilization for service providers and a poor quality-of-service (QoS) for users.

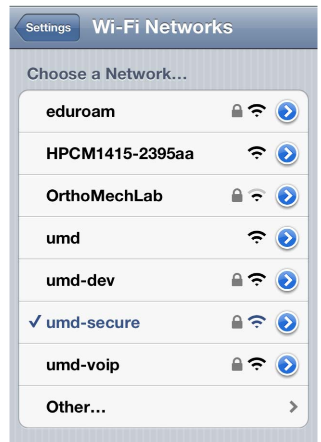


Fig. 1. Wi-Fi network selection.

Efficient resource utilization is an important issue in modern wireless access networks due to limited available resources such as signal power, temporal and spatial bandwidth. On one hand, the service provider attempts to maximize resource utilization such that the available resources can accommodate as many users as possible. On the other hand, due to the individual rationality and the selfish nature, a user aims to optimize his/her own utility. Therefore, a user's optimal strategy in such a resource-sharing scenario inevitably has to take into consideration the *negative network externality* [4], [5], i.e., the influence of other users' strategies on the user's own utility. Commonly referred in economics and business, the negative network externality is the effect that occurs when more users make the available resource less valuable. For example, the traffic congestion overloads the highway. Overwhelming customers degrade the quality-of-service in a restaurant. The negative network externality in these examples impairs the utilities of the users making the same decision.

In this paper, we firstly focus on how a user should choose

Manuscript received December 22, 2012; revised April 9 and June 12, 2013; accepted July 23, 2013. The associate editor coordinating the review of this paper and approving it for publication was P. Wang.

Y.-H. Yang, Y. Chen, and K.J.R. Liu are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, 20742, USA (e-mail: {yhyang, yan, kjrliu}@umd.edu).

C. Jiang is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA, and also with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, P. R. China (e-mail: chx.jiang@gmail.com).

C.-Y. Wang is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA, and also with GICE, National Taiwan University, Taipei, 106, Taiwan (R.O.C.) (e-mail: tomkywang@gmail.com).

Digital Object Identifier 10.1109/TWC.2013.090513.122033

1536-1276/13$31.00 © 2013 IEEE

one of the available wireless access networks considering the negative network externality. Wireless access network selection is an essential problem of resource utilization and has received great attention recently [6]–[17]. In [13], centralized approaches are investigated to provide congestion relief by explicit channel switching and network-directed roaming. A distributed access point selection algorithm based on no regret learning is proposed in [14]. The authors show that the algorithm can guarantee convergence to an equilibrium. The arrival and departure of the users in network selection problems are also considered in [16] and [17]. Another class of network selection approaches is based on game theory. Game theory has been recognized as an ideal tool to study the interactions among users [18], [19]. It has been widely used in wireless communications and networking for many different problems [19]–[23] including power control [20], cooperation stimulation [23], and security enforcement [24]. In [7], Mittal et al. consider users changing locations as strategies to obtain more resources and analyze the corresponding Nash equilibria (NE). In [12], the network selection is modelled as a congestion game, where players make decisions simultaneously to optimize the interference and throughput. Also, the congestion in the network selection game is similar to that in the channel selection game, e.g., [25]–[27]. In [25], an atomic congestion game in which resources are allowed to be reused among non-interfering users is considered. In [26] and [27], the authors investigated game theoretic solutions to the distributed channel selection problem in opportunistic spectrum access systems. A comprehensive review and comparison of existing decision-theoretic solutions including Markov decision process, game theory and stochastic control can be found in [28].

However, most of the existing works study the network selection problem under the scenario where users make decisions simultaneously. In this paper, we consider the problem under a different scenario where users make decisions *sequentially* and their optimal decisions involve the prediction of subsequent users' decisions due to the negative network externality. Sequential decisions considering the negative network externality effect are studied in the Chinese restaurant game [29]–[31], in which the equilibrium of grouping under the scenario of a fixed total number of players is characterized. In this work, we formulate the wireless access network selection problem as a stochastic game with negative network externality, where users arrive at and depart from networks in a probabilistic manner. The problem of finding the optimal decision rule is shown to be a multi-dimensional Markov Decision Process (MDP). Different from the conventional MDP [32], the multi-dimensional MDP has multiple potential functions and thus the dynamic programming (DP) [33] cannot be directly applied. We propose a modified value iteration algorithm to find the equilibrium for the multi-dimensional MDP. The analysis of the proposed algorithm shows that the strategy profile generated by the algorithm has a threshold structure, which enables us to save the storage space of the strategy profile from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$, where N^2 is the number of system states in the two-network scenario. Simulation results verify the analysis and demonstrate the efficiency and effectiveness of the proposed algorithm, i.e., while achieving the optimal strategy for the individual, the proposed algorithm attains similar performance of social welfare compared to the centralized method that maximizes the social welfare.

The second focus of this paper is the truthful mechanism design [34]–[38] for the network selection game. Mechanism design is to devise pricing and allocation rules satisfying the incentive compatibility [35], [36]. In the network selection game, users makes decisions relying on the system states which consist of the information provided by the networks, possibly owned by different operators with different interests. Therefore, the reported state may be untruthful if it is profitable to make a deceitful claim. In this work, we investigate the mechanism design problem with incentive compatibility constraints, which enforce the networks to report truthfully, while optimizing the utility of users. The formulated problem is a mixed integer programming problem which in general lacks an efficient solution. Exploiting the optimality of substructures, we propose a dynamic programming algorithm that can efficiently and optimally solve the problem in the two-network scenario. For the multi-network scenario, the proposed algorithm can outperform the heuristic greedy approach in a polynomial-time complexity. Finally, simulation results are shown to validate the analysis and demonstrate the effectiveness of the proposed algorithms.

The novelty and technical contribution of this work are summarized as follows. We formulate the distributed wireless access network selection problem as a multi-dimensional MDP, which, to the best of our knowledge, is new and has not been studied before. We propose a modified value iteration algorithm to search for an equilibrium. We also analyze the proposed algorithm and show that the resulting strategy profile has a threshold structure. We further propose an efficient dynamic programming algorithm to design a truthful mechanism which enforces the networks to truthfully reveal the state information.

The rest of the paper is organized as follows. The system model and the formulation of the wireless access network selection game is described in Section II. In Section III, we propose a modified value iteration algorithm for the multi-dimensional MDP. The threshold structure of the strategy profile generated by the proposed algorithm is analyzed in Section IV. In Section V, we describe the mechanism design problem for the network selection game and propose the dynamic programming algorithm. In Section VI, the performance of the proposed algorithms is evaluated using numerical simulation. Finally, Section VIII concludes the paper.

II. System Model and Problem Formulation

In this section, we describe in detail the system model and the problem formulation of the wireless access network selection problem. To better illustrate the idea, we first introduce some necessary notations including the probabilistic model and then characterize the (approximate) equilibrium. Note that as will be seen, the model is quite general and hence its application is not restricted to the network selection problem but can also be deployed in other problems with negative network externality.

A. System Model

The system under consideration comprises K wireless access networks and each network has a capacity of N users, i.e., a network can simultaneously serve at most N users. For the sake of notational conciseness, we consider that all the networks have the same capacity. The analysis can be easily extended to the system with networks of different capacity. We also assume that the networks have no buffer room for users, which means when a network is full, users cannot make request of connection to the network. Each user in network k obtains a utility $R_k(s_k)$ per unit time, where s_k is the current number of users in network k. The utility function is defined as the individual throughput, i.e., $R_k(s_k) = \log(1 + \frac{P_S/N_0}{(s_k-1)P_I/N_0+1}), \forall k$, which represents the achievable data rate under inter-user interference, where P_S/N_0 denotes the signal-to-noise power ratio, and P_I/N_0 is the interference-to-noise power ratio. The utility represents the quality-of-service (QoS) guaranteed by the network but restricted to the available resource such as the total transmission power and the bandwidth of radio frequency. The negative network externality is manifested in the decrease of the data rate as the number of users in the network increases due to a higher inter-user interference. Note that the utilities of users in the same network are assumed the same at each time slot since the network can provide the same QoS to each user by means of resource allocation, even though the instantaneous channel conditions of different users may be different. For example, centralized downlink power control algorithms [39], [40] can be applied by the network to attain a common signal to interference-plus-noise ratio (SINR) or to maximize the minimum SINR among the users.

The users with Poisson distributed arrival rate $\bar{\lambda}_0$ (users per second) have choices of connecting to one of the K networks. After a user makes his decision, he/she cannot switch to any of other networks and has to stay during a period of time with exponential distribution of parameter $\bar{\mu}$, which is assumed the same for all networks for simplicity. The users with arrival rate $\bar{\lambda}_k$ can only choose network k, for $k = 1, \ldots, K$. These users can be envisioned as either the users with certain deterministic behavior, or the users who can only have access to one specific network due to the geographical distribution. Note that incorporating this type of users only makes the system model more general since we can simply set these rates as zero if there are no such users.[1]

The system state $\mathbf{s} = (s_1, \ldots, s_K)$ takes its value from the state space $\mathcal{S} = \{(s_1, \ldots, s_K)|s_k = 0, 1, ..., N, k = 1, \ldots, K\}$, and represents the state that s_k users are in network k, for $k = 1, \ldots, K$. We consider a discrete time Markov system where a time slot has duration T (seconds). Then the arrival and departure probabilities $\lambda_k = \bar{\lambda}_k T e^{-\bar{\lambda}_k T}$ and $\mu = \bar{\mu} T e^{-\bar{\mu} T}$ can be approximated as $\lambda_k \approx \bar{\lambda}_k T, k = 0, \ldots, K$ and $\mu \approx \bar{\mu} T$ when T is sufficiently small [41]–[43]. Let $\mathcal{F}(\mathbf{s}) = \{k|s_k = N, k = 1, \ldots, K\}$ be the index set of the full networks which are serving the maximum number of users and thus cannot accept any more. The complement set of $\mathcal{F}(\mathbf{s})$ is denoted by $\bar{\mathcal{F}}(\mathbf{s}) = \{k|s_k < N, k = 1, \ldots, K\}$, i.e., the index set of the non-full networks. The strategy space of network selection is restricted in $\bar{\mathcal{F}}(\mathbf{s})$ when $\mathbf{s}$ is a boundary state, i.e., when $\sigma_{\mathbf{s}} \in \bar{\mathcal{F}}(\mathbf{s})$. We assume that the connection request from users arriving at the full networks will be rejected and the traffic then goes to other non-full networks. To model such a traffic transition, we therefore assume that the traffic immediately flows to the non-full network. For the two-network case, at most only one non-full network has room for those users, so the traffic goes to that non-full network. For the multi-network case, multiple non-full networks can accommodate those users. In order to provide a well-defined Markov system and to simplify the notation, we assume that the traffic goes to a specific network, i.e., $\min \bar{\mathcal{F}}(\mathbf{s})$, the network with the minimum index. Notice that if $\bar{\mathcal{F}}(\mathbf{s}) = \phi$, i.e., all networks are full, no connection request can be accepted. The network selection strategy when the user observes state $\mathbf{s}$ is denoted as $\sigma_{\mathbf{s}}$, which takes value in $\bar{\mathcal{F}}(\mathbf{s})$. We define $\sigma_{\mathbf{s}} = j$ if network j is chosen. The indicator function $I_k(\sigma_{\mathbf{s}})$ is then defined as: if $\sigma_{\mathbf{s}} = j$, $I_j(\sigma_{\mathbf{s}}) = 1$; otherwise $I_j(\sigma_{\mathbf{s}}) = 0$. We have the state transition probability of an arrival event as

$$P_{\text{sys}}(\mathbf{s} + \mathbf{e}_j|\mathbf{s}) = \begin{cases} \sum_{i \in \mathcal{F}(\mathbf{s})} \lambda_i + \lambda_j + I_j(\sigma_{\mathbf{s}})\lambda_0, & \text{if } j = \min \bar{\mathcal{F}}(\mathbf{s}), \\ \lambda_j + I_j(\sigma_{\mathbf{s}})\lambda_0, & \text{if } j \in \bar{\mathcal{F}}(\mathbf{s}) \backslash \{\min \bar{\mathcal{F}}(\mathbf{s})\}, \end{cases} \quad (1)$$

where $\mathbf{s}$ and $\mathbf{s} + \mathbf{e}_j$ denote the system states at the current time slot and the next time slot, and $\mathbf{e}_j$ is a standard basis vector whose j-th coordinate is 1 and other coordinates are 0. At system state $\mathbf{s}$, since the number of users in network j is s_j, the transition probability of a departure event is given by

$$P_{\text{sys}}(\mathbf{s} - \mathbf{e}_j|\mathbf{s}) = s_j\mu, \ j = 1, \ldots, K. \quad (2)$$

Furthermore, the probability that the system state remains the same is

$$P_{\text{sys}}(\mathbf{s}|\mathbf{s}) = \begin{cases} 1 - \sum_{j=0}^{K} \lambda_j - \sum_{j=1}^{K} s_j\mu, & \text{if } \bar{\mathcal{F}}(\mathbf{s}) \neq \phi, \\ 1 - \sum_{j=1}^{K} s_j\mu, & \text{if } \bar{\mathcal{F}}(\mathbf{s}) = \phi. \end{cases} \quad (3)$$

The duration of a time slot T should be chosen such that $\sum_{j=0}^{K} \lambda_j + KN\mu \leq 1$, i.e., $T \leq 1/(\sum_{j=0}^{K} \bar{\lambda}_j + KN\bar{\mu})$.

For instance, when $K = 2$, $0 \leq s_1 \leq N-1$, and $0 \leq s_2 \leq N-1$, the transition probability is given by

$$P_{\text{sys}}\{\mathbf{s}'|\mathbf{s} = (s_1, s_2)\} = \begin{cases} I_1(\sigma_{\mathbf{s}})\lambda_0 + \lambda_1, & \text{if } \mathbf{s}' = (s_1+1, s_2), \\ I_2(\sigma_{\mathbf{s}})\lambda_0 + \lambda_2, & \text{if } \mathbf{s}' = (s_1, s_2+1), \\ s_1\mu, & \text{if } \mathbf{s}' = (s_1-1, s_2), \\ s_2\mu, & \text{if } \mathbf{s}' = (s_1, s_2-1), \\ 1 - \lambda_0 - \lambda_1 - \lambda_2 - s_1\mu - s_2\mu, & \text{if } \mathbf{s}' = (s_1, s_2), \\ 0, & \text{otherwise}. \end{cases} \quad (4)$$

Similarly the corresponding transition probability for $s_1 = N$, $0 \leq s_2 \leq N-1$ or $0 \leq s_1 \leq N-1$, $s_2 = N$ can also be defined. Figure 2 depicts the state transition diagram

[1] More general types of users, such as users who can only connect to one of a subset of K networks, can be considered. Here for simplicity we only consider two types of users, i.e., users who have choices of connecting to any one of K networks, and users who can only choose one specific network.

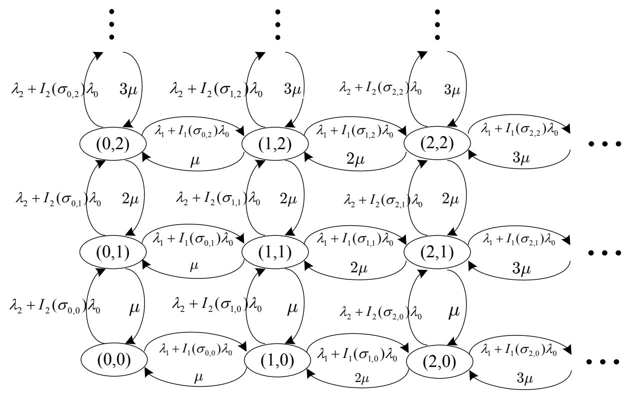


Fig. 2. State diagram of the 2-D Markov chain.

when $K = 2$. The dynamic of the two-network system can be described by a two-dimensional (2-D) Markov chain where the probability $P_{\text{sys}}(\mathbf{s}|\mathbf{s})$ is not shown in Figure 2 for conciseness.

B. Expected utility

The strategy profile $\boldsymbol{\sigma} = \{\sigma_{\mathbf{s}} | \forall \mathbf{s} \in \mathcal{S}\}$ is a mapping from the aggregate state space to the action space, i.e., $\boldsymbol{\sigma} : \{0, 1, ..., N\}^K \mapsto \{1, 2, \ldots, K\}$. Given a strategy profile $\boldsymbol{\sigma}$, we can obtain the system transition probability in (1) - (3). When a rational user arrives and observes system state $\mathbf{s}_0$, he/she makes the decision $\sigma_{\mathbf{s}_0} = \hat{k}$ which leads the user into the system state $\mathbf{s}_1 = \mathbf{s}_0 + \mathbf{e}_{\hat{k}}$. Then, the expected utility of the rational user is given by

$$V_{\hat{k}}(\mathbf{s}_1) = E\left[\sum_{t=1}^{\infty}(1-\mu)^{t-1}R_{\hat{k}}(\mathbf{s}_t)\middle|\, \mathbf{s}_1\right], \tag{5}$$

where $\mathbf{s}_t$ denotes the system state at time t. Since μ is the probability that the service is terminated in one time slot, then $(1-\mu)$ can be interpreted as the probability that the user stays in the network in one time slot. The value $(1-\mu)$ can also be regarded as the discounting factor for the future utility as shown later in (6). The strategy $\sigma_{\mathbf{s}_0} = \hat{k}$ determines which network the user will enter and thus which expected utility function the user will obtain. Denoted by $V_{\hat{k}}(\mathbf{s}_1)$, the expected utility function is the expected value of the discounted sum of the immediate utilities $R_{\hat{k}}(\mathbf{s}_t)$ accumulated from the next time slot. Notice that $\mathbf{s}_1 = \mathbf{s}_0 + \mathbf{e}_{\hat{k}}$ is uniquely determined by the user's strategy $\sigma_{\mathbf{s}_0}$, but the subsequent states $\mathbf{s}_t$, for $t \geq 2$, are stochastic and dependent on the arrival of other users, including users from user-arrival stream k, $1 \leq k \leq K$, and other rational users.

From the Bellman equation [32], the expected utility in (5) can be shown to satisfy the following recursive expression.

$$V_k(\mathbf{s}) = R_k(s_k) + (1-\mu)\sum_{\mathbf{s}'} P_k\left(\mathbf{s}'|\mathbf{s}\right) V_k(\mathbf{s}'), \tag{6}$$

where

$$P_k\left(\mathbf{s}'|\mathbf{s}\right) = \begin{cases} \sum_{i\in\mathcal{F}(\mathbf{s})}\lambda_i + \lambda_j + I_j(\sigma_{\mathbf{s}})\lambda_0, & \text{if } j = \min\bar{\mathcal{F}}(\mathbf{s}), \\ \lambda_j + I_j(\sigma_{\mathbf{s}})\lambda_0, & \text{if } j \in \bar{\mathcal{F}}(\mathbf{s})\backslash\left\{\min\bar{\mathcal{F}}(\mathbf{s})\right\}, \\ s_i\mu, & \text{if } \mathbf{s}' = \mathbf{s} - \mathbf{e}_i, \forall i \neq k, \\ (s_k - 1)\mu, & \text{if } \mathbf{s}' = \mathbf{s} - \mathbf{e}_k, \\ 1 - \sum_{j=0}^{K}\lambda_j - \sum_{j=1}^{K} s_j\mu + \mu, & \text{if } \mathbf{s}' = \mathbf{s}, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

which is the transition probability given that the user still stays in network k. The probability of transition from $\mathbf{s}$ to $\mathbf{s} - \mathbf{e}_k$ is $(s_k - 1)\mu$ since $s_k - 1$ users may leave the network. The transition probability from $\mathbf{s}$ to other states is similar to the definition of P_{sys} in (4).

C. Best Response of Rational Users

Due to the selfish nature, when observing the state $\mathbf{s}$, a rational user will choose the strategy $\sigma_{\mathbf{s}}$ to maximize his expected utility. Thus, the rational strategy $\sigma_{\mathbf{s}}$ has to satisfy

$$\sigma_{\mathbf{s}} = \arg\max_k V_k(\mathbf{s} + \mathbf{e}_k). \tag{8}$$

It can be seen that with the strategy profile in which the strategy of every state satisfies (8), no user can obtain a higher expected utility by unilateral deviation to any other strategy. Therefore, the strategy profile satisfying (6)-(8) is a Nash equilibrium of the stochastic game.

III. Modified Value Iteration Algorithm

The problem of finding the strategy profile satisfying (6)-(8) is not a conventional Markov Decision Process problem. In a conventional MDP problem [32], a single potential function is associated with each system state, and the optimal strategy can be obtained directly by optimizing the potential function. Such a problem can often be solved via the theory of dynamic programming (DP) [33]. However, in our model, multiple potential functions are related in a vector form:

$$\begin{bmatrix} V_1(\mathbf{s}) \\ V_2(\mathbf{s}) \\ \vdots \\ V_K(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} R_1(s_1) \\ R_2(s_2) \\ \vdots \\ R_K(s_K) \end{bmatrix} + (1-\mu)\begin{bmatrix} \mathbf{p}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{p}_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{p}_K \end{bmatrix}^T \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_K \end{bmatrix}, \tag{9}$$

where $\mathbf{0}$ denotes an all-zero vector, $\mathbf{p}_k$ and $\mathbf{v}_k$ are vectors comprising $P_k(\mathbf{s}'|\mathbf{s})$ and $V_k(\mathbf{s}')$ as elements, $k = 1, \ldots, K$. The transpose operator is denoted by $(\cdot)^T$.

The strategy $\sigma_{\mathbf{s}}$ is determined by comparing $V_k(\mathbf{s} + \mathbf{e}_k)$ for all k as in (8). Thus, DP cannot be directly applied in such a problem. It is important to point out that a user makes a decision after he arrives and observes the system state $\mathbf{s}$. The strategy leads the user into some network k and results in an expected utility $V_k(\mathbf{s} + \mathbf{e}_k)$. In subsequent time slots, the user cannot change from the network he/she is staying to any other network. The expected utility is affected by others' strategies through the transition probabilities as given in (6).

We can see that given the expected utilities $\{V_k\}_{k=1}^K$, the rational strategy profile $\boldsymbol{\sigma}$ should satisfy (8). On the other hand, given a strategy profile $\boldsymbol{\sigma}$, the expected utilities $\{V_k\}_{k=1}^K$

can be found by (6), where the transition probability $P_k(\mathbf{s}'|\mathbf{s})$ is a function of the strategy $\sigma_\mathbf{s}$. To obtain the optimal strategy profile $\boldsymbol{\sigma}^*$ satisfying (6)-(8), we propose a modified value iteration algorithm to iteratively solve the problem. At the n-th iteration, the rational strategy profile is given by

$$\sigma_\mathbf{s}^{(n+1)} = \arg\max_k V_k^{(n)}(\mathbf{s}+\mathbf{e}_k), \forall \mathbf{s} \in \mathcal{S}. \tag{10}$$

The expected utility functions can be obtained by solving

$$V_k^{(n+1)}(\mathbf{s}) = R_k(s_k) + (1-\mu) \sum_{\mathbf{s}' \in \mathcal{S}} P_k^{(n+1)}(\mathbf{s}'|\mathbf{s})\, V_k^{(n+1)}(\mathbf{s}'), \quad \forall \mathbf{s} \in \mathcal{S}, \forall k \in \{1,\ldots,K\}, \tag{11}$$

where the transition probability $P_k^{(n+1)}(\mathbf{s}'|\mathbf{s})$ is updated using the corresponding updated strategies, i.e.,

$$P_k^{(n+1)}(\mathbf{s}'|\mathbf{s}) = \begin{cases} \sum_{i \in \mathcal{F}(\mathbf{s})} \lambda_i + \lambda_j + I_j(\sigma_\mathbf{s}^{(n+1)})\lambda_0, & \text{if } \mathbf{s}' = \mathbf{s}+\mathbf{e}_j, j = \min \bar{\mathcal{F}}(\mathbf{s}) \\ \lambda_j + I_j(\sigma_\mathbf{s}^{(n+1)})\lambda_0, & \text{if } \mathbf{s}' = \mathbf{s}+\mathbf{e}_j, j \in \bar{\mathcal{F}}(\mathbf{s}) \setminus \{\min \bar{\mathcal{F}}(\mathbf{s})\} \\ s_j\mu, & \text{if } \mathbf{s}' = \mathbf{s}-\mathbf{e}_j, j \neq k \\ (s_k-1)\mu, & \text{if } \mathbf{s}' = \mathbf{s}-\mathbf{e}_k \\ 1 - \sum_{j \in \bar{\mathcal{F}}_\mathbf{s}} P_k(\mathbf{s}+\mathbf{e}_j|\mathbf{s}) - \sum_{j=1}^K P_k(\mathbf{s}-\mathbf{e}_j|\mathbf{s}), & \mathbf{s}' = \mathbf{s}, \\ 0, & \text{otherwise}, \end{cases} \tag{12}$$

The solution to (11) can be obtained through several approaches, one of which is the value iteration algorithm [32]. The algorithm first initializes $V_k^{(n+1)}(\mathbf{s})$ as an arbitrary value such as zero and iteratively updates it using (11). The iteration function is a contraction mapping so the convergence to a unique fixed point is guaranteed. Another approach is to consider (11) as K sets of linear systems, where each set has N^2 unknown variables corresponding to $\{V_k^{(n+1)}(\mathbf{s}), \forall \mathbf{s}\}$ and N^2 equations. Such linear systems can be solved by linear programming or matrix inversion.

In the next section, we will theoretically show that for $K=2$, the proposed algorithm results in a threshold structure of the strategy profile at each iteration, and such a threshold structure is also observed for general $K > 2$. However, the strategy profile may not converge but oscillates near the threshold due to the hard decision rule in (8). The non-convergence occurs when the rational strategy of the state near the threshold oscillates between different choices each time when the expected utility is updated. When such a situation happens, the expected utilities corresponding to different strategies are very close to each other. Hence, to solve this problem, we relax the hard decision rule by allowing a small region of tolerance for switching among the strategies [44], which leads to the soft decision rule as follows.

$$\sigma_\mathbf{s}^{(n+1)} = \begin{cases} \sigma_\mathbf{s}^{(n)}, \\ \quad \text{if } V_{\sigma_\mathbf{s}^{(n)}}^{(n)}(\mathbf{s}+\mathbf{e}_{\sigma_\mathbf{s}^{(n)}}) \geq \max_k V_k^{(n)}(\mathbf{s}+\mathbf{e}_k) - \epsilon, \\ \arg\max_k V_k^{(n)}(\mathbf{s}+\mathbf{e}_k), \\ \quad \text{if } V_{\sigma_\mathbf{s}^{(n)}}^{(n)}(\mathbf{s}+\mathbf{e}_{\sigma_\mathbf{s}^{(n)}}) < \max_k V_k^{(n)}(\mathbf{s}+\mathbf{e}_k) - \epsilon, \end{cases} \tag{13}$$

where $\epsilon > 0$ is a small constant. Table I summarizes the proposed modified value iteration algorithm for the multi-dimensional MDP. Notice that the algorithm stops when an equilibrium is found or all the strategy profiles are searched. By definition, when the algorithm obtains a solution, the resulting strategy profile is an ϵ-approximate NE [18], in which the strategy at each state has an expected utility that is at most ϵ less than that of any other strategy. Note that there may be multiple ϵ-approximate NEs especially for a larger ϵ when a larger region of tolerance is allowed for switching among the strategies.

TABLE I
MODIFIED VALUE ITERATION ALGORITHM

(i) Initialize: $V_k^{(0)}(\mathbf{s}) = 0$, $\forall k \in \{1,\ldots,K\}$, $\forall \mathbf{s} \in \mathcal{S}$. $T = \phi$.
(ii) **Loop** :
1. Update $\{\sigma_\mathbf{s}^{(n+1)}\}$ by (13).
If $\{\boldsymbol{\sigma}_\mathbf{s}^{(n+1)}\} = \{\boldsymbol{\sigma}_\mathbf{s}^{(n)}\}$, **then** stop loop.
else if $\{\boldsymbol{\sigma}_\mathbf{s}^{(n+1)}\} \in T$, **then**
choose a $\{\boldsymbol{\sigma}_\mathbf{s}\} \in \bar{T}$, and let $\{\boldsymbol{\sigma}_\mathbf{s}^{(n+1)}\} = \{\boldsymbol{\sigma}_\mathbf{s}\}$.
end if
$T = T \cup \{\boldsymbol{\sigma}_\mathbf{s}^{(n+1)}\}$.
2. Update $\{P_k^{(n+1)}(\mathbf{s}'|\mathbf{s})\}$ by (12).
3. Solve $\{V_k^{(n+1)}(\mathbf{s})\}$ in (11) by value iteration or linear programming.
Until $\bar{T} = \phi$ or $\{\boldsymbol{\sigma}_\mathbf{s}^{(n+1)}\} = \{\boldsymbol{\sigma}_\mathbf{s}^{(n)}\}$.

IV. THRESHOLD STRUCTURE OF STRATEGY PROFILE

In this section, we show that the strategy profile produced by the proposed modified value iteration algorithm in each iteration exhibits a threshold structure for two-network systems. With the assumption that $R_k(s_k)$, $k=1,2$, are non-increasing, the following lemma shows that $V_1(\mathbf{s})$ is non-decreasing and $V_2(\mathbf{s})$ is non-increasing along the line of $s_1+s_2=m$, $\forall m \in \{1,2,...,2N\}$.

Lemma 1: For $n \geq 0$*,*

$$V_1^{(n)}(\mathbf{s}) \geq V_1^{(n)}(\mathbf{s}+\mathbf{e}_1-\mathbf{e}_2), \tag{14}$$

$$V_2^{(n)}(\mathbf{s}) \leq V_2^{(n)}(\mathbf{s}+\mathbf{e}_1-\mathbf{e}_2). \tag{15}$$

Proof: We use induction to show that (14) and (15) hold for all $n \geq 0$.

i) Since $V_1^{(0)}(\mathbf{s})$ and $V_2^{(0)}(\mathbf{s})$ are initialized as zeros, (14) and (15) hold for $n=0$.

ii) We assume the induction hypothesis holds for some $n \geq 0$. Then it can be shown that (14) and (15) also hold for $(n+1)$ by analyzing the following difference. Let $\mathbf{s}' = \mathbf{s}+\mathbf{e}_1-\mathbf{e}_2$. For $0 \leq s_1 \leq N-2$ and $1 \leq s_2 \leq N-1$,

$$\begin{aligned} V_1^{(n+1)}(\mathbf{s}) - V_1^{(n+1)}(\mathbf{s}') &= R_1(s_1) - R_1(s_1+1) \\ &+ (1-\mu)\Big[\lambda_1\left(V_1^{(n)}(\mathbf{s}+\mathbf{e}_1) - V_1^{(n)}(\mathbf{s}'+\mathbf{e}_1)\right) \\ &+ \lambda_0\left(I_1(\sigma_\mathbf{s})V_1^{(n)}(\mathbf{s}+\mathbf{e}_1) - I_1(\sigma_{\mathbf{s}'})V_1^{(n)}(\mathbf{s}'+\mathbf{e}_1)\right) \\ &+ \lambda_2\left(V_1^{(n)}(\mathbf{s}+\mathbf{e}_2) - V_1^{(n)}(\mathbf{s}'+\mathbf{e}_2)\right) \\ &+ \lambda_0\left(I_2(\sigma_\mathbf{s})V_1^{(n)}(\mathbf{s}+\mathbf{e}_2) - I_2(\sigma_{\mathbf{s}'})V_1^{(n)}(\mathbf{s}'+\mathbf{e}_2)\right) \\ &+ (s_1-1)\mu V_1^{(n)}(\mathbf{s}-\mathbf{e}_1) - s_1\mu V_1^{(n)}(\mathbf{s}'-\mathbf{e}_1) \\ &+ s_2\mu V_1^{(n)}(\mathbf{s}-\mathbf{e}_2) - (s_2-1)\mu V_1^{(n)}(\mathbf{s}'-\mathbf{e}_2) \\ &+ (1-\lambda_0-\lambda_1-\lambda_2-s_1\mu-s_2\mu)\left(V_1^{(n)}(\mathbf{s}) - V_1^{(n)}(\mathbf{s}')\right)\Big]. \end{aligned} \tag{16}$$

Due to the fact that the utility function $R_1(s_1)$ is non-increasing in s_1 and the induction hypothesis which guarantees the non-negativeness of many differences of terms in (16), by rearranging a few terms, it suffices to discuss the following cases.

Case 1: $\sigma_{\mathbf{s}}^{(n)} = \sigma_{\mathbf{s}'}^{(n)} = 1$. Then, $V_1^{(n)}(\mathbf{s}+\mathbf{e}_1) - V_1^{(n)}(\mathbf{s}'+\mathbf{e}_1) \geq 0$ by the induction hypothesis.

Case 2: $\sigma_{\mathbf{s}}^{(n)} = \sigma_{\mathbf{s}'}^{(n)} = 2$. Then, $V_1^{(n)}(\mathbf{s}+\mathbf{e}_2) - V_1^{(n)}(\mathbf{s}'+\mathbf{e}_2) \geq 0$ by the induction hypothesis.

Case 3: $\sigma_{\mathbf{s}}^{(n)} = 1$ and $\sigma_{\mathbf{s}'}^{(n)} = 2$. Then, $V_1^{(n)}(\mathbf{s}+\mathbf{e}_1) - V_1^{(n)}(\mathbf{s}'+\mathbf{e}_2) = 0$.

Case 4: $\sigma_{\mathbf{s}}^{(n)} = 2$ and $\sigma_{\mathbf{s}'}^{(n)} = 1$. Then, $V_1^{(n)}(\mathbf{s}+\mathbf{e}_2) - V_1^{(n)}(\mathbf{s}'+\mathbf{e}_1) \geq 0$ by the induction hypothesis.

Therefore, we have $V_1^{(n+1)}(\mathbf{s}) - V_1^{(n+1)}(\mathbf{s}') \geq 0$, for $0 \leq s_1 \leq N-2$ and $1 \leq s_2 \leq N-1$. Next, it can be easily checked that the inequality still holds for the case of $s_1 = N-1$, $1 \leq s_2 \leq N-1$ as well as the case of $0 \leq s_1 \leq N-1$, $s_2 = N$. Similarly, $V_2^{(n)}(\mathbf{s}) \leq V_2^{(n)}(\mathbf{s}')$ can also be established. ■

The following lemma shows the difference of $V_1(\mathbf{s}+\mathbf{e}_1)$ and $V_2(\mathbf{s}+\mathbf{e}_2)$ is non-increasing along the line of $s_1+s_2 = m$, $\forall m \in \{1, 2, ..., 2N\}$.

Lemma 2: $V_1^{(n)}(\mathbf{s}+\mathbf{e}_1) - V_2^{(n)}(\mathbf{s}+\mathbf{e}_2) \geq V_1^{(n)}(\mathbf{s}'+\mathbf{e}_1) - V_2^{(n)}(\mathbf{s}'+\mathbf{e}_2)$*, where* $\mathbf{s}' = \mathbf{s}+\mathbf{e}_1-\mathbf{e}_2$.

Proof: It can be easily shown using Lemma 1. ■

Theorem 1: The strategy profile generated by the modified value iteration algorithm has a threshold structure for $K = 2$.

Proof: The soft decision rule in (13) for $K = 2$ can be rewritten as

$$\sigma_{\mathbf{s}}^{(n+1)} = \begin{cases} \sigma_{\mathbf{s}}^{(n)}, & \text{if } |V_1^{(n)}(\mathbf{s}+\mathbf{e}_1) - V_2^{(n)}(\mathbf{s}+\mathbf{e}_2)| \leq \epsilon, \\ 1, & \text{if } V_1^{(n)}(\mathbf{s}+\mathbf{e}_1) > V_2^{(n)}(\mathbf{s}+\mathbf{e}_2) + \epsilon, \\ 2, & \text{if } V_2^{(n)}(\mathbf{s}+\mathbf{e}_2) > V_1^{(n)}(\mathbf{s}+\mathbf{e}_1) + \epsilon. \end{cases} \quad (17)$$

If $\sigma_{\mathbf{s}}^{(n)} = 2$ and $\sigma_{\mathbf{s}}^{(n+1)} = 1$, i.e., the strategy of the current iteration is updated to be different from the one of the previous iteration, then we must have $V_1^{(n)}(\mathbf{s}+\mathbf{e}_1) > V_2^{(n)}(\mathbf{s}+\mathbf{e}_2) + \epsilon$. Lemma 2 implies that $V_1(\mathbf{s}'+\mathbf{e}_1) - V_2(\mathbf{s}'+\mathbf{e}_2)$ is non-increasing along the line of $s_1' + s_2' = s_1 + s_2$. Thus, for $\mathbf{s}' = \mathbf{s} - k\mathbf{e}_1 + k\mathbf{e}_2, k = 1, 2, \ldots, \min\{s_1, N-s_2\}$, we have

$$V_1(\mathbf{s}'+\mathbf{e}_1) - V_2(\mathbf{s}'+\mathbf{e}_2) \geq V_1(\mathbf{s}+\mathbf{e}_1) - V_2(\mathbf{s}+\mathbf{e}_2) > \epsilon > 0.$$

Therefore, $\sigma_{\mathbf{s}'}^{(n+1)} = 1$ for $\mathbf{s}' = \mathbf{s} - k\mathbf{e}_1 + k\mathbf{e}_2, k = 1, 2, ..., \min\{s_1, N-s_2\}$. Similarly, if $\sigma_{\mathbf{s}}^{(n)} = 1$ and $\sigma_{\mathbf{s}}^{(n+1)} = 2$, then $\sigma_{\mathbf{s}''}^{(n+1)} = 2$, for $\mathbf{s}'' = \mathbf{s} + k\mathbf{e}_1 - k\mathbf{e}_2, k = 1, 2, ..., \min\{N-s_1, s_2\}$. With the above discussion, the strategies along the line of $s_1+s_2 = m$, $\forall m \in \{1, 2, ..., 2N\}$ retain a threshold structure in each iteration. Since the initialization of the strategy profile exhibits a threshold structure trivially, the strategy profile obtained in each iteration of the algorithm has a threshold structure. ■

In a two-network system, the number of system states is N^2 and thus N^2 strategies are needed to be stored without the threshold structure. The storage space of each strategy is 1 bits. Now with such threshold structure on each line $s_1 + s_2 = m$, $m = 1, 2, ..., 2N$, we can simply store the threshold point on each line. Each threshold point requires the storage space of $\log N$ bits. Therefore, The storage of the strategy profile can be reduced from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$.

In this paper, we only provide the analysis for the two-network systems. The analysis for systems with more than two networks is difficult due to the lack of the optimality in a single potential function as in the admission control problem [45], [46]. However, it is observed from the simulation results in Section VI that the multi-network systems also possess the strategy profiles with threshold structures. The theoretic analysis of the threshold structure for the multi-network systems is important but out of the scope of this paper, and will serve as one of our future work.

V. Truthful Mechanism Design

In the above discussion, we have implicitly assumed the networks truthfully report their states s_k, and therefore the user can observe the true system state $\mathbf{s}$, by which he/she can make a decision to maximize his/her utility. However, without appropriate incentives, the networks may not truthfully report their states. Instead, a network may untruthfully report some state s_k' different from the true state s_k if profitable. In this section, we consider to enforce truth-telling as a dominant strategy for the networks by incorporating pricing rules into the wireless access network selection game.

A mechanism consists of pricing rules $\{P_k(\mathbf{s})\}$ and allocation rules $\{a_k(\mathbf{s})\}$, where $P_k(\mathbf{s})$ is denoted as the unit price of the expected rate $V_k(\mathbf{s})$ provided by network k at state $\mathbf{s}$, and $a_k(\mathbf{s})$ is denoted as the allocation probability, which is either 1 or 0, i.e., whether or not the user enters network k. The utility of network k is given by

$$U_k(\mathbf{s}) = V_k(\mathbf{s}+\mathbf{e}_k)P_k(\mathbf{s}) - c_k(\mathbf{s}+\mathbf{e}_k)a_k(\mathbf{s}), \quad (18)$$

where $c_k(\mathbf{s}+\mathbf{e}_k)$ is the cost per user. With the states reported from the networks, these rules determine the user allocation and the price the user has to pay, both as functions of the reports from networks. For example, if network k reports his state as s_k' and others report $s_{-k} = \{s_j : j \neq k\}$, his utility becomes $V_k(\mathbf{s}+\mathbf{e}_k)P_k(s_k', s_{-k}) - c_k(\mathbf{s}+\mathbf{e}_k)a_k(s_k', s_{-k})$. Notice that $V_k(\mathbf{s}+\mathbf{e}_k)$ and $c_k(\mathbf{s}+\mathbf{e}_k)$ are functions of true states that do not depend on the reports. Thus, the truth-telling or the incentive compatibility (IC) constraints are, $\forall s_k, s_k', s_{-k}$,

$$V_k(s_k+1, s_{-k})P_k(s_k, s_{-k}) - c_k(s_k+1, s_{-k})a_k(s_k, s_{-k}) \geq V_k(s_k+1, s_{-k})P_k(s_k', s_{-k}) - c_k(s_k+1, s_{-k})a_k(s_k', s_{-k}),$$

which means truth-telling is a dominant strategy for each network at each state. The mechanism also has to satisfy the individual rationality (IR) constraints, i.e., $\forall s_k, s_{-k}$,

$$V_k(s_k+1, s_{-k})P_k(s_k, s_{-k}) - c_k(s_k+1, s_{-k})a_k(s_k, s_{-k}) \geq 0, \quad (19)$$

which guarantees all networks would attend the mechanism.

In the previous sections, we study the network selection game with the focus of the interdependence between the users. In this section, we study the interplay among the networks. To this end, we assume that users' strategies are chosen based on the *ex ante* optimality [18], [35], i.e., the allocation rule is based on optimizing the expected objective over the state probability. The truthful mechanism design is to construct a set of pricing and allocation rules which optimize a specific objective while satisfying IC and IR constraints. For example, the mechanism design problem $\boldsymbol{\mathcal{P}}_p$ for minimizing the expected payment can be formulated as follows.

$$\boldsymbol{\mathcal{P}}_p: \quad \min_{\{P_k\},\{a_k\}} \quad \sum_{\mathbf{s}\in\mathcal{S}} \pi(\mathbf{s}) \sum_{k=1}^{K} P_k(\mathbf{s}) V_k(\mathbf{s}+\mathbf{e}_k) \tag{20}$$

$$\text{s.t.} \quad \text{(IC), (IR)}, \ a_k(\mathbf{s}) \in \{0,1\}, \forall \mathbf{s}, \forall k. \tag{21}$$

$$\sum_{k=1}^{K} a_k(\mathbf{s}) = 1, \forall \mathbf{s} \in \mathcal{S}, \tag{22}$$

Other mechanism design objectives such as the utility maximization $\boldsymbol{\mathcal{P}}_u$ can be formulated by substituting (20) with users' expected utility function as follows.

$$\max_{\{P_k\},\{a_k\}} \sum_{\mathbf{s}\in\mathcal{S}} \pi(\mathbf{s}) \sum_{k=1}^{K} \left[\lambda a_k(\mathbf{s}) V_k(\mathbf{s}+\mathbf{e}_k) - P_k(\mathbf{s}) V_k(\mathbf{s}+\mathbf{e}_k)\right]$$
$$\text{s.t. } (21), (22).$$

The unit cost $c_k(\mathbf{s}+\mathbf{e}_k)/V_k(\mathbf{s}+\mathbf{e}_k)$ is denoted as $w_k(\mathbf{s})$. The (IC) constraints become

$$P_k(s_k, s_{-k}) - w_k(s_k, s_{-k}) a_k(s_k, s_{-k}) \geq P_k(s'_k, s_{-k}) - w_k(s_k, s_{-k}) a_k(s'_k, s_{-k}), \forall s_k, s'_k, s_{-k}. \tag{23}$$

In the following, we need a monotonicity assumption for the unit cost, i.e., $w_k(s_k, s_{-k})$ is non-decreasing in s_k, i.e., $w_k(s_k, s_{-k}) \geq w_k(s'_k, s_{-k})$, if $s_k \geq s'_k$. Since $V_k(s_k, s_{-k})$ is non-increasing in s_k, the assumption holds when $c_k(s_k, s_{-k})$ is non-decreasing in s_k. For example, if the per-user cost is a constant in each network, i.e., $c_k(s_k, s_{-k}) = C_k$, then the assumption holds. The monotonicity of $w_k(s_k, s_{-k})$ leads to the threshold structure of $a_k(s_k, s_{-k})$ as in the following lemma.

Lemma 3: Under IC constraints, there exists a threshold value of s_k on the allocation rule $a_k(s_k, s_{-k})$, i.e., given s_{-k}, there exists $s_k^*(s_{-k}) \in \{-1, 0, 1, \ldots, N\}$, such that

$$a_k(s_k, s_{-k}) = \begin{cases} 1, & s_k \leq s_k^*(s_{-k}) \\ 0, & s_k > s_k^*(s_{-k}). \end{cases} \tag{24}$$

Proof: From (23), we have

$$P_k(s_k, s_{-k}) - P_k(s'_k, s_{-k}) \geq w_k(s_k, s_{-k}) \left[a_k(s_k, s_{-k}) - a_k(s'_k, s_{-k})\right]. \tag{25}$$

Interchanging s_k and s'_k, we also have

$$P_k(s'_k, s_{-k}) - P_k(s_k, s_{-k}) \geq w_k(s'_k, s_{-k}) \left[a_k(s'_k, s_{-k}) - a_k(s_k, s_{-k})\right]. \tag{26}$$

Combining the above two inequality leads to

$$\left[w_k(s_k, s_{-k}) - w_k(s'_k, s_{-k})\right] \left[a_k(s_k, s_{-k}) - a_k(s'_k, s_{-k})\right] \leq 0. \tag{27}$$

Thus, since $w_k(s_k, s_{-k})$ is non-decreasing in s_k, the allocation rule $a_k(s_k, s_{-k})$ has to be non-increasing in s_k. With this monotonicity and the fact that $a_k(s_k, s_{-k})$ can only have value of 0 or 1, we can conclude that there exists a threshold of $a_k(s_k, s_{-k})$ in s_k as described in (24). ■

Corollary 1: If $K = 2$, then $s_1^*(s_2)$ is non-decreasing in s_2, and $s_2^*(s_1)$ is non-decreasing in s_1.

Proof: Suppose $\exists s_2$ such that $s_1^*(s_2+1) < s_1^*(s_2)$. By Lemma 3, we have $a_1(s_1, s_2+1) = 0$, for $s_1 > s_1^*(s_2+1)$, which implies $a_2(s_1, s_2+1) = 1$, for $s_1 > s_1^*(s_2+1)$, due to the constraint that $a_1(\mathbf{s}) + a_2(\mathbf{s}) = 1$, $\forall \mathbf{s}$. Therefore, $a_2(s_1^*(s_2), s_2+1) = 1$, which implies $a_2(s_1^*(s_2), s_2) = 1$ by Lemma 3, but we also have $a_1(s_1^*(s_2), s_1) = 1$, which leads to a contradiction. ■

The following lemma shows that only adjacent IC constraints are necessary.

Lemma 4: Non-adjacent IC constraints are redundant.

Proof: Let us consider the two adjacent IC constraints as follows.

$$P_k(s_k, s_{-k}) - w_k(s_k, s_{-k}) a_k(s_k, s_{-k}) \geq P_k(s_k-1, s_{-k}) - w_k(s_k, s_{-k}) a_k(s_k-1, s_{-k}), \tag{28}$$

$$P_k(s_k-1, s_{-k}) - w_k(s_k-1, s_{-k}) a_k(s_k-1, s_{-k}) \geq P_k(s_k-2, s_{-k}) - w_k(s_k-1, s_{-k}).a_k(s_k-2, s_{-k}) \tag{29}$$

Adding (28) and (29), we have

$$\begin{aligned} &P_k(s_k, s_{-k}) - w_k(s_k, s_{-k}) a_k(s_k, s_{-k}) \\ &\geq P_k(s_k-2, s_{-k}) - w_k(s_k, s_{-k}) a_k(s_k-2, s_{-k}) \\ &\quad - w_k(s_k, s_{-k}) \left[a_k(s_k-1, s_{-k}) - a_k(s_k-2, s_{-k})\right] \\ &\quad + w_k(s_k-1, s_{-k}) \left[a_k(s_k-1, s_{-k}) - a_k(s_k-2, s_{-k})\right] \\ &\geq P_k(s_k-2, s_{-k}) - w_k(s_k, s_{-k}) a_k(s_k-2, s_{-k}). \end{aligned} \tag{30}$$

The last inequality is due to that $w_k(s_k, s_{-k})$ is increasing in s_k and $a_k(s_k, s_{-k})$ is decreasing in s_k. It shows that the non-adjacent IC constraints can be inferred from the adjacent ones. ■

Using the adjacent IC constraints, we can obtain the bounds for the payments, i.e., given an allocation rule $\{a_k(\mathbf{s})\}$, the incentive compatible payment rule $\{P_k(\mathbf{s})\}$ satisfies

$$\begin{aligned} &P_k(s_k, s_{-k}) + w_k(s_k, s_{-k}) \left[a_k(s_k-1, s_{-k}) - a_k(s_k, s_{-k})\right] \\ &\geq P_k(s_k-1, s_{-k}) \geq P_k(s_k, s_{-k}) \\ &\quad + w_k(s_k-1, s_{-k}) \left[a_k(s_k-1, s_{-k}) - a_k(s_k, s_{-k})\right] \end{aligned} \tag{31}$$

In the optimization problems $\boldsymbol{\mathcal{P}}_p$, we aim to minimize a linear combination of $P_k(s_k, s_{-k})$ with nonnegative coefficients. Clearly, the lower bound in (31) should be binding; otherwise, the objective function can always be better optimized by decreasing the non-binding $P_k(s_k, s_{-k})$. Hence, the payment rule can be expressed as

$$P_k(s_k, s_{-k}) = P_k(N, s_{-k}) + \sum_{r=s_k+1}^{N} w_k(r-1, s_{-k}) \left[a_k(r-1, s_{-k}) - a_k(r, s_{-k})\right]. \tag{32}$$

To minimize $P_k(s_k, s_{-k})$ while satisfying the IR constraint in (19), $P_k(N, s_{-k})$ should be set as 0. Substituting Lemma 3 into (32), we can conclude

$$P_k(s_k, s_{-k}) = \begin{cases} w_k(s_k^*, s_{-k}), & s_k \leq s_k^*, \\ 0, & s_k > s_k^*, \end{cases} \tag{33}$$

where s_k^* denotes $s_k^*(s_{-k})$ for notational simplicity.

From the IC and IR constraints, the pricing rule $\{P_k\}$ can be determined given the allocation rule $\{a_k\}$, which is specified by the thresholds $\{s_k^*\}$. Thus (33) simply means the pricing rule $\{P_k\}$ is also specified by the thresholds $\{s_k^*\}$. Using $\{s_k^*\}$ as optimization variables, the problem $\boldsymbol{\mathcal{P}}_p$ can be simplified as

$$\min_{\{s_k^*\}} \sum_{\mathbf{s}\in\mathcal{S}} \pi(\mathbf{s}) \sum_{k=1}^{K} P_k(\mathbf{s}) V_k(\mathbf{s}) \quad \text{s.t.} \quad (22), (24), (33). \tag{34}$$

With the simplification, however, the optimization problem is still difficult to be solved optimally since the optimization variables $\{s_k^*\}$ is discrete and the exhaustive search requires exponential-time complexity in N. Motivated by the optimal substructures in the two-network case, a dynamic programming algorithm is proposed for the above problem. The optimal solution to the primary problem can be broken down into solving the optimal solutions to its subproblems. The dynamic programming technique essentially performs recursive divide-and-conquer to tackle each of these subproblems. However, for the multi-network case, the proposed dynamic programming approach is suboptimal but the performance is satisfactory compared to the greedy method. Other traditional optimization algorithms such as branch-and-bound can be applied to optimally solve the mixed integer programming problem, but the computational complexity is prohibitively high (exponential in the number of states) since such an algorithm basically performs exhaustive tree search with certain pruning strategies. In general a mixed integer program does not have an efficient solution. In this paper, we aim to propose an algorithm that is able to achieve satisfactory performance with reasonable complexity (polynomial in the number of states).

A. Proposed Algorithm

Since the number of states is N^K, the exhaustive search over all possible allocation rules requires complexity of $\mathcal{O}(K^{N^K})$. Such an exponential complexity is formidably high even for a moderate N. In this subsection, we propose a polynomial time algorithm based on dynamic programming to search for the thresholds $\{s_k^*\}$. Let $f_k^{\mathrm{DP}}(\{s_i : i \in \mathcal{I}\}|\{s_j : j \in \mathcal{J}\})$ denote the optimal value of a set of system states specified by $(\{s_i : i \in \mathcal{I}\}|\{s_j : j \in \mathcal{J}\})$, where the set $\mathcal{J}$ consists of coordinates with coordinate j being fixed as s_j. The set $\mathcal{I}$ consists of the coordinates with ranges, where coordinate i ranges from 1 to s_i. The set $\mathcal{I}$ has k coordinates, i.e., the considered set of system states is k-dimensional. The optimal value function f_k^{DP} can be computed using lower-dimensional optimal value functions. The recursive calculation is described by the following equations. For $k = 2, \ldots, K$,

$$f_k^{\mathrm{DP}}(\{s_i : i \in \mathcal{I}\}|\{s_j : j \in \mathcal{J}\}) = \min_{i \in \mathcal{I}} \Big\{ f_k^{\mathrm{DP}}(s_i - 1, s_{-i}|\{s_j : j \in \mathcal{J}\}) + f_{k-1}^{\mathrm{DP}}(s_{-i}|\{s_j : j \in \mathcal{J} \cup \{i\}\}) \Big\},$$

where $s_{-i} = \{s_l : l \neq i, l \in \mathcal{I}\}$. (35)

$$a_{i^*}(s_{i^*}, s'_{-i^*}, s_j, s_{-j}) = 0, \forall s'_{-i^*} \preceq s_{-i^*}, \tag{36}$$

$$i^* = \arg\min_{i \in \mathcal{I}} \Big\{ f_k^{\mathrm{DP}}(s_i - 1, s_{-i}|\{s_j : j \in \mathcal{J}\}) + f_{k-1}^{\mathrm{DP}}(s_{-i}|\{s_j : j \in \mathcal{J} \cup \{i\}\}) \Big\}, \tag{37}$$

where $s'_{-i^*} \preceq s_{-i^*}$ denotes $s'_{-i^*} \in \{s'_l : s'_l \leq s_l, l \neq i^*, l \in \mathcal{I}\}$. The boundary condition is

$$f_1^{\mathrm{DP}}(s_i|s_{-i}) = f_1^{\mathrm{DP}}(s_i - 1|s_{-i}) \frac{w_i(s_i, s_{-i})}{w_i(s_i - 1, s_{-i})} + \pi(s_i, s_{-i}) V_i(s_i + 1, s_{-i}) w_i(s_i, s_{-i}), \tag{38}$$

$$a_{-i^*}(s_{i^*}, s'_{-i^*}, s_j, s_{-j}) = 1, \forall s'_{-i^*} \leq s_{-i^*}, \tag{39}$$

where i^* is the minimizer in (37) when $k = 2$. Notice that (38) is equivalent to $f_1^{\mathrm{DP}}(s_i|s_{-i}) = \sum_{r=0}^{s_i} \pi(r, s_{-i}) V_i(r + 1, s_{-i}) w_i(s_i, s_{-i})$, but the recursive form in (38) is more efficient in computation with the price of using more storage space. The proposed algorithm is to evaluate $f_K^{\mathrm{DP}}(N, \ldots, N)$ with $\mathcal{I} = \{1, \ldots, K\}$ and $\mathcal{J} = \phi$ by using (35)-(39). The following proposition shows the optimality of the solution obtained by the proposed algorithm when $K = 2$. The proof is omitted due to space limitation.

TABLE II
DYNAMIC PROGRAMMING ALGORITHM FOR MECHANISM DESIGN

(i) Initialization: obtain $\{V_k^{(0)}(\mathbf{s})\}$ and $\{\pi^{(0)}(\mathbf{s})\}$ using Table I.
(ii) **Loop**:
1. With initial $\mathcal{I} = \{1, \ldots, K\}$, $\mathcal{J} = \phi$, evaluate $f_K^{(n)}(N, \ldots, N)$ using (35)-(39) to obtain $\{a_k^{(n+1)}(\mathbf{s})\}$ and $\{P_k^{(n+1)}(\mathbf{s})\}$.
2. Calculate $\{V_k^{(n+1)}(\mathbf{s})\}$ and $\{\pi^{(n+1)}(\mathbf{s})\}$.
Until $\{a_k^{(n+1)}(\mathbf{s})\}$ and $\{P_k^{(n+1)}(\mathbf{s})\}$ converge.

Proposition 1: For $K = 2$, the proposed algorithm optimally solves $\boldsymbol{\mathcal{P}}_p$ in $\mathcal{O}(N^2)$.

For $K \geq 3$, the solution obtained by the proposed algorithm may be sub-optimal since monotonicity of allocation thresholds in Corollary 1 only holds when $K = 2$. However, it will be shown in Section VI that the proposed algorithm still outperforms the heuristic greedy method. For a general K, the computational complexity of the proposed algorithm can be shown to be $\mathcal{O}(N^K)$, which is polynomial in N.

Given the expected rate $\{V_k(\mathbf{s})\}$ and the stationary probability $\{\pi(\mathbf{s})\}$, the proposed dynamic programming can efficiently find solutions of the allocation rule $\{a_k(\mathbf{s})\}$ and the pricing rule $\{P_k(\mathbf{s})\}$ to the problem $\boldsymbol{\mathcal{P}}_p$. However, $\{V_k(\mathbf{s})\}$ and $\{\pi(\mathbf{s})\}$ depend on $\{a_k(\mathbf{s})\}$ since the state transition probability depends on $\{a_k(\mathbf{s})\}$. Therefore, we propose to iteratively update $\{V_k(\mathbf{s})\}$, $\{\pi(\mathbf{s})\}$, and $\{a_k(\mathbf{s})\}$. The proposed mechanism design algorithm for the network selection game is summarized in Table II. In the numerical simulation, we observed that the iterative algorithm exhibits very fast convergence. The typical number of iterations to converge is between 5 to 8.

The proposed algorithm can be easily modified to solve $\boldsymbol{\mathcal{P}}_u$ by replacing the $\min$ in (35) and (37) with the $\max$, and changing the boundary condition in (38) to be $f_1(s_i|s_{-i}) = \sum_{r=0}^{s_i} \pi(r, s_{-i}) V_k(r, s_{-i})(\lambda - w(s_i, s_{-i}))$.

VI. Numerical Simulation

In this section, we use numerical simulation to verify the analysis and evaluate the performance of the proposed modified value iteration algorithm as the rational strategy. The proposed method is compared with the following schemes. We first define the social welfare given a strategy profile $\boldsymbol{\sigma}$ as $SW^{\boldsymbol{\sigma}} = \sum_{\mathbf{s} \in \mathcal{S}} \pi^{\boldsymbol{\sigma}}(\mathbf{s}) \sum_{k=1}^{K} s_k R_k(s_k)$, where $\pi^{\boldsymbol{\sigma}}(\mathbf{s})$ is the stationary probability at system state $\mathbf{s}$. The centralized method is to exhaustively search through all the possible strategy profiles and choose the one that achieves the largest social welfare, i.e., $\boldsymbol{\sigma}^{\mathrm{cent}} = \arg\max_{\boldsymbol{\sigma}} SW^{\boldsymbol{\sigma}}$. Thus, the centralized method requires a computational complexity of $\mathcal{O}(K^{|\mathcal{S}|})$, which is exponentially increasing in the number of system states and is impossible to be used in practice. The myopic strategy is obtained by choosing the largest

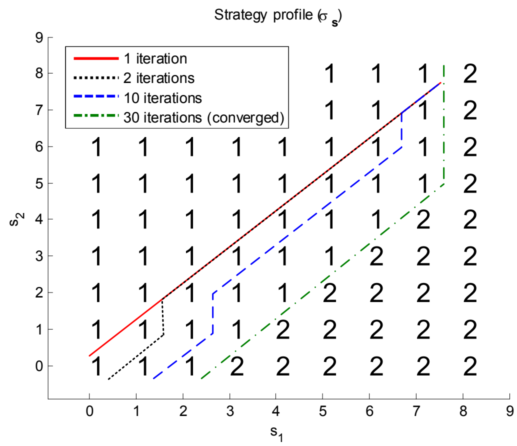


Fig. 3. The threshold structure of the strategy profile during iterations of the proposed algorithm.

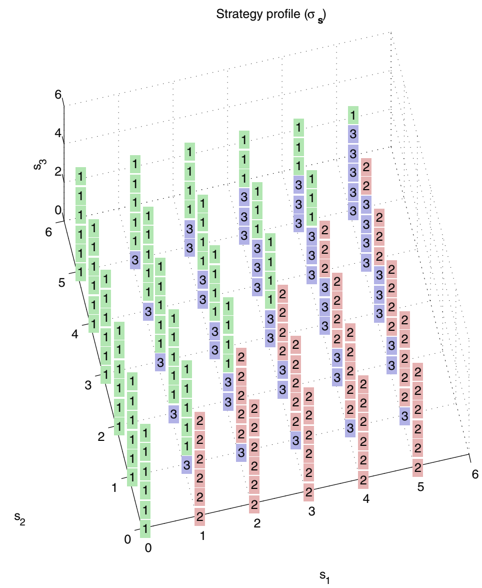


Fig. 4. The threshold structure of the strategy profile for a three-network system.

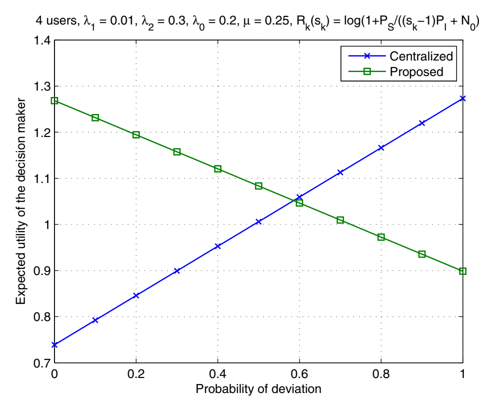


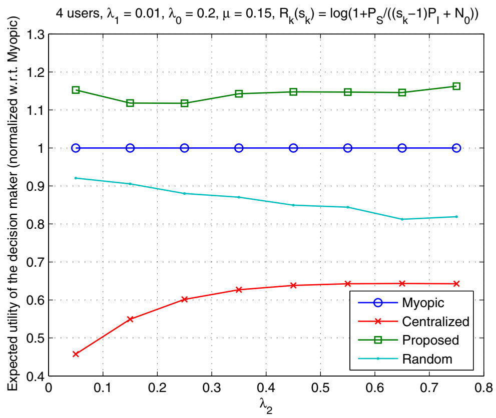
Fig. 5. Comparison of the proposed method and the centralized method for the decision maker's expected utility versus probability of deviation.

immediate utility after making the decision, i.e., $\sigma_{\mathbf{s}}^{\text{myop}} = \arg\max_{k\in\{1,\ldots,K\}} R_k(s_k+1)$. In current cellular systems, the cell selection is done by choosing the base-station with the highest detected SNR. Such an approach is similar to the myopic strategy since it only concerns about the immediate utility. Finally, the random strategy is to randomly make the decision with equal probability, i.e., $Pr\left\{\sigma_{\mathbf{s}}^{\text{rand}} = k\right\} = \frac{1}{|\bar{\mathcal{F}}(\mathbf{s})|}, \forall k \in \bar{\mathcal{F}}(\mathbf{s})$, where $|\cdot|$ denotes the cardinality of a set. In the following simulation, the performance of the random strategy is obtained by averaging the performance of 1000 instances for each set of parameters.

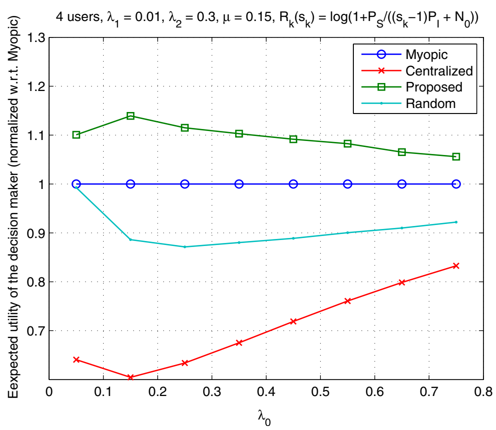
The algorithm analysis in Section IV shows that there exists a threshold structure of the strategies along each line of $s_1+s_2=m, \forall m \in \{1,2,...,2N\}$. We verify the analysis by numerical simulation in Figure 3, which illustrates the strategy profile computed by the proposed algorithm in a two-network system where $P_s/N_0 = 50$, $P_I/N_0 = 10$, $T = 0.08$ (sec), $\bar{\lambda}_0 = 0.5$ (users/sec), $\bar{\lambda}_1 = 0.125$ (users/sec), $\bar{\lambda}_2 = 2.5$ (users/sec), $\bar{\mu} = 1.25$ (users/sec), $\epsilon = 0.05$ and $N = 8$. The x-axis (y-axis) denotes s_1 (s_2), i.e., the number of users in network 1 (network 2). The number marked at the coordinate $\mathbf{s} = (s_1, s_2)$ denotes the computed strategy $\sigma_{\mathbf{s}}$, which is either 1 or 2 in this scenario. This figure shows the strategy profile converges in 30 iterations. The green (dot-dash) line is drawn in between different strategies to emphasize the threshold. The threshold lines of certain iterations (1, 2, and 10) are also shown in the figure to illustrate the evolution of the strategy profile during the iterations of the proposed algorithm. It is observed that at each iteration, the threshold structure of the strategies always exists along the diagonal lines as the analysis in Section IV. In the rest of simulations, instead of specifying the arrival rates and the time slot duration, we consider the parameters as transition probabilities since the relative values of these probabilities directly influence the resulting performance. Figure 4 shows the converged strategy profile of a three-network system, where $P_s/N_0 = 50$, $P_I/N_0 = 10$, $\lambda_0 = 0.1$, $\lambda_1 = 0.1$, $\lambda_2 = 0.2$, $\lambda_3 = 0.3$, $\mu = 0.1$, $\epsilon = 0.05$ and $N = 5$. It is observed that the strategy profile also has a threshold structure.

Figure 5 validates the individual rationality of the proposed method in a two-network system, where the parameters are set to be $P_s/N_0 = 50$, $P_I/N_0 = 10$, $\lambda_0 = 0.2$, $\lambda_1 = 0.01$, $\lambda_2 = 0.3$, $\mu = 0.25$, $\epsilon = 0.05$, and $N = 4$. The decision maker's expected utility, defined as $E[V_{\sigma_{\mathbf{s}}}(\mathbf{s}+\mathbf{e}_{\sigma_{\mathbf{s}}})]$, is evaluated versus the probability of deviation p_d. For computational tractability of the centralized method, the number of users N is set to be 4. Note that the time slot duration is chosen to ensure that $\lambda_0+\lambda_1+\lambda_2+2N\mu \leq 1$ but the relative values of these probabilities are retained. The user at state $\mathbf{s}$ deviates from the given strategy $\sigma_{\mathbf{s}}$ with probability p_d. The decision maker's expected utility can only be impaired if he deviates from the strategy profile generated by the proposed method. However, by deviating from the centralized strategy that maximizes the social welfare, the user can possibly obtain higher expected utility (about 70% performance improvement in Figure 5). Clearly, the individual rationality is not satisfied for the centralized strategy.

Figure 6(a) and 6(b) show the comparison of the decision maker's expected utility with different strategy profiles in a two-network system where $P_s/N_0 = 50$, $P_I/N_0 = 10$,

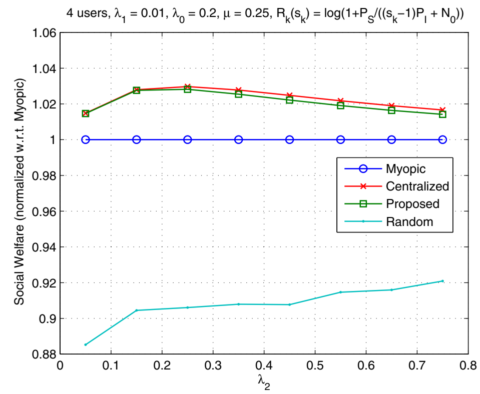


(a) The decision maker's expected utility versus λ_2.

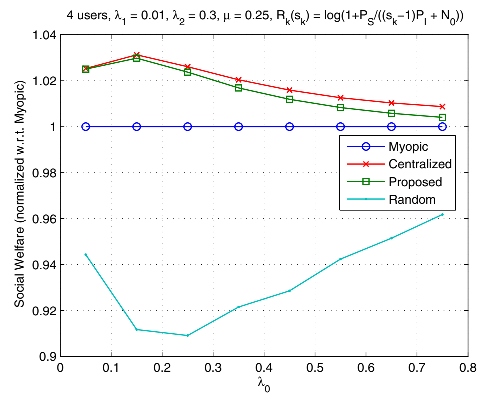


(b) The decision maker's expected utility versus λ_0.

Fig. 6. Comparison of different strategies for the decision maker's expected utility.



(a) The social welfare versus λ_2.



(b) The social welfare versus λ_0.

Fig. 7. Comparison of different strategies for the social welfare.

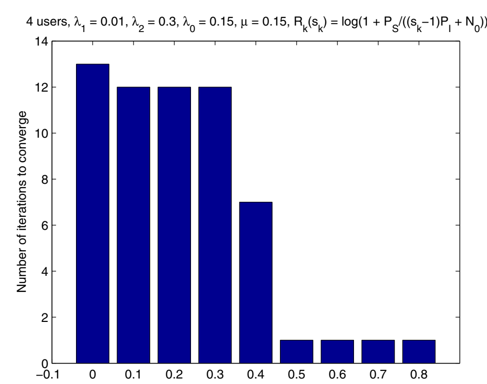


Fig. 8. The impact of ϵ on the number of iterations for the strategy profile to converge.

$\lambda_1 = 0.01$, $\mu = 0.15$, $\epsilon = 0.5$, and $N = 4$. We use the myopic strategy as the baseline by normalizing the performance of other methods with that of the myopic strategy. In Figure 6(a), $\lambda_0 = 0.2$ and λ_2 is varied from 0.05 to 0.75. In Figure 7(b), $\lambda_2 = 0.3$ and λ_0 is varied from 0.05 to 0.75. It can be seen that the proposed method performs the best among all the schemes since the decision maker optimizes his expected utility by choosing network to his best advantage. The myopic strategy always has performance 1 due to the normalization. The random strategy is worse than the myopic method which exploits the information of the immediate utility. The centralized method performs the worst because it maximizes the social welfare and results in sacrificing the decision maker's expected utility.

In Figure 7(a) and 7(b), we compare the social welfare performance of the strategy profiles generated by different approaches in a two-network system where the parameters are $P_s/N_0 = 50$, $P_I/N_0 = 10$, $\lambda_0 = 0.2$, $\mu = 0.25$, $\epsilon = 0.05$ and $N = 4$. In Figure 7(a), $\lambda_1 = 0.01$ and λ_2 is varied from 0.05 to 0.75. In Figure 7(b), $\lambda_2 = 0.3$ and λ_0 is varied from 0.05 to 0.75. The performance of each method is normalized by the myopic one. It can be seen that the proposed method performs similar to that of the centralized method which maximizes the social welfare. Figure 8 shows the impact of ϵ on the number of iterations for the strategy profile to converge using the proposed modified value iteration algorithm. It can be seen that when ϵ increases, it requires smaller number of iterations to converge since the region of tolerance for switching among the strategy profile is larger, and possibly more ϵ-approximate NEs are available.

Figure 9 and 10 show the performance comparison for different mechanism designs when $K = 2$ and $K = 3$, respectively. The exhaustive search is to search over all possible allocation rules and find out the one with the optimal objective value. The greedy algorithm is characterized by the

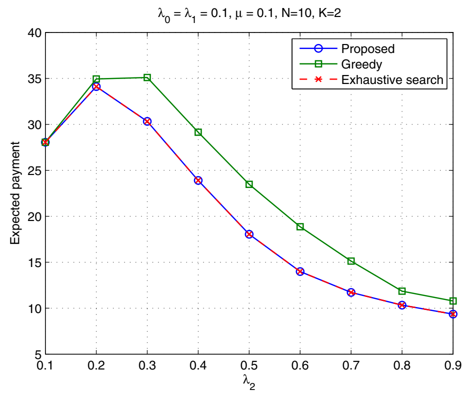


Fig. 9. Comparison of different mechanism designs for the expected payment versus λ_2 when $K = 2$.

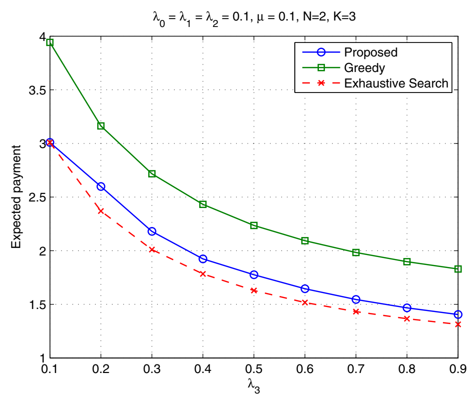


Fig. 10. Comparison of different mechanism designs for the expected payment versus λ_3 when $K = 3$.

following recursive formula.

$$f_k^{\mathrm{G}}(\{s_i, i \in \mathcal{I}\}|\{s_j, j \in \mathcal{J}\}) = \min_{i\in\mathcal{I}} \left\{ f_{k-1}^{\mathrm{G}}(s_{-i}|\{s_j, j \in \mathcal{J} \cup \{i\}\}) \right\}, \quad (40)$$

$$a_{i^*}(s_{i^*}, s'_{-i^*}, s_j, s_{-j}) = 0, \forall s'_{-i^*} \preceq s_{-i^*}, \text{ where } i^* = \arg\min_{i\in\mathcal{I}} \left\{ f_{k-1}^{\mathrm{G}}(s_{-i}|s_j, j \in \mathcal{J} \cup \{i\}) \right\}. \quad (41)$$

The boundary condition is

$$f_1^{\mathrm{G}}(s_i|s_{-i}) = f_1^{\mathrm{G}}(s_i - 1|s_{-i}) \frac{w_i(s_i, s_{-i})}{w_i(s_i - 1, s_{-i})} + \pi(s_i, s_{-i}) V_i(s_i + 1, s_{-i}) w_i(s_i, s_{-i}), \quad (42)$$

$$a_{-i^*}(s_{i^*}, s'_{-i^*}, s_j, s_{-j}) = 1, \forall s'_{-i^*} \leq s_{-i^*}, \text{ where } i^* \text{ is the minimizor in (41) when } k = 2. \quad (43)$$

The greedy algorithm is to evaluate $f_K^{\mathrm{G}}(N, \ldots, N)$ with $\mathcal{I} = \{1, \ldots, K\}$ and $\mathcal{J} = \phi$ by using (40)-(43). With a similar analysis, the computational complexity of the greedy algorithm can be shown to be $\mathcal{O}(N^K)$. Compared with the proposed DP algorithm, the greedy method is a heuristic approach which makes a local optimal decision according to lower dimensional results. We can see more clearly by considering the case $K = 2$, i.e.,

$$f_2^{\mathrm{G}}(s_1, s_2) = \min \left\{ f_1^{\mathrm{G}}(s_2|s_1), f_1^{\mathrm{G}}(s_1|s_2) \right\}, \quad (44)$$

$$(a_1(s_1, s'_2), a_2(s_1, s'_2)) = (0, 1), \forall s'_2 \leq s_2, \text{ if } f_1^{\mathrm{G}}(s_2|s_1) > f_1^{\mathrm{G}}(s_1|s_2), \quad (45)$$

$$(a_1(s'_1, s_2), a_2(s'_1, s_2)) = (1, 0), \forall s'_1 \leq s_1, \text{ if } f_1^{\mathrm{G}}(s_2|s_1) \leq f_1^{\mathrm{G}}(s_1|s_2). \quad (46)$$

For example, when evaluating $f_2^{\mathrm{G}}(N, N)$, if $f_1^{\mathrm{G}}(s_2 = N|s_1 = N)$ is larger than $f_1^{\mathrm{G}}(s_1 = N|s_2 = N)$, then state (N, N) is allocated to network 1. Due to Lemma 3, the states $\{(s_1, N), \forall s_1 \leq N\}$ are all allocated to network 1. Since the unallocated states so far are $\{(s_1, s_2), 0 \leq s_1 \leq N, 0 \leq s_2 \leq N - 1\}$, we can then evaluate $f_2^{\mathrm{G}}(N, N - 1)$, and so on. In Figure 9, we can see that the proposed DP algorithm can achieve the same performance as the exhaustive search when $K = 2$, but requires only a polynomial time complexity. The greedy algorithm has a worse performance since it makes a local optimal decision to determine the thresholds of allocation rules. In Figure 10, different mechanism design approaches are compared for $K = 3$. It can be seen that the proposed DP algorithm still outperforms the greedy method. As discussed in Section V, for a general K the proposed DP algorithm may not achieve the global optimum. However, with much lower complexity compared to the exhaustive search, the proposed algorithm can achieve reasonably good results and thus can serve as an approximate approach.

VII. Discussion

Although we focus on the wireless access network selection problem in this paper, we should notice that the model described in this work is very general and can be applied into many other problems. A closely related scenario is the cell selection problem in cellular networks [47]–[49]. When a mobile station desires to inform the cellular system whether it is on the air, it registers to a base station which corresponds to a cellular cell. In most current cellular systems, the cell selection process is simply accomplished by a local signal-to-noise ratio (SNR)-based strategy, which is to detect the SNR of each cell and choose the cell with the largest SNR [48]. However, such a simple strategy does not take into account the strategies of others, i.e., the negative network externality. The QoS experienced by a mobile station will be degraded if the limited resources are shared with a large number of users. The utilization of system resources will also be degraded since such a strategy results in cellular cells with unbalanced load.

It can be seen that the cell selection problem has the same structure with the wireless access network selection problem. Mobile stations sequentially choose one cellular cell (corresponding to a base station) to register based on the obtained information about each available cell. The utility of a mobile station is determined by the expected throughput during the period it stays in the cell. Furthermore, the instantaneous throughput of a mobile station in a certain cell is affected by the crowdedness of the cell due to the limited bandwidth and the delay caused by the scheduling overhead. Thus, a rational mobile station should choose a cellular cell in consideration of other mobile stations' decisions to avoid the crowdedness.

VIII. Conclusion

In this paper, we have studied the wireless access network selection problem as a stochastic game with negative network externality, where a user decides which network to connect to by considering subsequent users' decisions. The problem is shown to be a multi-dimensional MDP. We propose a modified value iteration algorithm to obtain the optimal strategy profile for each selfish user. The analysis of the proposed algorithm shows that the resulting strategy profile exhibits a threshold structure along each diagonal line. Such a threshold structure can be used to save the storage space of the strategy profile from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$ in the two-network scenario. Simulation results are shown to validate the analysis and demonstrate that rational users will not deviate from the strategy profile obtained by the proposed algorithm. For the expected utility of the decision maker, the proposed method is superior to other approaches. Moreover, its social welfare performance is shown to be similar to that of the centralized strategy which maximizes the social welfare.

We further investigated truth-telling enforcing mechanism design in the wireless access network selection problem. The mechanism design captures the incentive compatibility and individual rationality constraints while optimizing the utility of users. The formulated problem as a mixed integer program in general does not have an efficient solution. By exploiting the optimal substructures, a dynamic programming algorithm is proposed to optimally solve the mixed integer programming problem in the two-network scenario. For the multi-network scenario, the proposed algorithm can outperform the heuristic greedy approach in a polynomial-time complexity. Finally, simulation results substantiate the optimality in the two-network case and also demonstrate the effectiveness of the proposed algorithm in the multi-network scenario.

REFERENCES

[1] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sept. 2008.

[2] M. Ahmed, "Call admission control in wireless networks: a comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 7, no. 1, pp. 49–68, 2005.

[3] D. Gao, J. Cai, and K. N. Ngan, "Admission control in IEEE 802.11e wireless LANs," *IEEE Netw.*, vol. 19, no. 4, pp. 6–13, 2005.

[4] W. H. Sandholm, "Negative externalities and evolutionary implementation," *Rev. Economic Studies*, vol. 72, no. 3, pp. 885–915, July 2005.

[5] G. Fagiolo, "Endogenous neighborhood formation in a local coordination model with negative network externalities," *J. Economic Dynamics Control*, vol. 29, no. 1-2, pp. 297–319, 2005.

[6] J. Hou and D. C. O'Brien, "Vertical handover-decision-making algorithm using fuzzy logic for the integrated radio-and-OW system," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 176–185, Nov. 2006.

[7] K. Mittal, E. M. Belding, and S. Suri, "A game-theoretic analysis of wireless access point selection by mobile users," *Comput. Commun.*, vol. 31, no. 10, pp. 2049–2062, 2008.

[8] D. Niyato and E. Hossain, "Dynamics of network selection in heterogeneous wireless networks: an evolutionary game approach," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 2008–2017, May 2009.

[9] Q. Song and A. Jamalipour, "Network selection in an integrated wireless LAN and UMTS environment using mathematical modeling and computing techniques," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 42–48, June 2005.

[10] A. Argento, M. Cesana, N. Gatti, and I. Malanchini, "A game theoretical study of access point association in wireless mesh networks," *Comput. Commun.*, vol. 35, no. 5, pp. 541–553, 2012.

[11] D. Charilas, O. Markaki, D. Nikitopoulos, and M. Theologou, "Packet-switched network selection with the highest QoS in 4G networks," *Comput. Netw.*, vol. 52, no. 1, pp. 248–258, 2008.

[12] M. Cesana, N. Gatti, and I. Malanchini, "Game theoretic analysis of wireless access network selection: models, inefficiency bounds, and algorithms," in *Proc. 2008 Intl. Conf. Performance Evaluation Methodologies Tools*, pp. 6:1–6:10.

[13] A. Balachandran, P. Bahl, and G. Voelker, "Hot-spot congestion relief in public-area wireless networks," in *Proc. 2002 IEEE Workshop Mobile Comput. Syst. Appl.*, pp. 70–80.

[14] L. Chen, "A distributed access point selection algorithm based on no-regret learning for wireless access networks," in *Proc. 2010 IEEE VTC – Spring*, pp. 1–5.

[15] M. Satyanarayanan, "Pervasive computing: vision and challenges," *IEEE Personal Commun.*, vol. 8, no. 4, pp. 10–17, Aug 2001.

[16] J. Konorski, "Multihomed wireless terminals: MAC configuration and network selection games," in *Proc. 2011 Intl. Conf. Inf. Netw.*, pp. 224–229.

[17] F. Xu, C. Tan, Q. Li, G. Yan, and J. Wu, "Designing a practical access point association protocol," in *Proc. 2010 IEEE INFOCOM*.

[18] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.

[19] B. Wang, Y. Wu, and K. J. R. Liu, "Game theory for cognitive radio networks: an overview," *Comput. Netw.*, vol. 54, no. 14, pp. 2537–2561, Oct. 2010.

[20] Z. Han and K. J. R. Liu, "Noncooperative power-control game and throughput game over wireless networks," *IEEE Trans. Commun.*, vol. 53, no. 10, pp. 1625–1629, Oct. 2005.

[21] B. Wang, Z. Han, and K. J. R. Liu, "Distributed relay selection and power control for multiuser cooperative communication networks using Stackelberg game," *IEEE Trans. Mobile Comput.*, vol. 8, no. 7, pp. 975–990, July 2009.

[22] K. J. R. Liu and B. Wang, *Cognitive Radio Networking and Security: A Game Theoretical View*. Cambridge University Press, 2010.

[23] Y. Chen and K. J. R. Liu, "Indirect reciprocity game modelling for cooperation stimulation in cognitive networks," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 159–168, Jan. 2011.

[24] L. Xiao, Y. Chen, W. Lin, and K. J. R. Liu, "Indirect reciprocity security game for large-scale wireless networks," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 4, pp. 1368–1380, Aug. 2012.

[25] C. Tekin, M. Liu, R. Southwell, J. Huang, and S. Ahmad, "Atomic congestion games on graphs and their applications in networking," *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, pp. 1541–1552, Oct. 2012.

[26] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in cognitive radio networks: global optimization using local interaction games," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 2, pp. 180–194, Apr. 2012.

[27] ——, "Opportunistic spectrum access in unknown dynamic environment: a game-theoretic stochastic learning solution," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1380–1391, Apr. 2012.

[28] Y. Xu, A. Anpalagan, Q. Wu, L. Shen, Z. Gao, and J. Wang, "Decision-theoretic distributed channel selection for opportunistic spectrum access: strategies, challenges and solutions," *IEEE Commun. Surveys Tuts.*, 2013.

[29] C.-Y. Wang, Y. Chen, and K. J. R. Liu, "Chinese restaurant game," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 898–901, Dec. 2012.

[30] ——, "Sequential Chinese restaurant game," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 571–584, Feb. 2013.

[31] C. Jiang, Y. Chen, Y.-H. Yang, C.-Y.Wang, and K. J. R. Liu, "Dynamic Chinese restaurant game in cognitive radio networks," in *Proc. 2013 IEEE INFOCOM*.

[32] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 1994.

[33] D. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, 2007.

[34] R. B. Myerson, "Optimal auction design," *Mathematics Operations Research*, vol. 6, no. 1, pp. 58–73, 1981.

[35] V. Krishna, *Auction Theory*. Academic Press/Elsevier, 2009.

[36] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, *Algorithmic Game Theory*. Cambridge University Press, 2007.

[37] E. J. Friedman and D. C. Parkes, "Pricing WiFi at starbucks: issues in online mechanism design," in *Proc. 2003 ACM Conf. Electron. Commerce*, pp. 240–241.

[38] R. Vohra, *Mechanism Design: A Linear Programming Approach*. Cambridge University Press, 2011.

[39] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Trans. Veh. Technol.*, vol. 41, no. 1, pp. 57–62, 1992.

[40] T.-H. Lee, J.-C. Lin, and Y. Su, "Downlink power control algorithms for cellular radio systems," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 89–94, 1995.

[41] E. Parzen, *Stochastic Processes*. Holden-Day, 1962.

[42] M. Lopez-Benitez and F. Casadevall, "Discrete-time spectrum occupancy model based on Markov chain and duty cycle models," in *Proc. 2011 IEEE DySPAN*, pp. 90–99.

[43] J.-L. Wang, Y.-H. Xu, Z. Gao, and Q.-H. Wu, "Discrete-time queuing analysis of opportunistic spectrum access: single user case," *Frequenz*, vol. 65, no. 11-12, pp. 335–341, 2011.

[44] J. Razavilar, K. J. R. Liu, and S. Marcus, "Jointly optimized bit-rate/delay control policy for wireless packet networks with fading channels," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 484–494, Mar. 2002.

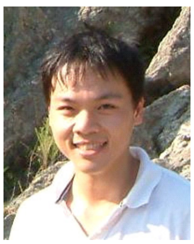
[45] G. Koole, "Monotonicity in Markov reward and decision chains: theory and applications," *Found. Trends. Stoch. Syst.*, pp. 1–76, 2007.

[46] E. Cil, E. Ormeci, and F. Karaesmen, "Effects of system parameters on the optimal policy structure in a class of queueing control problems," *Queueing Syst.*, vol. 61, pp. 273–304, 2009.

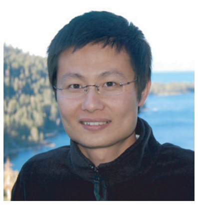
[47] A. Sang, X. Wang, M. Madihian, and R. Gitlin, "Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems," *Wireless Netw.*, vol. 14, pp. 103–120, 2008.

[48] D. Amzallag, R. Bar-Yehuda, D. Raz, and G. Scalosub, "Cell selection in 4G cellular networks," in *Proc. 2008 IEEE INFOCOM*, pp. 700–708.

[49] L. Gao, X. Wang, G. Sun, and Y. Xu, "A game approach for cell selection and resource allocation in heterogeneous wireless networks," in *Proc. 2011 IEEE SECON*, pp. 530–538.



Yu-Han Yang (S'06) received his B.S. in electrical engineering in 2004, and two M.S. degrees in computer science and communication engineering in 2007, from National Taiwan University, Taipei, Taiwan. He is currently pursuing the Ph.D. degree in University of Maryland, College Park, USA. His research interests include wireless communication and signal processing. He received Class A Scholarship from the ECE department, National Taiwan University in Fall 2005 and Spring 2006. He is a recipient of Study Abroad Scholarship from Taiwan (R.O.C.) government in 2009-2010.



Yan Chen (S'06-M'11) received the Bachelor's degree from University of Science and Technology of China in 2004, the M. Phil degree from Hong Kong University of Science and Technology (HKUST) in 2007, and the Ph.D. degree from University of Maryland College Park in 2011. From 2011 to 2013, he is a Postdoctoral research associate in the Department of Electrical and Computer Engineering at University of Maryland College Park.

Currently, he is a Principal Technologist at Origin Wireless Communications. He is also affiliated with Signal and Information Group of University of Maryland College Park. His current research interests are in social learning and networking, behavior analysis and mechanism design for network systems, multimedia signal processing and communication.

Dr. Chen received the University of Maryland Future Faculty Fellowship in 2010, Chinese Government Award for outstanding students abroad in 2011, University of Maryland ECE Distinguished Dissertation Fellowship Honorable Mention in 2011, and was the Finalist of A. James Clark School of Engineering Deans Doctoral Research Award in 2011.

Chunxiao Jiang (S'09-M'13) received his B.S. degree in information engineering from Beijing University of Aeronautics and Astronautics (Beihang University) in 2008 and the Ph.D. degree from Tsinghua University (THU), Beijing in 2013, both with the highest honors. During 2011-2012, he visited the Signals and Information Group (SIG) at Department of Electrical & Computer Engineering (ECE) of University of Maryland (UMD), supported by China Scholarship Council (CSC) for one year. Dr. Jiang is currently a research associate in ECE department of UMD with Prof. K. J. Ray Liu, and also a post-doctor in EE department of THU. His research interests include the applications of game theory and queuing theory in wireless communication and networking and social networks.

Dr. Jiang received the Beijing Distinguished Graduated Student Award, Chinese National Fellowship and Tsinghua Outstanding Distinguished Doctoral Dissertation in 2013.

Chih-Yu Wang (S'07-M'13) received the B.S. degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan. in 2007. He has been a visiting student in University of Maryland, College Park in 2011. He received the Ph.D. degree in the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan, in 2012. His research interests mainly are applications of game theory in wireless networking and social networking.

K. J. Ray Liu (F'03) was named a Distinguished Scholar-Teacher of University of Maryland, College Park, in 2007, where he is Christine Kim Eminent Professor of Information Technology. He leads the Maryland Signals and Information Group conducting research encompassing broad areas of signal processing and communications with recent focus on cooperative and cognitive communications, social learning and network science, information forensics and security, and green information and communications technology.

Dr. Liu is the recipient of numerous honors and awards including IEEE Signal Processing Society Technical Achievement Award and Distinguished Lecturer. He also received various teaching and research recognitions from University of Maryland including university-level Invention of the Year Award; and Poole and Kent Senior Faculty Teaching Award, Outstanding Faculty Research Award, and Outstanding Faculty Service Award, all from A. James Clark School of Engineering. An ISI Highly Cited Author, Dr. Liu is a Fellow of IEEE and AAAS.

Dr. Liu is President of IEEE Signal Processing Society where he has served as Vice President — Publications and Board of Governor. He was the Editor-in-Chief of *IEEE Signal Processing Magazine* and the founding Editor-in-Chief of *EURASIP Journal on Advances in Signal Processing*.