Congestion-Aware Distributed Network Selection for Integrated Cellular and Wi-Fi Networks

Man Hon Cheung, Fen Hou, *Member, IEEE*, Jianwei Huang, *Fellow, IEEE*, and Richard Southwell

Abstract— Intelligent network selection plays an important role in achieving an effective data offloading in the integrated cellular and Wi-Fi networks. However, previously proposed network selection schemes mainly focused on offloading as much data traffic to Wi-Fi as possible, without systematically considering the *Wi-Fi network congestion* and the *ping-pong effect*, both of which may lead to a poor overall user quality of experience. Thus, in this paper, we study a more practical network selection problem by considering both the impacts of the *network congestion* and *switching penalties*. More specifically, we formulate the users' interactions as a Bayesian network selection game (NSG) under the incomplete information of the users' mobilities. We prove that it is a Bayesian potential game and show the existence of a pure Bayesian Nash equilibrium that can be easily reached. We then propose a *distributed* network selection (DNS) algorithm based on the network congestion statistics obtained from the operator. Furthermore, we show that computing the optimal *centralized* network allocation is an NP-hard problem, which further justifies our distributed approach. Simulation results show that the DNS algorithm achieves the highest user utility and a good fairness among users, as compared with the on-the-spot offloading and cellular-only benchmark schemes.

Index Terms—Mobile data offloading, cellular and Wi-Fi integration, Bayesian potential game, network selection.

I. INTRODUCTION

With the proliferation of mobile video and web applications, consumer demands for wireless data services are growing rapidly, to a point that the cellular network capacity is pushed to its limit. According to Cisco's forecast, mobile data traffic will increase to 30.6 exabytes per month by 2020, which corresponds to an 8-fold increase between 2015 and 2020 globally [2]. The cellular network capacity, however, is growing at a much slower pace, and cannot keep up with the explosive growth in data traffic. A cost-effective and timely solution for alleviating the cellular network congestion problem is to use some complementary technologies, such as Wi-Fi or small cells, to offload the cellular traffic. In fact, Cisco predicted that the percentage of offloaded traffic will grow and exceed that of the cellular traffic, reaching an offloading ratio of 55% of the global mobile traffic by 2020 [2]. Owing to the existing popularity of Wi-Fi usage and deployment[1], we will focus on the network selection in the integrated cellular and Wi-Fi networks in this paper.

Through the current ongoing standardization efforts, such as the access network discovery and selection function (ANDSF) and Hotspot 2.0 [3], [4], the cellular and Wi-Fi networks are becoming more tightly coupled with each other. More specifically, under this cellular and Wi-Fi integration, the Wi-Fi networks would usually be owned and managed by the cellular operator, who ensures a seamless connectivity for the users. Also, the same operator will be responsible for making all the network selections silently in the background, so a user does not need to know whether he is connected to the cellular or Wi-Fi network. Furthermore, the operator will ensure that all functionality and services are consistently available regardless of whether the user is on cellular or Wi-Fi.

Intelligent network selection plays a critical role in the integrated cellular and Wi-Fi networks, to achieve an effective mobile data offloading and improve the users' quality of experience (QoE). One popular choice that is used by many smartphones by default is the on-the-spot offloading (OTSO) scheme, where the device simply offloads its data traffic to a Wi-Fi network whenever possible, and only uses the cellular network if no Wi-Fi exists (or the Wi-Fi interface is turned off). The OTSO scheme is simple to implement but has two possible drawbacks. First, under the OTSO policy, devices that are in close proximity may choose the same Wi-Fi network, hence experience the *network congestion* and achieve low throughput, especially during the peak hours in some densely populated areas. In other words, this users' *herd behaviour* without any *coordination* leads to the Wi-Fi network congestion [5]. Second, a user may incur a *switching penalty* in the forms of switching time and a switching cost when it switches between different networks. The switching time corresponds to the delay during handoff, and the switching cost accounts for the additional power consumption and QoE disruption [6]. Without taking into account this switching penalty, a network selection policy may result in the *ping-pong effect* [4] with too frequent network switching, which leads to a throughput reduction and faster battery degradation.

Although *network congestion* and *switching penalty* are two important factors in the design of an effective *intelligent*

M. H. Cheung is affiliated with both the University of Macau and the Chinese University of Hong Kong; E-mail: mhcheung@ie.cuhk.edu.hk. F. Hou is with the Department of Electrical and Computer Engineering, University of Macau, Macau; E-mail: fenhou@umac.mo. J. Huang is with the Network Communications and Economics Lab (NCEL), Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong, China; E-mail: jwhuang@ie.cuhk.edu.hk. R. Southwell is with the Department of Mathematics, University of York, York, United Kingdom; E-mail: richard.southwell@york.ac.uk.

[1]From Cisco's data, 64.2 million public Wi-Fi hotspots have already installed since 2015 [2].

network selection algorithm, most prior related literature, including [7]–[11], neglected the effects of these two factors. Balasubramanian *et al.* in [7] proposed that a user can perform data offloading by making predictions of future Wi-Fi availability using the past mobility history. Lee *et al.* in [8] described the on-the-spot offloading (OTSO) scheme that most smartphones are using today by default. Ristanovic *et al.* in [9] considered an energy-efficient offloading for delay-tolerant applications. They proposed to extract typical users' mobility profiles for the prediction of Wi-Fi availabilities. Im *et al.* in [10] considered the cost-throughput-delay tradeoff in user-initiated Wi-Fi offloading. Given the predicted future usage and the availability of Wi-Fi, the proposed system decides which application should offload its traffic to Wi-Fi at a given time, while taking into account the cellular budget constraint of the user. Moon *et al.* in [11] implemented a new transport layer to handle network disruption and delay for the development of delay-tolerant Wi-Fi offloading apps, by scheduling multiple flows to meet their deadlines with the maximal Wi-Fi usage. Furthermore, although the studies in [12]–[15] took the network congestion into account, they did not consider the effect of the switching penalty. Aryafar *et al.* in [12] studied the network selection dynamics in heterogeneous wireless networks under two classes of throughput models. They characterized the Pareto-efficiency of the equilibria and proposed a network selection algorithm with hysteresis mechanism. Following the work in [12], Monsef *et al.* in [13] first considered a client-centric network selection model for autonomous user decision, and characterized the convergence time and conditions. They further studied a hybrid client-network model, where a user is allowed to switch network if this decision is in line with the network controller's potential function. Mahindra *et al.* in [14] considered the practical implementation of the intelligent network selection in LTE and Wi-Fi networks. The system consists of an interface assignment algorithm that dynamically assigns user flows to interfaces and an interface switching service that performs seamless interface switching for HTTP-based flows. Hu *et al.* in [15] proposed an adaptive network selection algorithm based on the attractor selection mechanism for the users to dynamically select the suitable access points. Both the offloading effectiveness and traffic delay were considered as the performance metrics. In summary, the network selection problem considering the network congestion and switching penalty in data offloading has not been explored in the literature.

In this paper, we jointly consider both the network congestion and switching penalty and address the practical considerations of user mobility, and location, user, and time dependent Wi-Fi availabilities. For the user mobility, we assume that the operator only has the statistical information about the users' mobility patterns, which capture their daily movement habits [16]. We also consider several general assumptions on the Wi-Fi availabilities. First, we assume that the Wi-Fi availability is *location-dependent*, because Wi-Fi access points (APs) are only available at some limited locations due to their smaller coverages. Second, it may be *time-dependent* due to the access policies of the administrators of the Wi-Fi APs. For example, some Wi-Fi APs may be configured in the open access mode

when the owner is away, but in the closed access mode when the owner is back. Third, it may be *user-dependent*, as users who have subscribed to different data plans or Wi-Fi services (e.g., Skype Wi-Fi) can have different privileges to access different Wi-Fi networks. Given these practical considerations with heterogeneous users and networks, the network selection problem is very challenging to tackle.

Due to the coupling of the users' decisions in causing the network congestion, we apply the non-cooperative game theory to study this congestion-aware network selection problem. More specifically, with the statistical information on users' mobility patterns, we formulate the users' network selections over a period of multiple time slots as a *Bayesian game* [17]. In general, it is difficult to characterize the existence and convergence of the Bayesian Nash equilibrium. Nevertheless, we are able to show that the formulated game is a *Bayesian potential game* [18], which enables us to design a distributed network selection (DNS) with some nice convergence properties. It should be noted that convergence is important for congestion-aware network selections, where users switch networks based on the experienced network congestion levels. Without convergence guarantees, the system may result in oscillations. In addition, as a benchmark, we show that computing the socially optimal solution that maximizes the users' aggregate utilities in a centralized setting is an NP-hard problem.

In summary, the main contributions of our work are as follows:

- *Practical modeling*: We study the users' network selection problem by taking into account the practical issues of network congestion, switching penalties, and statistical information of the users' various possible mobility patterns.
- *NP-hard centralized network allocation benchmark*: We show that maximizing the users' aggregate utilities is an NP-hard problem, which motivates us to consider the distributed setting.
- *Distributed network selection algorithm*: We formulate the users' network selection interactions as a Bayesian game. We show that it is a potential game, derive its closed-form exact potential function, and propose a practical DNS algorithm with nice convergence properties.
- *Load balancing*: Simulation results show that the proposed DNS scheme achieves a good fairness and improves the user utility of the cellular-only and OTSO schemes by 66.7% under a medium switching cost. We also show that the OTSO scheme performs reasonably well with a low switching cost and a low Wi-Fi availability.

The rest of the paper is organized as follows. We first describe the system model in Section II. We study the centralized network allocation and the distributed network selection game in Sections III and IV, respectively. We present the simulation results in Section V and conclude the paper in Section VI.

II. SYSTEM MODEL

In this section, we discuss the system model for the network selection in the integrated cellular and Wi-Fi networks. More
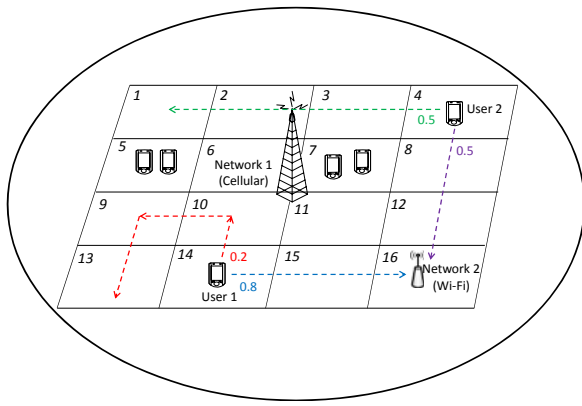
Fig. 1. An example of the integrated cellular and Wi-Fi network, where $\mathcal{N} = \{0, 1, 2\}$ is the set of networks and $\mathcal{L} = \{1, \ldots, 16\}$ is the set of locations. In each time slot, a user can remain idle (i.e., choose the auxiliary network 0), access the cellular network (i.e., network 1) or the Wi-Fi network (i.e., network 2) if available, as the Wi-Fi availability is user, location, and time dependent. Each user may have *multiple* possible mobility patterns, where the operator has *incomplete* information on their probability distributions. We consider the users' network selections across a time period of multiple time slots by taking into account the *user mobility*, *network congestion*, and *switching penalties*.

specifically, we describe the networks setting in Section II-A and a user's network availability and mobility pattern in Section II-B. We present his action as a network-time routes in Section II-C and his utility function in Section II-D.

A. Network Setting

As shown in Fig. 1, we consider an integrated cellular and Wi-Fi system, where the Wi-Fi networks are tightly integrated with the cellular network in terms of the radio frequency coordination and network management [19]. Let $\mathcal{N} = \{0, 1, \ldots, N\}$ be the set of $N + 1$ networks, where network $n = 1$ corresponds to the cellular network and network $n \in \mathcal{N}_{\text{wifi}} = \{2, \ldots, N\}$ corresponds to a Wi-Fi network. We introduce an auxiliary *idle network* $n = 0$ to model the situation that the user chooses to remain idle and is not actively using any networks. The network parameters are described as follows.

Definition 1 (Network Parameters): Each network $n \in \mathcal{N}$ is associated with:

- *Network capacity* $\mu[n]$: The maximum total amount of data rate that network n can serve the users in each time slot.
- *Switching cost* $c[n, n']$: The cost incurred by a user when he switches from network $n \in \mathcal{N}$ to network $n' \in \mathcal{N}$. It can account for additional power consumption and QoE disruption [6] during network switching.
- *Switching time* $\delta[n, n']$: The delay incurred by a user when he switches from network $n \in \mathcal{N}$ to network $n' \in \mathcal{N}$, which is the total number of time slots required to tear down the old connection of network n and setup the new connection of network n'. It corresponds to the delay during handoff between different wireless networks.

To account for the fact that there is no network switching when a user keeps using network $n \in \mathcal{N}$, we have $c[n, n] = 0$

and $\delta[n, n] = 0$, $\forall n \in \mathcal{N}$. For the idle network, we make the following additional practical assumptions:

Assumption 1 (Idle network): (a) $\mu[0] = 0$; (b) The switching time through the idle network satisfies

$$\delta[n, n'] + \delta[n', n''] \geq \delta[n, 0] + \delta[0, n''], \ \forall n, n', n'' \in \mathcal{N} \backslash \{0\}. \tag{1}$$

(c) The switching cost through the idle network satisfies

$$c[n, n'] + c[n', n''] > c[n, 0] + c[0, n''], \ \forall n, n', n'' \in \mathcal{N} \backslash \{0\}. \tag{2}$$

Assumption 1(a) implies that a user cannot receive any data during the idle state. Assumption 1(b) accounts for the fact that an idle state requires less time to "setup" or "tear down" than switching through a third network n'. Assumption 1(c) captures the additional power and signalling overhead during handovers that involves one more network n'.[2]

B. User Setting

Let $\mathcal{I} = \{1, \ldots, I\}$ be the set of users, $\mathcal{L} = \{1, \ldots, L\}$ be the set of locations, and $\mathcal{T} = \{1, \ldots, T\}$ be the set of time slots. We define a user's network availability[3] and mobility pattern as follows.

Definition 2 (User's Network Availability and Mobility Pattern): A user $i \in \mathcal{I}$ is associated with:

- *User, location, and time dependent network availabilities*[4] $\mathcal{M}[i, l, t] \subseteq \mathcal{N}$: The set of networks available for user $i \in \mathcal{I}$ at location $l \in \mathcal{L}$ and time $t \in \mathcal{T}$.[5]
- *Mobility pattern* $\boldsymbol{\theta}_i = (l[i, t] \in \mathcal{L}, \forall t \in \mathcal{T}) \in \boldsymbol{\Theta}_i$: The locations of user i in the period of T time slots due to his mobility, where $l[i, t]$ is the position of user i at time t, and $\boldsymbol{\Theta}_i$ is the set of all possible mobility patterns of user i given his initial location at time $t = 1$. Each user may have *multiple* mobility patterns. As an example, in Fig. 1, for a total of $T = 4$ slots, user i has two possible mobility patterns: $\boldsymbol{\theta}_1^{(1)} = (14, 15, 16, 16)$ and $\boldsymbol{\theta}_1^{(2)} = (14, 10, 9, 13)$, so $\boldsymbol{\Theta}_1 = \{\boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_1^{(2)}\}$. We also refer to $\boldsymbol{\theta}_i$ as the *type* of user i.
- *Prior Probability on Mobility Pattern* $p(\boldsymbol{\theta}_i) \geq 0$: The probability that user i chooses mobility pattern $\boldsymbol{\theta}_i$.[6] We have $\sum_{\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i} p(\boldsymbol{\theta}_i) = 1$. As an example in Fig. 1, we have $p(\boldsymbol{\theta}_1^{(1)}) = 0.8$ and $p(\boldsymbol{\theta}_1^{(2)}) = 0.2$.

---

[2]It is possible to use strict inequalities in both (1) and (2). However, since the switching time is an integer in this paper, it is more practical to consider an inequality in (1).

[3]Our modeling on network availability is quite general as it allows each location to have more than one Wi-Fi access points (APs) and each AP to cover more than one location. Thus, there can be overlapping coverage areas of different networks. Also, it is a straightforward extension to consider multiple cellular networks (e.g., deployed by different mobile operators), and assume that each location can be covered by an arbitrary number of these networks.

[4]If a network's quality frequently falls below a tolerable threshold due to fading and shadowing, then the users can simply assume that the network is not available in practice.

[5]For the rest of the paper, we will assume that the idle network is available for all the users at all possible locations and time slots, so that network $0 \in \mathcal{M}[i, l, t]$, $\forall i \in \mathcal{I}, l \in \mathcal{L}, t \in \mathcal{T}$.

[6]Thus, $p(\boldsymbol{\theta}_i)$ is a system parameter on user i's mobility, which can be collected from the mobile device automatically in the background.

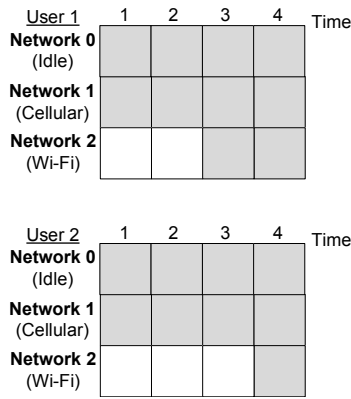Fig. 2. Heterogeneous network availabilities of user 1 and user 2 shown in Fig. 1 under mobility patterns $\boldsymbol{\theta}_1 = (14, 15, 16, 16)$ and $\boldsymbol{\theta}_2 = (4, 8, 12, 16)$. Here, a grey block means that the network is available (i.e., network $n \in \mathcal{N}[i, t]$) and a white block means that the network is not available.

We refer to this general mobility setting as the *random mobility pattern case*. It includes the special case of *deterministic* mobility pattern, where the user knows his own mobility pattern accurately.

Given user i's mobility pattern and the network availabilities, we can compute his available set of networks at time t as[7]

$$\mathcal{N}[i, t] = \mathcal{M}\big[i, l[i, t], t\big]. \tag{3}$$

An example of $\mathcal{N}[i, t]$ is given in Fig. 2.

C. Network-Time Route as Action

After describing a user's network availability and mobility, we define his *action* as his network selections across multiple time slots, which is referred to as *network-time route* define below. Let $\tilde{\mathcal{R}}_i$ be the set of all possible network-time routes of user i. Given user i's mobility pattern $\boldsymbol{\theta}_i$, we let $\mathcal{R}_i(\boldsymbol{\theta}_i) \subseteq \tilde{\mathcal{R}}_i$ be the set of all *feasible* network-time route of user i define as follows.

Definition 3 (Feasible Network-Time Route): Given user i's mobility pattern $\boldsymbol{\theta}_i$, his *feasible network-time route* is a sequence

$$\boldsymbol{r}_i = \Big((n_i^1, t_i^1), (n_i^2, t_i^2), \ldots, (n_i^{Q_i}, t_i^{Q_i})\Big) \in \tilde{\mathcal{R}}_i, \tag{4}$$

which indicates user i's network selections in all time slots, except those time slots when user i is in the middle of network switching and is not associated with any network. It satisfies the following conditions:

1) Causality: $1 = t_i^1 < t_i^2 < \ldots < t_i^{Q_i} \leq T$.
2) Eligibility: $n_i^q \in \mathcal{N}[i, t_i^q]$, for each $q \in \{1, \ldots, Q_i\}$.
3) Switching time: $t_i^{q+1} - t_i^q = \delta[n_i^q, n_i^{q+1}] + 1$, for each $q \in \{1, \ldots, Q_i - 1\}$.

Condition 1) accounts for the fact that time is always increasing. Condition 2) ensures that user i is eligible to select the networks according to their availabilities as defined in (3). Condition 3) ensures that the time difference between successive elements in the sequence of network-time route is

[7]As a result, we will focus on $\mathcal{N}[i, t]$, instead of $\mathcal{M}[i, l, t]$, for the rest of the paper.
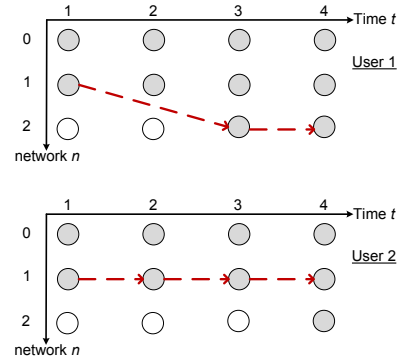

Fig. 3. The network-time routes chosen by the two users when switching time $\delta[1, 2] = 1$.

consistent with the switching time between the corresponding networks. More specifically, when $n_i^q = n_i^{q+1}$, it means that user i keeps using the same network at the next time slot. Without involving any network switching, we have $t_i^{q+1} - t_i^q = 1$ since $\delta[n, n] = 0, \forall n \in \mathcal{N}$. When $n_i^q \neq n_i^{q+1}$, it means that user i switches from network n_i^q to n_i^{q+1}. Therefore, user i can use network n_i^{q+1} after finishing the switching process, which takes switching time of $\delta[n_i^q, n_i^{q+1}]$.

To facilitate the introduction of the user's utility function in the next subsection, we define the network-time points of a feasible network-time route as the network-time selections along it.

Definition 4 (Network-time points): Given a feasible route $\boldsymbol{r}_i \in \mathcal{R}_i(\boldsymbol{\theta}_i)$ in (4), we define its network-time points as the set

$$\mathcal{V}(\boldsymbol{r}_i) = \Big\{(n_i^1, t_i^1), (n_i^2, t_i^2), \ldots, (n_i^{Q_i}, t_i^{Q_i})\Big\}. \tag{5}$$

The set can also be represented as the $Q_i - 1$ network-time point pairs

$$\mathcal{E}(\boldsymbol{r}_i) = \Big\{\big((n_i^q, t_i^q), (n_i^{q+1}, t_i^{q+1})\big) : q = 1, \ldots, Q_i - 1\Big\}, \tag{6}$$

which are the consecutive pairs of network-time points visited by user i in route $\boldsymbol{r}_i$.

An example of the feasible network-time routes is shown in Fig. 3. In this example, we have $\boldsymbol{r}_1 = \big((1, 1), (2, 3), (2, 4)\big)$, so $\mathcal{V}(\boldsymbol{r}_1) = \{(1, 1), (2, 3), (2, 4)\}$ and $\mathcal{E}(\boldsymbol{r}_1) = \Big\{\big((1, 1), (2, 3)\big), \big((2, 3), (2, 4)\big)\Big\}$. The pair of network-time points $\big((1, 1), (2, 3)\big)$ means that user 1 accesses network 1 at time slot 1, and switches to network 2 at time slot 3 after taking one time slot of switching time. The pair of network-time points $\big((2, 3), (2, 4)\big)$ denotes that user 1 accesses network 2 at time slot 3, and keeps using network 2 at time slot 4. The corresponding network selections of users at 1 and 2 at different locations and time slots are illustrated in Fig. 4.

D. Utility Function

For the design of a practical congestion-aware network selection algorithm, a user's utility function should take both the network congestion and the negative impact of ping-pong effect into account. Let $\omega[(n, t), \boldsymbol{r}, \boldsymbol{\theta}] = |j \in \mathcal{I} : (n, t) \in$
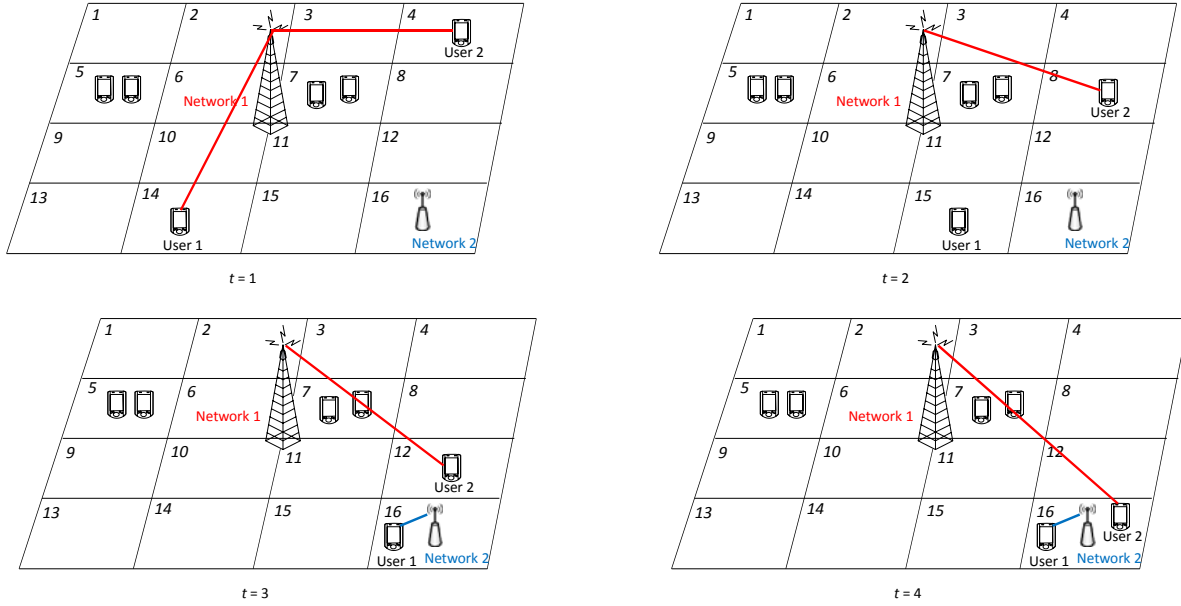
Fig. 4. The network selections of users 1 and 2 at different locations in the four time slots.

$\mathcal{V}(\boldsymbol{r}_j), \boldsymbol{r}_j \in \mathcal{R}_j(\boldsymbol{\theta}_j)|$ be the *network congestion level*, which counts the number of network-time routes chosen in action profile $\boldsymbol{r} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_I)$ that pass through the network-time point (n, t) under mobility patterns $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I)$. In other words, it denotes the total number of users accessing network n at time t. For a route $\boldsymbol{r}_i \in \mathcal{R}_i(\boldsymbol{\theta}_i)$ in (4), user i's utility function consists of two parts:

- Total throughput: The summation of user i's achieved throughput over all the network-time points along route $\boldsymbol{r}_i$ (i.e., set $\mathcal{V}(\boldsymbol{r}_i)$ in (5)). At each network-time point (n, t), the achieved throughput of user i is the network capacity $\mu[n]$ divided by the network congestion level $\omega[(n, t), \boldsymbol{r}, \boldsymbol{\theta}]$.
- Total switching cost: The summation of the switching costs for network switching of every network-time point pairs in route $\boldsymbol{r}_i$ (i.e., set $\mathcal{E}(\boldsymbol{r}_i)$ in (6)). More specifically, for each network-time point pair $\boldsymbol{e} = \left((n_i^q, t_i^q), (n_i^{q+1}, t_i^{q+1})\right) \in \mathcal{E}(\boldsymbol{r}_i)$, we define its switching cost as

$$g[\boldsymbol{e}] = g\left[(n_i^q, t_i^q), (n_i^{q+1}, t_i^{q+1})\right] = c[n_i^q, n_i^{q+1}]. \quad (7)$$

Overall, given the action profile $\boldsymbol{r}$ and mobility patterns $\boldsymbol{\theta}$ of all users, if $\boldsymbol{r}_i \in \mathcal{R}_i(\boldsymbol{\theta}_i)$, user i's *utility function* is

$$U_i(\boldsymbol{r}, \boldsymbol{\theta}) = \sum_{(n,t) \in \mathcal{V}(\boldsymbol{r}_i)} \frac{\mu[n]}{\omega[(n, t), \boldsymbol{r}, \boldsymbol{\theta}]} - \sum_{\boldsymbol{e} \in \mathcal{E}(\boldsymbol{r}_i)} g[\boldsymbol{e}]. \quad (8)$$

III. Centralized Network Allocation

One natural formulation is to consider the centralized network allocation that maximizes the users' aggregate utilities. However, we show that this would be an NP-hard problem even in the special case of deterministic mobility pattern,

where the users' mobility patterns are known.[8]

We first formally define a centralized socially optimal network allocation as follows.

Definition 5 (Socially Optimal Network Allocation): Given the deterministic mobility patterns $\boldsymbol{\theta}$, an action profile $\boldsymbol{r}^*$ is a socially optimal network allocation if it maximizes the social welfare:

$$\boldsymbol{r}^* = \underset{\boldsymbol{r}_i \in \mathcal{R}_i(\boldsymbol{\theta}_i), \forall i \in \mathcal{I}}{\arg\max} \ \mathbb{W}(\boldsymbol{r}, \boldsymbol{\theta}) \triangleq \sum_{i \in \mathcal{I}} U_i(\boldsymbol{r}, \boldsymbol{\theta}), \quad (9)$$

where the social welfare $\mathbb{W}(\boldsymbol{r}, \boldsymbol{\theta})$ is defined as the users' aggregate utility.

With Assumption 1 in Section II-A, we can show that there always exists a socially optimal network allocation that each network-time point is chosen by at most one user.

Lemma 1: Under Assumption 1 and the given deterministic mobility patterns $\boldsymbol{\theta}$, there always exists a socially optimal network allocation $\boldsymbol{r}^*$ such that the congestion level $\omega[(n, t), \boldsymbol{r}^*, \boldsymbol{\theta}] \leq 1$, for all network-time points $(n, t) \in \mathcal{N} \times \mathcal{T}$.

Due to space limitation, the proof of Lemma 1 is given in [20]. With Lemma 1, we can show that the problem of finding a socially optimal network allocation is NP-hard.

Theorem 1: The problem of finding a socially optimal network allocation in (9) is NP-hard.

The proof of Theorem 1 is given in [20]. Thus, solving the centralized network allocation problem is *infeasible* for practical system with a potentially large number of users, networks, and mobility patterns. Moreover, it is more practical to study the scenario that the users are autonomously selecting

---

[8]It should be noted that in the general case with random mobility patterns, we will consider the maximization of the users' aggregate expected utilities, which involves solving a number of network allocation problems under the deterministic mobility pattern in the form of problem (9). In other words, if problem (9) is NP-hard, then the problem under the random mobility case is also NP-hard.

the networks themselves, rather than the operator controlling their network choices. This motivates us to formulate the *distributed* network selection problem as a non-cooperative game, as we discuss next.

IV. DISTRIBUTED NETWORK SELECTION GAME

In this section, we formulate the users' network selections with incomplete information as a distributed network selection game (NSG). We first describe the non-cooperative game formulation in Section IV-A. We then show that it is a Bayesian potential game with the finite improvement property and derive its exact potential function in closed-form in Section IV-B. Finally, we proposed a distributed network selection (DNS) algorithm to coordinate the users' decisions in Section IV-C.

A. Network Selection Game with Incomplete Information

In practice, each user may only have *incomplete* information of the types (i.e., mobility patterns) of all the users (even including himself) at the beginning of the time period. More specifically, we assume that the users' types $\boldsymbol{\theta}$ follow a known *prior* probability distribution $p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I)$.[9] In addition, the utility functions, available actions, possible types, and the prior distributions of user types are assumed to be public information. We then formulate the network selection game as a *Bayesian game* [17] as follows.

Definition 6 (Network Selection Game): A network selection game is a tuple $\Omega = (\mathcal{I}, \tilde{\mathcal{R}}, \boldsymbol{\Theta}, p, \boldsymbol{U})$ defined by:

- Players: The set of users $\mathcal{I}$.
- Actions: The set of action profiles of all the users is $\tilde{\mathcal{R}} = \tilde{\mathcal{R}}_1 \times \ldots \times \tilde{\mathcal{R}}_I$.
- Types: The set of type space of all the users is $\boldsymbol{\Theta} = \boldsymbol{\Theta}_1 \times \ldots \times \boldsymbol{\Theta}_I$, where $\boldsymbol{\Theta}_i$ is the set of possible mobility patterns of user i.
- Prior information: The common prior probability $p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I)$ over the types of all users.
- Utilities: The vector $\boldsymbol{U} = (U_i, \forall i \in \mathcal{I})$ contains the utility functions of the users defined in (8).

Under this Bayesian game setting, user i's *strategy* is a mapping $s_i : \Theta_i \to \tilde{\mathcal{R}}_i$, which specifies user i's action for each possible type. In other words, $\boldsymbol{s}_i(\boldsymbol{\theta}_i)$ specifies the network-time route that user i should choose given his mobility pattern $\boldsymbol{\theta}_i$. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I) \in \boldsymbol{\Theta}$ be the users' type profile. So $\boldsymbol{s}(\boldsymbol{\theta}) = (\boldsymbol{s}_1(\boldsymbol{\theta}_1), \ldots, \boldsymbol{s}_I(\boldsymbol{\theta}_I))$ is the action profile of all the users given that their type profile is $\boldsymbol{\theta}$. The *expected utility* of user i under strategy profile $\boldsymbol{s}$ is

$$EU_i(\boldsymbol{s}) = \sum_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} U_i(\boldsymbol{s}(\boldsymbol{\theta}), \boldsymbol{\theta}) \, p(\boldsymbol{\theta}). \quad (10)$$

Let $\boldsymbol{s}_{-i} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{i-1}, \boldsymbol{s}_{i+1}, \ldots, \boldsymbol{s}_I)$ denote the strategies of all the users except user i. A strategy profile can be written as $\boldsymbol{s} = (\boldsymbol{s}_i, \boldsymbol{s}_{-i})$. Let $\mathcal{S}_i$ be the strategy space of user i and

let $\mathcal{S} = \mathcal{S}_1 \times \ldots \times \mathcal{S}_I$ be the strategy space of all the users. We define the Bayesian Nash equilibrium [17] as follows.

Definition 7 (Bayesian Nash equilibrium): The strategy profile $\boldsymbol{s}^*$ is a pure strategy Bayesian Nash equilibrium (BNE)[10] if

$$\boldsymbol{s}_i^* = \arg\max_{\boldsymbol{s}_i \in \mathcal{S}_i} EU_i(\boldsymbol{s}_i, \boldsymbol{s}_{-i}^*), \forall\, i \in \mathcal{I}. \quad (11)$$

B. Bayesian Potential Game

In general, it is difficult to establish the analytical results of the BNE. Nevertheless, in this subsection, we are able to show that Ω is a *Bayesian potential game* [18], which exhibits the finite improvement property. It thus implies the existence of and the convergence to the BNE.

First, we present the definition of a Bayesian potential game [18].

Definition 8 (Bayesian Potential Game): Bayesian Game Ω is a *Bayesian potential game* if there exists an exact potential function $\Phi(\boldsymbol{r}, \boldsymbol{\theta})$ such that

$$U_i(\boldsymbol{r}_i, \boldsymbol{r}_{-i}, \boldsymbol{\theta}) - U_i(\boldsymbol{r}_i', \boldsymbol{r}_{-i}, \boldsymbol{\theta}) = \Phi(\boldsymbol{r}_i, \boldsymbol{r}_{-i}, \boldsymbol{\theta}) - \Phi(\boldsymbol{r}_i', \boldsymbol{r}_{-i}, \boldsymbol{\theta}),$$
$$\forall\, \boldsymbol{r}_i, \boldsymbol{r}_i' \in \mathcal{R}_i(\boldsymbol{\theta}_i), \boldsymbol{\theta} \in \Theta, i \in \mathcal{I}.[11] \quad (12)$$

In other words, Bayesian potential game is a Bayesian game, where the change in the value of the utility function is equal to the change in the value of the potential function when the action profile changes.

Theorem 2: Game Ω is a Bayesian potential game with the exact potential function given by

$$\Phi(\boldsymbol{r}, \boldsymbol{\theta}) = \sum_{(n,t) \in \mathcal{N} \times \mathcal{T}} \sum_{q=1}^{\omega[(n,t), \boldsymbol{r}, \boldsymbol{\theta}]} \frac{\mu[n]}{q} - \sum_{i \in \mathcal{I}} \sum_{\boldsymbol{e} \in \mathcal{E}(\boldsymbol{r}_i)} g[\boldsymbol{e}]. \quad (13)$$

The proof of Theorem 2 is given in Appendix A.

1) Properties of the Bayesian Potential Game: Before stating the convergence properties of the users' interactions, let us recall some definitions [22].

Definition 9 (Better and best response updates): Starting from a strategy profile $\boldsymbol{s} = (\boldsymbol{s}_i, \boldsymbol{s}_{-i})$, a *better response update* is an event where a single user i changes his strategy from $\boldsymbol{s}_i \in \mathcal{S}_i$ to $\boldsymbol{s}_i' \in \mathcal{S}_i$, and increases his expected utility as a result, i.e., $EU_i(\boldsymbol{s}_i', \boldsymbol{s}_{-i}) > EU_i(\boldsymbol{s}_i, \boldsymbol{s}_{-i})$.

A *best response update* is a special type of better response update, where the newly selected strategy $\boldsymbol{s}_i'$ maximizes user i's expected utility among user i's all possible better response updates.

Definition 10 (Finite improvement property): A Bayesian game possesses the *finite improvement property* (FIP) when asynchronous[12] better response updates always converge to a BNE (defined in Definition 7) within a finite number of

---

[9]It is possible that this prior information on mobility patterns can be obtained from the mobile operator. However, as we will discuss in Section IV-C, the actual implementation of the DNS algorithm (i.e., Algorithm 1) only requires the aggregate network usage statistics from the operator, instead of the detailed users' mobility information, so there are no privacy concerns in the proposed algorithm.

[10]Alternatively, we may define the BNE based on each user i's *ex-interim* expected utility $EU_i(\boldsymbol{s}_i, \boldsymbol{s}_{-i}^*|\boldsymbol{\theta}_i)$ [17], where user i knows his own type $\boldsymbol{\theta}_i$. In this case, the strategy profile $\boldsymbol{s}^*$ is a pure strategy BNE if $\boldsymbol{s}_i^* = \arg\max_{\boldsymbol{s}_i \in \mathcal{S}_i} EU_i(\boldsymbol{s}_i, \boldsymbol{s}_{-i}^*|\boldsymbol{\theta}_i), \forall\, \boldsymbol{\theta}_i \in \Theta_i, i \in \mathcal{I}$. However, Theorem I in [21] showed that such an alternative definition is equivalent to the BNE defined in (11), where user i knows nothing about any user's actual type (including his own type).

[12]Asynchronous updates imply that there will be no two users updating their strategies at the same time.

steps, irrespective of the initial strategy profile or the users' updating order.

The FIP implies that better response updating always leads to a pure strategy BNE, which implies the existence of pure strategy BNE. As a result, it ensures the efficient convergence of the users' network selection to an equilibrium point and thus the stability of the system.

Theorem 3: Game Ω possesses the finite improvement property.

The proof of Theorem 3 is given in Appendix B.

C. Algorithm Design

With the nice FIP just derived, we propose a distributed network selection (DNS) algorithm (i.e., Algorithm 1) for the users to make their network selection decisions autonomously. This algorithm relies on the network information obtained from the operator in Algorithm 2. Different from the Bayesian game setting discussed above that requires the prior information of the users' mobility, we design the DNS algorithm in a way that it only requires the users to report their network selection statistics. Thus, it eliminates the privacy concern of asking users to reveal their mobility information.

1) Algorithm 1: To initialize, a user inputs his destination on the mobile device (line 1). Based on his current location (e.g., obtained from global positioning system (GPS)) and mobility history, the device can compute the user's possible trajectories and the corresponding probabilities (line 2). The device then queries the subscription plan (line 3) and network parameters (line 4) on behalf of the users to determine the available network resources.

Next, we describe the best response update of user i:

- *Network information collection*: Let Γ_i be the set of time slots (line 6) that user i updates his strategy, which corresponds to a network-time route for each of his possible type $\boldsymbol{\theta}_i$ (line 7). Based on the user, location, and time availability of the networks, the device computes user i's available networks (line 8) and his set of feasible network-time routes (line 9).[13]

- *Best response computation*: Let $p_{-i}[q, (n, t)]$ be the probability that the network-time point (n, t) would be occupied by q users (i.e., the congestion level is q) given a set $\mathcal{I}\backslash\{i\}$ of users, where $q = 0, \ldots, I - 1$. With this information on the probability mass function (pmf) of congestion level in each network at different time from the operator in Algorithm 2 (line 10), each user performs a best response update to maximize his expected utility[14] $u_i(\boldsymbol{r}_i|\boldsymbol{\theta}_i)$ for choosing route $\boldsymbol{r}_i$ given that his type is $\boldsymbol{\theta}_i$ (line 11), where the first and second terms on the right hand side of (14) represent the expected throughput and switching cost for choosing route $\boldsymbol{r}_i$, respectively. Each best response update can be computed by applying a shortest path algorithm [23], such as Bellman-Ford

---

Algorithm 1 *Distributed network selection (DNS) algorithm for user $i \in \mathcal{I}$.*

Initialization
1: User's input: Destination.
2: Compute the set Θ_i of possible trajectories and the corresponding probability $p(\boldsymbol{\theta}_i) \geq 0$ for each $\boldsymbol{\theta}_i \in \Theta_i$ based on the user's current location and mobility history, such that $\sum_{\boldsymbol{\theta_i} \in \Theta_i} p(\boldsymbol{\theta}_i) = 1$.
3: Query user i's subscription plan $\mathcal{M}[i, l, t] \ \forall l \in \mathcal{L}, t \in \mathcal{T}$ from the operator's database.
4: Query network parameters from the operator's database: network capacity $\mu[n], \ \forall n \in \mathcal{N}$, switching time $\delta[n, n']$, and switching cost $c[n, n'], \ \forall n, n' \in \mathcal{N}, n \neq n'$.

Planning Phase: Bayesian Network Selection Game
5: **repeat**
6:   **if** $\tau \in \Gamma_i$
7:     **for** $\boldsymbol{\theta_i} \in \Theta_i$
8:       Compute network availabilities $\mathcal{N}[i, t] := \mathcal{M}[i, l[i, t], t]$, $\forall t \in \mathcal{T}$ for trajectory $\boldsymbol{\theta}_i = (l[i, t], \forall t \in \mathcal{T})$.
9:       Determine set $\mathcal{R}_i(\boldsymbol{\theta}_i)$ of feasible network-time routes from $\mathcal{N}[i, t]$ and switching time $\delta[n, n'], \forall n, n' \in \mathcal{N}$ by Definition 3.
10:      Obtain pmf $p_{-i}[q, (n, t)]$ for $q = 0, \ldots, I - 1$ for all $(n, t) \in \mathcal{N} \times \mathcal{T}$ from the operator (see Algorithm 2).
11:      Perform a best response update for type $\boldsymbol{\theta_i}$, by identifying a route $\boldsymbol{r}_i \in \mathcal{R}_i(\boldsymbol{\theta}_i)$ that maximizes user i's expected utility given that his type is $\boldsymbol{\theta_i}$:

$$u_i(\boldsymbol{r_i}|\boldsymbol{\theta_i}) := \sum_{(n,t) \in \mathcal{V}(\boldsymbol{r}_i)} \sum_{q=0}^{I-1} \left(\frac{\mu[n]}{q+1}\right) p_{-i}[q, (n, t)] - \sum_{\boldsymbol{e} \in \mathcal{E}(\boldsymbol{r}_i)} g[e]. \tag{14}$$

12:      Update the strategy $\boldsymbol{s}_i(\boldsymbol{\theta}_i) := \boldsymbol{r}_i$.
13:    **end for**
14:    Compute the network selection statistics under strategy $\boldsymbol{s}_i$:

$$p_i[n, t] := \sum_{\boldsymbol{\theta_i} \in \Theta_i} p(\boldsymbol{\theta}_i) \times \mathbf{1}_{(n,t) \in \mathcal{V}(\boldsymbol{s}(\boldsymbol{\theta_i}))}, \ \forall (n, t) \in \mathcal{N} \times \mathcal{T}, \tag{15}$$

where $\mathbf{1}_{\{.\}}$ is the indicator function.
15:    Report the network selection statistics $\boldsymbol{p}_i := (p_i[n, t], \ \forall (n, t) \in \mathcal{N} \times \mathcal{T})$ to the operator.
16: **end if**
17: **until** $\tau \geq \tau^{\max}$.

Network Selection Phase
18: User i determines his actual trajectory $\boldsymbol{\theta}_i$.
19: Select networks in different time slots based on action $\boldsymbol{s}_i(\boldsymbol{\theta_i})$.

---

algorithm, and can be performed in an asynchronous fashion by the users until reaching the iteration limit[15] of $\tau^{\max}$, which represents the maximum number of iterations that the algorithm will run (line 17).

- *Information exchange*: Each user needs to report his individual network selection statistics to the operator, so that the operator can calculate the aggregate network congestion statistics by Algorithm 2. Let $p_i[n, t]$ be the probability that user i would access network-time point (n, t) under strategy $\boldsymbol{s}_i$. After the strategy $\boldsymbol{s}_i$ is determined, the device computes $p_i[n, t]$ by summing the probability $p(\boldsymbol{\theta}_i)$ that network-time point (n, t) is chosen in the action $\boldsymbol{s}_i(\boldsymbol{\theta}_i)$ in (15) (line 14) and reports the net-

---

[13]The operator can obtain the Wi-Fi network information if the Wi-Fi APs are deployed by the operator. For other Wi-Fi APs, it is possible to obtain this information from other operators under roaming agreements.

[14]For notational simplicity, we define $u_i(\boldsymbol{r}_i|\boldsymbol{\theta}_i)$ as the expected utility in the algorithm, which is actually equal to $EU_i(\boldsymbol{s}_i, \boldsymbol{s}_{-i}|\boldsymbol{\theta}_i)$.

[15]According to FIP, the algorithm will always converge in a finite number of iterations. However, we add this iteration limit to ensure that the convergence does not take too long, thus allows the tradeoff between the performance and convergence time. In our simulation in Section V, we choose $\tau^{\max} = 20$.

Algorithm 2 *Information update algorithm for the mobile operator.*

    Initialization

1: Establish subscription plan database for users to retrieve: $\mathcal{M}[i,l,t] \ \forall i \in \mathcal{I}, l \in \mathcal{L}, t \in \mathcal{T}$.
2: Establish network parameter database for users to retrieve: Network capacity $\mu[n], \forall n \in \mathcal{N}$, switching time $\delta[n,n']$, and switching cost $c[n,n'], \forall n, n' \in \mathcal{N}, n \neq n'$.
3: Allocate memory for the users' network selection statistics $\boldsymbol{p}_i, \forall i \in \mathcal{I}$.
4: Synchronize the clock timer $\tau := 1$ with all the users.
    Information Update for Bayesian NSG in Algorithm 1
5: **repeat**
6:     **if** network congestion update request is received from user i
7:         Calculate the pmf $p_{-i}[q,(n,t)]$ of the congestion level $q = 0, \ldots, I-1$ for all user $i \in \mathcal{I}$ and $(n,t) \in \mathcal{N} \times \mathcal{T}$, and update in the database:

$$p_{-i}[q,(n,t)] := \sum_{\mathcal{M} \in \mathbb{M}_{-i}(q)} \Big(\prod_{m \in \mathcal{M}} p_m[n,t] \Big)\Big(\prod_{k \in \mathcal{I} \setminus \mathcal{M} \cup \{i\}} (1 - p_k[n,t]) \Big). \tag{16}$$

8:     **end if**
9:     Set $\tau := \tau + 1$.
10: **until** $\tau \geq \tau^{\max}$.

---

work selection statistics $\boldsymbol{p}_i = (p_i[n,t], \forall (n,t) \in \mathcal{N} \times \mathcal{T})$ of all the network-time points to the operator (line 15).

- *Network selection*: After that, once user i' actual trajectory $\boldsymbol{\theta}_i$ is determined (line 18), he will choose the network-times points based on his action $\boldsymbol{s}_i(\boldsymbol{\theta_i})$ (line 19).

2) Algorithm 2: Then, we describe how the operator can compute the aggregate network congestion statistics $p_{-i}[q,(n,t)]$ in Algorithm 2. Based on the information $p_j[n,t]$ obtained from other users $j \in \mathcal{I} \setminus \{i\}$, the operator compute $p_{-i}[q,(n,t)]$ in (16) by counting the probabilities of the network selection with congestion level q (excluding user i). We define $\mathbb{M}_{-i}(q)$ as the set of all possible subsets of set $\mathcal{I} \setminus \{i\}$ with q users. As an example, for $\mathcal{I} = \{1,2,3,4\}$, we have $\mathbb{M}_{-3}(2) = \big\{\{1,2\}, \{1,4\}, \{2,4\}\big\}$.

3) Computational Complexity: We are interested in understanding how long it takes to converge to a pure strategy BNE in the planning phase in Algorithm 1. Theorem 4 ensures that each best response update (i.e., lines 7-13 in Algorithm 1) can be computed efficiently in polynomial time.

Theorem 4: Each best response update of user i can be computed in $\mathcal{O}(|\Theta_i| N^3 T^3)$ time.

The proof of Theorem 4 is given in [20]. For the number of best response updates required for convergence, we will study it in the next section through numerical examples.

V. PERFORMANCE EVALUATIONS

In this section, we study the performance of our DNS algorithm by comparing it with two benchmark schemes over various system parameters. More specifically, to evaluate the users' level of satisfaction and to understand the network selections of these schemes, we evaluate the user utility, fairness, and the amount of network switching. We also show the impact of the prior probabilities of the mobility patterns on the network selections.

A. Parameters and Settings

For each set of system parameters, we run the simulations 10000 times with randomized network settings and users' mobility patterns in MATLAB and show the average value. Unless specified otherwise, the cellular (LTE category 5) network capacity $\mu[1]$ and the Wi-Fi network capacity $\mu[n], n \in \mathcal{N}_{\text{wifi}}$ are normally distributed random variables with means equal to 300 Mbps [24] and 54 Mbps[16] [25], respectively, with standard deviations equal to 5 Mbps. The probability that a Wi-Fi connection is available at a particular location p^{wifi} is equal to 0.5. We consider a two-minute duration, where the duration of a time slot $\Delta t = 10$ seconds, so $T = 12$. We assume that the network switching time is equal to one (i.e., $\delta[n,n'] = 1, \forall n, n' \in \mathcal{N}, n \neq n'$). For the switching cost not involving the idle network, we assume that they are the same that $c[n,n'] = c^{switch}, \forall n, n' \in \mathcal{N} \setminus \{0\}, n \neq n'$. However, the switching cost involving the idle network is halved (i.e., $c[0,n] = c[n,0] = c^{switch}/2, \forall n \in \mathcal{N}$) such that (2) in Assumption 1(c) is satisfied. We consider that all the Wi-Fi networks are available to all the users within its coverage all the time.

In our performance evaluation, we compare our DNS scheme against two benchmark schemes:

- *Cellular-only* scheme: The users use the cellular network all the time to avoid network switching, hence there is no data offloading.
- *On-the-spot offloading* (OTSO) scheme: It is an offloading policy commonly used in most mobile devices today [26], where the data traffic is offloaded to Wi-Fi network whenever Wi-Fi is available. Otherwise, if Wi-Fi networks are not available soon, the cellular connection will be used.

B. Performance Evaluations of Deterministic Mobility Patterns

In this subsection, we first evaluate the schemes in the deterministic mobility pattern case. Here, the users' deterministic mobility patterns are generated based on the same location transition matrix $P = [p(l' \mid l)]_{L \times L}$, where $p(l' \mid l)$ is the probability that a user will move to location l' given that it is currently at location l.[17] All user move around $L = 16$ possible locations in a four by four grid (similar to that in Fig. 1). The probability that the user stays at a location is $p(l \mid l) = 0.6, \forall l \in \mathcal{L}$. Moreover, it is equally likely for the user to move to any one of the neighbouring locations. Take location 7 in Fig. 1 as an example, the probability that the user will move to locations 3, 6, 8, or 11 is equal to $(1 - 0.6)/4 = 0.1$. For the edge location 12, however, the probability of moving to any one of its three neighbouring locations is $(1 - 0.6)/3 = 0.133$.

---

[16]Besides the difference in wireless communication standards, we consider a mean cellular network capacity much higher than the mean Wi-Fi network capacity, because the scale of the cellular network is larger and covers the users in multiple locations, while a Wi-Fi AP covers users in one particular location only.

[17]In our simulations, we just use the location transition matrix as a way to generate the users' mobility patterns. We want to clarify that it does not matter how to generate the users' locations as long as they are given as system parameters in this paper.
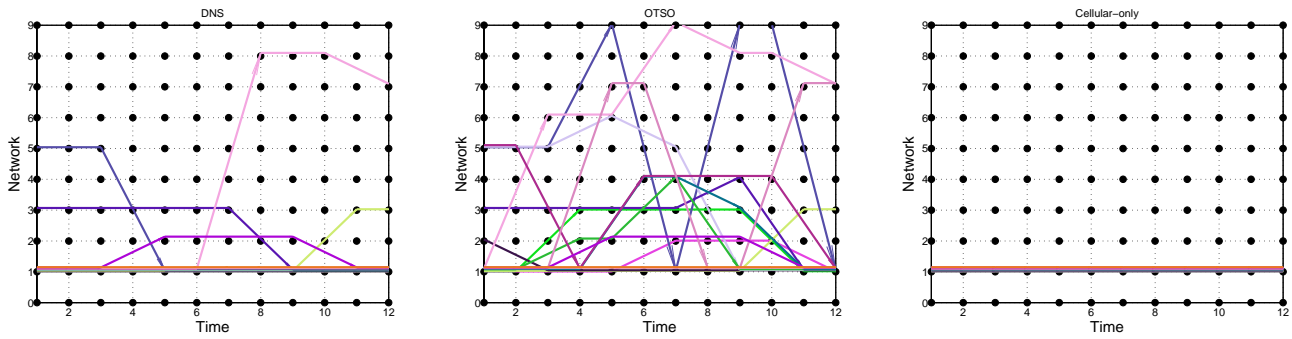
Fig. 5.   Illustration of the network selection of DNS, OTSO, and cellular-only schemes with $I = 15$ users and $c^{switch} = 400$.
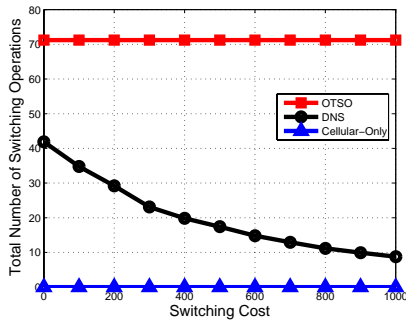

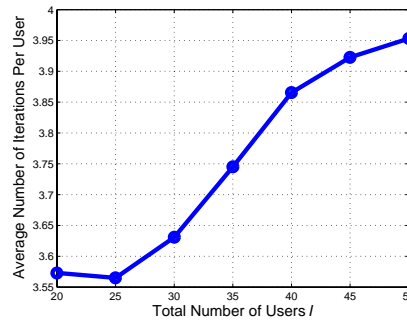Fig. 6.   The total number of switching operations versus switching cost c^{switch} for $I = 30$.

Fig. 7.   The average number of best response update iterations per user for convergence in the DNS scheme with $c^{switch} = 400$.
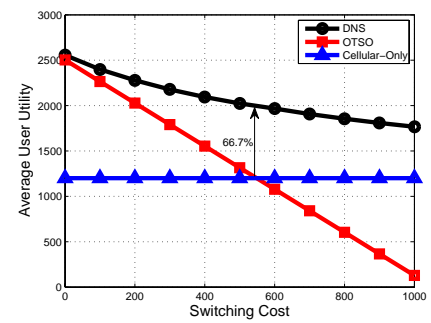
Fig. 8.   The average user utility versus switching cost c^{switch} for $I = 30$.

1) Illustration of Network Selection Schemes: In Fig. 5, we illustrate the network selections under DNS, OTSO, and cellular-only schemes for $I = 15$ users and a switching cost $c^{switch} = 400$. We can see that the OTSO scheme prefers Wi-Fi networks, so it results in a lot of network switching. On the other extreme, the cellular-only scheme uses only the cellular network, so there is no network switching.

2) Network Switching and Scalability: (**Summary of observations**) We first show that the DNS scheme is able to *adaptively* choose the number of switching operations based on the switching cost. We also show that the DNS scheme is *scalable* by considering the number of best response updates for convergence.

In Fig. 6, we plot the total number of network switching operations against the switching cost c^{switch} for $I = 30$. We can see that the performances of both the cellular-only and OTSO schemes are static as they are independent of c^{switch}. However, the DNS scheme responds to the increasing switching cost by decreasing the number of switching.

In Theorem 4, we have established that each best response update can be computed in polynomial time. In Fig. 7, we continue with the evaluation of the convergence speed of Algorithm 1 by counting the average number of best response updates per user required for convergence with respect to different I with $c^{switch} = 400$. We observe that Algorithm 1 scales well with the increasing user population. In particular, each user only needs to perform 3.57 and 3.95 best response updates on average for $I = 20$ and $I = 50$, respectively, before the strategy profile converges to a pure strategy BNE.

3) Average User Utility: (**Summary of observations**) In this subsection, we study the impact of various system parameters on the user utility. Overall, we find that the DNS scheme achieves the highest utility by taking into account both the *ping-pong effect* and *Wi-Fi network congestion*. The results also reveal that the OTSO performs well under a low switching cost and a low Wi-Fi availability.

In Fig. 8, we plot the average user utility against the switching cost c^{switch} for $I = 30$. First, we observe that the proposed DNS scheme achieves the highest user utility compared with OTSO and cellular-only schemes. More specifically, the DNS scheme improves the utility of these two schemes by 66.7% when $c^{switch} = 550$. In addition, for the DNS scheme, we see that its utility decreases gradually with c^{switch}. This is because DNS is aware of the increasing switching cost and thus reduces the number of switching operations (as shown in Fig. 6), which results in a milder reduction in utility. For the OTSO scheme, as it is unaware of the switching cost, the average user utility experiences a heavy reduction when c^{switch} is large. For the cellular-only scheme, since it does not perform any network switching, the user utility is independent of c^{switch}.

In Fig. 9, we plot the average user utility against the number of users I for $c^{switch} = 400$. In general, when I increases, the congestion level increases, so the average utility under all three schemes decrease. We observe that the DNS scheme results in the highest user utility, which suggests that it achieves a good load balancing across the networks. For the cellular-only scheme, since it does not access any available Wi-Fi network capacity, the average user utility is significantly low.
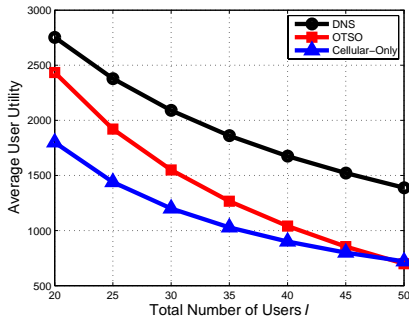
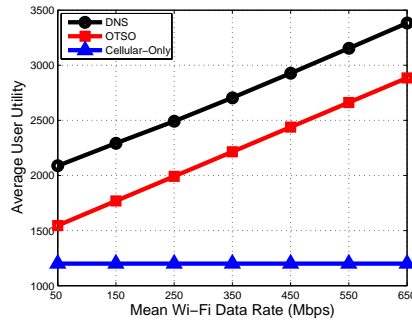Fig. 9. The average user utility versus number of users I for $c^{switch} = 400$.

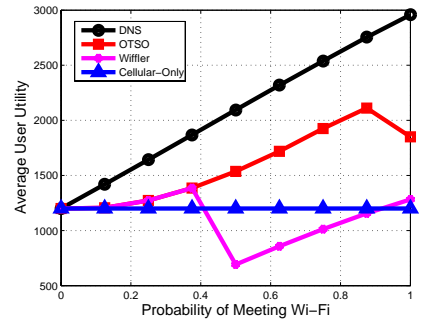Fig. 10. The average user utility versus the mean Wi-Fi data rate for $I = 30$ and $c^{switch} = 400$.

Fig. 11. The average user utility versus the probability of meeting Wi-Fi for $I = 30$ and $c^{switch} = 400$.

In Fig. 10, we plot the average user utility against the mean Wi-Fi data rate for $I = 30$ and $c^{switch} = 400$. We observe that the result is intuitive, where the utility under both the DNS and OTSO schemes increases with the mean Wi-Fi data rate. Also, the DNS scheme achieves the highest user utility among the three schemes.

Furthermore, we aim to study the impact of the probability of meeting Wi-Fi p^{wifi} on different schemes. Here, we compare with an additional *Wiffler* scheme [7], which is a prediction-based offloading scheme that operates as follows. Let ζ be the estimated amount of data that can be transferred using Wi-Fi by the deadline. If Wi-Fi is available in the current location, then Wi-Fi will be used immediately. If Wi-Fi is not available, the user needs to check whether the condition $\zeta \geq \theta k$ is satisfied, where k is the remaining size of the file to be transferred, and $\theta > 0$ is the conservative coefficient that tradeoffs the amount of data offloaded with the completion time of the file transfer. If this condition is satisfied, meaning that the estimated data transfer using Wi-Fi is large enough, then the user will stay idle and wait for the Wi-Fi connection. Otherwise, the user will use the cellular connection. Here, we set $\theta = 1$ as suggested in [7] and consider $k = 0.5$ in the simulation.

In Fig. 11, we plot the average user utility against the probability of meeting Wi-Fi p^{wifi} for $I = 30$ and $c^{switch} = 400$. For the DNS scheme, we can see that the utility increases with p^{wifi}, as the users experience a lower level of network congestion when more Wi-Fi networks are available. For the OTSO scheme, we observe a similar trend from small to medium p^{wifi}. Surprisingly, it experiences a drop in utility when p^{wifi} is above 0.9. The reason is that when p^{wifi} is high such that Wi-Fi coverage is almost ubiquitous, all the users would use the Wi-Fi networks all the time, making the Wi-Fi networks very congested but leaving the cellular network with almost no user. Thus, the average user utility at $p^{\text{wifi}} = 1$ corresponds to the average throughput obtained from the Wi-Fi networks only (i.e., excluding the cellular network) minus the total switching cost. For the cellular-only scheme, since it is independent of the Wi-Fi availability, the average user utility is independent of p^{wifi}. For the Wiffler scheme, when $p^{\text{wifi}} < 0.5$, it is the same as the OTSO scheme, which prefers to use Wi-Fi network when it is available, but the cellular network otherwise. However, when $p^{\text{wifi}} \geq 0.5$, it becomes

a Wi-Fi only scheme, which it will remain idle (instead of using the cellular network) when Wi-Fi is not available. Thus, its user utility increases with p^{wifi} when more Wi-Fi networks are available.

4) Fairness: (**Summary of observations**) In this subsection, we study the fairness of the network resource allocation and show that the DNS scheme achieves a high degree of fairness. In addition, the fairness of the OTSO scheme decreases sharply under a high switching cost.

In Fig. 12, we evaluate the degree of fairness among the users by plotting the Jain's fairness index [27] defined as $\left(\sum_{i \in \mathcal{I}} U_i(\boldsymbol{r}, \boldsymbol{\theta})\right)^2 / \left(I \sum_{i \in \mathcal{I}} U_i(\boldsymbol{r}, \boldsymbol{\theta})^2\right)$ against c^{switch} for $I = 30$. Since the users under the cellular-only scheme always have the same utility, its fairness index is always equal to one. Furthermore, we notice that the fairness indices of both the DNS and OTSO schemes decrease with c^{switch}. For the DNS scheme, as c^{switch} increases, the users switch networks less often (as shown in Fig. 6). In this way, the utilities among the users at a larger c^{switch} are less balanced than that at a smaller c^{switch}, so the degree of fairness decreases with c^{switch}. For the OTSO scheme, although its network selection is independent of c^{switch}, the increase in c^{switch} widens the disparity in utilities among the users with different number of network switching. Moreover, we observe in Fig. 12 that the DNS scheme achieves a higher fairness index than the OTSO scheme, and the fairness index of the DNS scheme decreases much slowly than that of the OTSO scheme. It suggests that the adaptive DNS scheme results in a fairer resource allocation than the static OTSO scheme.

C. Performance Evaluations of Random Mobility Patterns

In this subsection, we evaluate our proposed DNS scheme under the random mobility pattern case. Here, we assume that the cellular network capacity $\mu[1]$ and the Wi-Fi network capacity $\mu[n], n \in \mathcal{N}_{\text{wifi}}$ are normally distributed random variables with means equal to 100 Mbps and 50 Mbps, respectively, and standard deviations equal to 5 Mbps. The probability of meeting Wi-Fi $p^{\text{wifi}} = 0.9$. The switching penalties are the same as that in Section V-B. We consider a one-minute duration, so $T = 6$ for $\Delta t = 10$ seconds. There are $I = 8$ users moving around $L = 5$ possible locations on a
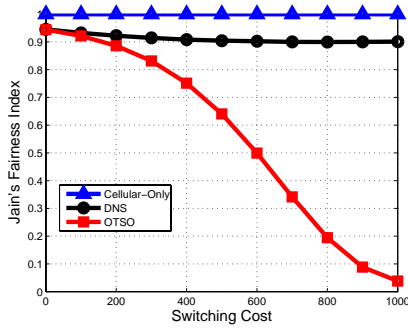
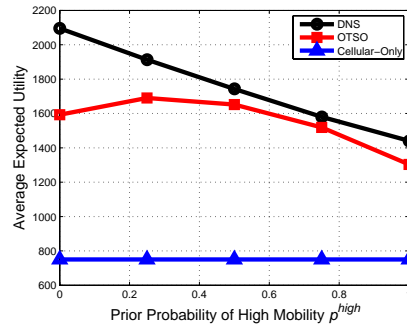Fig. 12. The Jain's fairness index versus switching cost c^{switch} for $I = 30$.

Fig. 13. The average user utility versus the prior probability p^{high} of high mobility for $I = 8$ and $c^{switch} = 10$.
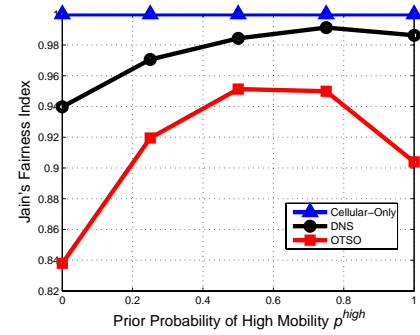
Fig. 14. The Jain's fairness index versus the prior probability p^{high} of high mobility for $I = 8$ and $c^{switch} = 10$.

straight road.[18] For each set of system parameters, we run the simulations 1000 times with randomized network settings and users' mobility patterns in MATLAB and show the average value.

For each user in the random mobility pattern case here, we consider that there are two possible mobility patterns that are generated with different characteristics:

- High mobility: With a prior probability p^{high}, the user will *frequently* move across L locations. In the simulation, we assume that the user has a total probability of 0.9 in moving to one of his neighboring locations and a probability of 0.1 in staying at his current location.
- Low mobility: With a prior probability $1 - p^{high}$, the user moves much less frequently. In the simulation, we assume that the user has a total probability of 0.1 in moving to one of his neighboring locations and a probability of 0.9 in staying at his current location.

1) Impact of Prior Distribution of Mobility Patterns:
(**Summary of observations**) Consistent with the observations under the deterministic mobility case, we see that the DNS achieves the highest expected utility and a high level of fairness under the random mobility case.

In Fig. 13, we plot the average expected utility against the prior probability p^{high} of high mobility when switching cost $c^{switch} = 10$. First, we can see that both the utilities under the DNS and OTSO schemes decrease with p^{high}, because of the higher total switching cost when the users switch networks more often under a high mobility. Nevertheless, the DNS scheme results in a higher expected user utility. For the cellular-only scheme, since the users select the cellular network regardless of their mobility, the user utility is independent of p^{high}.

In Fig. 14, we plot the Jain's fairness index [27] $\left(\sum_{i \in \mathcal{I}} EU_i(\boldsymbol{s})\right)^2 / \left(I \sum_{i \in \mathcal{I}} EU_i(\boldsymbol{s})^2\right)$ against p^{high} for $c^{switch} = 10$. We can see that the DNS scheme achieves a higher degree of fairness than the OTSO scheme. Moreover, for both the DNS and OTSO schemes, there is an increase in fairness when p^{high} increases from a small value to a medium value. We observe that it is due to the larger percentage drop

---

[18]Due to the relatively higher complexity to execute the DNS algorithm under the random mobility pattern case (especially the need to run 1000 times to have a good estimation of the average performance), we consider a smaller scale of simulation in this subsection.

in the expected utility for high-utility users, which increases the fairness. However, for the OTSO scheme, there is a further drop in fairness when $p^{high} > 0.5$, because the expected utilities of some users (not necessarily the high-utility users) decrease, which leads to a reduction in fairness.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we studied the intelligent network selection problem with the objective of achieving an effective data offloading for cellular and Wi-Fi integration. In particular, we focused on understanding the impact of *network congestion* and *switching penalty* due to the herd behaviour and ping-pong effect, respectively, which were not systematically considered in the literature. As a benchmark, we formulated the centralized user utility maximization problem and showed that it is an NP-hard problem, which motivated us to consider a distributed approach. More specifically, with the statistical information of the user mobility, we formulated the users' interactions as a Bayesian network selection game, proved that it is a potential game, and proposed a distributed network selection (DNS) algorithm with provably nice convergence properties. Compared with the on-the-spot offloading (OTSO) and cellular-only schemes, our simulation results showed that the proposed DNS algorithm results in the highest user utility and a good fairness by avoiding Wi-Fi network congestion and costly network switching. In addition, we showed that the OTSO scheme performs especially well under a low switching cost and a low Wi-Fi availability.

In this work, we considered the static setting where each user knows the network conditions and the statistical information of his possible mobility patterns. For the future work, we plan to consider a dynamic setting where a user needs to make online network selections, while considering the time-varying network conditions and mobility patterns. Moreover, we have remarked that the complexity of implementing the DNS algorithm in the random mobility pattern case can be high. Thus, it is important to design a low-complexity DNS algorithm to converge to an approximate equilibrium of the game, while still taking into account the network congestion, switching penalty, and user mobility that we considered in this paper. In addition, it is interesting to analyze the performance of the proposed scheme under the framework of stochastic geometry [28].

APPENDIX

A. Proof of Theorem 2

In the proof, we want to show that the utility function in (8) and the potential function in (13) satisfy (12). First, starting from the original action profile $\boldsymbol{r} = (\boldsymbol{r}_i, \boldsymbol{r}_{-i})$, we define a new action profile $\boldsymbol{r}'$, where $\boldsymbol{r}'_j = \boldsymbol{r}_j$ if $j \neq i$ and $\boldsymbol{r}'_j \neq \boldsymbol{r}_j$ if $j = i$. In other words, only user i changes its action from $\boldsymbol{r}_i$ to $\boldsymbol{r}'_i$ in the new action profile $\boldsymbol{r}' = (\boldsymbol{r}'_i, \boldsymbol{r}_{-i})$.

Next, we define an partition of set $\mathcal{N} \times \mathcal{T}$, which consists of four non-overlapping sets of the network-time points

$$
\begin{aligned}
\mathcal{B}^{(1)} &= \{(n,t) : (n,t) \in \mathcal{V}(\boldsymbol{r}_i), (n,t) \notin \mathcal{V}(\boldsymbol{r}'_i)\}, \\
\mathcal{B}^{(2)} &= \{(n,t) : (n,t) \notin \mathcal{V}(\boldsymbol{r}_i), (n,t) \in \mathcal{V}(\boldsymbol{r}'_i)\}, \\
\mathcal{B}^{(3)} &= \{(n,t) : (n,t) \in \mathcal{V}(\boldsymbol{r}_i), (n,t) \in \mathcal{V}(\boldsymbol{r}'_i)\}, \\
\mathcal{B}^{(4)} &= \{(n,t) : (n,t) \notin \mathcal{V}(\boldsymbol{r}_i), (n,t) \notin \mathcal{V}(\boldsymbol{r}'_i)\},
\end{aligned}
\tag{17}
$$

where $\mathcal{B}^{(1)} \cup \mathcal{B}^{(2)} \cup \mathcal{B}^{(3)} \cup \mathcal{B}^{(4)} = \mathcal{N} \times \mathcal{T}$. As a result, considering the difference in congestion level in network-time point (n,t) between action profiles $\boldsymbol{r}$ and $\boldsymbol{r}'$, we have

$$
\omega[(n,t),\boldsymbol{r},\boldsymbol{\theta}] - \omega[(n,t),\boldsymbol{r}',\boldsymbol{\theta}] = \begin{cases} 1, & \text{if } (n,t) \in \mathcal{B}^{(1)}, \\ -1, & \text{if } (n,t) \in \mathcal{B}^{(2)}, \\ 0, & \text{if } (n,t) \in \mathcal{B}^{(3)} \cup \mathcal{B}^{(4)}. \end{cases}
\tag{18}
$$

For example, in the first line in (18), we have one more user (i.e., user i) choosing the network-time point $(n,t) \in \mathcal{B}^{(1)}$ in the action profile $\boldsymbol{r} = (\boldsymbol{r}_i, \boldsymbol{r}_{-i})$ than in $\boldsymbol{r}' = (\boldsymbol{r}'_i, \boldsymbol{r}_{-i})$, since users other than i choose the same action profile $\boldsymbol{r}_{-i}$.

Let $A \triangleq \sum_{\boldsymbol{e} \in \mathcal{E}(\boldsymbol{r}'_i)} g[\boldsymbol{e}] - \sum_{\boldsymbol{e} \in \mathcal{E}(\boldsymbol{r}_i)} g[\boldsymbol{e}]$. As a result, we have

$$
\begin{aligned}
&\Phi(\boldsymbol{r}_i, \boldsymbol{r}_{-i}, \boldsymbol{\theta}) - \Phi(\boldsymbol{r}'_i, \boldsymbol{r}_{-i}, \boldsymbol{\theta}) \\
&= \left(\sum_{j \in \mathcal{I}} \sum_{\boldsymbol{e} \in \mathcal{E}(\boldsymbol{r}'_j)} g[\boldsymbol{e}] - \sum_{j \in \mathcal{I}} \sum_{\boldsymbol{e} \in \mathcal{E}(\boldsymbol{r}_j)} g[\boldsymbol{e}] \right) \\
&\quad + \sum_{(n,t) \in \mathcal{N} \times \mathcal{T}} \left(\sum_{q=1}^{\omega[(n,t),\boldsymbol{r},\boldsymbol{\theta}]} \frac{\mu[n]}{q} - \sum_{q=1}^{\omega[(n,t),\boldsymbol{r}',\boldsymbol{\theta}]} \frac{\mu[n]}{q} \right) \\
&= A + \sum_{(n,t) \in \mathcal{B}^{(1)} \cup \mathcal{B}^{(2)} \cup \mathcal{B}^{(3)} \cup \mathcal{B}^{(4)}} \left(\sum_{q=1}^{\omega[(n,t),\boldsymbol{r},\boldsymbol{\theta}]} \frac{\mu[n]}{q} - \sum_{q=1}^{\omega[(n,t),\boldsymbol{r}',\boldsymbol{\theta}]} \frac{\mu[n]}{q} \right) \\
&= A + \sum_{(n,t) \in \mathcal{B}^{(1)}} \left(\sum_{q=1}^{\omega[(n,t),\boldsymbol{r},\boldsymbol{\theta}]} \frac{\mu[n]}{q} - \sum_{q=1}^{\omega[(n,t),\boldsymbol{r}',\boldsymbol{\theta}]} \frac{\mu[n]}{q} \right) \\
&\quad + \sum_{(n,t) \in \mathcal{B}^{(2)}} \left(\sum_{q=1}^{\omega[(n,t),\boldsymbol{r},\boldsymbol{\theta}]} \frac{\mu[n]}{q} - \sum_{q=1}^{\omega[(n,t),\boldsymbol{r}',\boldsymbol{\theta}]} \frac{\mu[n]}{q} \right) \\
&= A + \sum_{(n,t) \in \mathcal{B}^{(1)}} \frac{\mu[n]}{\omega[(n,t),\boldsymbol{r},\boldsymbol{\theta}]} - \sum_{(n,t) \in \mathcal{B}^{(2)}} \frac{\mu[n]}{\omega[(n,t),\boldsymbol{r}',\boldsymbol{\theta}]} \\
&= A + \sum_{(n,t) \in \mathcal{B}^{(1)} \cup \mathcal{B}^{(3)}} \frac{\mu[n]}{\omega[(n,t),\boldsymbol{r},\boldsymbol{\theta}]} - \sum_{(n,t) \in \mathcal{B}^{(2)} \cup \mathcal{B}^{(3)}} \frac{\mu[n]}{\omega[(n,t),\boldsymbol{r}',\boldsymbol{\theta}]} \\
&= A + \sum_{(n,t) \in \mathcal{V}(\boldsymbol{r}_i)} \frac{\mu[n]}{\omega[(n,t),\boldsymbol{r},\boldsymbol{\theta}]} - \sum_{(n,t) \in \mathcal{V}(\boldsymbol{r}'_i)} \frac{\mu[n]}{\omega[(n,t),\boldsymbol{r}',\boldsymbol{\theta}]} \\
&= U_i(\boldsymbol{r}_i, \boldsymbol{r}_{-i}, \boldsymbol{\theta}) - U_i(\boldsymbol{r}'_i, \boldsymbol{r}_{-i}, \boldsymbol{\theta}).
\end{aligned}
\tag{19}
$$

Here, the first equality is due to the definition in (13). The second equality is due to $\boldsymbol{r}'_j = \boldsymbol{r}_j$ for $j \neq i$ and $\mathcal{B}^{(1)} \cup \mathcal{B}^{(2)} \cup \mathcal{B}^{(3)} \cup \mathcal{B}^{(4)} = \mathcal{N} \times \mathcal{T}$. The third equality is due to the fact that

$$
\sum_{q=1}^{\omega[(n,t),\boldsymbol{r},\boldsymbol{\theta}]} \frac{\mu[n]}{q} - \sum_{q=1}^{\omega[(n,t),\boldsymbol{r}',\boldsymbol{\theta}]} \frac{\mu[n]}{q} = 0, \text{ for } (n,t) \in \mathcal{B}^{(3)} \cup \mathcal{B}^{(4)}.
\tag{20}
$$

The fourth equality is due to the algebraic manipulation based on (18). The fifth equality is due to $\omega[(n,t),\boldsymbol{r},\boldsymbol{\theta}] = \omega[(n,t),\boldsymbol{r}',\boldsymbol{\theta}]$ if $(n,t) \in \mathcal{B}^{(3)}$ from (18). The sixth equality is due to $\mathcal{V}(\boldsymbol{r}_i) = \mathcal{B}^{(1)} \cup \mathcal{B}^{(3)}$ and $\mathcal{V}(\boldsymbol{r}'_i) = \mathcal{B}^{(2)} \cup \mathcal{B}^{(3)}$. The last equality is due to the definition in (8). ∎

B. Proof of Theorem 3

First, we define a function

$$
\Psi(\boldsymbol{s}) \triangleq \sum_{\boldsymbol{\theta} \in \Theta} \Phi(\boldsymbol{s}(\boldsymbol{\theta}), \boldsymbol{\theta}) p(\boldsymbol{\theta}).
\tag{21}
$$

We can show that

$$
\begin{aligned}
EU_i(\boldsymbol{s}) - EU_i(\boldsymbol{s}') &= \sum_{\boldsymbol{\theta} \in \Theta} \left(U_i(\boldsymbol{s}(\boldsymbol{\theta}), \boldsymbol{\theta}) - U_i(\boldsymbol{s}'(\boldsymbol{\theta}), \boldsymbol{\theta}) \right) p(\boldsymbol{\theta}) \\
&= \sum_{\boldsymbol{\theta} \in \Theta} \left(\Phi(\boldsymbol{s}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \Phi(\boldsymbol{s}'(\boldsymbol{\theta}), \boldsymbol{\theta}) \right) p(\boldsymbol{\theta}) = \Psi(\boldsymbol{s}) - \Psi(\boldsymbol{s}'),
\end{aligned}
\tag{22}
$$

where the first equality is due to the definition in (10). The second equality is due to (12) and Theorem 2. Thus, $\Psi(\boldsymbol{s})$ is the *potential function* of game Ω. From [29], every finite game with a potential function has the FIP. ∎

REFERENCES

[1] M. H. Cheung, R. Southwell, and J. Huang, "Congestion-aware network selection and data offloading," in *Proc. of IEEE CISS*, Princeton, NJ, Mar. 2014.

[2] Cisco Systems, "Cisco visual networking index: Global mobile data traffic forecast update, 2015-2020," White Paper, Feb. 2016.

[3] Alcatel-Lucent and British Telecommunications, "Wi-Fi roaming: Building on ANDSF and Hotspot2.0," White Paper, 2012.

[4] 4G Americas, "Integration of cellular and Wi-Fi networks," White Paper, Sept. 2013.

[5] M. Paolini, "Taking Wi-Fi beyond offload: Integrated Wi-Fi access can differentiate service and generate new revenues," White Paper, 2012.

[6] R. Southwell, J. Huang, and X. Liu, "Spectrum mobility games," in *Proc. of IEEE INFOCOM*, Orlando, FL, Mar. 2012.

[7] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Proc. of ACM MobiSys*, San Francisco, CA, June 2010.

[8] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, "Mobile data offloading: How much can WiFi deliver?" in *Proc. of ACM CoNEXT*, Philadelphia, PA, Nov. 2010.

[9] N. Ristanovic, J.-Y. Le Boudec, A. Chaintreau, and V. Erramilli, "Energy efficient offloading of 3G networks," in *Proc. of IEEE MASS*, Valencia, Spain, Oct. 2011.

[10] Y. Im, C. Joe-Wong, S. Ha, S. Sen, T. T. Kwon, and M. Chiang, "AMUSE: Empowering users for cost-aware offloading with throughput-delay tradeoffs," in *Proc. of IEEE INFOCOM*, Turin, Italy, Apr. 2013.

[11] Y. Moon, D. Kim, Y. Go, Y. Kim, Y. Yi, S. Chong, and K. Park, "Practicalizing delay-tolerant mobile apps with Cedos," in *Proc. of ACM MobiSys*, Florence, Italy, May 2015.

[12] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "RAT selection games in HetNets," in *Proc. of IEEE INFOCOM*, Turin, Italy, Apr. 2013.

[13] E. Monsef, A. Keshavarz-Haddad, E. Aryafar, J. Saniie, and M. Chiang, "Convergence properties of general network selection games," in *Proc. of IEEE INFOCOM*, Hong Kong, China, Apr. 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2687383, IEEE Journal on Selected Areas in Communications

13

[14] R. Mahindra, H. Viswanathan, K. Sundaresan, M. Y. Arslan, and S. Rangarajan, "A practical traffic management system for integrated LTE-WiFi networks," in *Proc. of ACM MobiCom*, Maui, HI, Sept. 2014.

[15] Z. Hu, Z. Lu, Z. Li, and X. Wen, "Adaptive network selection based on attractor selection in data offloading," in *Proc. of IEEE WCNC*, Doha, Qatar, Apr. 2016.

[16] J. Ghosh, M. J. Beal, H. Q. Ngo, and C. Qiao, "On profiling mobility and predicting locations of wireless users," in *Proc. of ACM REALMAN*, Florence, Italy, May 2006.

[17] Y. Shoham and K. Leyton-Brown, *Multi Agent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2008.

[18] G. Facchini, F. van Megen, P. Borm, and S. Tijs, "Congestion models and weighted Bayesian potential games," *Springer Theory and Decision*, vol. 42, no. 2, pp. 193–206, Mar. 1997.

[19] Ericsson, "Wi-Fi in heterogeneous networks: An integrated approach to delivering the best user experience," White Paper, Nov. 2012.

[20] M. H. Cheung, F. Hou, J. Huang, and R. Southwell, "Congestion-aware distributed network selection for integrated cellular and Wi-Fi networks," http://arxiv.org/abs/1703.00216, 2017.

[21] J. C. Harsanyi, "Games with incomplete information played by "Bayesian" players part II: Bayesian equilibrium points," *Management Science*, vol. 14, no. 5, pp. 320–334, Jan. 1968.

[22] B. Vocking and R. Aachen, "Congestion games: Optimization in competition," in *Proc. of 2nd Algorithms and Complexity Workshop*, Durham, Sept. 2006.

[23] M. A. Weiss, *Data Structures and Algorithm Analysis in C*, 2nd ed. Menlo Park, CA: Addison-Wesley, 2003.

[24] Wikipedia, "E-UTRA." [Online]. Available: https://en.wikipedia.org/wiki/E-UTRA.

[25] "IEEE 802.11," http://standards.ieee.org/getieee802/download/802.11-2007.pdf, 2007.

[26] S. Rayment and J. Bergstrom, "Achieving carrier-grade Wi-Fi in the 3GPP world," Ericsson Review, 2012.

[27] R. K. Jain, D. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Eastern Research Lab, Tech. Report DEC-TR-301, Sept. 1984.

[28] W. Bao and B. Liang, "Stochastic geometric analysis of user mobility in heterogeneous wireless networks," *IEEE J. on Selected Areas in Commun.*, vol. 33, no. 10, pp. 2212–2225, Oct. 2015.

[29] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, May 1996.

Fen Hou (M'10) is an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Macau. She received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, Canada, in 2008. She worked as a postdoctoral fellow in the Electrical and Computer Engineering at the University of Waterloo and in the Department of Information Engineering at the Chinese University of Hong Kong from 2008 to 2009 and from 2009 to 2011, respectively. Her research interests include resource allocation and scheduling in broadband wireless networks, protocol design and QoS provisioning for multimedia communications in broadband wireless networks, Mechanism design and optimal user behavior in mobile crowd sensing networks and mobile data offloading. She is the recipient of IEEE GLOBECOM Best Paper Award in 2010 and the Distinguished Service Award in IEEE MMTC in 2011. Dr. Fen Hou served as the co-chair in ICCS 2014 Special Session on Economic Theory and Communication Networks, INFOCOM 2014 Workshop on Green Cognitive Communications and Computing Networks (GCCCN), IEEE Globecom Workshop on Cloud Computing System, Networks, and Application (CCSNA) 2013 and 2014, ICCC 2015 Selected Topics in Communications Symposium, and ICC 2016 Communication Software Services and Multimedia Application Symposium, respectively. She currently serves as the vice-chair (Asia) in IEEE ComSoc Multimedia Communications Technical Committee (MMTC) and an associate editor for IET Communications as well.

Jianwei Huang (S'01-M'06-SM'11-F'16) is an Associate Professor and Director of the Network Communications and Economics Lab (ncel.ie.cuhk.edu.hk), in the Department of Information Engineering at the Chinese University of Hong Kong. He received the Ph.D. degree from Northwestern University in 2005, and worked as a Postdoc Research Associate at Princeton University during 2005-2007. Dr. Huang is the co-recipient of 8 Best Paper Awards, including IEEE Marconi Prize Paper Award in Wireless Communications in 2011. He has co-authored six books, including the textbook on "Wireless Network Pricing." He received the CUHK Young Researcher Award in 2014 and IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2009. Dr. Huang has served as an Associate Editor of IEEE/ACM Transactions on Networking, IEEE Transactions on Cognitive Communications and Networking, IEEE Transactions on Wireless Communications, and IEEE Journal on Selected Areas in Communications - Cognitive Radio Series. He has served as the Chair of IEEE ComSoc Cognitive Network Technical Committee and Multimedia Communications Technical Committee. He is an IEEE Fellow, a Distinguished Lecturer of IEEE Communications Society, and a Thomson Reuters Highly Cited Researcher in Computer Science.

Man Hon Cheung received the B.Eng. and M.Phil. degrees in Information Engineering from the Chinese University of Hong Kong (CUHK) in 2005 and 2007, respectively, and the Ph.D. degree in Electrical and Computer Engineering from the University of British Columbia (UBC) in 2012. Currently, he is a postdoctoral fellow in the Department of Electrical and Computer Engineering at the University of Macau. He worked as a postdoctoral fellow in the Department of Information Engineering in CUHK. He received the IEEE Student Travel Grant for attending *IEEE ICC 2009*. He was awarded the Graduate Student International Research Mobility Award by UBC, and the Global Scholarship Programme for Research Excellence by CUHK. He serves as a Technical Program Committee member in *IEEE ICC*, *Globecom*, *WCNC*, and *WiOpt*. His research interests include the design and analysis of wireless network protocols using optimization theory, game theory, and dynamic programming, with current focus on mobile data offloading, mobile crowdsensing, and network economics.

Richard Southwell did his BSc in Theoretical Physics at the University of York, MSc in Mathematics at the University of York, and Ph.D. in Mathematics at the University of Sheffield. After working as a research associate in the amorphous computing project, he moved to Hong Kong to work as a researcher at the Network Communications and Economics Lab (NCEL) in the Information Engineering Department at the Chinese University of Hong Kong. Later he became an Assistant Professor at Institute for Interdisciplinary Information Sciences (IIIS) in Tsinghua University, Beijing. Then he moved back to Hong Kong and again worked as a researcher in NCEL, as well as at the Department of Management Science in the City University of Hong Kong. He is currently working at the York Centre for Complex Systems Analysis, in connection with the Department of Mathematics. His research interests include graph theory, game theory, complex systems, projective geometry, topology, and dynamics. Currently, he is using partial differential based equations to model marine ecology.