Data Offloading in Mobile Cloud Computing: A Markov Decision Process Approach

Dongqing Liu*†, Lyes Khoukhi†, Abdelhakim Hafid*
*Department of Computer Science and Operational Research
University of Montreal, QC, Canada
Email: ahafid@iro.umontreal.ca
†Environment and Autonomous Networks Lab (ERA)
University of Technology of Troyes, France
Email: {dongqing.liu, lyes.khoukhi}@utt.fr

***Abstract*—In this paper, we study mobile data offloading problem under the architecture of mobile cloud computing (MCC), where mobile data can be delivered by cellular, WiFi and Device-to-Device (D2D) communication networks. In order to minimize the overall cost for data delivery task, it is crucial to reduce cellular network usage while satisfying delay requirements. In the proposed model, a portion of the cellular data traffic is offloaded through WiFi and D2D networks. We formulate the data offloading problem as a finite horizon Markov Decision Process (FHMDP). We solve the problem using hybrid offloading algorithm for delay sensitive and delay tolerant applications. The simulation results show that the proposed offloading scheme can achieve minimal total cost compared with other three offloading schemes.**

I. Introduction

Global mobile traffic is growing dramatically in recent years. This is because the number of smart mobile phones and data-heavy mobile applications, such as video streaming and cloud backup, is increasing rapidly. According to a report from Cisco, global mobile data traffic grew 74% in 2015, while mobile network (cellular) connection speeds only grew 20% [1]. Moreover, it is forecasted that cloud applications will account for 84% mobile data traffic in 2017, compared with 74% by the end of 2012 [2]. The growing speed of mobile traffic will push the current cellular network to the limit. The Quality of Experience (QoE) of mobile services will not be guaranteed without the high-speed and stable network connections between mobile service subscribers (MSs) and mobile network operator (MNO) [3]. However, it is impractical to keep extending the current cellular network infrastructure to improve QoE, given the corresponding expensive investment. In order to cope with this problem, mobile data offloading technology can be an alternative solution. Offloading is considered as a promising technique to move data traffic from cellular network to other wireless networks; indeed, it represents a complementary wireless technology to transfer data originally targeted to flow through cellular network [4]. When MNO sends mobile data to MS, it will be able to choose from many wireless networks instead of only cellular network. In addition to cellular network, current target wireless networks include WiFi and D2D networks.

WiFi offloading has become a conventional solution to reduce mobile data traffic in cellular network. The WiFi Access Points (APs) covered in cellular network can be used efficiently to reduce data traffic [5, 6]. The authors in [7] show that about 65% of data traffic in cellular network can be offloaded to WiFi network. The result is based on the assumption that most of the time, MS stays at home/office. However, when MS moves around in cellular network, the WiFi connection time is reduced greatly. The temporal coverage decreases to 11%, when MSs move around. Although WiFi APs can provide better transfer speed than cellular network, their coverage area is much smaller than cellular network [8–11].

Another mobile offloading method is based on D2D communication network [12], where mobile devices can connect with each other directly. The data transfer uses the strategy called store-carry-forward. In this strategy, some mobile users can store data in the buffer (called mobile helpers, MHs), carry the data when they are moving, and forward the data to MS when they can connect with each other [13]. This strategy requires a well designed mobility prediction model and is widely used in D2D network [14]. When MNO wants to deliver data to a set of MSs, it can first send the data to some MHs. Then, MHs will help transfer data to MSs using opportunistic connections. With more than half a billion mobile devices and connections added in 2015 [1], D2D network is becoming an important data delivery scheme. Unlike WiFi AP, MH, which carries mobile data, can move randomly. However, the transfer speed of D2D network is low and the mobility patterns of MHs or MSs are difficult to predict.

In this paper, we consider a hybrid offloading model for a single MS in MCC. In MCC, both cellular and WiFi networks receive mobile data from cloud and then transfer the data to mobile users [2, 15]. In this model, MS moves around in the coverage area of seamless cellular network. Since the coverage area of WiFi AP is limited, there will be many WiFi APs that MS can access during the data delivery process. We assume that both the base station and WiFi APs are located at some fixed sites. In this cellular network, many mobile users (including MS and MH) can receive data directly from WiFi or cellular network. MH can also transfer data to other MSs. In

978-1-4673-8999-0/17/$31.00 ©2017 IEEE

the hybrid offloading model, MNO can deliver mobile data to MSs with three methods in MCC: (1) WiFi AP based mobile data delivery: MNO can send data directly to the mobile user if the user location is covered by WiFi. Since the transfer speed and stability can be guaranteed, this will be the best situation MNO can expect; (2) MH based mobile data delivery: It is based on MHs willing to share their mobile device resources to help MNO with data delivery process. MHs get rewards from MNO in return; the coverage area of MHs can be quite small. The success of this delivery method is based on two factors (a) MHs and common mobile users request same kinds of data, and (b) they are near to each other; and (3) cellular network based mobile data delivery: When a mobile user cannot receive the desired data by the first two methods, MNO will send data to MSs using this method.

The main contributions of our paper can be summarized as follows:

- We propose a hybrid offloading model under the concept of MCC, where mobile data can be delivered through three wireless networks, namely cellular, WiFi and D2D networks. MNO decides when to use which network to transfer data in order to minimize the total cost.
- We formulate the decision problem in a hybrid wireless network as a finite horizon Markov Decision Process (FHMDP) problem, and propose an offloading algorithm that can tackle different delay requirements (i.e. loose and tight delay tolerant).
- The simulation results show that our algorithm achieves the best cost, compared to three offloading schemes.

The rest of this paper is organized as follows. Section II describes the hybrid mobile data offloading model. Section III defines the formulation for finite horizon MDP. Section IV proposes an offloading algorithm based on value iteration. Section V evaluates the performance of the proposed offloading scheme. Section VI concludes the paper.

II. System Model

In MCC, mobile devices can access mobile services in two ways, i.e., through cellular network and WiFi AP, as shown in Figure 1. For cellular network, mobile devices are connected to the base station, which is connected to the cloud. For WiFi network, mobile devices are connected to the WiFi AP, which is connected to the cloud. All the mobile devices can connect to the cloud through cellular base station or WiFi AP. Since the relatively high data transmission cost for cellular network may prevent some mobile users from using MCC for cloud applications, we use WiFi network whenever possible to reduce the total cost. Moreover, we consider using D2D network to further reduce the overall cost when WiFi is not available.

Our model aims to offload mobile data as much as possible using WiFi and D2D networks, in order to reduce the total cost. As a side effect, this will reduce congestion in cellular network, since the connection requests to cellular network will decrease, due to alternative wireless networks. It is worth noting that the locations of WiFi APs and the base station are stationary, while MHs are moving around in the coverage area of base station. MHs can be considered as supplementary to WiFi APs because of their mobility.

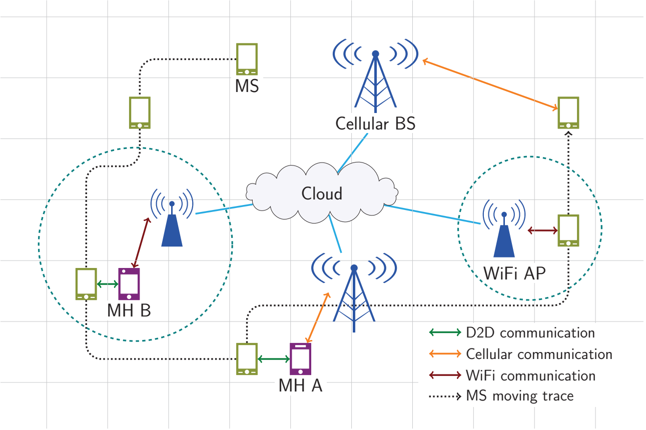


Fig. 1. Mobile data offloading in Mobile Cloud Computing

Our idea is to make use of delay tolerance of mobile data and mobility of mobile users to seek opportunities of using WiFi and D2D networks. The possibility of offloading cellular traffic to WiFi and D2D networks depends greatly on the delay requirements (i.e. tolerant or not) of mobile data. If the data is delay tolerant, then MNO can defer transmission to increase the possibility of using other networks to implement the transmission task. Otherwise (i.e. delay sensitive data), MNO is unlikely to offload mobile data from cellular network. In order to maximize the offloading probability and satisfy delay constraints, MNO has to make offloading decisions for each mobile user in cellular network. In this paper, we propose a finite horizon Markov Decision Process (FHMDP) to formulate this problem.

We consider a scenario where MNO delivers mobile data from cloud to a single MS in the coverage area of seamless cellular network. In Figure 1, there are two kinds of mobile users: MS and MH. The dotted line shows the movement of mobile users and the full line shows the data transmission process. When MS moves around, it receives data from three wireless networks: (1) WiFi based data transmission: MS can receive data from WiFi AP directly; (2) cellular network based data transmission: MS can receive data directly from the cellular base station (Cellular BS); (3) D2D based data transmission: When MS meets MH A/B that stores the requested data by MS, MS receives data from MH A/B. The objective of using three wireless networks is to minimize the cellular network usage, while satisfying delay constraint for MS.

III. Problem Formulation

In this section, we formulate the mobile data offloading problem in MCC as a FHMDP problem. For each offloading problem, we assume that data of size K needs to be transferred before deadline D, and L is the number of grids (or locations) that MS can move around before D. The system state for a

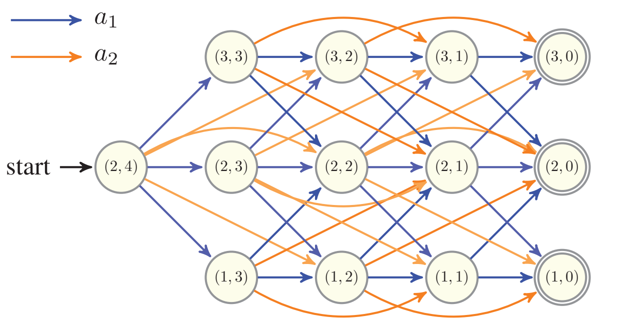


Fig. 2. A sample reduced state transition for MS, where the first digit is the location of MS l, and the second digit is data size k. Action a_1 can transfer 1 data size each time, while a_2 can transfer 2 data size each time. The state with double circle is the terminal state, where $k = 0$, such as states $(1, 0)$, $(2, 0)$ and $(3, 0)$.

single MS and multiple MHs is defined as $s = (l, u, k, \mathcal{H})$, where $l \in \mathcal{L} = \{1, 2, \cdots, L\}$ is the location of MS, u denotes data type and $k \in \mathcal{K} = \{0, 1, \cdots, K\}$ is the data size to be transmitted. The set $\mathcal{H}$ includes the locations of MHs. MNO needs to choose a wireless network at each decision epoch $d \in \mathcal{D} = \{1, 2, \cdots, D\}$. If $k = 0$ when $d \leq D$, the data offloading process is completed.

We consider that cellular network can provide seamless coverage to all the grids. We classify the grids by available WiFi or D2D networks: (1) $\mathcal{L}_d^1 = \{l \in \mathcal{L}$: *l can only access cellular network*$\}$; (2) $\mathcal{L}_d^2 = \{l \in \mathcal{L}$: *l can access WiFi network*$\}$; (3) $\mathcal{L}_d^3 = \{l \in \mathcal{L}$: *l can access D2D network*$\}$; (4) $\mathcal{L}_d^4 = \{l \in \mathcal{L}$: *l can access WiFi and D2D networks*$\}$. Since the locations of MHs may be different at each decision epoch, $\mathcal{L}_d^1, \mathcal{L}_d^2, \mathcal{L}_d^3$ and $\mathcal{L}_d^4$ change over time and satisfy the following Eqs. (1) and (2).

$$\mathcal{L}_d^1 \cup \mathcal{L}_d^2 \cup \mathcal{L}_d^3 \cup \mathcal{L}_d^4 = \mathcal{L} \tag{1}$$

$$\mathcal{L}_d^1 \cap \mathcal{L}_d^2 \cap \mathcal{L}_d^3 \cap \mathcal{L}_d^4 = \varnothing \tag{2}$$

The state variable u represents the mobile data type. We consider that each data type has a corresponding delay requirement (e.g. loose delay or tight delay). $\mathcal{U}$ represents the set of data types. To simplify the model, we consider two data types, each of which has different QoS requirements. More specifically, data types (e.g. VoIP) that are delay-sensitive belong to set $\mathcal{U}^1$; the other types (e.g. software update) are in set $\mathcal{U}^0$. Thus, $\mathcal{U} = \mathcal{U}^0 \cup \mathcal{U}^1$.

There are four actions corresponding to four offloading decisions for MS. At each decision epoch, MNO selects one of the actions for data transmission. Formally, action $a \in \mathcal{A} = \{1, 2, 3, 4\}$: (1) $a = 1$ means that MS will wait for a chance to receive data from WiFi or D2D network; (2) $a = 2$ means that MS will receive data from cellular network; (3) $a = 3$ means that MS will receive data from WiFi network; (4) $a = 4$ means that MS can receive data from D2D network. Notice that $a = 3$ is available when MS is in WiFi coverage and $a = 4$ is available when MS can access MH. In our model, we also consider the influence of different mobile data types. For delay sensitive data, we cannot use D2D network to transmit it because of its slow data rate. The action available at location l is defined as $a \in \mathcal{A}(l, u) \subseteq \mathcal{A}$.

We define the available actions according to the location of mobile users and the type of mobile data.

$$\mathcal{A}(l, u) = \begin{cases} \{1, 2\}, & l \in \mathcal{L}_d^1,\ u \in \mathcal{U} \\ \{1, 2, 3\}, & l \in \mathcal{L}_d^2,\ u \in \mathcal{U} \\ \{1, 2, 4\}, & l \in \mathcal{L}_d^3,\ u \in \mathcal{U}^0 \\ \{1, 2, 3, 4\}, & l \in \mathcal{L}_d^4,\ u \in \mathcal{U}^0 \end{cases} \tag{3}$$

We define the action cost function according to the action taken at each time slot (i.e. the period between two decision epochs). The transition cost $c_d(s, a)$ is equal to the action cost function $cost(a)$.

$$c_d(s, a) = cost(a) = \nu_a \chi_a \tag{4}$$

where ν_a and χ_a are the network data rate and the price of data size unit for action a (e.g., $\nu_1 = 0$ and $\chi_1 = 0$ for waiting action). The action cost χ_2, χ_3 and χ_4 are incurred by the usage of cellular, WiFi, and D2D networks at each time slot, respectively. The benefit of mobile data offloading is based on the fact that $\chi_3 < \chi_2$ and $\chi_4 < \chi_2$. It means that the price to send data using cellular network is higher than that using WiFi and D2D networks. The total cost of transmitting data of size K is the sum of cost units incurred at each time slot during the total transmission time.

There may be some data transmission tasks that cannot be completed before the deadline. For failed data transmissions (i.e. $k > 0$ when $d > D$), we set the penalty cost function in (5). It is based on the data type u and the remainder size k of the data transmission process.

$$c_{D+1}(s) = penalty(k, u) = k^{(u+1)} \tag{5}$$

The memoryless mobility pattern of mobile users is defined in Eq. (6). The new location l' depends only on the past location l and has no relation with data type and data size. We design a two dimensions movement pattern. Every mobile user (including MSs and MHs), at each decision epoch, can stay where it is with probability μ, called stable factor. Alternatively, it can move randomly to a neighboring location with probability $\rho_i, i \in \{1, 2, 3, 4\}$, where i represents one of four possible moving directions(i.e., north, south, east and west). The stable factor μ and the probability of moving direction ρ_i satisfy Eq. (7).

$$P(l'|l) = \begin{cases} \mu, & l' = l \\ \rho_i, & otherwise \end{cases} \tag{6}$$

where

$$\mu + \sum \rho_i = 1, \quad i \in \{1, 2, 3, 4\} \tag{7}$$

Since MSs and MHs may randomly move before the deadline, we are interested in the situation where they can meet (connect) with each other at some other location. The probability that MS m can connect with MH n at decision epoch d and location l is defined as $P_d^m(l) * P_d^n(l)$. $P_d^m(l)$ is defined in Eq. (8); it represents the probability for MS m stays in location l during decision epoch d. l_m is the initial location of MS before the offloading process.

$$P_d^m(l) = \begin{cases} 1, & \text{if } P_d^m(l) = P_0^m(l_m) \\ \sum\limits_{l' \in \mathcal{L}} P_{d-1}^m(l') \cdot P(l|l'), & \text{otherwise} \end{cases} \tag{8}$$

The system state transition probability is the probability that the system state will go into s' in the next decision epoch if action a is taken at current state s. Since the movement of MS from location l to location l' or MH from h to h' is independent of k, u and transmission action a, we have

$$P(s'|s,a) = P(l'|l) \cdot \prod_{h \in \mathcal{H}} P(h'|h) \cdot P(k'|l,u,k) \tag{9}$$

where

$$P(k'|l,u,k) = \begin{cases} 1, & k' = k - \nu_a \ and \ a \in A(l,u) \\ 0, & otherwise \end{cases} \tag{10}$$

$P(l'|l)$ is the probability that MS will move from location l to location l' and $P(h'|h)$ is the probability that MS will move from location h to location h'. $P(k'|l,u,k)$ indicates that data size k' in next decision epoch is based on current data size k, location l and data type u. An illustrated MS state transition graph is shown in Fig. 2. The terminal states are those with $k = 0$.

IV. Hybrid Offloading Algorithm

In this section, we propose an algorithm, called hybrid offloading algorithm, to compute the optimal offloading policy, according to the movement of each mobile user. Since the offloading decision is based on the location and data type of mobile users, we first consider how these two parameters affect the offloading decision process. A policy π is a set of decisions at each state and decision epoch. It is defined as $a = \pi_d(s)$ in FHMDP. The policy space for π is denoted by Π. We aim to find the best policy π, which can minimize the overall cost for transmitting data of size K before deadline D. The objective function is defined as follows.

$$\min_{\pi \in \Pi} E_s^{\pi} \left[\sum_{d=1}^{D} c_d(s, \pi_d(s)) + c_{D+1}(s) \right] \tag{11}$$

We solve problem (11) using the hybrid offloading algorithm (*Algorithm 1*), based on value iteration method [16]. Before presenting our offloading algorithm, we define the value function as follows.

$$V_d^*(s) = \min_{a \in \mathcal{A}(l,u)} Q_d(s,a) \tag{12}$$

where

$$\begin{aligned} Q_d(s,a) &= \sum_{s' \in S} P(s'|s,a)[c_d(s,a,s') + V_{d+1}^*(s')] \\ &= \sum_{s' \in S} P(s'|s,a) cost(a) + \sum_{(l',u,k',\mathcal{H}') \in S} P(s'|s,a) V_{d+1}^*(s') \\ &= \nu_a \chi_a + \sum_{(l',u,k',\mathcal{H}') \in S} P(l'|l) \cdot \prod_{h \in \mathcal{H}} P(h'|h) \\ &\quad \cdot P(k'|l,u,k) V_{d+1}^*(s') \\ &= \nu_a \chi_a + \sum_{l',h' \in \mathcal{L}} P(l'|l) \cdot \prod_{h \in \mathcal{H}} P(h'|h) \\ &\quad \cdot V_{d+1}^*(l', u, (k - \nu_a), \mathcal{H}') \end{aligned} \tag{13}$$

Notice that:(1) the first equation in Eq. (13) shows that $Q_d(s,a)$ consists of current cost caused by taking action a and the future cost when s evolves into s'; (2) the following equations are derived by Eqs. (4), (9) and (10), respectively.

Our hybrid offloading algorithm consists of two phases: offloading planning phase and offloading running phase. In the planning phase, an optimal FHMDP policy is generated by value iteration, which is based on Eq. (14), where the current iteration value $V_d^n(s)$ is calculated by the last iteration value $V_d^{n-1}(s)$.

$$V_d^n(s) = \min_{a \in \mathcal{A}(l,u)} \sum_{s' \in S} P(s'|s,a)[c_d(s,a,s') + V_{d+1}^{n-1}(s')] \tag{14}$$

In the offloading phase, MNO takes action from the optimal policy according to current systems state. If data type $u \in \mathcal{U}^0$, MNO takes action according to the optimal policy. However, it does not guarantee data delivery before the deadline. If data type $u \in \mathcal{U}^1$, MNO first checks whether current data size can be transmitted using cellular network before deadline (line 20). If the response is yes, MNO will take action according to the optimal policy. Otherwise, MNO will transmit data using cellular network in order to complete data delivery before deadline. Notice that the function $\kappa(u)$ (line 20) is used to control the delay sensitivity of different data types; a higher delay sensitivity u leads to a higher function value. Meanwhile, $\kappa(u)$ can cancel out the prediction error of the memoryless mobility model.

V. Performance Evaluation

In this section, we demonstrate the performance of our proposed data offloading scheme in MCC. We use three metrics to evaluate the offloading performance:

- *Total cost.* The total network cost spent by MS during the data transmission process.
- *Completion time.* The total time that is actually used for data transmission.
- *Offloading ratio.* The percentage of cellular traffic that MO transmits through WiFi or D2D networks.

Algorithm 1 Hybrid Delayed Tolerant Offloading Algorithm

```
Planning Phase
Initialize data type $u$ and $\mathcal{H}$ with locations of MHs
Initialize $V_d^0(s)$ with Eqs. (4) and (5)
repeat
  for $d \in \mathcal{D}$ do
    for $l \in \mathcal{L}$ do
      for $k \in \mathcal{K}$ do
        compute $V_d^n(s)$ using Eq. (14)
        compute $rsd_d^n(s) = \|V_d^n(s) - V_d^{n-1}(s)\|$
      end for
    end for
  end for
until $rsd_d^n(s) < \epsilon$
return  Best policy $\pi_d^*(s)$
Running Phase
Set $d := 1$ and $k := K$
while $d \leq D$ and $k > 0$ do
  Get the locations of MS and MHs
  Set action $a := \pi_d^*(s)$
  if $k > \nu_2 \times (D - d) \times \kappa(u)$ then
    $k := k - \nu_2$
  else
    $k := k - \nu_a$
  end if
  $d := d + 1$
end while
```

We compare our delayed offloading scheme (we name $D4$) with three other schemes: (1) optimal delayed WiFi offloading scheme ($D3$) [17]: prediction based cellular data offloading uses WiFi network; (2) non-delayed WiFi offloading scheme ($ND3$) [18]: data transmission is switched between WiFi and cellular network; and (3) non-delayed WiFi/D2D offloading scheme ($ND4$): WiFi network is used whenever available; D2D communication is used when inequation $k < \nu_2 * (D-d)$ is satisfied, where k is the portion of data size that is not transmitted yet, d is current decision epoch and D is the deadline; otherwise, cellular network is used. Meanwhile, non-offloading scheme (NO) is used for comparison.

In our simulation, the time slot between two sequential decision epochs is set to be 10 seconds. At each decision epoch, MS or MH chooses a moving direction as Eq.(6). The stable factor μ is 0.4. We test our model with different mobility traces generated randomly according to our memoryless mobility pattern. We run each simulation in 10^4 different network settings (i.e. the locations of WiFi APs and MHs are generated randomly) and show the average value. The data rates of celluar, WiFi and D2D networks are 16 Mbps, 24 Mbps and 8 Mbps, respectively. The unit costs of celluar, WiFi and D2D networks are 1, 0.1 and 0.2, respectively [13].

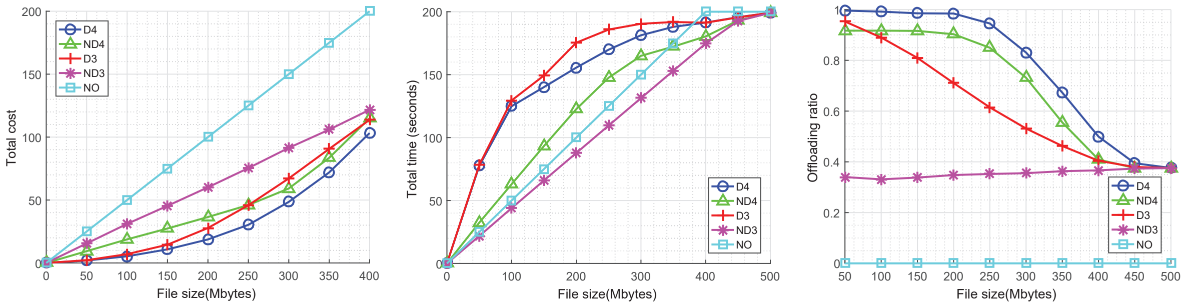
Notice that since MHs can carry multiple types of data and many MSs may request mobile data at the same time, only a subset of MSs can benefit from MHs. This is quite different from WiFi and cellular networks, under which the availability is not a big problem. Also, the WiFi connection is not available all the time. The probability of WiFi connection (resp. MH connection) is set to 0.8 (resp. 0.5).

Figure 3(a) shows the variation of the total cost with the data size. We observe that our proposed scheme $D4$ outperforms the three other offloading schemes by achieving the lowest cost for any data size. We also observe that when $K \leq 250$ Mbytes, $D3$ outperforms $ND4$; it is not the case when $K > 250$ Mbytes. This can be explained by the fact that $D3$ can use delayed WiFi offloading to offload more data than $ND4$ when $K \leq 250$ Mbytes, while $ND4$ can use low cost D2D to offload more data than $D3$ when $K > 250$ Mbytes.

Figure 3(b) shows the variation of the completion time with the data size. As expected, the completion time increases with the data size. We observe that $ND3$ and $ND4$ achieve lower completion time compared to $D3$ and $D4$. This is because that the objective of $D3$ and $D4$ is to reduce the usage of cellular network and seek for opportunities to use WiFi and D2D networks (in opposition to $ND3$ and $ND4$). We also observe that $D4$ outperforms $D3$ because it makes use of D2D network. However, $ND3$ outperforms $ND4$ for any data size. This is because that the data rate of D2D network is lower than cellular and WiFi networks. Using D2D network can increase the completion time in non-delay schemes (e.g. $ND3$ outperforms $ND4$), while decrease the completion time in delayed schemes (e.g. $D4$ outperforms $D3$).

Fig. 3(c) shows that the offloading ratio reduces with the increase of data size except for $ND3$. This is because that $ND3$ transmits data based on the location l, without considering current data size k and decision epoch d. Thus, the offloading ratio of $ND3$ cannot change with the data size. Moreover, $D3$ drops rapidly with the increase of data size, while $D4$ and $ND4$ drop slowly. This is because $D4$ and $ND4$ use alternative D2D network to offload data except for WiFi. Notice that $D4$ has the highest offloading ratio. We then observe that the offloading ratios for delayed and non-delayed schemes are the same when $K \geq 500$ Mbytes. This is because that 500 Mbytes cannot be transmitted in 200 seconds under the setting used in our simulations. Since all the offloading schemes try to complete the data delivery task before deadline, they will use WiFi network wherever possible and cellular network when WiFi network is not available, which is the offloading policy for $ND3$. It means that all the other offloading schemes (i.e., $D4$, $ND4$, and $D3$) degenerate to the policy used by $ND3$. Then MO has to extend the deadline in order to increase the offloading ratio. Notice that, although $ND4$ does not use delay based policy, it can offload more data than $D3$ when $K > 76$ Mbytes. The benefit comes from the setting where the number of MHs is large enough. However, the advantage of delay policy shows up when $K < 76$ Mbytes.

We conclude that our scheme D4 achieves the lowest cost while satisfying data transmission deadlines. It achieves the minimal transmission cost and outperforms D3 in transmission time with different data size and deadline. D4 uses almost all the deadline in order to wait for offloading opportunities.



(a) Total cost versus data size K.

(b) Total time versus data size K.

(c) Offloading ratio versus data size K.

Fig. 3. Performance comparison for different schemes, with same deadline $D = 200$ seconds and different data size K. Here, $\nu_2 = 16$ Mbps, $\nu_3 = 24$ Mbps, and $\nu_4 = 8$ Mbps.

VI. Conclusion

In this paper, we studied the mobile data offloading problem in MCC. We considered the hybrid offloading network, where data traffic originally transferred by cellular network can be offloaded to WiFi and D2D networks. Given that the WiFi network has high connection speed but low coverage area and D2D network can help offloading data where WiFi network is not deployed, the hybrid offloading model can effectively reduce data traffic in cellular network. In order to support delay tolerant data transmission, we introduced a hybrid offloading algorithm for delay sensitive and delay tolerant applications by making use of different data types. The simulation results show that our proposed scheme can achieve a minimal total cost as compared to existing offloading schemes.

References

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

[2] Y. Xu and S. Mao, "A survey of mobile cloud computing for rich media applications," *IEEE Wireless Communications*, vol. 20, no. 3, pp. 46–53, 2013.

[3] A. Karamoozian, A. Hafid, M. Boushaba, and M. Afzali, "QoS-aware resource allocation for mobile media services in cloud environment," *2016 13th IEEE Annual Consumer Communications Networking Conference (CCNC)*, no. Mcc, pp. 732–737, 2016.

[4] F. Rebecchi, M. Dias de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, "Data offloading techniques in cellular networks: a survey," *Communications Surveys & Tutorials, IEEE*, vol. 17, no. 2, pp. 580–603, 2015.

[5] J. Lee, Y. Yi, S. Chong, and Y. Jin, "Economics of WiFi offloading: Trading delay for cellular capacity," *Wireless Communications, IEEE Transactions on*, vol. 13, no. 3, pp. 1540–1554, 2014.

[6] M. Rebai, L. Khoukhi, H. Snoussi, and F. Hnaien, "Optimal placement in hybrid vanets-sensors networks," in *2012 Wireless Advanced (WiAd)*, June 2012, pp. 54–57.

[7] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can WiFi deliver?" *IEEE/ACM Transactions on Networking (TON)*, vol. 21, no. 2, pp. 536–550, 2013.

[8] A. Aijaz, H. Aghvami, M. Amani, and A. H. Aghvami, "A survey on mobile data offloading: technical and business perspectives," *IEEE Wireless Communications*, vol. 20, no. April, pp. 104–112, 2013.

[9] L. Khoukhi, A. El Masri, A. Sardouk, A. Hafid, and D. Gaiti, "Toward fuzzy traffic adaptation solution in wireless mesh networks," *IEEE Transactions on Computers*, vol. 63, no. 5, pp. 1296–1308, 2014.

[10] M. A. Togou, A. Hafid, and L. Khoukhi, "Scrp: Stable cds-based routing protocol for urban vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1298–1307, 2016.

[11] L. Khoukhi and S. Cherkaoui, "Experimenting with fuzzy logic for qos management in mobile ad hoc networks," *Int. J. Comput. Science Netw. Security*, vol. 8, no. 8, pp. 372–386, 2008.

[12] Y. Li, M. Qian, D. Jin, P. Hui, Z. Wang, and S. Chen, "Multiple mobile data offloading through disruption tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1579–1596, 2014.

[13] X. Zhuo, W. Gao, G. Cao, and S. Hua, "An incentive framework for cellular traffic offloading," *Mobile Computing, IEEE Transactions on*, vol. 13, no. 3, pp. 541–555, 2014.

[14] A. Nadembega, A. S. Hafid, and R. Brisebois, "Mobility prediction model-based service migration procedure for follow me cloud to support qos and qoe," in *Communications (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–6.

[15] M. Othman, S. A. Madani, S. U. Khan *et al.*, "A survey of mobile cloud computing application models," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 393–413, 2014.

[16] A. Kolobov, "Planning with Markov decision processes: An AI perspective," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 6, no. 1, pp. 1–210, 2012.

[17] M. H. Cheung and J. Huang, "DAWN: Delay-Aware Wi-Fi offloading and network selection," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1214–1223, 2015.

[18] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*. ACM, 2010, pp. 209–222.