Congestion-Aware Network Selection and Data Offloading

Man Hon Cheung, Richard Southwell, and Jianwei Huang

***Abstract*—Wi-Fi offloading is a cost-effective approach to provide an immediate capacity relief to the congested areas in a cellular network. However, previously proposed schemes mainly focus on alleviating the cellular congestion by offloading the data traffic to Wi-Fi as much as possible, but without systematic considerations of the network congestion, switching penalty, and pricing in the networks. In this paper, we consider the network selection and data offloading problem in an integrated cellular Wi-Fi system by incorporating the practical considerations of (i) user mobility, (ii) location, user, and time dependent Wi-Fi availabilities, (iii) network dependent switching time and switching cost for changing network connections, and (iv) usage-based pricing into our modeling. We formulate the interactions of the users' congestion-aware network selection decisions across multiple time slots as a non-cooperative network selection game (NSG), where the strategy of each user corresponds to a route on a graph. We prove that the NSG is equivalent to a *congestion game*, which implies that the game has the finite improvement property. As a result, when the players repeatedly perform better response updates, the system is guaranteed to converge to a pure Nash equilibrium. Simulation results show that our proposed NSG scheme achieves a better load balancing than two static heuristic schemes.**

I. Introduction

Today's cellular networks are heavily loaded mainly due to the huge amount of mobile video data traffic [1]. According to Cisco's forecast, mobile data traffic will increase by 11-fold between 2013 and 2018 globally [2]. The network capacity, however, is growing at a much slower pace, and cannot keep up with the explosive growth in data traffic [2]. In recent years, the mobile operators (MOs) started to realize and demonstrate that *Wi-Fi offloading*, where data traffic originally targeted towards the cellular network is offloaded to Wi-Fi networks, is a cost-effective approach for providing immediate capacity relief to the congested areas of their mobile data networks [3]–[6].

Although Wi-Fi can alleviate the traffic congestion in cellular networks, using it as a stand-alone capacity-offload solution limits its abilities to provide the mobile users (MUs) with a seamless and secure mobile broadband experience [7], [8]. When an MU decides to *manually* switch to a Wi-Fi network, the MO may lose visibility of the MU's activities. Thus, it is difficult to maintain a quality of service (QoS) guarantee to the MUs. To avoid this, the MO can tightly integrate the Wi-Fi networks with the cellular networks through different recent IEEE and 3rd Generation Partnership Project (3GPP) standards. For example, in the IEEE 802.11u standard [9], the network discovery and selection functionality will advertise the network information related to the access network type, roaming consortium, and venue information through management frames. With this information, the access network discovery and selection function (ANDSF) [10], [11] of the 3GPP standard assists an MU to choose a suitable Wi-Fi network by providing it with a list of preferred access networks and the corresponding connection policies. However, these standards do not clearly specify the detailed network selection and data offloading policies.

Previous works on data offloading mainly aim to reduce the traffic load in the cellular network or cellular data plan consumption by offloading as much data traffic through the Wi-Fi network as possible, but they do not systematically consider the network congestion, switching penalty, and network pricing. Balasubramanian *et al.* in [3] proposed that an MU can perform data offloading by making predictions of future Wi-Fi availability using the past mobility history. Lee *et al.* in [4] described the on-the-spot offloading (OTSO) scheme that most smartphones are using by default. The OTSO scheme adopts a simple offloading policy that an MU offloads its data traffic to a Wi-Fi network whenever possible. Ristanovic *et al.* in [12] considered energy-efficient offloading for delay-tolerant applications. They showed that the proposed offloading algorithms can offload a significant amount of traffic from the cellular network and extend the battery lifetime. Im *et al.* in [13] 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 MU. For the MUs in a congested area with a Wi-Fi coverage, such data offloading schemes may result in a congested Wi-Fi network [7], and thus a poor quality of service for the users. These previously considered schemes neglect the switching penalty, such as the delay and additional energy consumption, which MUs may incur when they change their network connections. In addition, these schemes do not take into account the network pricing, which may lead to large usage fees for the MUs. In short, the network selection problem considering the network congestion, switching penalty, and network pricing in data offloading has not been explored by previous literature.

As a result, in the design of our congestion-aware network selection algorithm, we aim to address the practical considerations of (i) user mobility, (ii) location, user, and time dependent

The authors are with the Network Communications and Economics Lab (NCEL), Department of Information Engineering, the Chinese University of Hong Kong; E-mails: {mhcheung, jwhuang}@ie.cuhk.edu.hk, and richardsouthwell254@gmail.com. This work is supported by the General Research Funds (Project Number CUHK 412713 and CUHK 412511) established under the University Grant Committee of the Hong Kong Special Administrative Region, China.

Wi-Fi availabilities, (iii) network dependent switching time and switching cost, and (iv) usage-based pricing. For (i), we suppose that each MU is moving among different locations over multiple time slots according to a predetermined trajectory. For (ii), first, 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 MUs who have subscribed for different data plans or Wi-Fi services (e.g., Skype Wi-Fi) can have different privileges to access different Wi-Fi networks. Moreover, some MO-owned Wi-Fi hotspots are reserved for high-priority users, while other Wi-Fi hotspots may be open for free public access. For (iii), we assume that an MU may incur a *switching time* and a *switching cost* when it decides to switch between different networks. The switching time may correspond to the delay during handoff between different wireless networks, and the switching cost may account for practical issues such as the additional power consumption and QoS disruption [14]. For (iv), we consider the usage-based pricing in wireless networks, which is widely adopted to alleviate the network congestion nowadays. Given these practical considerations with heterogeneous MUs and networks, the network selection problem in the integrated cellular Wi-Fi system [7], [8] is very challenging to solve.

In this paper, we consider the congestion-aware network selection problem to address the above modeling challenges. We formulate the network selection interactions of the MUs in multiple time slots as a non-cooperative network selection game (NSG), where the strategy of each user corresponds to a route on a network-time graph. We prove that the NSG is equivalent to a *congestion game* [14]–[16]. This implies that the game has the finite improvement property, meaning that when the players repeatedly perform better response updates, the system is guaranteed to reach a pure Nash equilibrium (NE) [14]–[16]. The proof of this result is similar to the one used in [14], where we mapped a spectrum mobility game into the framework of congestion game. However, our result here is more general than the one in [14], because the players in this work do not have the same set of available resources, and the switching time and switching cost are *network dependent*.

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

- To the best of our knowledge, this is the first paper that studies the interactions of *multiple* heterogeneous MUs in terms of network selection and data offloading.
- *Practical modeling*: We consider the details of (i) user mobility, (ii) location, user, and time dependent Wi-Fi availabilities, (iii) network dependent switching time and switching cost, and (iv) usage-based pricing in our model.
- *Game-theoretic analysis*: We show that the NSG possesses the finite improvement property, which guarantees the convergence to a pure NE. We also quantify the computation time of a better response update.
- *Load balancing*: Simulation results show that the proposed NSG scheme achieves a better load balancing performance than the cellular-only and OTSO schemes.

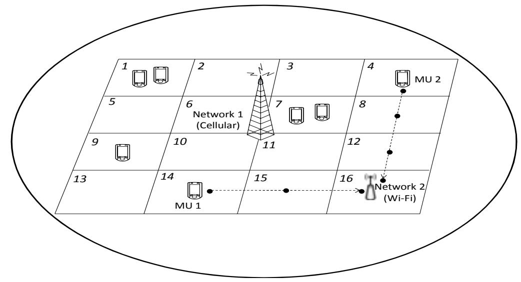

Fig. 1. An example of the network setting, where $\mathcal{N} = \{1, 2\}$ is the set of networks. We assume that the cellular network (i.e., network 1) is available at all the locations, while the Wi-Fi network (i.e., network 2) is location, time, and user dependent. The MUs are moving within a set of locations $\mathcal{L} = \{1, \ldots, 16\}$ over multiple time slots.

The rest of the paper is organized as follows. The system model is described in Section II. We consider the network selection game in Section III. Simulation results are given in Section IV, and we conclude the paper in Section V.

II. System Model

As shown in Fig. 1, we consider an integrated cellular 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 [8]. The networks are indexed by $n \in \mathcal{N} = \{1, \ldots, N\}$, where $n = 1$ denotes the cellular network, and $n \in \mathcal{N}_{\text{wifi}} = \{2, \ldots, N\}$ denotes a Wi-Fi network. We consider a slotted system indexed by time slot $t \in \mathcal{T} = \{1, \ldots, T\}$ of length Δt. Let $\mu^{(n)}$ be the capacity of network $n \in \mathcal{N}$, and let $x^{(n,t)}$ be the background traffic of network $n \in \mathcal{N}$ at time $t \in \mathcal{T}$. The background cellular traffic load $x^{(1,t)}$ can be generated from users who can only gain access to the cellular network, but not the Wi-Fi network (e.g., MUs who do not move into the Wi-Fi coverage areas, or MUs with devices not supporting Wi-Fi). The background Wi-Fi traffic load $x^{(n,t)}$ of network $n \in \mathcal{N}_{\text{wifi}}$ can be generated by users who can only access the Wi-Fi network $n \in \mathcal{N}_{\text{wifi}}$, but not the cellular network (e.g., desktop computers).

Let $\mathcal{I} = \{1, \ldots, I\}$ be the set of mobile users (MUs), and let $\mathcal{L} = \{1, \ldots, L\}$ be the set of locations. Let $\mathcal{M}(i, l, t) \subseteq \mathcal{N}$ be the set of networks available for user $i \in \mathcal{I}$ at location $l \in \mathcal{L}$ and time $t \in \mathcal{T}$. That is, we model the user, location, and time dependent network availabilities. We assume that the cellular network is available to all the MUs at all possible locations all the time, so that network $1 \in \mathcal{M}(i, l, t), \forall\, i \in \mathcal{I}, l \in \mathcal{L}, t \in \mathcal{T}$. Let $(k_i(t), \forall\, t \in \mathcal{T})$ be the *trajectory* of MU i, where $k_i(t) \in \mathcal{L}$ is the position of user $i \in \mathcal{I}$ at time $t \in \mathcal{T}$. As an example, in Fig. 1, the trajectories of MU 1 and MU 2 are given by $(14, 15, 16, 16)$ and $(4, 8, 12, 16)$ for a total of $T = 4$ slots, respectively. In general, given the starting point, destination, path, and moving speed of an MU, it is possible to estimate its trajectory for the next few minutes with a reasonable degree of accuracy. So in this paper, we assume that the trajectory of each MU for the near future is known. For simplicity, we define $\mathcal{N}(i, t) = \mathcal{M}(i, k_i(t), t)$ to be the set of networks available to MU $i \in \mathcal{I}$ at time

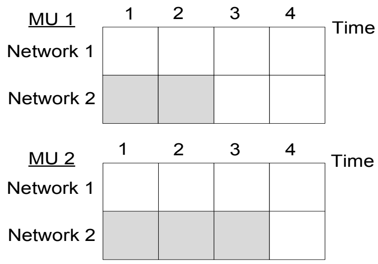

Fig. 2. Heterogeneous network availabilities of MU 1 and MU 2 shown in Fig. 1, where a white block means that the network is available (i.e., network $n \in \mathcal{N}(i,t)$) and a gray block means that the network is not available.

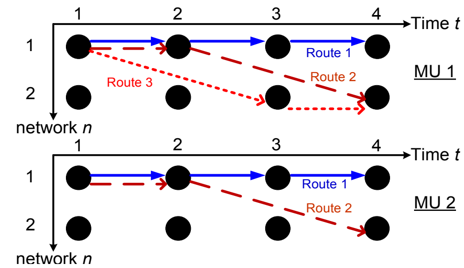

Fig. 3. The set of feasible routes of the two MUs with switching time = 1. In this example, MUs 1 and 2 have three and two feasible routes, respectively.

$t \in \mathcal{T}$. Due to the difference in the mobility patterns and Wi-Fi availabilities of the MUs, their network availabilities at different time are different in general. In Fig. 2, we show an example of the heterogeneous network availabilities of MU 1 and MU 2 shown in Fig. 1.

In this work, we define the *utility* of an MU at a particular time as its received throughput. Let m be the number of MUs who have chosen network $n \in \mathcal{N}$ at time $t \in \mathcal{T}$. The network throughput at that time is equal to $\frac{\mu^{(n)}}{m+x^{(n,t)}}$, which is inversely proportional to the congestion level in the network [17]. We suppose that the MO adopts the commonly used *usage-based pricing*, where the required payment of an MU is directly proportional to its data usage. Let $\gamma(n)$ be the price per unit of usage in network $n \in \mathcal{N}$. By considering a saturated traffic case, where each MU always has data to send, the *payment* of an MU for using network $n \in \mathcal{N}$ at time $t \in \mathcal{T}$ is equal to $\gamma(n)\frac{\mu^{(n)}}{m+x^{(n,t)}}\Delta t$. Overall, the *surplus* (i.e., utility minus payment) of an MU in using network $n \in \mathcal{N}$ at time $t \in \mathcal{T}$ with m MUs in the network is given by

$$\sigma^{(n,t)}(m) = \big(a - \gamma(n)\Delta t\big)\frac{\mu^{(n)}}{m + x^{(n,t)}}, \tag{1}$$

where $a \geq 0$ is the the scaling weight between the utility and payment. Here, we assume that $a \geq \gamma(n)\Delta t$, $\forall\, n \in \mathcal{N}$. Since a rational MU would avoid a network that results in a negative surplus for itself, we will not consider networks that do not satisfy this constraint.

III. Network Selection Game

In this section, we formulate the network selection problem as a non-cooperative network selection game (NSG), where each rational MU aims to maximize its payoff. We analyze the convergence properties of the game.

In the NSG, which we will soon define, players must decide which networks to use in the near future. This is complicated because switching networks may take several time slots. We use a *directed graph* $\mathcal{G}$ to represent how MUs can switch between networks as time passes by. The directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ has a vertex set $\mathcal{V} = \mathcal{N} \times \mathcal{T}$ (which represents the different networks at different times), and an edge set $\mathcal{E} = \mathcal{E}^{stick} \cup \mathcal{E}^{switch}$ which represent the different ways MUs can stick or switch networks over time. Vertex $(n,t) \in \mathcal{V}$ of $\mathcal{G}$ represents network n at time t. If $[(n,t),(n',t')] \in \mathcal{E}$ is an edge pointing from vertex (n,t) to vertex (n',t'), then an MU i can change from using n at t, to using n' at t', provided $n' \in \mathcal{N}(i,t')$. Here

$$\mathcal{E}^{stick} = \Big\{\big[(n,t),(n,t+1)\big] : (n,t) \in \mathcal{V}, (n,t+1) \in \mathcal{V}\Big\} \tag{2}$$

is the set of edges which represent the actions of keep using the same network. An example of such edge is $\big[(1,1),(1,2)\big]$ in route 2 of MU 1 in Fig. 3. The set of all edges corresponding to switching between different networks is $\mathcal{E}^{switch} = \bigcup_{n,n' \in \mathcal{N}, n \neq n'} \mathcal{F}^{n,n'}$, where

$$\begin{aligned}\mathcal{F}^{n,n'} = \Big\{\big[(n,t),(n',t+s_{n,n'}+1)\big] : (n,t) \in \mathcal{V},\\ (n',t+s_{n,n'}+1) \in \mathcal{V}, n \neq n'\Big\}\end{aligned} \tag{3}$$

is the set of all edges that represent switchings from network n to network n'. Here $s_{n,n'} \geq 0$ is the amount of *switching time* that an MU must wait in order to switch from network n to a network n'. For example, consider the edge $\big[(1,2),(2,4)\big] \in \mathcal{F}^{1,2}$ from route 2 of MU 1 in Fig. 3. This corresponds to MU 1 using network 1 at the end of time slot 2, and then spending $s_{1,2} = 1$ time slot switching to network 2.

A *route* in a directed graph is a sequence of linked vertices (i.e., each vertex is linked to its successor in the sequence), which represents the *strategy* or *network usage plan* of an MU. Let us define the set of all feasible routes $\mathcal{R}_i$ of MU $i \in \mathcal{I}$ to be the set of all routes r_i through the directed graph $\mathcal{G}$, which starts at a vertex of the form $(n,1) : n \in \mathcal{N}$, ends at a vertex of the form $(n',T) : n' \in \mathcal{N}$, and has the property that each vertex (n,t) within the route r_i is such that $n \in \mathcal{N}(i,t)$. In other words, $\mathcal{R}_i$ is the strategy set of player i in the NSG, which represents all the feasible network usage plans of MU i. Let $\mathcal{V}(r_i)$ and $\mathcal{E}(r_i)$ denote the set of vertices and edges within the route $r_i \in \mathcal{R}_i$ chosen by player i. For example, in Fig. 3, route 2 of MU 1 is $r_1 = \big((1,1),(1,2),(2,4)\big)$, where $\mathcal{V}(r_1) = \{(1,1),(1,2),(2,4)\}$ and $\mathcal{E}(r_1) = \big\{\big((1,1),(1,2)\big),\big((1,2),(2,4)\big)\big\}$.

In the *strategy profile* $\boldsymbol{r} = (r_1, \ldots, r_I) \in \mathcal{R}_1 \times \ldots \times \mathcal{R}_I$, each player i picks a route (or strategy) $r_i \in \mathcal{R}_i$. The *payoff* of MU i under the strategy profile $\boldsymbol{r}$ is given by

$$\rho_i(\boldsymbol{r}) = \sum_{v \in \mathcal{V}(r_i)} \sigma^v\Big(m^v(\boldsymbol{r})\Big) - \sum_{e \in \mathcal{E}(r_i)} g(e), \tag{4}$$

where the surplus $\sigma^v(\cdot) = \sigma^{(n,t)}(\cdot)$ is defined in (1), and $m^v(\boldsymbol{r}) = \Big|\{j \in \mathcal{I} : v \in \mathcal{V}(r_j)\}\Big|$ is the number of MUs that choose vertex $v \in \mathcal{V}$ under the strategy profile $\boldsymbol{r}$. As a result, the payoff of each MU depends on the level of congestion in the chosen networks. $g(e)$ is the cost of traversing an edge $e = [(n,t),(n',t')] \in \mathcal{E}$, which is equal to the *switching cost* $c_{n,n'} \geq 0$ for switching from network n to n' as

$$g(e) = g\Big([(n,t),(n',t')]\Big) = c_{n,n'}. \qquad (5)$$

To sum up, the payoff in (4) represents the total surpluses minus the total switching costs along route $\boldsymbol{r}$. As an example, assume that the cellular price $\gamma(1) = \gamma$ and Wi-Fi is free of charge such that $\gamma(n) = 0, \forall\, n \in \mathcal{N}_{\text{wifi}}$. In Fig. 3, when MU 1 chooses route 2 and MU 2 chooses route 1, the payoff $\rho_1(\boldsymbol{r})$ of player 1 is equal to $\frac{(a-\gamma\Delta t)\mu^{(1)}}{2+x^{(1,1)}} + \frac{(a-\gamma\Delta t)\mu^{(1)}}{2+x^{(1,2)}} + \frac{a\mu^{(2)}}{1+x^{(2,4)}} - c_{1,2}$.

Formally, a NSG is a tuple

$$\Omega = \big(\mathcal{I}, \mathcal{N}, \mathcal{T}, (\mathcal{N}(i,t))_{i\in\mathcal{I},t\in\mathcal{T}}, (\sigma^{(n,t)})_{n\in\mathcal{N},t\in\mathcal{T}}, (c_{n,n'}, s_{n,n'})_{n,n'\in\mathcal{N}}\big),$$

where each player i chooses a feasible route $r_i \in \mathcal{R}_i$ and receives a payoff as described by (4). Notice that Ω is a static game where each player chooses a pure strategy (i.e., a route) once with the complete information of the game.

Before we state our results about the convergence properties of the NSG, let us recall some definitions [15], [17].

Definition 1: A *better response update* is an event where a player i changes its strategy (route) and increases its payoff as a result. Starting from a strategy profile $\boldsymbol{r} = (r_i, \boldsymbol{r}_{-i})$, player i's change of strategy from r_i to $r' \in \mathcal{R}_i$ is a better response update if and only if $\rho_i(r', \boldsymbol{r}_{-i}) > \rho_i(r_i, \boldsymbol{r}_{-i})$.

Definition 2: A strategy profile $\boldsymbol{r}^*$ is a *pure Nash equilibrium* (NE) if

$$\rho_i(r_i^*, \boldsymbol{r}_{-i}^*) \geq \rho_i(r_i, \boldsymbol{r}_{-i}^*), \forall\, r_i \in \mathcal{R}_i, \forall\, i \in \mathcal{I}. \qquad (6)$$

In other words, $\boldsymbol{r}^*$ is a pure NE if and only if no player can perform a better response update from $\boldsymbol{r}^*$.

Definition 3: A game possesses the *finite improvement property* when the better response updates always converge to a pure NE within a finite number of steps, independent of the initial strategy profile or the players' updating order.

Theorem 1: Every NSG possesses the finite improvement property.

Theorem 1 and Definition 3 together imply the *existence* of and *convergence* to a pure NE. To prove Theorem 1, we show that each NSG is equivalent[1] to a *congestion game* [14]–[16]. In a congestion game, players choose resources. The payoff a player receives from using a given resource depends on the number of users choosing the same resource. Congestion games always have the finite improvement property, and hence so do our games. Full details can be found in [19].

A natural question is how long it takes for a strategy profile to converge to a pure NE. Theorem 2 ensures that each better response update can be computed in polynomial time.

Theorem 2: A player in a NSG can compute a better response update in $O(I^2)$ time.

[1]When we say two games are *equivalent* we mean that they are *weakly isomorphic* as defined in [18]

Algorithm 1 *Network Selection Game (NSG) Scheme.*

Initialization
for $i \in \mathcal{I}$
  Obtain MU i's trajectory $(k_i(t), \forall\, t \in \mathcal{T})$ based on its starting point, destination, path, and moving speed.
  Retrieve MU i's network availabilities $\mathcal{M}(i,l,t)$ $\forall\, l \in \mathcal{L}, t \in \mathcal{T}$ from the database of the MO.
  Retrieve $\mu^{(n)}$, $\forall\, n \in \mathcal{N}$, $x^{(n,t)}$, $\forall\, n \in \mathcal{N}, t \in \mathcal{T}$, and $s_{n,n'}, c_{n,n'}$, $\forall\, n, n' \in \mathcal{N}$.
  Set $\mathcal{N}(i,t) := \mathcal{M}(i, k_i(t), t)$, $\forall\, t \in \mathcal{T}$.
  Setup $\mathcal{R}_i$ from $\mathcal{N}(i,t)$, $\forall\, t \in \mathcal{T}$ and $s_{n,n'}$, $\forall\, n, n' \in \mathcal{N}$.
  Randomly pick route $r_i \in \mathcal{R}_i$.
end for
Planning Phase: Network Selection Game
while $\boldsymbol{r}$ is not a pure NE
  for $i \in \mathcal{I}$
    Perform a better response update: Find a route $r_i \in \mathcal{R}_i$ that increases player i's payoff as described by (4).
  end for
end while
Announce route r_i to MU i, $\forall\, i \in \mathcal{I}$
Network Selection Phase
for each MU $i \in \mathcal{I}$
  Select network n in each time slot t based on the network usage plan r_i.
end for

The proof of Theorem 2 is given in [19]. In order to prove Theorem 2, we show that the problem of finding a better response update can be reformulated as a shortest path problem. For the number of better response updates required for convergence, we will study in Section IV.

Our results lead us to propose a centralized network selection game scheme in Algorithm 1 for the MO to make the network selection decisions for the MUs. In the planning phase, the MO performs better response updates for the MUs until reaching a pure NE. In the network selection phase, each MU selects and switches networks according to its network usage plan (i.e., route) at the pure NE.

IV. Performance Evaluations

In this section, we evaluate the performance of the NSG scheme (i.e., Algorithm 1) by comparing it with the cellular-only scheme and the OTSO scheme. We show the impact of various system parameters on the average payoff of each MU, the total number of switching operations, and fairness under these three schemes. We also study the convergence rate and differentiated services of the NSG scheme.

For each experiment, we ran with 1000 different network trials in MATLAB and show the average value. Unless specified otherwise, we consider that cellular (LTE category 5) data rate $\mu^{(1)}$ and Wi-Fi (IEEE 802.11g) data rate $\mu^{(n)}, n \in \mathcal{N}_{\text{wifi}}$ are normally distributed random variables with means equal to 300 Mbps [20] and 54 Mbps [21], respectively, and standard deviations equal to 5 Mbps. The probability that a Wi-Fi connection is available at a particular location is equal to 0.9. We consider a two-minute duration, where the duration of a time slot $\Delta t = 10$ seconds, so $T = 12$. We assume that the switching time and cost among all the networks are the same, where $s_{n,n'} = 1$ and $c_{n,n'} = c^{\textit{switch}}, \forall\, n, n' \in \mathcal{N}, n \neq n'$. We consider that all the Wi-Fi networks are available to all

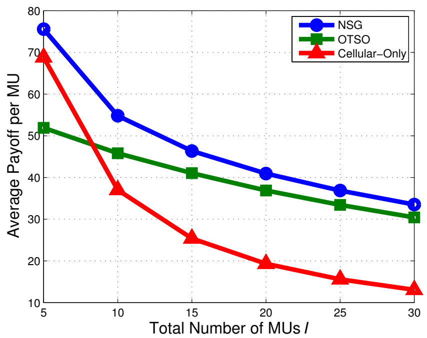

Fig. 4. The average payoff versus the total number of MUs I for $c^{switch} = 1$.

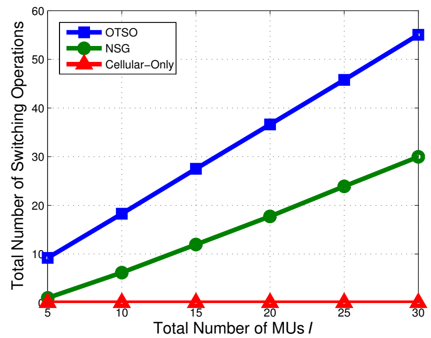

Fig. 5. Number of switching operations versus the total number of MUs I for $c^{switch} = 1$.

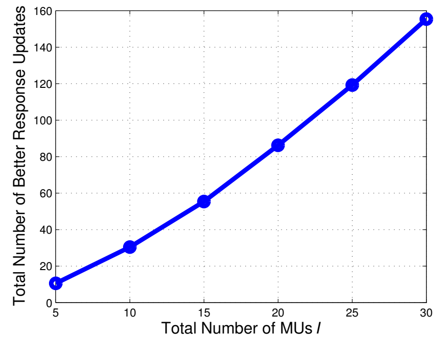

Fig. 6. Number of better response updates for convergence in the NSG scheme for $c^{switch} = 1$.

the users within its coverage all the time with a zero cost, so $\gamma(n) = 0, \forall n \in \mathcal{N}_{\text{wifi}}$. The cellular price $\gamma(1)$ is US \$3/Gbyte. We assume that the background traffic $x^{(n,t)} \in \{0, 1, 2\}, \forall n \in \mathcal{N}, t \in \mathcal{T}$, and it is a constant over time in a network. We assume that the scaling weight $a = 0.015$.

Moreover, we consider that the MUs are moving around $L = 16$ possible locations in a four by four grid (similar to that in Fig. 1), and their trajectories are generated from the same location transition matrix $P = [p(l' \,|\, l)]_{L\times L}$, where $p(l' \,|\, l)$ is the probability that an MU will move to location l' given that it is currently at location l. We assume that the probability that the MU stays at a location is $p(l \,|\, l) = 0.6, \forall l \in \mathcal{L}$. Moreover, it is equally likely for the MU to move to any one of the neighbouring locations. So for an example in Fig. 1, at location 7, the probability that the MU will move to locations 3, 6, 8, or 11 is equal to $(1 - 0.6)/4 = 0.1$. For the edge location 12, the probability of moving to any one of its three neighbouring locations is $(1 - 0.6)/3 = 0.133$. For the *cellular-only* scheme, all the MUs use the cellular network in all the time slots, so there is no network switching. For the *on-the-spot offloading* (OTSO) scheme [4], which is a static policy commonly used in the handsets, the data traffic is offloaded to the Wi-Fi network whenever Wi-Fi is available. Otherwise, the cellular connection will always be used.

First, we consider the impact of the number of MUs I in the system when switching cost $c^{switch} = 1$ and cellular price $\gamma(1) =$ US \$3/Gbyte. In Fig. 4, we plot the average payoff per MU against I. In general, when I increases, the level of congestion in the system increases, so the average payoffs under all three schemes decrease with I. Notice that when I is small, the cellular network is lightly loaded, so the cellular-only scheme, which does not incur any switching penalty, achieves a higher payoff than the OTSO scheme. However, when I is large, the OTSO scheme has a better performance, as it offloads a large amount of data traffic to the Wi-Fi network, and reduces the network congestion in the cellular network. We observe that the NSG scheme (which involves the MUs picking actions according to a pure NE generated by Algorithm 1) has the highest average payoff, which suggests that it achieves a good load balancing across the networks. In Fig. 5, we plot the total number of switching operations against I. As we can see, the number of network switchings increases linearly with I under both the NSG and OTSO schemes, and there is no switching under the cellular-only scheme.

In Theorem 2, we have established that each better response update can be computed in polynomial time. In Fig. 6, we continue with the evaluation of the convergence speed of Algorithm 1 by counting the total number of better response updates required for convergence with respect to different I for $c^{switch} = 1$ and $\gamma(1) =$ US \$3/Gbyte. We observe that Algorithm 1 scales well with the increasing MU population. In particular, each MU only needs to perform 2.11 and 5.18 better response updates for $I = 5$ and $I = 30$, respectively, before the strategy profile converges to a pure NE.

Then, we study the effects of varying the switching cost c^{switch} for $I = 30$. In Fig. 7, we plot the total number of switching operations against c^{switch}. We can see that both the cellular-only and OTSO schemes are static in that they are independent of c^{switch}. However, the NSG scheme is adaptive to the increasing switching cost by decreasing the number of switchings. In Fig. 8, we evaluate the degree of fairness among the MUs by plotting the Jain's fairness index [22] defined as $\left(\sum_{i\in\mathcal{I}} \rho_i(\boldsymbol{r})\right)^2 / \left(I \sum_{i\in\mathcal{I}} \rho_i(\boldsymbol{r})^2\right)$ against c^{switch}. Since the MUs under the cellular-only scheme always have the same payoff, its fairness index is always equal to one. Besides, we notice that the fairness indices of both the NSG and OTSO schemes decrease with c^{switch}. For the NSG scheme, as c^{switch} increases, the MUs switch networks less often as shown in Fig. 7. In this way, the payoffs among the MUs 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 payoffs among the MUs with different number of switchings. Moreover, we observe in Fig. 8 that the NSG scheme achieves a higher fairness index than the OTSO scheme, and the difference in fairness index increases with c^{switch}. It suggests that the adaptive NSG scheme results in a fairer payoff allocation than the static OTSO scheme.

Finally, we study the impact of the location, user, and time dependent Wi-Fi availability on the average payoff of the MUs. In particular, we suppose there are three *priority classes*: MUs in class 1 can access all the networks at all times, provided that a network is available at the location. MUs in class 2 can access the cellular network all the time. However, even

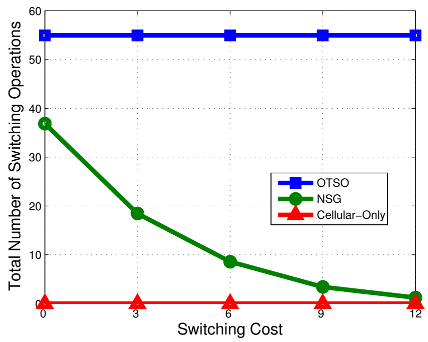

Fig. 7. The total number of switching operations versus switching cost c^{switch} for $I = 30$.

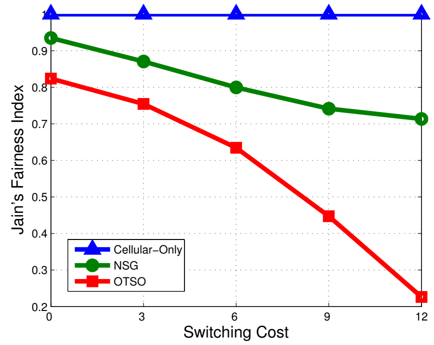

Fig. 8. The Jain's fairness index versus switching cost c^{switch} for $I = 30$.

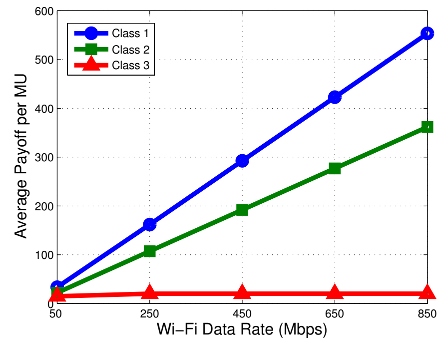

Fig. 9. The average payoff in three priority classes under the NSG scheme versus Wi-Fi data rate $\mu^{(n)}, n \in \mathcal{N}_{\text{wifi}}$ for $I = 30$.

when they are at a location with a Wi-Fi hotspot, they can only access it 50% of the time. MUs in class 3 can only access the cellular network, but not any of the Wi-Fi networks. In Fig. 9, we plot the average payoff of the MUs under the NSG scheme versus the Wi-Fi data rate $\mu^{(n)}, n \in \mathcal{N}_{\text{wifi}}$ for $I = 30$, $\mu^{(1)} = 300$ Mbps, $c^{switch} = 1$, and $\gamma(1) =$ US $6/Gbyte. As we can see, by increasing the data rate from around 50 Mbps (for IEEE 802.11g) to 850 Mbps (for IEEE 802.11ac), the average payoffs of the MUs in classes 1 and 2 increase, but the average payoff of the MUs in class 3 is roughly constant. Moreover, we observe that MUs with a higher Wi-Fi availability (class 1 here) can achieve a higher average payoff, so the NSG scheme can provide differentiated services for different priority classes.

V. Conclusions and Future Work

In this paper, we studied the congestion-aware network selection and data offloading problem in an integrated cellular Wi-Fi system, which is a practical network resource allocation problem related to the new IEEE and 3GPP standards. The challenge is to design a network selection algorithm that incorporates the practical considerations of (i) user mobility, (ii) location, user, and time dependent Wi-Fi availabilities, (iii) network dependent switching time and switching cost, and (iv) usage-based pricing. Simulation results showed that our proposed scheme based on the network selection game achieves a good load balancing. The results also revealed that the on-the-spot offloading scheme performs well under a low switching cost and a high cellular traffic load, while the cellular-only scheme is effective only under a low cellular traffic load. To the best of our knowledge, this is the first paper that studies the network selection and data offloading problem with *multiple heterogeneous users*. In the future work, we will consider the more general case with random mobility patterns, where the operator only has the statistical information of the possible trajectories of users.

References

[1] Ericsson, "Ericsson mobility report," White Paper, Nov. 2013.

[2] Cisco Systems, "Cisco visual networking index: Global mobile data traffic forecast update, 2013-2018," White Paper, Feb. 2014.

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

[4] 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.

[5] L. Gao, G. Iosifidis, J. Huang, and L. Tassiulas, "Economics of mobile data offloading," in *Proc. of IEEE International Workshop on Smart Data Pricing*, Turin, Italy, Apr. 2013.

[6] G. Iosifidis, L. Gao, J. Huang, and L. Tassiulas, "An iterative double auction for mobile data offloading," in *Proc. of IEEE WiOpt*, Tsukuba Science City, Japan, May 2013.

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

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

[9] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 9: Interworking with External Networks*, IEEE Std. 802.11u, 2011.

[10] 3GPP, "Architecture enhancements for non-3GPP accesses," 3rd Generation Partnership Project (3GPP), TS 23.402. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/23402.htm.

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

[12] 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.

[13] 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.

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

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

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

[17] X. Chen and J. Huang, "Distributed spectrum access with spatial reuse," *IEEE J. on Selected Areas in Commun.*, vol. 31, no. 3, pp. 593–603, Mar. 2013.

[18] J. Gabarro, A. Garcia, and M. Serna, "On the complexity of game isomorphism," *Mathematical Foundations of Computer Science*, vol. 4708, pp. 559–571, Aug. 2007.

[19] M. H. Cheung, R. Southwell, and J. Huang, "Congestion-aware network selection and data offloading," Tech. Rep., Feb. 2014. [Online]. Available: http://jianwei.ie.cuhk.edu.hk/publication/AppendixIntegratedOffloading.pdf.

[20] Wikipedia, "4G." [Online]. Available: http://en.wikipedia.org/wiki/4G.

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

[22] 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.