A Reinforcement Learning Approach for Cost- and Energy-Aware Mobile Data Offloading

Cheng Zhang*, Bo Gu†, Zhi Liu‡, Kyoko Yamori§‡, and Yoshiaki Tanaka¶‡
*Department of Computer Science and Communications Engineering, Waseda University, Tokyo, 169-0072 Japan
†Department of Information and Communications Engineering, Kogakuin University, Tokyo, 192-0015 Japan
‡Global Information and Telecommunication Institute, Waseda University, Tokyo, 169-8555 Japan
§Department of Management Information, Asahi University, Mizuho-shi, 501-0296 Japan
¶Department of Communications and Computer Engineering, Waseda University, Tokyo, 169-8555 Japan
Email: cheng.zhang@akane.waseda.jp

***Abstract*—With rapid increases in demand for mobile data, mobile network operators are trying to expand wireless network capacity by deploying WiFi hotspots to offload their mobile traffic. However, these network-centric methods usually do not fulfill interests of mobile users (MUs). MUs consider many problems to decide whether to offload their traffic to a complementary WiFi network. In this paper, we study the WiFi offloading problem from MU's perspective by considering delay-tolerance of traffic, monetary cost, energy consumption as well as the availability of MU's mobility pattern. We first formulate the WiFi offloading problem as a finite-horizon discrete-time Markov decision process (FDTMDP) with known MU's mobility pattern and propose a dynamic programming based offloading algorithm. Since MU's mobility pattern may not be known in advance, we then propose a reinforcement learning based offloading algorithm, which can work well with unknown MU's mobility pattern. Extensive simulations are conducted to validate our proposed offloading algorithms.**

***Index Terms*—WiFi, mobile data offloading, reinforcement learning, energy-aware**

I. Introduction

The mobile data traffic is growing rapidly. According to the investigation of Cisco Systems [1], the mobile data traffic is expected to reach 24.3 exabytes per month by 2019, while it is only 2.5 exabytes per month at the end of 2014. On the other hand, the growth rate of the mobile network capacity is far from catching up with the growth of mobile traffic demand, which has become a big problem for wireless mobile network operators (MNOs). Even though 5G technology is promising for providing huge wireless network capacity [2], it takes time for development and it is also costly. Economic methods such as time-dependent pricing [3][4][5] have been proposed to change users' usage pattern, which is not user-friendly. Up to now, the best practice for increasing mobile network capacity is to deploy complementary network (such as WiFi and femtocells), which can be quickly deployed and cost-efficient. Then, part of mobile users' (MUs)' traffic demand can be offloaded from MNOs' cellular network to the complementary WiFi network.

A mobile device automatically changing its connection type (such as from cellular network to WiFi network) is called *vertical handover* [6]. Mobile data offloading is facilitated by new standards such as Hotspot 2.0 [7] and the 3GPP Access Network Discovery and Selection Function (ANDSF) standard [8], with which information of network (such as price and network load) can be broadcasted to MU in real-time. Then MU can make offloading decision intelligently based on the real-time network information.

There are quite a number of works related to WiFi offloading problem. However, previous works either considered WiFi offloading problem from network providers' perspective without considering MU's quality of service (QoS) [9][10], or studied WiFi offloading from MU's perspective [11][12][13][14], but did not take the energy consumption as well as cost problems into consideration. An important assumption was that MU's mobility pattern was predictable, on which many previous works [13][14] were based.

In this paper, we study WiFi offloading problem from MU's perspective. MU's target is to minimize its total cost under usage based pricing, while taking monetary cost, preference for energy consumption, availability of MU's mobility pattern and application's delay tolerance into consideration.

First, a general user offloading scenario is considered, the cost- and energy-aware WiFi offloading problem is modeled as a finite-horizon discrete-time Markov decision process (FDTMDP) under the assumption that MU's mobility pattern is known in advance. We propose a dynamic programming based algorithm to solve the FDTMDP problem. However, MU's mobility pattern could not always be got in advance.

Second, to deal with the case of unknown MU's mobility pattern, an online reinforcement learning [15] based algorithm is proposed. And optimal decision is made through learning when MU's mobility pattern is unknown.

To the best of our knowledge, this is the first paper that studies the mobile data offloading problem from user's perspective with unknown mobile probability of the user, while considering the cost and energy consumption problem at the same time.

The rest of this paper is organized as follows. Section II describes the related work. Section III illustrates the system model. Section IV formulates the user's WiFi offloading problem as discrete-time finite-horizon Markov decision process and proposed a dynamic programming based algorithm. Section V proposes the reinforcement learning

based algorithm to solve the user's WiFi offloading problem. Section VI illustrates the simulation and results. Finally, we conclude this paper in Section VII.

II. Related Work

Mobile data offloading has been widely studied in the past. Gao et al. in [9] studied the cooperation among one MNO and multiple access point owners (APOs) by utilizing Nash bargaining theory, while the case of multiple MNOs and multiple APOs is studied in [10], where double auction were adopted. The aforementioned papers [9][10] considered mobile data offloading market from perspective of network without considering MUs' experience directly.

On the other hand, papers [11][12][13][14] had considered offload delay-tolerant traffic from MUs' perspective. In [11], Balasubramanian et al. implemented a prototype system called *Wiffler* to leverage delay-tolerant traffic and fast switching to 3G. Im et al. in [13] not only took a MU's throughput-delay tradeoffs into account, but also considered MU's 3G budget explicitly. MU's mobility pattern was predicted by second-order Markov chain. In [14], Cheung studied the problem of offloading delay-tolerant application for each user. A Markov decision process was formulated to minimize total data usage payment under a usage-based pricing. Even though an algorithm with low complexity was proposed and has been shown effective, one important assumption was that MU' mobility pattern was known in advance.

The above literature does not consider energy consumption problem when offloading traffic from cellular network to complementary network. Actually, the battery life has always been a concern for smartphones. [16][17] have studied how to design an energy-efficient framework for mobile data offloading. However, MU's throughput-delay trade off and budget constraint are not considered in these work. While it was shown in [12] that WiFi data offloading saved 55% of battery power due to much higher data rate WiFi can provide, it was verified in [17] that WiFi networks could consume more energy than cellular network when WiFi throughput was lower. In order to clarify the contradiction, it is necessary to consider energy consumption to establish a cost- and energy-aware mobile data offloading scheme.

Reinforcement learning has been utilized to tackle the challenge in the traffic offloading problem. In [18], Chen et al. used Q-learning to minimize energy consumption of heterogenous cellular network, which was also from the perspective of network without considering energy consumption of MU's device, let alone the offloading cost of MU.

Different from aforementioned papers, in this paper, we consider MU's monetary cost and energy consumption in our mobile data offloading problem from the perspective of users. Furthermore, we do not assume any MU's mobility pattern in advance.

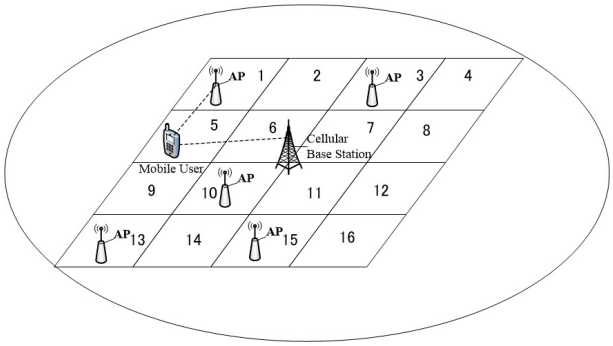

Fig. 1. System scenario.

III. System Model

Since the cellular network coverage is rather high, it is assumed that MU is always in a cellular network, but not always can access WiFi access points (APs). The WiFi APs are always deployed at home, stations, shopping malls and so on. Therefore, we assume that WiFi access is location-dependent (see Fig. 1). MU may have to wait for WiFi connection. We consider a slotted time system as $t \in \mathcal{T} = \{1, ..., T\}$, where T is the deadline for a data transmission. MU can move in L possible locations, which denoted as $\mathcal{L} = \{1, ..., L\}$. In a location $l \in \mathcal{L}$ at time $t \in \mathcal{T}$, if there is no WiFi AP, MU can choose to use the cellular network to transfer data immediately, or not to use network (idle), expecting to encounter with a location with WiFi AP in the future within the deadline T. For MU, how to make decision in location l at time t, depends on MU's tradeoff among total monetary cost, energy consumption and time left for data transferring. MU's mobility can be modelled by a Markovian model as in [13][14]. Therefore, the MU's decision making problem can be modelled as a finite-horizon Markov decision process.

We define the system *state* as in Eq. (1)

$$\boldsymbol{s} = \{l, b\} \tag{1}$$

where $l \in \mathcal{L} = \{1, ..., L\}$ is the MU's location index, which can be got from GPS. $\mathcal{L}$ is the location set of MU. $b \in \mathcal{B} \subseteq [0, B]$ denotes the remaining size of delay-tolerant traffic to be transferred within T time slot, where B is the total size.

MU's *action* a at each decision epoch is to determine whether to transmit data through WiFi (if WiFi is available), or cellular network, or just keep idle. Therefore, MU's action set is denoted as in Eq. (2)

$$\mathcal{A} = \{idle, WiFi, cellular\} \tag{2}$$

At each epoch t, three factors affect MU's decision.

- (1) *the monetary cost*: it is the payment from MU to network service provider, we assume that the network service provider adopts usage based price $p(l, a)$. Please note that this price is also dependent on location l and action a. If WiFi is available at location l, MU may choose action $\{idle\}$, $\{WiFi\}$, or $\{cellular\}$. Hence, $p(l, \{idle\})$ equals to 0, while $p(l, \{WiFi\})$ and $p(l, \{cellular\})$ are the

prices set by MNO and APO, respectively. We denote $\gamma(l,a)$ as throughput in bps at location l with action a, which is also dependent on the location and action of MU. Obviously, $\gamma(l,\{idle\})$ is equal to 0. We define the monetary cost $c_t(\boldsymbol{s},a)$ as in Eq. (3)

$$c_t(\boldsymbol{s},a) = \min\{b,\gamma(l,a)\}p(l,a) \tag{3}$$

- (2) *the energy consumption*: it is the energy consumed when transmitting data through WiFi or cellular network. We denote MU's awareness of energy as in Eq. (4)

$$\xi_t(\boldsymbol{s},a) = \theta_t\varepsilon(l,a)\min\{b,\gamma(l,a)\} \tag{4}$$

  where $\varepsilon(l,a)$ is energy consumption in joule/bits at location l with action a. It has been shown in [17] that $\varepsilon(l,a)$ was a decreasing function of throughput. θ_t is MU's preference for energy consumption. θ_t is the weight on energy consumption set by MU. For example, if MU can soon charge his smartphone, he may set θ_t to a small value, or if MU is in an urgent status and could not charge in a short time, he may set a large value for θ_t to save energy consumption. We evaluate the effect of different θ_t in Section VI.
- (3) *the penalty*: if the data transmission is not finished in deadline T, the penalty for MU is defined as Eq. (5).

$$\hat{c}_{T+1}(\boldsymbol{s}) = \hat{c}_{T+1}(l,b) = g(b) \tag{5}$$

  where $g(k)$ is a non-negative non-decreasing function. $T+1$ means that the penalty is calculated after deadline T.

The probabilities associated with different state changes are called transition probabilities. We denote *transition probability* as in Eq. (6)

$$\Pr(\boldsymbol{s}'|\boldsymbol{s},a) \tag{6}$$

Eq. (6) shows the probability of state $\boldsymbol{s}'$ if action a is chosen at state $\boldsymbol{s}$. It is assumed that the remaining size is independent of location change, we have

$$\begin{aligned}\Pr(\boldsymbol{s}'|\boldsymbol{s},a) &= \Pr((l',b')|(l,b),a)\\ &= \Pr(l'|l)\Pr(b'|(l,b),a)\end{aligned} \tag{7}$$

where

$$\Pr(b'|(l,b),a) = \begin{cases} 1 & \text{if } b' = [b-\gamma(l,a)]^+ \text{ and a}\in\{WiFi, cellular\} \\ 0 & \text{otherwise}\end{cases} \tag{8}$$

$[x]^+$ is equal to $\max\{x,0\}$. MU's probability from l to l' is denoted as $Pr(l'|l)$, which is assumed as known (see Assumption 1).

Assumption 1: MU's mobile probability from current place to the next place is known in advance.

MU's mobility pattern can be derived from MU's historical data, which has been widely studied in the literature, such as [13].

MU's *policy* is defined as in Eq. (9)

$$\pi = \left\{\phi_t(l,b),\ \forall\ t\in\mathcal{T}, l\in\mathcal{L}, b\in\mathcal{B}\right\} \tag{9}$$

where $\phi_t(l,b)$ is a function mapping from state $\boldsymbol{s}=(l,b)$ to an decision action at t. The set of π is denoted as Π. If policy π is adopted, the state is denoted as $\boldsymbol{s}^\pi$.

The objective of MU is to minimize the expected total cost (include the monetary cost and the energy consumption) from $t=1$ to $t=T$ and penalty at $t=T+1$ by a optimal π^* (see Eq. (10))

$$\min_{\pi\in\Pi} E^\pi_{s_1}\left[\sum_{t=1}^{T} r_t(\boldsymbol{s}^\pi,a) + \hat{c}_{T+1}(\boldsymbol{s}^\pi)\right] \tag{10}$$

where $r_t(\boldsymbol{s},a)$ is the sum of the monetary cost and the energy consumption as in Eq. (11)

$$r_t(\boldsymbol{s},a) = c_t(\boldsymbol{s},a) + \xi_t(\boldsymbol{s},a) \tag{11}$$

IV. Dynamic Programming Based Algorithm

In this section, Eq. (10) is a standard problem of finite-horizon discrete-time Markov decision process (MDP). We propose a dynamic programming based algorithm to solve the problem.

For a MDP problem, it is important to identify the *optimality equation* (or *Bellman equation*) [19]. Denote $\mathcal{V}_t(\boldsymbol{s})$ as the minimal expected total cost of the MU from t to $T+1$ at state $\boldsymbol{s}$. The Bellman equation is defined as in Eq. (13).

$$\mathcal{V}_t(\boldsymbol{s}) = \min_{a\in\mathcal{A}}\left\{Q_t(\boldsymbol{s},a)\right\} \tag{12}$$

where for $l\in\mathcal{L}$, $b\in\mathcal{B}$ and $a\in\mathcal{A}$, we have

$$\begin{aligned}&Q_t(\boldsymbol{s},a)\\ &= r_t(\boldsymbol{s},a) + \sum_{l'\in\mathcal{L}}\sum_{b'\in\mathcal{B}}\Pr(\boldsymbol{s}'|\boldsymbol{s},a)\mathcal{V}_{t+1}(\boldsymbol{s}')\\ &= \underbrace{c_t(\boldsymbol{s},a)+\xi_t(\boldsymbol{s},a)}_{\text{cost for the current } t} + \underbrace{\sum_{l'\in\mathcal{L}}\sum_{b'\in\mathcal{B}}\Pr((l',b')|(l,b),a)\mathcal{V}_{t+1}(l',b')}_{\text{expected future cost start from } t+1}\\ &= \min\{b,\gamma(l,a)\}p(l,a) + \theta_t\varepsilon(l,a)\min\{b,\gamma(l,a)\}\\ &+ \sum_{l'\in\mathcal{L}}\Pr(l'|l)\mathcal{V}_{t+1}(l',[b-\gamma(l,a)]^+)\end{aligned} \tag{13}$$

Based on the Bellman equation Eq. (13), we propose Algorithm 1. In the optimal policy calculation phase, optimal policy is calculated by backward induction from epoch T to 1, where $\sigma>0$ is the granularity of the total data B. Then, MU's offloading data policy is decided in each slot in offloading data transmission phase. It is obvious that the time complexity of Algorithm 1 is $O(TLB/\sigma)$.

Theorem 1: The policy $\pi^* = \left\{\phi^*_t(l,b),\ \forall\ t\in\mathcal{T}, l\in\mathcal{L}, b\in\mathcal{B}\right\}$ generated in Algorithm 1 is the problem (10)'s optimal solution.

Proof: It is obvious by the principle of optimality in [19].

Q.E.D

Algorithm 1: Dynamic Programming Based Algorithm

Optimal Policy Calculation Phase
Set $\mathcal{V}_{T+1}(l,b)$,$\forall\ l \in \mathcal{L}$, $k \in \mathcal{B}$ by Eq. (5)
Set t:=T
while $t \geq 1$:
  for $l \in \mathcal{L}$:
    Set b :=0
    while $b \leq B$:
      Calculate $Q_t(\boldsymbol{s}, a)$, $\forall\ a \in \{WiFi, cellular\}$ using Eq. (13)
      Set $\phi_t^*(l,b) := \arg\min_{a\in\mathcal{A}}\{Q_t(\boldsymbol{s},a)\}$
      Set $\mathcal{V}_t(l,b) := Q_t(\boldsymbol{s}, \phi_t^*(l,b))$
      Set b:=$b + \sigma$
    end while
  end for
  Set t:=$t-1$
end while
The optimal policy π^* is generated for the following offloading data transmission phase

Offloading Data Transmission Phase
Set $t := 1$, $b := B$
while $t \leq T$ and $b > 0$:
  l is determined from GPS
  Set action $a := \phi_t^*(l,b)$ according to π^* (the optimal policy)
  if $a \in \{WiFi, Cellular\}$
    Transmit $\gamma(l, \{Cellular\})$ bits data to the cellular network, or offload $\gamma(l, \{WiFi\})$ to the WiFi network
    Set b :=$[b - \gamma(l,a)]^+$
  end if
  Set $t := t + 1$
end while

V. Reinforcement Learning Based Offloading Algorithm

It is assumed in Assumption 1 that MU's mobility pattern is known, then transition probability (see Eq. (6)) also can be calculated in advance for optimal policy calculation in Algorithm 1. However, the MU's mobility pattern may not be easily gotten or not be so correct. Therefore, one key question may be asked is as follows: *How does MU set policy to solve the problem (10) if MU's mobility pattern is unknown?*

In order to solve the problem (10) with unknown MU mobility pattern, we propose a reinforcement learning based algorithm in this section.

In reinforcement learning algorithm, an agent makes optimal decision by acquiring knowledge of the unknown environment through learning. We adopt the temporal difference (TD) learning algorithm [15], which requires no model for the environment and are fully incremental. Specifically, a Q-learning algorithm [18] is employed. The optimal policy can be obtained from optimal Q-value, $Q_t^*(\boldsymbol{s}, a)$, which is shown in Eq. (14)

$$\phi_t^* = \arg\min_{a\in\mathcal{A}} Q_t^*(\boldsymbol{s}, a) \tag{14}$$

In order to learn $Q_t^*(\boldsymbol{s}, a)$, we use the following update rule in Eq. (15)

$$Q_{t+1}(\boldsymbol{s}, a) = Q_t(\boldsymbol{s}, a) + \alpha_t \delta_t \tag{15}$$

where $\alpha_t \in (0, 1]$ is the learning rate and

$$\delta_t = r_t(\boldsymbol{s}, a) + \lambda \min_{a'\in\mathcal{A}} Q_t(\boldsymbol{s}', a') - Q_t(\boldsymbol{s}, a) \tag{16}$$

is the TD at t. $\lambda \in (0, 1]$ is the discounting rate. It is obvious that transition probability (or MU's mobility pattern) is no longer needed in Q-learning.

Assumption 2: The learning rate α_t is assumed to satisfy the following equations.

$$\sum_{t=1}^{T} \alpha_t = \infty \text{ and } \sum_{t=1}^{T} \alpha_t^2 < \infty \tag{17}$$

We propose reinforcement learning based algorithm as shown in Algorithm 2. It is obvious that the time complexity is $O(TB/\sigma)$, which is much less than dynamic programming based algorithm in Section IV. For the convergence of the proposed Q-learning algorithm in Algorithm 2, we presents Theorem 2 as follows.

Theorem 2: Under ϵ-greedy[1] policy, the proposed Q-learning algorithm in Algorithm 2 converges to the optimal solution with probability one.

Proof: By [15], this is obvious since that the learning rate α_t satisfies Assumption 2.

Q.E.D

Algorithm 2: Reinforcement Learning Based Algorithm

Set $t := 1$, $b := B$,
while $t \leq T$ and $b > 0$:
  l is determined from GPS
  $rnd \leftarrow$ random number in [0,1]
  if $rnd < \epsilon$:
    Choose action a randomly
  else:
    Choose action a based on Eq. (14)
  end if
  if $a \in \{WiFi, Cellular\}$
    Transmit $\gamma(l, \{Cellular\})$ bits data to the cellular network, or offload $\gamma(l, \{WiFi\})$ to the WiFi network
  end if
  Calculate $r_t(\boldsymbol{s}, a)$ by Eq. (11), and set $\boldsymbol{s}' = (l, [b - \gamma(l,a)]^+)$
  Set δ_t:=$r_t(\boldsymbol{s}, a) + \lambda \min_{a'\in\mathcal{A}} Q_t(\boldsymbol{s}', a') - Q_t(\boldsymbol{s}, a)$
  Set $Q_{t+1}(\boldsymbol{s}, a) := Q_t(\boldsymbol{s}, a) + \alpha_t \delta_t$
  Set b :=$[b - \gamma(l,a)]^+$
  Set $t := t + 1$
end while

VI. Performance Evaluation

In this section, the performance of our dynamic programming based algorithm and reinforcement learning based algorithm are evaluated by comparing them with DAWN [14] in terms of the total payment, the energy consumption, and the probability of completing file transfer. We developed a simulator by Python 2.7, which can be downloaded from our website [20].

A four by four grid as shown in Fig. 1 is used. Therefore, L is 16. Four WiFi APs are randomly deployed in L locations. The cellular usage price is assumed as US \$10/Gbyte, while the WiFi usage price is US \$1/Gbyte. $p(l|l) = 0.6$ means that the probability that MU stays in the same place from time t to t' is 0.6. And MU moves to the neighbour location with equal probability, which can be calculated as $p(l'|l) =$

[1] ϵ-greedy means that instead of selecting action based on action-value estimates all the time, to select an action at random with a small probability ϵ,

TABLE I
ENERGY VS. THROUGHPUT.

Throughput (Mbps)	Energy (joule/Mpbs)
11.257	0.7107
16.529	0.484
21.433	0.3733

$(1 - 0.6)/(\text{number of neighbors})$. For example, if l is 11, there are 4 neighbours {7,10,12,15} for location 11. Then the probability to the neighbour location from t to t' is $p(l'|l) = (1 - 0.6)/4 = 0.1$. WiFi throughput $\gamma(l, \{WiFi\})$ is assumed as 15 Mbps[2], while cellular throughput $\gamma(l, \{Cellular\})$ is 10 Mbps[3]. The throughput standard deviation of both WiFi and cellular network is assumed to 5 Mbps. σ in Algorithm 1 is assumed as 1 Mbits. Time for each epoch is 5 seconds. The penalty function is $g(b) = b^2$ [14]. Both the learning rate α_t and discounting rate λ in Algorithm 2 is set to 0.1.

For the energy consumption is a decreasing function of throughput, we have the sample data from [21] (see Table I). We then fit the sample data by a two order polynomial function in [0,25] as shown in Fig. 2. Please note that the energy consumption of cellular and WiFi may be different for the same throughput, but we assume they are the same and use the same fitting function as in Fig. 2. MU's energy preference θ_t is time-dependent, but we assume that $\theta_t = \theta$ is not time-dependent in our simulation.

Firstly, we compare the performance of our proposed

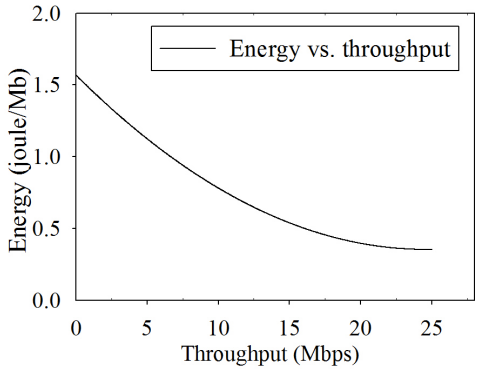

Fig. 2. Energy consumption (joule/Mb) vs. Throughput (Mbps).

dynamic programming (DP) based algorithm with DAWN algorithm in [14]. In Fig. 3, we showed that the energy consumption is decreasing with MU's energy preference θ. Large θ means that MU are concerned with energy consumption, while $\theta = 0$ means MU do not care about energy consumption. Obviously, the proposed DP algorithm is equal to the DAWN algorithm when $\theta = 0$. Fig. 3 shows that energy consumption by DAWN algorithm is the highest. MU can save energy by setting θ greater than 0 according to

[2]Even though WiFi can achieve much higher throughput, we tested many times by a iPhone 5s on one of biggest Japanese wireless carrier's public WiFi AP. The average throughput is 15 Mbps.

[3]We also tested by a iPhone 5s on one of biggest Japanese wireless carrier's cellular network. We use the classic value 10 Mbps for cellular throughput.

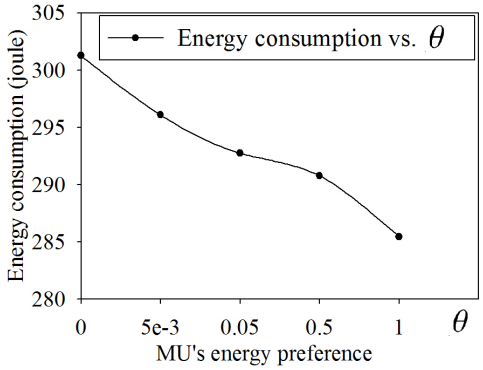

Fig. 3. Energy consumption (joule) vs. MU's energy preference.

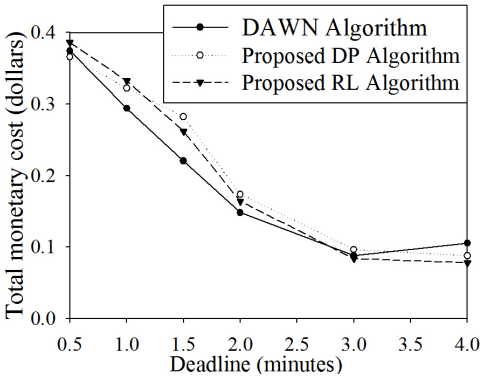

Fig. 4. Monetary cost vs. Deadline with θ=0.5.

his/her preference.

In Fig. 4, we plot total monetary cost of the DAWN algorithm, our proposed DP and RL algorithm with different deadline for data transfer. It is shown that our proposed DP algorithm cost a little bit more than the DAWN algorithm. And our proposed RL algorithm cost a little less than DP algorithm when deadline is long. The reason is that it costs MU much more when some cheap and energy-consuming WiFi APs are eliminated in our proposed DP algorithm. This means that MU faces the trade-off between monetary cost and energy consumption cost. If MU is concerned with energy consumption cost, he/she may face much higher monetary cost.

In Fig. 5, it is shown that the probability of completing data transfer almost the same between the DAWN algorithm and our proposed DP algorithm. When deadline is 0.5 minutes, our proposed DP algorithm is a little better than the DAWN algorithm.

Secondly, we we compare the performance of our proposed reinforcement learning (RL) based algorithm with the DP algorithm and the DAWN algorithm when MU's mobile probability is unknown. In offloading data transmission phase of Algorithm 1, MU moves randomly to the next location when MU's mobility pattern is unknown. In Fig. 6, it is shown that the probability of completing data transfer become rather bad when MU's mobility pattern is unknown. But our proposed

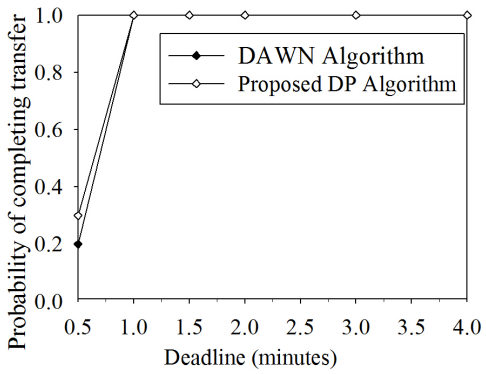

Fig. 5. Probability of completing transfer vs. Deadline with θ=0.5.

RL algorithm performance much better.

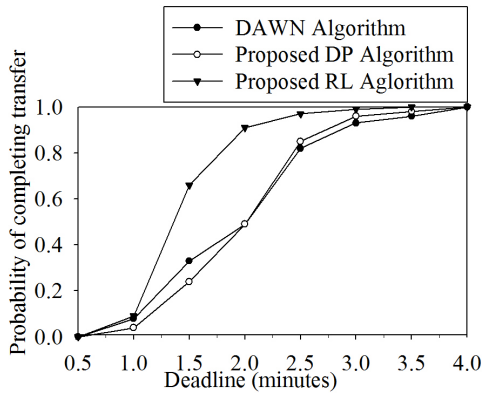

Fig. 6. Probability of completing transfer vs. Deadline with θ=0.5.

VII. Conclusion

In this paper, we study WiFi offloading problem from MU's perspective. MU's target is to minimize its total cost under usage based pricing, while taking monetary cost, preference for energy consumption, availability of MU's mobility pattern and application's delay tolerance into consideration.

A general user offloading scenario is considered, the cost- and energy-aware WiFi offloading problem is modeled as a finite-horizon Markov decision process under the assumption that MU's mobility pattern is known in advance. We propose a dynamic programming algorithm to solve the MDP problem. When MU's mobility pattern could not be got in advance, an online reinforcement learning based algorithm is proposed. Simulation results show that MU can tradeoff the monetary cost and energy consumption by setting different energy consumption preference. It is also shown that reinforcement learning based algorithm works well even when MU's mobility pattern is unknown, while dynamic programming based algorithm's performance is rather bad with unknown MU's mobility pattern.

In the future, we will consider to use MU's history information to improve performance of reinforcement learning based algorithm. While a single MU is considered in this paper, it will be interesting to consider multiple MUs in the system.

References

[1] Cisco Systems, "Cisco visual networking index: Global mobile data traffic forecast update, 2014-2019," Feb. 2015.

[2] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 71–78, March 2014.

[3] C. Zhang, B. Gu, K. Yamori, S. Xu, and Y. Tanaka, "Duopoly competition in time-dependent pricing for improving revenue of network service providers," *IEICE Trans. Commun.*, vol. E96-B, no. 12, pp. 2964–2975, Dec. 2013.

[4] C. Zhang, B. Gu, Z. Liu, K. Yamori, and Y. Tanaka, "Oligopoly competition in time-dependent pricing for improving revenue of network service providers considering different qos functions," *Proc. 15th Asia-Pacific Network Operations and Management Symposium (APNOMS 2015), Busan, Korea*, pp. 273–278, Aug. 2015.

[5] C. Zhang, B. Gu, K. Yamori, S. Xu, and Y. Tanaka, "Oligopoly competition in time-dependent pricing for improving revenue of network service providers with complete and incomplete information," *IEICE Trans. Commun.*, vol. E98-B, no. 1, pp. 30–32, Jan. 2015.

[6] J. Márquez-Barja, C. T. Calafate, J.-C. Cano, and P. Manzoni, "Review: An overview of vertical handover techniques: Algorithms, protocols and tools," *Comput. Commun.*, vol. 34, no. 8, pp. 985–997, June 2011.

[7] Cisco Systems, "The future of hotspots: Making wi-fi as secure and easy to use as cellular," *White Paper*, 2011.

[8] Alcatel and British Telecommunications, "Wi-fi roaming building on andsf and hotspot2.0," *White Paper*, 2012.

[9] L. Gao, G. Iosifidis, J. Huang, L. Tassiulas, and D. Li, "Bargaining-based mobile data offloading," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1114–1125, June 2014.

[10] G. Iosifidis, L. Gao, J. Huang, and L. Tassiulas, "A double-auction mechanism for mobile data-offloading markets," *IEEE/ACM Trans. Netw.*, vol. 22, no. 4, pp. 1271–1284, Aug. 2014.

[11] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3g using wifi," in *Proc. 8th international conference on Mobile systems, applications, and services (MobiSys 2010)*, June 2010, pp. 209–222.

[12] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can wifi deliver?" *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 536–550, April 2013.

[13] Y. Im, C. Joe-Wong, S. Ha, S. Sen, T. . Kwon, and M. Chiang, "AMUSE: Empowering users for cost-aware offloading with throughput-delay tradeoffs," in *Proc. IEEE Conference on Computer Communications (INFOCOM 2013)*, April 2013, pp. 435–439.

[14] M. H. Cheung and J. Huang, "DAWN: Delay-aware wi-fi offloading and network selection," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1214 – 1223, June 2015.

[15] P. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.

[16] N. Ristanovic, J.-Y. L. Boudec, A. Chaintreau, and V. Erramilli, "Energy efficient offloading of 3g networks," in *Proc. 2011 IEEE 8th International Conference on Mobile Adhoc and Sensor Systems (MASS 2011)*, Oct. 2011, pp. 202–211.

[17] A. Y. Ding, B. Han, Y. Xiao, P. Hui, A. Srinivasank, M. Kojo, and S. Tarkoma, "Enabling energy-aware collaborative mobile data offloading for smartphones," in *Proc. 10th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON 2013)*, June 2013, pp. 487–495.

[18] X. Chen, C. Wu, Y. Zhou, and H. Zhang, "A learning approach for traffic offloading in stochastic heterogeneous cellular networks," in *Proc. 2015 IEEE International Conference on Communications (ICC 2015)*, June 2015, pp. 3347–3351.

[19] R. Bellman, *Dynamic Programming*, Princeton University Press, 1957.

[20] C. Zhang, "Python based numerical simulator," Oct. 2015. [Online]. Available: https://sites.google.com/site/abbottzhang2015/offloading

[21] A. Murabito, "A comparison of efficiency, throughput, and energy requirements of wireless access points," *Report of InterOperability Laboratory, University of New Hampshire*, March 2009. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf