

Win-Coupon: An Incentive Framework for 3G Traffic Offloading

Xuejun Zhuo*, Wei Gao[†], Guohong Cao[†], Yiqi Dai*

* Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China. Email: zhuoxj07@mails.tsinghua.edu.cn; dyq@mail.tsinghua.edu.cn

[†]The Pennsylvania State University, University Park, State College, PA 16802, USA. Email: {weigao, gcao}@cse.psu.edu

Abstract—3G networks are currently facing severe traffic overload problems caused by excessive demands of mobile users. Offloading part of the 3G traffic through other forms of networks, such as Delay Tolerant Networks (DTNs), WiFi hotspots, and Femtocells, is a promising solution. However, since these networks can only provide intermittent and opportunistic connectivity to mobile users, utilizing them for 3G traffic offloading may result in a non-negligible delay. As the delay increases, the users' satisfaction decreases. In this paper, we investigate the tradeoff between the amount of traffic being offloaded and the users' satisfaction. We provide a novel incentive framework to motivate users to leverage their delay tolerance for 3G traffic offloading. To minimize the incentive cost given an offloading target, users with high delay tolerance and large offloading potential should be prioritized for traffic offloading. To effectively capture the dynamic characteristics of users' delay tolerance, our incentive framework is based on reverse auction to let users proactively express their delay tolerance by submitting bids. We further take DTN as a case study to illustrate how to predict the offloading potential of the users by using stochastic analysis. Extensive trace-driven simulations verify the efficiency of our incentive framework for 3G traffic offloading.

I. INTRODUCTION

The recent commercialization and popularization of 3G networks provide smartphone users with ubiquitous Internet access, and significantly enhance mobile user's capability of accessing web content. However, the explosive growth of user population and their demands for bandwidth-eager multimedia content raise big challenges to the 3G network infrastructure. A huge amount of 3G data traffic has been generated by mobile users, which exceeds the capacity of 3G network and hence deteriorates the network quality provided to mobile users [1]. To address such challenges, the most straightforward solution is to increase the capacity of 3G networks, which however is expensive and inefficient. In contrast, it would be desirable if part of 3G traffic can be effectively offloaded.

Some recent research efforts have been focusing on offloading 3G traffic to other forms of networks, such as DTNs, WiFi hotspots, and Femtocells [2], [3], [4], and they generally focus on maximizing the amount of 3G traffic that can be offloaded. In most cases, due to user mobility, these networks available for 3G traffic offloading only provide intermittent and opportunistic network connectivity to the users, and the traffic offloading hence results in non-negligible data downloading

delay¹. In general, more offloading opportunities may appear by requesting the mobile users to wait for a longer time before actually downloading the data from the 3G networks, but this will also make the users become more impatient and hence reduce their satisfaction.

In this paper, we focus on investigating the tradeoff between the amount of traffic being offloaded and the users' satisfaction, and further propose a novel incentive framework to motivate users to leverage their delay tolerance for 3G offloading. Users are provided with incentives, i.e., receiving discount for their service charge if they are willing to wait longer for data downloading. During the delay, part of the 3G data traffic may be opportunistically offloaded to other networks mentioned above, and the user is assured to receive the remaining part of the data via 3G network when the delay period ends.

The major challenge of designing such an incentive framework is to minimize the incentive cost of 3G operator which includes the total discount provided to the mobile users, subject to an expected amount of traffic being offloaded. To achieve this goal, two important factors should be taken into account, i.e., the *delay tolerance* and *offloading potential* of the users. The users with high delay tolerance and large offloading potential should be prioritized in 3G traffic offloading.

First, with the same period of delay, the users with higher delay tolerance require less discount to compensate their satisfaction loss. In order to effectively capture the dynamic characteristics of the users' delay tolerance, we propose an incentive mechanism based on reverse auction which is proved to conduct a justified pricing. In our mechanism, the users act as sellers to proactively send bids, which include the delay that they are willing to experience and the discount that they want to obtain for this delay. Such discount requested by users is called "*coupon*" in the rest of the paper. The 3G operator then acts as the buyer to buy the delay tolerance from the users.

Second, with the same period of delay, users with larger offloading potential are able to offload more data traffic. For example, the offloading potential of the user who requests popular data is large, because it can easily retrieve the data pieces from other contacted peer users during the delay period. To effectively capture the offloading potential of the users, we

¹Without loss of generality, we consider that most of the 3G traffic is generated in the downlink.

take DTN as a case study and propose an accurate prediction model by using stochastic analysis.

The optimal auction outcome is determined by considering both the delay tolerance and offloading potential of the users to achieve the minimum incentive cost, given an offloading target. The auction winners set up contracts with the 3G operator for the delay they wait and the coupon they earn, and the other users directly download data via 3G at the original price. More specifically, the contribution of the paper is three-fold:

- We propose a novel incentive framework to motivate users to leverage their delay tolerance for 3G offloading.
- We design an incentive mechanism based on reverse auction, named *Win-Coupon*, which dynamically determines the traffic offloading solution based on the delay tolerance and offloading potential of the users. We formally prove the three desirable properties of Win-Coupon: 1) truthfulness, 2) individual rationality, and 3) polynomial time computability.
- Taking DTNs as a case study, we provide an accurate model using stochastic analysis to predict users' offloading potential based on their data access and mobility patterns.

The rest of the paper is organized as follows. In Section II we briefly review the existing work. Section III provides an overview of our approach and the related background knowledge. Section IV describes the details of our incentive framework, and proves its desirable properties. Section V evaluates the performance of Win-Coupon through trace-driven simulations and Section VI concludes the paper.

II. RELATED WORK

To deal with the traffic overload problem in 3G networks, some studies propose to utilize DTNs to conduct offloading. Whitbeck et al. [5] design a framework, called Push-and-Track, which includes multiple strategies to determine how many copies should be injected by 3G and to whom, and then leverages DTNs to offload 3G traffic. Han et al. [2] provide three simple algorithms to exploit DTNs to facilitate data dissemination among mobile users, in order to reduce the overall 3G traffic. Many research efforts have focused on how to improve the performance of data access in DTNs. In [6], the authors provide theoretical analysis to the stationary and transient regimes of data dissemination. Some later works [7] [8] disseminate data among the mobile users by exploiting their social relations. Data replication is another way to facilitate data access in DTNs, which has been studied in [9] [10]. Being orthogonal with how to improve the performance of data access in DTNs, in this paper, we propose an accurate model to capture the expected volume of 3G traffic that can be offloaded to DTNs to facilitate our incentive framework design.

Public WiFi can also be utilized for 3G traffic offloading. In [4], the authors measure the offloading potential of the public WiFi based on city wide vehicular traces. Compared to the vehicle based high mobility scenario in [4], Lee et al. [3] consider a more general mobile scenario, and present a quantitative study on the performance of on-the-spot and delayed offloading by using WiFi. The prediction of future WiFi availability is important to the offloading scheme design,

and has been studied in [11] [12]. In [11], the authors propose to enable mobile users to schedule their data transfers when higher WiFi transmission rate can be achieved based on the prediction. In [12], a Lyapunov framework based algorithm, called SALSA, is proposed to optimize the energy-delay trade-off of the mobile devices with both 3G and WiFi interfaces.

All the existing offloading studies have not considered the satisfaction loss of the users when a longer delay is caused by traffic offloading. To motivate users to leverage their delay tolerance for 3G offloading, we propose an auction based incentive framework. Auction has been widely used in network design. Applying auction in the spectrum leasing is one of the most practical applications. FCC has already auctioned the unused spectrum in the past decade [13], and there are a large amount of works on wireless spectrum auctions [14] [15] [16]. Moreover, auction has also been applied for designing incentive mechanism to motivate selfish nodes to forward data for others [17] [18]. However, none of them has applied auction techniques to 3G traffic offloading.

III. OVERVIEW

A. The Big Picture

In this section, we give an overview of the Win-Coupon framework. By considering the users' delay tolerance and offloading potential, Win-Coupon uses a reverse auction based incentive mechanism to motivate users to help 3G offloading. Fig. 1 illustrates the main idea. The 3G operator acts as the buyer, who offers coupons to users in exchange for them to wait for some time and opportunistically offload the traffic. When users request data, they are motivated to send bids along with their request messages to the 3G operator. Each bid includes the information of how long the user is willing to wait and how much coupon he wants to obtain as a return for the extra delay. Then, the 3G operator infers users' delay tolerance. In addition, users' offloading potential should also be considered when deciding the auction outcome. Based on the priori information collected, such as users' data access and mobility patterns, the 3G operator predicts users' offloading potential.

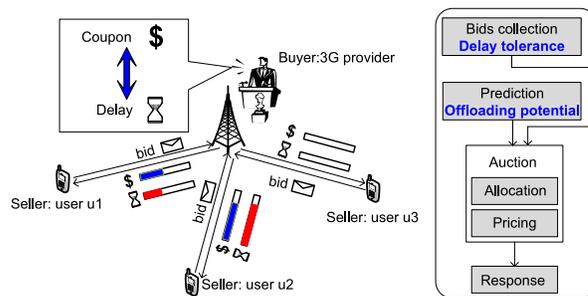


Fig. 1. The main idea of Win-Coupon

The optimal auction outcome is to minimize the 3G operator's incentive cost subject to a given offloading target according to the bidders' delay tolerance and offloading potential. The auction contains two main steps: *allocation* and *pricing*. In the allocation step, the 3G operator decides which bidders are the winners and how long they need to wait. In the pricing step, the

3G operator decides how much to pay for each winner. Finally, the 3G operator returns the bidders with the auction outcome which includes the assigned delay and the value of coupon for each bidder. The winning bidders (e.g. user $u1$ and $u2$ shown in Fig. 1) obtain the coupon, and are assured to receive the data via 3G when their promised delay is reached. For example, suppose p is the original data service charge, if user $u1$ obtains the coupon with value c in return for delay t , it only needs to pay $p - c$ for the data service. During the delay period, $u1$ may retrieve some data pieces from other intermittently available networks, e.g., by contacting other peers which cache the data or moves into the wireless range of APs. Once delay t passes, the 3G network pushes the remaining data pieces to $u1$ to assure the promised delay. The losing bidders (e.g. user $u3$ shown in Fig. 1) immediately download data via 3G at the original price.

B. User Delay Tolerance

With the increase of downloading delay, the user's satisfaction decreases accordingly, the rate of which reflects the user's delay tolerance. To flexibly model users' delay tolerance, we introduce a *satisfaction function* $S(t)$, which is a monotonically decreasing function of delay t , and represents the price that the user is willing to pay for the data service with the delay. The satisfaction function is determined by the user himself, his requested data, and various environmental factors. We assume that each user has an upper bound of delay tolerance for each data. Once the delay reaches the bound, the user's satisfaction becomes zero, indicating that the user is not willing to pay for the data service. Fig. 2 shows an example of the satisfaction function $S(t)$ of a specific user for a specific data, where t_{bound} is the upper bound of the user's delay tolerance. p is the original charge for the data service, and the satisfaction curve represents the user's expected price for the data as the delay increases. For example, with delay t_1 the user is only willing to pay p_1 instead of p . $p - p_1$ is the satisfaction loss caused by delay t_1 .



Fig. 2. User's satisfaction function

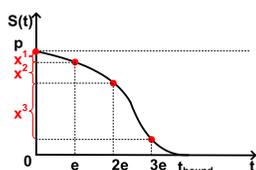


Fig. 3. User's private value

C. Auctions

In economics, auction is a typical method to determine the value of a commodity that has an undetermined and variable price. It has been widely applied to many fields. Most auctions are *forward auction* which involves a single seller and multiple buyers, and the buyers send bids to compete for obtaining the commodities sold by the seller. In this paper, our incentive mechanism is based on *reverse auction* [19] which involves a single buyer and multiple sellers, and the buyer decides its purchase based on the bids sent by the sellers. To begin with, we introduce some notations.

Bid (b_i): It is submitted by bidder i to express i 's valuation on the resource for sale, which is not necessarily true.

Private value (x_i): It is the true valuation made by bidder i for the resources; i.e., the true price that bidder i wants to obtain for selling the resource. This value is only known by the bidder.

Market-clearing price (p_i): It is the price actually paid by the buyer to bidder i . This price cannot be less than the bids submitted by i .

Utility (u_i): It is the residual worth of the sold resource for bidder i , namely the difference between i 's market-clearing price p_i and private value x_i .

The bidders in the auction are assumed to be rational and risk neutral. A common requirement for auction design is the so-called *individual rationality*.

Definition 1: An auction is with individual rationality if all bidders are guaranteed to obtain non-negative utility.

The rational bidders decide their bidding strategy in order to maximize their utility. Let \mathcal{N} denote the set of all bidders. The concept of *weakly dominant strategy* is defined as follows.

Definition 2: $b_i = \beta_i$ is a weakly dominant strategy for user i if and only if: $u_i(\beta_i, \beta_{-i}) \geq u_i(\beta'_i, \beta_{-i})$, $\forall \beta'_i \neq \beta_i$.

Here $\beta_{-i} = \{\beta_1, \beta_2, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_{|\mathcal{N}|}\}$ denotes the set of strategies of all other bidders except for bidder i . We can see a weakly dominant strategy maximizes i 's utility regardless of the strategies chosen by all other bidders. If for every bidder, truthfully setting its bid to its private value is a weakly dominant strategy, the auction is *truthful* (strategyproof).

Definition 3: An auction is truthful if each bidder, say i , has a weakly dominant strategy, in which $b_i = x_i$.

The truthfulness eliminates the expensive overhead for bidders to strategize against other bidders and prevents the market manipulation. Also, it assures the efficient allocation by encouraging bidders to reveal their true private values. Vickrey-Clarke-Groves (VCG) [20], [21], [22] is the most well-studied auction format, due to its truthful property. However, VCG only ensures truthfulness when the optimal allocation can be found, and it usually cannot assure the truthfulness when applied to non-optimal approximate algorithms [23]. Unfortunately, the allocation problem in Win-Coupon is NP-hard. We aim to solve this challenge and design a novel incentive mechanism which possesses three important properties: 1) truthfulness, 2) individual rationality, and 3) polynomial time computability.

IV. MAIN APPROACH OF WIN-COUPON

In this section, we illustrate the details of Win-Coupon. In the reverse auction based Win-Coupon, the buyer is the 3G operator who pays coupon in exchange for longer delay of the users. The sellers are the 3G users who sell their delay tolerance to win coupon. The right side of Fig. 1 shows the flow chart of Win-Coupon. At first, the 3G operator collects the bids to derive the delay tolerance of the bidders, and predicts their offloading potential. Then, based on the derived information, a reverse auction is conducted, which includes two main steps: allocation and pricing. Finally, the 3G operator returns the auction outcome to the bidders.

In the rest of this section, we first introduce the bidding process. Then we present the main algorithms of auction and prove its desirable properties. Finally, we use DTN as a case study to show how to predict bidders' offloading potential.

A. Bidding

To obtain coupon, the users attach bids with their data requests to reveal their delay tolerance. For each user, the upper bound t_{bound} of its delay tolerance can be viewed as the resources that it wants to sell. The user can divide t_{bound} into multiple time units, and submit multiple bids $\mathbf{b} = \{b^1, b^2, \dots, b^l\}$ to indicate the value of coupon it wants to obtain for each additional time unit of delay, where l equals $\lfloor \frac{t_{bound}}{e} \rfloor$, and e is the length of one time unit. By receiving these bids, the 3G operator knows that the user wants to obtain coupon with value no less than $\sum_{k=1}^{k_i} b^k$ by waiting for k_i time units. The length of time unit e can be flexibly determined by the 3G operator. Shorter time unit results in larger bids with more information, which increases the performance of the auction, but it also induces more communication overhead and higher computational complexity. To simplify the presentation, in the rest of the paper delay t is normalized by time unit e .

As shown in Fig. 2, $p - S(t)$ is the satisfaction loss of the user due to delay t . Then, $p - S(t)$ represents the private value of the user to the delay, namely the user wants to obtain the coupon with value no less than $p - S(t)$ for delay t . Thus, the private value of the user to each additional time unit of delay is $\mathbf{x} = \{x^1, x^2, \dots, x^l\}$, where x^k ($k \in \{1, \dots, l\}$), equals $S(k-1) - S(k)$. For example, as shown in Fig. 3, the user wants to obtain the coupon with value no less than x^1 if it waits for one time unit, $x^1 + x^2$ for two time units, and $x^1 + x^2 + x^3$ for three time units. Generally, the user can set its bids with any value at will, however we will prove that the auction in Win-Coupon is truthful, which guarantees that the users would bid their private value; that is, $b^k = x^k$, for all k .

B. Auction Algorithms

Win-Coupon is run periodically in each auction round. Usually, the auction would result in an extra delay for the bidders to wait for the auction outcome. However, different from other long-term auctions, such as the FCC-style spectrum leasing, the auction round in our scenario is very short, since hundreds of users may request 3G data service at the same time. Also, because the bidders who are willing to submit bids are supposed to have a certain degree of delay tolerance, the extra delay caused by auction can be neglected. Next, we describe two main steps of the auction: allocation and pricing.

1) *Allocation*: In traditional reverse auction, the allocation solution is purely decided by the bids; i.e., the bidders who bid the lowest price win the game. However, in our scenario, besides the bids which express the bidders' delay tolerance, the offloading potential of the bidders should also be considered. Let $\{t_1, t_2, \dots, t_{|\mathcal{N}|}\}$ represent the allocation solution, where t_i denotes the length of delay that 3G operator wants to buy from bidder i . Note that since each bidder is asked to wait for integer multiples of time unit, t_i is an integer. If t_i equals zero,

bidder i loses the game. The allocation problem in Win-Coupon can be formulated as follows:

Definition 4: The allocation problem is to determine the optimal solution $\{t_1, t_2, \dots, t_{|\mathcal{N}|}\}$ which minimizes the total incentive cost, subject to a given offloading target.

$$\min_{t_i} \sum_{i \in \mathcal{N}} \sum_{k=1}^{t_i} b_i^k \quad (1)$$

$$s.t. \quad \sum_{i \in \mathcal{N}} V_i^d(t_i) \geq v_0 \quad (2)$$

$$\forall i, t_i \in \{0, 1, 2, \dots, l_i\}. \quad (3)$$

In Eq.(1), $\sum_{k=1}^{t_i} b_i^k$ denotes the value of the coupon that the 3G operator needs to pay bidder i in exchange for its delay t_i . $V_i^d(t)$ in Eq.(2) denotes the expected volume of traffic that can be offloaded, if bidder i downloads data d and is willing to wait for delay t . We will provide the details on how to predict $V_i^d(t)$ in Section IV-C. We assume that within a short auction round, each bidder only requests one data item, so that each i is mapped to a single d . Thus, this constraint ensures that the total expected volume of offloaded traffic is no less than the offloading target v_0 . Eq.(3) ensures that the delay that each bidder i waits does not exceed l_i , the maximum number of time units that i is willing to wait.

It is easy to prove that our allocation problem can be reduced to the 0-1 knapsack problem, under the assumption that $l_i = 1$, for all i . The 0-1 knapsack problem is proved to be NP-hard, and thus our problem is also NP-hard. Next, we transform the original problem, and derive the optimal solution of the new problem by dynamic programming.

We transform the original problem by replacing constraint (2) with $\sum_{i \in \mathcal{N}} \lfloor V_i^d(t_i) M \rfloor \geq \lfloor v_0 M \rfloor$, where $M = 10^n$ is a common scalar. With a larger M , the optimal solution of the new problem becomes closer to that of the original problem, and the former converges to the latter when M is large enough. On the other hand, larger M increases the computational complexity of the algorithm. In the following, we define $\hat{V}_i^d(t_i) = \lfloor V_i^d(t_i) M \rfloor$, and $\hat{v}_0 = \lfloor v_0 M \rfloor$.

Let T_i^v denote the minimum time units of delay that bidder i needs to wait to offload v volume of traffic, and C_i^v denote the corresponding value of coupon that i requests. Note that here and in the rest of this section, traffic volume v is scaled by M without specification. Then, we have:

$$T_i^v = \arg \min_k \{ \hat{V}_i^d(k) \geq v \} \quad (4)$$

$$C_i^v = \sum_{k=1}^{T_i^v} b_i^k \quad (5)$$

We use $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{|\mathcal{N}|}\}$ to denote the bid set including all the bids sent by the bidders in set \mathcal{N} , and use $\mathcal{B}_i = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_i\}$ to denote the bid set including all the bids sent by the first i bidders in \mathcal{N} . Assume only the first i bidders join the auction, we define $C_{\mathcal{B}_i}^v$ to be the minimal incentive cost incurred to achieve a given offloading target v with the bid set \mathcal{B}_i , and define $\mathcal{T}_{\mathcal{B}_i}^v = \{t_1, t_2, \dots, t_i\}$ to be

the corresponding optimal allocation solution. Our allocation algorithm is illustrated in Algorithm 1 with $\mathcal{T}_B^{\hat{v}_0}$ giving the optimal allocation solution. In Algorithm 1, line 4 to 8 update $\mathcal{T}_{B_i}^v$, $C_{B_i}^v$ to include a new bidder at each iteration. Line 6 searches for the optimal allocation solution $\mathcal{T}_{B_i}^v$ to obtain minimal $C_{B_i}^v$. The complexity of the algorithm is $O(|\mathcal{N}|\hat{v}_0^2)$.

2) *Pricing*: The VCG-style pricing is generally used in forward auction, which involves single seller with limited resources for sale, and multiple buyers. The bidders who have the highest bid win the game, and each winning bidder pays the ‘‘opportunity cost’’ that its presence introduces to others. It is proved that this pricing algorithm provides bidders with the incentives to set their bids truthfully. Based on the basic idea, in our pricing algorithm, the 3G operator also pays bidder i the coupon with value equal to the ‘‘opportunity cost’’ exerted to all the other bidders due to i ’s presence. Given the offloading target \hat{v}_0 , let $c1 = C_{B \setminus \{b_i\}}^{\hat{v}_0}$ denote the total value of coupons requested by all the bidders under the optimal allocation solution without the presence of i . Let $c2 = (C_B^{\hat{v}_0} - \sum_{k=1}^{t_i} b_i^k)$ denote the total value of coupons requested by all the bidders except for i under the current optimal allocation solution. Then, i ’s ‘‘opportunity cost’’ is defined as the difference between $c1$ and $c2$. Thus, i ’s market-clearing price can be derived as:

$$p_i = c1 - c2 = C_{B \setminus \{b_i\}}^{\hat{v}_0} - (C_B^{\hat{v}_0} - \sum_{k=1}^{t_i} b_i^k). \quad (6)$$

The pricing algorithm is illustrated in Algorithm 2, and the computational complexity of the algorithm is $O(A|\mathcal{N}|\hat{v}_0^2)$, where A is the number of winning bidders.

Algorithm 1: Win-Coupon-Allocation (\mathcal{N}, \mathcal{B})

```

1 for  $v = 0$  to  $\hat{v}_0$  do
2    $\mathcal{T}_{B_i}^v = \{T_i^v\}$ ;
3    $C_{B_i}^v = C_i^v$ ;
4 for  $i = 2$  to  $|\mathcal{N}|$  do
5   for  $v = 0$  to  $\hat{v}_0$  do
6      $s^* = \arg \min_{s \in [0, v]} \{C_{B_{i-1}}^s + C_i^{v-s}\}$ ;
7      $\mathcal{T}_{B_i}^v = \mathcal{T}_{B_{i-1}}^{s^*} \cup \{T_i^{v-s^*}\}$ ;
8      $C_{B_i}^v = C_{B_{i-1}}^{s^*} + C_i^{v-s^*}$ ;
9 return  $\mathcal{T}_B^{\hat{v}_0}, C_B^{\hat{v}_0}$ ;
```

Algorithm 2: Win-Coupon-Pricing ($\mathcal{N}, \mathcal{B}, \mathcal{T}_B^{\hat{v}_0}, C_B^{\hat{v}_0}$)

```

1 for  $i = 1$  to  $|\mathcal{N}|$  do
2   if  $i$  is the winning bidder then
3     Win-Coupon-Allocation( $\mathcal{N} \setminus \{i\}, \mathcal{B} \setminus \{b_i\}$ );
4      $p_i = C_{B \setminus \{b_i\}}^{\hat{v}_0} - (C_B^{\hat{v}_0} - \sum_{k=1}^{t_i} b_i^k)$ ;
5   else
6      $p_i = 0$ ;
7 return  $p_i$ , for all  $i$ ;
```

3) *Properties*: In Section IV-B1, IV-B2, we have shown that Win-Coupon can be solved in polynomial time, if a suitable scalar M is selected. Next, we prove that Win-Coupon also has the properties: truthfulness and individual rationality.

Theorem 1: In Win-Coupon, for each bidder, say i , setting its bids truthfully, i.e., $\mathbf{b}_i = \mathbf{x}_i$, is a weakly dominant strategy.

Proof: We assume that when bidder i sets its bids truthfully, i.e., $\mathbf{b}_i = \mathbf{x}_i$, 3G operator would buy delay t_i from it, and its market-clearing price is $p_i = C_{B \setminus \{b_i\}}^{\hat{v}_0} - (C_B^{\hat{v}_0} - \sum_{k=1}^{t_i} b_i^k)$. Then, the utility obtained by i is $u_i = C_{B \setminus \{b_i\}}^{\hat{v}_0} - (C_B^{\hat{v}_0} - \sum_{k=1}^{t_i} b_i^k) - \sum_{k=1}^{t_i} x_i^k$. Now, suppose that bidder i sets its bids untruthfully, i.e., $\mathbf{b}_i' \neq \mathbf{x}_i$. Then, the length of delay t_i' that 3G operator would buy from i falls into two cases: 1) $t_i' = t_i$ and 2) $t_i' \neq t_i$.

In case 1), the market-clearing price paid to bidder i would become $p_i' = C_{B \setminus \{b_i'\}}^{\hat{v}_0} - (C_B^{\hat{v}_0} - \sum_{k=1}^{t_i'} b_i'^k)$. Due to the sub-problem optimality in deriving the incentive cost $C_{B \setminus \{b_i'\}}^{\hat{v}_0}$, $C_B^{\hat{v}_0} = C_{B \setminus \{b_i'\}}^{\hat{v}_0 - \hat{V}_i^d(t_i')} + \sum_{k=1}^{t_i'} b_i'^k$. Then we have $p_i' = C_{B \setminus \{b_i'\}}^{\hat{v}_0} - C_{B \setminus \{b_i'\}}^{\hat{v}_0 - \hat{V}_i^d(t_i')}$, where $C_{B \setminus \{b_i'\}}^{\hat{v}_0}$ and $C_{B \setminus \{b_i'\}}^{\hat{v}_0 - \hat{V}_i^d(t_i')}$ are independent of the bids sent by bidder i . Therefore, if $t_i' = t_i$, then $p_i' = p_i$, which is unaffected and the utility of bidder i has no change.

In case 2), similarly the market-clearing price paid to bidder i would be changed to $p_i' = C_{B \setminus \{b_i'\}}^{\hat{v}_0} - C_{B \setminus \{b_i'\}}^{\hat{v}_0 - \hat{V}_i^d(t_i')}$. Then, the new utility obtained by i equals $u_i' = C_{B \setminus \{b_i'\}}^{\hat{v}_0} - C_{B \setminus \{b_i'\}}^{\hat{v}_0 - \hat{V}_i^d(t_i')} - \sum_{k=1}^{t_i'} x_i^k$. The utility gain obtained by bidder i by setting $\mathbf{b}_i' \neq \mathbf{b}_i$ can be calculated as:

$$\begin{aligned} \Delta u_i &= u_i' - u_i = (C_{B \setminus \{b_i'\}}^{\hat{v}_0} - C_{B \setminus \{b_i'\}}^{\hat{v}_0 - \hat{V}_i^d(t_i')} - \sum_{k=1}^{t_i'} x_i^k) \\ &\quad - (C_{B \setminus \{b_i\}}^{\hat{v}_0} - C_{B \setminus \{b_i\}}^{\hat{v}_0 - \hat{V}_i^d(t_i)} - \sum_{k=1}^{t_i} x_i^k) \\ &= (C_{B \setminus \{b_i\}}^{\hat{v}_0 - \hat{V}_i^d(t_i)} + \sum_{k=1}^{t_i} x_i^k) - (C_{B \setminus \{b_i'\}}^{\hat{v}_0 - \hat{V}_i^d(t_i')} + \sum_{k=1}^{t_i'} x_i^k). \end{aligned}$$

When bidder i sets its bids truthfully as $\mathbf{b}_i = \mathbf{x}_i$, Buying delay with length t_i from it is the optimal solution of the 3G operator to minimize the incentive cost. Therefore, keeping other settings unchanged, the solution with buying delay t_i' instead of t_i from bidder i leads to larger incentive cost. Thus we have $(C_{B \setminus \{b_i\}}^{\hat{v}_0 - \hat{V}_i^d(t_i)} + \sum_{k=1}^{t_i} x_i^k) < (C_{B \setminus \{b_i'\}}^{\hat{v}_0 - \hat{V}_i^d(t_i')} + \sum_{k=1}^{t_i'} b_i'^k)$. Since $C_{B \setminus \{b_i'\}}^{\hat{v}_0 - \hat{V}_i^d(t_i')}$ is independent of \mathbf{b}_i , and $\mathbf{b}_i = \mathbf{x}_i$, we have $C_{B \setminus \{b_i\}}^{\hat{v}_0 - \hat{V}_i^d(t_i)} + \sum_{k=1}^{t_i} b_i^k = C_{B \setminus \{b_i'\}}^{\hat{v}_0 - \hat{V}_i^d(t_i')} + \sum_{k=1}^{t_i'} x_i^k$. Thus $\Delta u_i < 0$, under this case, bidder i also cannot obtain higher utility by setting $\mathbf{b}_i \neq \mathbf{x}_i$. ■

Theorem 2: In Win-Coupon, all bidders are guaranteed to obtain non-negative utility.

Proof: We have proved that for each bidder, say i , if it participates the auction game, setting its bids truthfully as $\mathbf{b}_i = \mathbf{x}_i$, is a weakly dominant strategy. The utility that i obtains equals $u_i = C_{B \setminus \{b_i\}}^{\hat{v}_0} - C_{B \setminus \{b_i\}}^{\hat{v}_0 - \hat{V}_i^d(t_i)} - \sum_{k=1}^{t_i} x_i^k = C_{B \setminus \{b_i\}}^{\hat{v}_0} - C_B^{\hat{v}_0}$, where t_i is the optimal length of delay that the 3G operator would buy from i to minimize the incentive cost. Since $C_{B \setminus \{b_i\}}^{\hat{v}_0}$ is the incentive cost incurred by the solution with 3G operator buying delay with length of 0 instead of t_i from bidder i , we have $C_{B \setminus \{b_i\}}^{\hat{v}_0} \geq C_B^{\hat{v}_0}$. Therefore, Win-Coupon guarantees that all bidders would obtain non-negative utility. ■

4) *Reserve Price*: In forward auction, the seller has the option to declare a *reserve price* for its resources. The reserve price means that the seller would rather withhold the resources if the bids are too low (lower than the reserve price). In Win-Coupon, to guarantee the 3G operator obtaining non-negative

profit, we also provide it with the option to set a reserve price to indicate the highest incentive cost it is willing to pay for offloading one traffic unit. If the value of coupon asked by the bidders exceeds the reserve price, the 3G operator would rather not trade with them. Suppose that the 3G operator sets a reserve price c_0 , which means that it is willing to spend at most c_0 for offloading one traffic unit. Adding the reserve price c_0 can be understood as adding a virtual bidder in the auction round. The bids sent by the bidder is $\{c_0, c_0, \dots, c_0\}$, and it can offload one traffic unit per one time unit of delay.

C. Predicting the Bidders' Offloading Potential

By motivating users to wait for some time, part of the 3G traffic can be offloaded to other intermittently available networks. One such example is DTN which generally coexists with 3G networks, and does not rely on any infrastructure. Mobile users can share data via DTNs by contacting each other. In urban area with higher user density, mobile users have more chances to contact other users who have their requested data. Large data requests such as video clips tend to drain most of the 3G resource, and such requests can also tolerate some delay. By offloading them via DTNs, the 3G payload can be significantly reduced. In this section, we illustrate how to predict the potentials of the users to offload their traffic via DTNs. Note that we use "user" and "node" interchangeably.

1) *Models*: Due to high node mobility, large data items such as video are hard to be completely transmitted when two nodes contact. In [24], it has been proved that the Random Linear Network Coding (RLNC) techniques can significantly improve the data transmission efficiency, especially when the transmission bandwidth is limited. Thus, in our model, RLNC is adopted to encode the original data into a set of coded packets. As long as the requester collects enough number of any linearly independent coded packets of its requested data, the data can be reconstructed. A desirable property of RLNC is that besides the data source, any node can independently generate new coded packets from its received ones to further diversify the coded packets. Due to page limit, we omit the details of RLNC and suggest interested readers to refer to [25].

In the following analysis to simplify the presentation, we assume that the contact process between each node pair follows i.i.d. Poisson distribution with rate λ , and exactly one packet can be transmitted when two nodes contact. Our analysis based on these assumptions can be extended to more general cases such as node pairs follow contact processes other than Poisson, and they can transmit arbitrary number of packets during a contact.

2) *The Basic Idea of Prediction*: We describe the rationale of prediction in one auction round. The starting time of this round is denoted by t_0 . The objective of the prediction is to calculate the expected volume of traffic $V_i^d(t)$ that can be offloaded to DTNs, if node i requests data item d and is willing to wait for delay t . By using RLNC, data item d has been encoded into a set of coded packets, and any s linear independent packets can be used to reconstruct d . We say that a node retrieves an *innovative* coded packet, if the

packet is linearly independent to all the coded packets cached in the node. We make an assumption that when a node contacts another node which has cached some coded packets of its requested data, it can always retrieve an innovative packet with very high probability. This assumption has been commonly used in prior works [24], [26]. In practice, if the size of the finite field to generate the coding coefficients is large enough, the probability is very close to 1.

Node i can retrieve one packet by contacting a node which has some coded packets of data item d , until it has collected all s packets. We use variable T_r ($1 \leq r \leq s$) to represent the time that node i retrieves r packets of d , and let $F_{T_r}(t)$ denote the Cumulative Distribution Function (CDF) of T_r . Thus, $V_i^d(t)$ can be computed as follows:

$$V_i^d(t) = h \int_0^t R(t)(1 - F_{T_s}(t))dt \quad (7)$$

where h is the size of one coded packet. $1 - F_{T_s}(t)$ is the probability that node i has not received all s packets at time t . $R(t)$ represents the receiving rate of node i at time t . Due to the i.i.d Poisson contact processes with rate λ between node pairs, $R(t)$ equals $\lambda N_d(t)$, where $N_d(t)$ denotes the total number of nodes that has at least one packet of data d at time t . Next, we describe how to calculate $N_d(t)$ and $F_{T_s}(t)$.

3) *Calculation of $N_d(t)$* : Based on nodes' interests to data d , all the nodes in the network except for node i can be divided into two classes: \mathcal{D} and \mathcal{I} , where \mathcal{D} contains all the non-interesters and \mathcal{I} contains all the interesters. The interesters include both the nodes which are downloading the data, and those which have already downloaded the data. To facilitate our analysis, we further divide class \mathcal{I} into $s + 2$ subclasses: $\mathcal{I}_0, \mathcal{I}_1, \dots, \mathcal{I}_s, \mathcal{I}_E$, based on the nodes' current downloading progress of data d . Specifically, \mathcal{I}_j ($j \in [0, s]$) includes all the nodes in the network other than node i which have already downloaded j packets of data d , and \mathcal{I}_E includes all the nodes which have finished data downloading before and already deleted the data from their buffer.

Based on our description of Win-Coupon in Section IV, each node in class \mathcal{I} has a promised delay. When the delay ends, the 3G operator would automatically push the remaining data packets to the node. For the nodes which lose the auction or choose to directly download data without bidding, their waiting delay is zero. To characterize the different waiting delays of the nodes, we further decompose each class \mathcal{I}_j ($j \in [0, s - 1]$) into $g + 1$ subclasses $\mathcal{I}_{j1}, \mathcal{I}_{j2}, \dots, \mathcal{I}_{jg}, \mathcal{I}_{j\infty}$, where g denotes the maximal remaining delay of the current downloading nodes. \mathcal{I}_{jk} ($j \in [0, s - 1], k \in [1, g]$) includes the nodes in class \mathcal{I}_j whose remaining delay is k time slots. For the new requesters which transit from class \mathcal{D} to class \mathcal{I} after time t_0 , we assume they prefer waiting a long delay to retrieve the complete data d via DTNs. Such new requesters in class \mathcal{I}_j are classified into the subclass $\mathcal{I}_{j\infty}$. Under this assumption, the derived $V_i^d(t)$ is a lower bound of the actual value, due to the following reason. If the delays of the new requesters are limited, after the delay, the 3G operator would directly push the traffic to them, which

potentially increases the data copies in the network, and results in a larger $V_i^d(t)$.

Next, we analyze how the defined network states vary with time. Let $N_{\mathcal{C}}(t)$ denote the number of nodes in class \mathcal{C} at time $t > t_0$. For example, $N_{\mathcal{I}_{jk}}(t)$ represents the number of nodes in class \mathcal{I}_{jk} at time t . We model the class transition process as two types: *active transition* and *passive transition*.

- **Active transition:** A node would actively transit from one class to another class by three ways: 1) The node is in class \mathcal{D} , and transits to class $\mathcal{I}_{0\infty}$ by generating a request for data d ; 2) The node is in class \mathcal{I}_j ($j \in [0, s-1]$) and transits to class \mathcal{I}_{j+1} by retrieving a packet from a contacted node; 3) The node is in class \mathcal{I}_s , and transits to class \mathcal{I}_E by deleting d from its buffer. The active transition processes are marked as the black arrows in Fig. 4.

- **Passive transition:** A node would passively transit from class \mathcal{I}_{jk} ($j \in [0, s-1], k \in [2, g]$) to class $\mathcal{I}_{j(k-1)}$, and transit from class \mathcal{I}_{j1} ($j \in [0, s-1]$) to class \mathcal{I}_s , when one time slot passes. Note that the latter transition is caused by the 3G operator pushing the remaining traffic to the node when its promised delay ends. The passive transition processes are marked as the blue dotted arrows in Fig. 4.

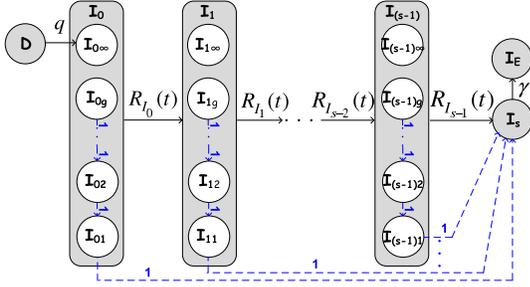


Fig. 4. Class transition processes

In the following, we use ODEs to first analyze the active transition process. We assume that there are qdt portion of the nodes in class \mathcal{D} that transit to class $\mathcal{I}_{0\infty}$ between time t and $t+dt$, where dt is infinitesimal, and q is the query rate decided by the popularity of data d . As a node in class \mathcal{I}_j ($j \in [0, s-1]$) contacts another node in class $\mathcal{I}_{j'}$ ($j' \in [1, s]$), the former node retrieves a packet from the latter and transit into class \mathcal{I}_{j+1} . Let $R_{\mathcal{I}_j}(t)$ ($j \in [0, s-1]$) denote the receiving rate of the node in class \mathcal{I}_j at time t , and we have:

$$R_{\mathcal{I}_0}(t) = \lambda \left(\sum_{y=1}^s N_{\mathcal{I}_y}(t) \right) \quad (8)$$

$$R_{\mathcal{I}_j}(t) = \lambda \left(\sum_{y=1}^s N_{\mathcal{I}_y}(t) - 1 \right), \forall j \in [1, s-1] \quad (9)$$

where 1 in Eq.(9) represents the node itself, since the node cannot retrieve new packet from itself. After a node has completely downloaded data d , it may delete it from its local buffer. We assume that there are γdt portion of the nodes in class \mathcal{I}_s that delete data d and transit to class \mathcal{I}_E , between time t and $t+dt$. Given all the initial value of the number of nodes in each

classes at the starting time, $N_{\mathcal{I}_{jk}}(t)$ ($j \in [0, s], k \in [1, g] \cup \infty$) can be computed by solving the following ODEs.

$$\frac{d(N_{\mathcal{I}_{0\infty}}(t))}{dt} = N_{\mathcal{D}}(t)q - N_{\mathcal{I}_{0\infty}}(t)R_{\mathcal{I}_0}(t) \quad (10)$$

$$\frac{d(N_{\mathcal{I}_{0k}}(t))}{dt} = -N_{\mathcal{I}_{0k}}(t)R_{\mathcal{I}_0}(t), \forall k \in [1, g] \quad (11)$$

$$\frac{d(N_{\mathcal{I}_{jk}}(t))}{dt} = N_{\mathcal{I}_{(j-1)k}}(t)R_{\mathcal{I}_{j-1}}(t) - N_{\mathcal{I}_{jk}}(t)R_{\mathcal{I}_j}(t), \quad \forall j \in [1, s-1], k \in [1, g] \cup \infty \quad (12)$$

$$\frac{d(N_{\mathcal{I}_s}(t))}{dt} = \sum_{\forall k} N_{\mathcal{I}_{(s-1)k}}(t)R_{\mathcal{I}_{s-1}}(t) - N_{\mathcal{I}_s}(t)\gamma. \quad (13)$$

Eq.(10) characterizes the varying rate of $N_{\mathcal{I}_{0\infty}}(t)$ which is composed of two parts: 1) $N_{\mathcal{D}}(t)q$ nodes transit to this class from class \mathcal{D} by generating a request for d , 2) $N_{\mathcal{I}_{0\infty}}(t)R_{\mathcal{I}_0}(t)$ nodes transit from the class to class $N_{\mathcal{I}_{1\infty}}(t)$ by retrieving a packet from its contacted node. Eq.(11) depicts the varying rate of $N_{\mathcal{I}_{0k}}(t)$. $N_{\mathcal{I}_{0k}}(t)R_{\mathcal{I}_0}(t)$ nodes transit from class \mathcal{I}_{0k} to class \mathcal{I}_{1k} by retrieving a packet from others. Eq.(12) shows the varying rate of $N_{\mathcal{I}_{jk}}(t)$ ($j \in [1, s-1], k \in [1, g] \cup \infty$), which also consists of two parts: 1) $N_{\mathcal{I}_{(j-1)k}}(t)R_{\mathcal{I}_{j-1}}(t)$ nodes join the class from class $\mathcal{I}_{(j-1)k}$, 2) $N_{\mathcal{I}_{jk}}(t)R_{\mathcal{I}_j}(t)$ nodes leave from the class to class $\mathcal{I}_{(j+1)k}$. Eq.(13) shows the varying rate of $N_{\mathcal{I}_s}(t)$, where the first term denotes the number of nodes that join the class from class $\mathcal{I}_{(s-1)k}$ ($k \in [1, g] \cup \infty$), and the second term denotes the number of nodes which delete the data and transit to class \mathcal{I}_E .

Besides the active transition, the passive transition would happen at the end of each time slot. At the end of each time slot, we update the number of nodes in each class as follows:

$$N_{\mathcal{I}_{jk}}(t) = N_{\mathcal{I}_{j(k+1)}}(t^-), \forall j \in [0, s-1], k \in [1, g-1] \quad (14)$$

$$N_{\mathcal{I}_s}(t) = N_{\mathcal{I}_s}(t^-) + \sum_{j=0}^{s-1} N_{\mathcal{I}_{j1}}(t^-). \quad (15)$$

The number of nodes in the rest of the classes which are not listed in Eq.(14) and (15) remains the same. Also, at the end of each time slot, the maximal delay of the existing downloading nodes would minus 1 (i.e., $g = g-1$). By combining the active and passive transition processes, the network state at any time t ($t > t_0$) can be derived. Thus, we can calculate $N_d(t)$, the number of nodes which has at least one packet of data d at time t , as $N_d(t) = \sum_{j=1}^s N_{\mathcal{I}_j}(t)$.

4) *Calculation of $F_{T_s}(t)$:* The derivative of $F_{T_r}(t)$ ($r \in [2, s]$) is represented as follows by using ODEs:

$$\frac{dF_{T_r}(t)}{dt} = \frac{Pr(T_r \leq t+dt) - Pr(T_r \leq t)}{dt} \quad (16)$$

$$= \frac{R(t)dt(Pr(T_{r-1} \leq t) - Pr(T_r \leq t))}{dt} \quad (17)$$

$$= R(t)(F_{T_{r-1}}(t) - F_{T_r}(t)), \forall r \in [2, s]. \quad (18)$$

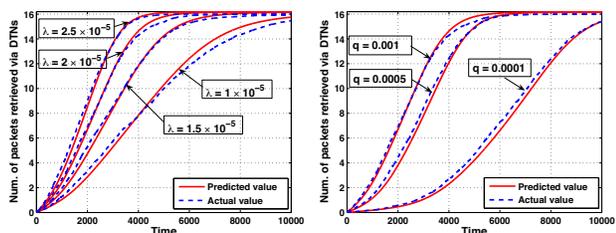
We ignore the probability that node i receives more than one packets during a very short time interval dt . Thus, the probability that T_r , the time for node i receives r packets, is between the range of $[t, t+dt]$ equals the probability that node

i exactly receives $r - 1$ packets before time t , and receives the r th packet during time t to $t + dt$. Thus, we derive Eq.(17) from Eq.(16). Similarly, we also derive $\frac{dF_{T_1}(t)}{d(t)} = R(t)(1 - F_{T_1}(t))$. Therefore, given the initial values that $F_{T_r}(t_0) = 0$ ($r \in [1, s]$), $F_{T_s}(t)$ can be derived by solving the following ODEs:

$$\frac{dF_{T_1}(t)}{d(t)} = R(t)(1 - F_{T_1}(t)) \quad (19)$$

$$\frac{dF_{T_r}(t)}{d(t)} = R(t)(F_{T_{r-1}}(t) - F_{T_r}(t)), \forall r \in [2, s]. \quad (20)$$

5) *Numerical Results*: To verify the accuracy of our prediction model and analyze the impacts of the system parameters, we numerically solve the ODEs and compare the prediction results to the actual values derived from the Monte-Carlo simulations. In the simulations, we generate 300 nodes following i.i.d. Poisson contact process, and one data item with 16 packets and query rate $q = 0.001$. The same set of parameters is imported to the ODEs. We focus on the number of downloaded packets along time t on a specific node, and compare the results derived in the simulation with that from solving the ODEs. The results given by the simulation are averaged over 200 runs. Fig. 5(a) shows the results with different contact rate λ . We can see that the prediction results are very close to the values given by the simulations, which verifies the accuracy of our prediction model. The larger the contact rate is, the earlier the node collects all 16 packets. We further compare the results when the query rate q varies, as shown in Fig. 5(b). The prediction also achieves results close to that of the simulations. As the query rate increases, the node collects more packets from other peers as time passes. This implies that if a node requests a popular item, its offloading potential is large.



(a) With different contact rate λ (b) With different data query rate q
Fig. 5. Comparison of predicted results and actual results

V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of Win-Coupon through trace-driven simulations. First, we introduce the simulation setup. Then, we evaluate the performance of Win-Coupon under different system parameters: the number of bidders, the reserve price, and the users' delay tolerance. The scale of the trace, in terms of the number of users and their contact frequencies, is generally smaller than that of the realistic scenarios. In our evaluation, this results in a long auction round for the 3G operator to collect enough bids, as well as long downloading delay experienced by the users. To reveal the faithful parameter values under real scale, we further

generate a large-scale trace by replicating the nodes in the original trace 10 times. The evaluation results of Win-Coupon on the large-scale trace are given at the end of this section.

A. Simulation Setup

Our performance evaluation are conducted on the UCSD trace [27], which records the contact history of 275 HP Jornada PDAs carried by students over 77 days. Based on the trace, we generate 50 data items, and each contains 8 packets. The query rate q and delete rate γ for each data item are randomly generated within the range of $[1.0 \times 10^{-8}, 1.0 \times 10^{-6}]$ and $[1.0 \times 10^{-8}, 5.0 \times 10^{-8}]$. When nodes request data, they can choose to attach bids with the request message based on their satisfaction function. In the simulations, we model the user satisfaction function as: $S(t) = p - at^b$, where p is the original data service charge, and we assume that all the data items have the same charge $p = 0.8$. a determines the scale of $S(t)$, and a smaller a results in higher delay tolerance. b determines the shape of $S(t)$. When $b > 1$, $b = 1$ and $b < 1$, $S(t)$ is a concave, linear and convex function respectively. In the simulations, we randomly generate parameter a and b within the range of $[0.08, 0.32]$ and $[0.8, 1.2]$ for each node to each data unless specified differently. In the simulations, the trace for the first five days is used for warmup, during which some nodes can directly download data without bidding. The presented results are averaged over 10 runs. To evaluate Win-Coupon, the following performance metrics are used:

- *Offloaded traffic*: The total amount of traffic in terms of the number of packets that are actually offloaded.
- *Allocated coupon*: The total incentive cost spent by the 3G operator for offloading purpose.
- *Average downloading delay*: The average time a bidder spends to get the complete data after he sends the data request.

B. Simulation results

1) *Impact of the number of bidders*: First, we evaluate the performance of Win-Coupon for different number of bidders. The results are shown in Fig. 6(a) 6(b) 6(c). The number of bidders is set to 40, 80, and 120 by varying the length of one auction round. The reserve price is set to 0.25, i.e., the 3G operator is willing to pay at most 0.25 for offloading one traffic unit. As can be seen from Fig. 6(a), the actual offloaded traffic by adopting Win-Coupon keeps close to the offloading target, until a certain upper bound reaches. The bound represents the upper limit of the traffic that can be offloaded by fully utilizing the delay tolerance and the offloading potential of the bidders given the reserve price. More traffic can be potentially offloaded if more bidders participate in the auction.

As can be seen from Fig. 6(b) 6(c), with the increase of offloading target, the allocated coupon and the average downloading delay increase accordingly, until reaching the offloaded traffic bound. The total value of coupon allocated by the 3G operator is strictly controlled by the reserve price which is marked as the black dotted line in Fig. 6(b). With the same amount of traffic that is actually offloaded, the increase in the number of bidders results in less allocated coupon and shorter

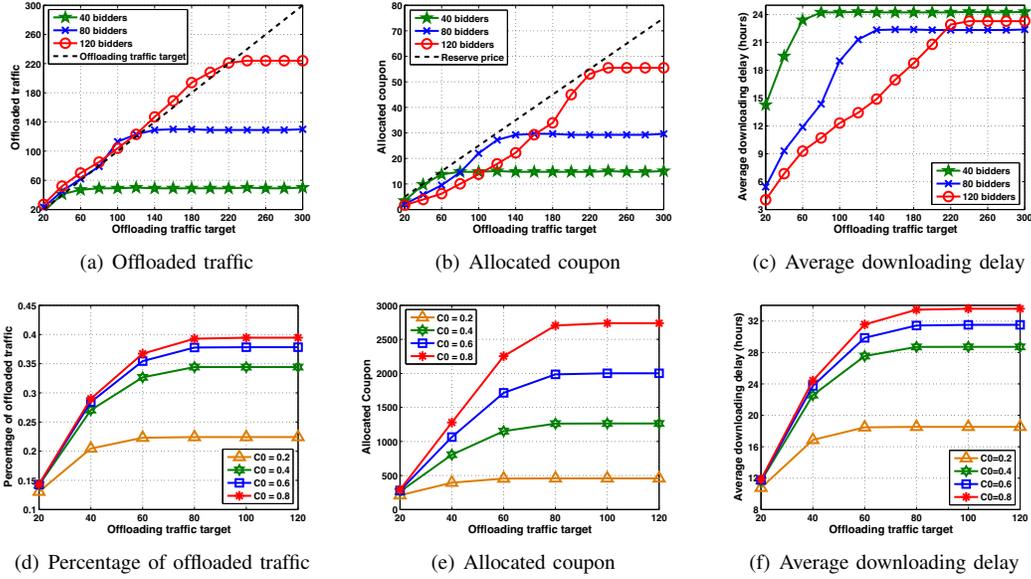


Fig. 6. Performance of Win-Coupon under different number of bidders and reserve prices

average delay. For example, when the number of bidders set to 40, 80 and 120, the 3G operator spends 9.8, 5.8, and 3.9 to actually offload 40 units of traffic, and the average downloading delay is 19.5, 9.3, and 6.9 hours. The reason behind this phenomenon is that when more bidders participate in the auction, it is more likely to have more bidders with high delay tolerance or large offloading potential. To offload the same amount of traffic, the bidders with high delay tolerance request less coupon to compensate their satisfaction loss, and the bidders with large offloading potential need to wait for shorter delay. Thus, the incentive cost and the average delay decrease when more bidders participate in the auction.

2) *Impact of the reserve price*: To evaluate the impact of reserve price on performance of Win-Coupon, we pick the contact history of three days (6th day – 8th day) from the trace to run the simulations. We fix the auction round to be one hour, and set the reserve price to 0.2, 0.4, 0.6, and 0.8 respectively. The results are shown in Fig. 6(d) 6(e) 6(f). As can be seen, with the increase of reserve price, more traffic can be offloaded. This is because the higher reserve price indicates larger willingness of the 3G operator to pay for offloading one unit of traffic, and then potentially motivates more users for offloading. When the reserve price is set to 0.8, almost 40% of the traffic can be offloaded as shown in Fig. 6(d). However, higher reserve price results in higher incentive cost as shown in Fig. 6(e). To balance this tradeoff, the 3G operator can set the reserve price appropriately according to its budget. Also, as shown in Fig. 6(f), the average downloading delay increases as the reserve price increases, since more users are selected as the winning bidder and are motivated to wait.

3) *Impact of the delay tolerance of the users*: To evaluate the impact of users' delay tolerance, we generate three scenarios with high, middle and low delay tolerance, by randomly setting the parameter a within the range of $[0.04, 0.16]$, $[0.08, 0.32]$, and $[0.16, 0.64]$. The reserve price is set to 0.4,

and other settings remain the same as that used in the last subsection. The simulation results are shown in Fig. 7. As can be seen from Fig. 7(a) 7(c), when the delay tolerance becomes larger, more traffic can be offloaded, and the average downloading delay increases. When the offloading target is set to 100, more than 50% of the traffic is offloaded to DTNs in the scenario of high delay tolerance.

Fig. 7(b) 7(d) show the value of allocated coupon and the percentage of winning bidders with different offloading targets. When the offloading target is as low as 20, as the users' delay tolerance gets higher, the incentive cost of the 3G operator drops, since less coupon is requested by the bidders. Also, fewer bidders win the game in the scenarios with higher delay tolerance, since more traffic can be offloaded by each winning bidder. As the offloading target further increases, the amount of traffic that is actually offloaded almost remains the same in the low delay tolerance scenario. This is because the users in this scenario are not willing to wait longer and the traffic being offloaded is bounded. Then, both the allocated coupon and the percentage of winning bidders in this scenario remain the same. However, in the scenarios with middle and high delay tolerance, as the offloading target increases, more traffic can be offloaded by better exploiting users' delay tolerance, then both the two metrics increase accordingly.

4) *Large-scale trace*: In the previous simulations on the UCSD trace, the impacts of the system parameters on the performance of Win-Coupon are presented, but the duration of one auction round and the downloading delay of the bidders are set to be long. This is due to the small scale of the trace. In reality, however, the network scale would be much larger. It is easier for the 3G operator to collect enough bids and for the users to contact more peers, within a short period of time. To validate Win-Coupon for more practical use, we generate a large-scale trace including 2750 nodes by replicating the nodes in the UCSD trace 10 times. The contact rate

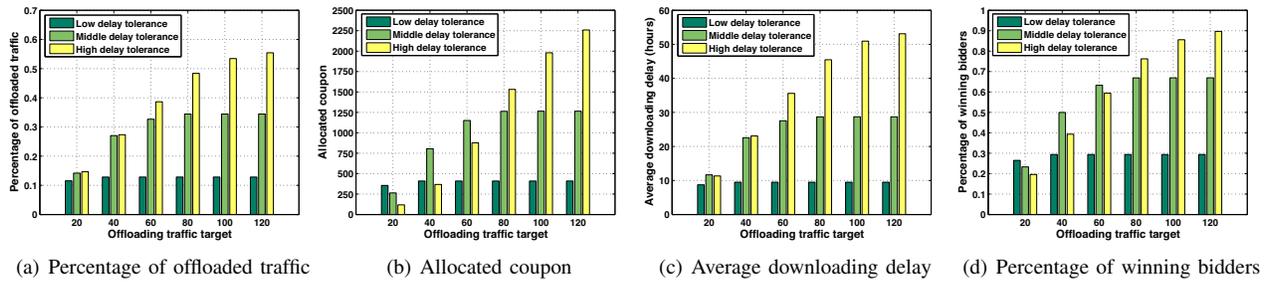


Fig. 7. Performance of Win-Coupon under different bidders' delay tolerance

between the nodes in the same copy remains the same as the original trace, and the contact rate between the nodes in the different copies is set to the average aggregated contact rate derived in the original trace. The parameter a is set randomly within the range of $[1, 1.6]$, and the duration of one auction round is reduced to 10 minutes. Fig. 8 shows the evaluation results. Similar trend is captured as in previous simulations. As the offloading traffic target increases, more traffic can be actually offloaded. When the offloading traffic target reaches 120, almost 40% of the traffic can be offloaded, which provides comparable performance to that derived from the UCSD trace. More importantly, the average delay is only 4.38 hours. As can be seen, both the duration of one auction round and the average delay becomes reasonable for practical use when the network scale becomes large.

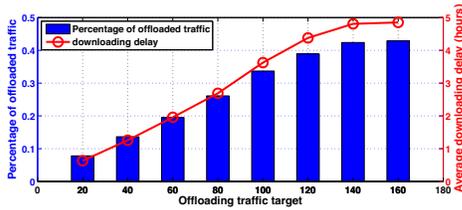


Fig. 8. Performance of Win-Coupon on large-scale trace

VI. CONCLUSION

In this paper, we propose a novel incentive framework for 3G traffic offloading. The basic idea is to motivate the mobile users with high delay tolerance and large offloading potential to offload their traffic to other intermittently available networks such as DTN. To capture the dynamic characteristics of users' delay tolerance, we design an incentive mechanism based on reverse auction. Our mechanism is proved to guarantee truthfulness, individual rationality, and low computational complexity. Moreover, we take DTN as a case study, and design an accurate model to predict the offloading potential of the users by using stochastic analysis. Extensive trace-driven simulations validate the efficiency and practical use of our incentive framework in 3G traffic offloading.

REFERENCES

- [1] M. Reardon, "Cisco predicts wireless-data explosion." [Online]. Available: http://news.cnet.com/8301-30686_3-10449758-266.html
- [2] B. Han, P. Hui, V. Kumar, M. V. Marathe, G. Pei, and A. Srinivasan, "Cellular traffic offloading through opportunistic communications: A case study," in *Proc. of ACM CHANTS*, 2010.
- [3] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, "Mobile data offloading: How much can wifi deliver?" in *Proc. of ACM CoNEXT*, 2010.

- [4] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3g using wifi," in *Proc. of ACM MOBISYS*, 2010.
- [5] J. Whitbeck, Y. Lopez, J. Leguay, V. Conan, and M. D. Amorim, "Relieving the wireless infrastructure: When opportunistic networks meet guaranteed delays," in *Proc. of IEEE WoWMoM*, 2011.
- [6] C. Boldrini, M. Conti, and A. Passarella, "Modelling Data Dissemination in Opportunistic Networks," in *Proc. of ACM CHANTS*, 2008.
- [7] P. Costa, C. Mascolo, M. Musolesi, and G. Picco, "Socially Aware Routing for Publish Subscribe in Delay-Tolerant Mobile Ad Hoc Networks," *IEEE JSAC*, vol. 26, no. 5, pp. 748–760, 2008.
- [8] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in Delay Tolerant Networks: A Social Network Perspective," in *Proc. of ACM MOBIHOC*, 2009.
- [9] X. Zhuo, Q. Li, G. Cao, Y. Dai, B. Szymanski, and T. L. Porta, "Social-Based Cooperative Caching in DTNs: A Contact Duration Aware Approach," in *Proc. of IEEE MASS*, 2011.
- [10] X. Zhuo, Q. Li, W. Gao, G. Cao, and Y. Dai, "Contact Duration Aware Data Replication in Opportunistic Mobile Networks," in *Proc. of IEEE ICNP*, 2011.
- [11] A. J. Nicholson and B. D. Noble, "Breadcrumbs: Forecasting mobile connectivity," in *Proc. of ACM MOBICOM*, 2008.
- [12] M. R. Ra, J. Paek, A. B. Sharma, R. Govindan, M. H. Krieger, and M. J. Neely, "Energy-delay tradeoffs in smartphone applications," in *Proc. of ACM MOBISYS*, 2010.
- [13] P. Cramton, "Spectrum Auctions," *Handbook of Telecommunications Economics*, pp. 605–639, 2002.
- [14] P. Xu, S. Wang, and X. Li, "SALSA: Strategyproof Online Spectrum Admissions for Wireless Networks," *IEEE Trans on Computers*, pp. 1691–1702, 2010.
- [15] X. Zhou, S. Gandhi, S. Suri, and H. Zheng, "Ebay in the Sky: Strategy-Proof wireless spectrum auctions," in *Proc. of ACM MOBICOM*, 2008.
- [16] S. Gandhi, C. Buragohain, L. Cao, H. Zheng, and S. Suri, "A General Framework for Wireless Spectrum Auctions," in *Proc. of IEEE DySPAN*, 2007.
- [17] L. Anderegg and S. Eidenbenz, "Ad hoc-VCG: A Truthful and Cost-Efficient Routing protocol for Mobile Ad hoc Networks with Selfish Agents," in *Proc. of ACM MOBICOM*, 2003.
- [18] W. Wang, X. Li, and Y. Wang, "Truthful Multicast in Selfish Wireless Networks," in *Proc. of ACM MOBICOM*, 2004.
- [19] D. P. Bertsekas, D. A. Castanon, and H. Tsaknakis, "Reverse Auction and the Solution of Inequality Constrained Assignment Problems," *SIAM J. on Optimization*, pp. 268–299, 1993.
- [20] W. Vickrey, "Counterspeculation, Auction and Competitive Sealed Tenders," *Journal of Finance*, pp. 8–37, 1961.
- [21] E. Clarke, "Multipart pricing of public goods," *Public Choice*, pp. 17–33, 1971.
- [22] T. Groves, "Incentives in Teams," *Econometrica*, pp. 617–631, 1973.
- [23] N. Nisan and A. Ronen, "Algorithmic Mechanism Design," *Games and Economic Behavior*, vol. 35, pp. 166–196, 2001.
- [24] Y. Lin, B. Li, and B. Liang, "Stochastic Analysis of Network Coding in Epidemic Routing," *IEEE JSAC*, vol. 26, no. 5, pp. 794–808, 2008.
- [25] P. Chou, Y. Wu, and K. Jain, "Practical Network Coding," in *Proc. of Annual Allerton Conf. on Comm., Control, and Comput.*, 2003.
- [26] C. Zhang, Y. Fang, and X. Zhu, "Throughput-delay tradeoffs in large-scale manets with network coding," in *Proc. of IEEE INFOCOM*, 2009.
- [27] M. McNett and G. Voelker, "Access and mobility of Wireless PDA users," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 9, no. 2, pp. 40–55, 2005.