Optimal Mobile Data Offloading Policy through Delay Tolerant Networks with Selfish Nodes

Zhen Liu, Yahui Wu, and Su Deng

Abstract-Mobile data access is suffering the curse of the computationally enhanced increase of smart phones, which overloads the traditional cellular network. This paper tries to offload the cellular network traffic through Delay Tolerant Networks (DTN) formed by the short-range communication technologies in these smart phones (e.g., WiFi, Bluetooth). In particular, the content providers deliver L initial replicas of the same data through cellular network to L users. Then those users further propagate the information among all the subscribed users through opportunistic communications. However, phones are carried by different users who may be selfish and not want to help others. If the initial replicas are delivered to these selfish nodes, other users may not get data timely. This paper proposes a method to select the proper number of initial replicas and an optimal policy using the Markov process to distribute the limited initial replicas. Simulation results show the accuracy of the model, and numerical results show that our optimal policy can get better performance.

Index Terms—Cellular traffic offloading, delay tolerant networks, selfish users, markov process, optimal policy.

I. INTRODUCTION

With the increasing of smart phones, such as the iPhone, Blackberry, and Android devices, more and more users become eager to engage with the mobile applications. Applications related to social networks [1], mobile phone sensing [2], and content distribution [3] are just a few of the examples. Therefore, mobile data traffic is growing at an unprecedented rate, which causes many problems to network providers. Many researchers from networking and financial sectors [4] forecast that mobile users will consume 7GB of traffic per month, which is 5.4 times more than that of today's average per month by 2014. In addition, the total mobile data traffic throughout the world will reach about 3.6 exabytes per month, which is about 39 times more than that in 2009 and the compound annual increasing rate is about 108%. Therefore, it is very important to find new ways to manage such increasing mobile data traffic.

This paper focuses on the offloading problem of cellular network through the short range communication technologies (e.g., Bluetooth). The basic idea is that: the content provider only delivers limited replicas (initial replicas) of the data through traditional cellular network, and other users who do not get data from the content provider directly will obtain data from the users that have data. If the number of the initial replicas is smaller than the number of users, the traffic is offloaded.

However, the end-to-end path between two users may not exist when they want to communicate with each other, so this policy can only provide intermittent and opportunistic network connectivity and these users form a Delay Tolerant Network (DTN) [5]. Any two nodes can communicate with each other only when they come into the transmission range of each other in DTN. In order to provide communication service in such challenging networks, nodes in DTN adopt a *store-carry-forward* communication strategy. This strategy exploits the opportunistic connectivity and node mobility to relay and carry data. Specially, when the next hop is not available for the current node, it will *store* the data in its buffer, *carry* the data along the movement, and *forward* until a new communication opportunity arises.

There are many challenges in the process that using DTN to offload cellular network traffic. In this paper, we mainly consider two of them. The first one is how to select the proper number of initial replicas, and the second one is how to select the proper users to carry the initial replicas.

Recently, many data dissemination methods [7]-[9] have been proposed in DTN and can be used in different situations. A detail survey can be found in [10]. However, most of them need certain prior knowledge of the network, so it is hard to use in many cases. Even though some methods use the on-line learning methods to get the knowledge, this learning process may be slowly, that is, there may be not enough time to learn. Therefore, in this paper, we use the Epidemic Routing (ER) method, in which each node receiving the data carries it as it moves, and then forwards it to all new nodes it encounters.

In this paper, we explore the optimal mobile data offloading policy using the Markov process. The main contributions of this paper can be summarized as follows:

- We are the first to propose an optimal replicas distribution method based on the selfish nature of users using the Markov process.
- We propose a method to select proper number of initial replicas.
- We check the accuracy of our model through some simulations.

II. OPTIMAL OFFLOADING POLICY

A. Problem Description

Consider a network combined with N mobile phones (by

Manuscript received March 5, 2013, revised July 13, 2013.

The authors are with Science and Technology on Information System Engineering Laboratory, College of Information System and Management, National University of Defense Technology, ChangSha, HuNan, China (e-mail: {hero12251976, yahui_wu, sudeng}@ 163.com).

abuse of language, they also can be seen as consumers, users or nodes) and some Base Stations (BS). Every phone can connect with the cellular network through a BS. A snapshot of the network can be seen in Fig. 1. The thick lines indicate the connection through cellular network, and the thin lines show the opportunistic link between smart phones through Bluetooth. Suppose a content provider creates a message Dat the initial time, and it wants to propagate D to all users in the network. Therefore, all of the users are subscribers. To limit the network traffic, one message can have at most Lreplicas. In some cases, consumers (people carried the phone) cannot tolerate too much delay, so the content provider must finish their service with limited time. The task of this paper is to select the proper value of L and to distribute the limited replicas efficiently.



Fig. 1. A snapshot of a small fraction of the network.

Every mobile phone can connect with the cellular network through the BS. At the initial time, the content provider may release the L replicas to any one of the BS or some of them, and the BS received message distribute these replicas cooperatively. That is, they select L initial consumers to carry the initial replicas cooperatively and this equals to the case that there are only one BS. The non-cooperative case between BS will be our future work. Every consumer got message becomes a service and may forward D to other consumers through the opportunistic link. So users may receive data from users which are not in the initial selected set, and the data dissemination process can be seen as a multi-hop process. The opportunistic link exists between two consumers only when they come into the transmission range of each other, which means a communication contact, so the mobility rule of the consumers is critical. In this paper, we assume that the occurrence of contacts between two nodes follows a Poisson distribution. Therefore, the inter-meeting time between two contacts conforms to the exponential distribution. This assumption has been used in wireless communications many years. At present, some works shows that this assumption is only an approximation to the message transmission process, and they reveal that nodes encounter with each other according to the power law distribution [11]. However, they also recognize that if you consider long traces, the tail of the distribution is not heavy tailed, but exponential. In addition, the work in [12] shows that individual inter-meeting time can be shaped to be exponential by choosing an appropriate domain size with respect to a given time scale. Moreover, the work in [13] describes the inter-meeting time of individual human in mobile social networks by the exponential process and validates this model experimentally on real trace data. There are also some works which model the inter-meeting time

between vehicles and find that it follows the exponential distribution by analyzing a large number of real car mobility traces [14], [15]. In fact, the phones are carried by peoples who may be walking or driving a car. Therefore, we can use the datasets in social networks or vehicles for the analysis. According to the description above, the assumption for the inter-meeting time is rational in our application, and we can assume that the inter-meeting time between two consumers follows an exponential distribution with the parameter denoted by λ .

As described in Section I, there are two classes of the selfish behaviors. To describe the social selfishness, we divided the *N* nodes into *K* classes. The number of nodes in *i* class is N_i . Therefore, we have the following equation,

$$N = \sum_{j=1}^{K} N_j \tag{1}$$

Due to the impact of the individual selfishness, nodes in *i* class forward data towards each other with probability p_i . Similarly, due to the impact of the social selfishness, nodes in *i* class forward data towards nodes in *j* class with probability p_{ij} . Both p_i and p_{ij} belongs to the range [0, 1]. As shown in [6], one node is more willing to forward data to nodes that belongs to the same community. Therefore, we have $p_i > p_{ij}$ in many cases. However, the relation between p_i and p_{ij} does not have any impact on deriving our theoretical model. Therefore, we do not make any assumption about the relation between them.

B. Data Dissemination Process

Given *L* initial replicas, we use the Markov process to model the data dissemination process through opportunistic links. Let symbol $M_i(t)$ denote the number of nodes that have message *D* in *i* class at time *t*, so the state of the network at time *t* can be denoted as $M(t)=(M_1(t), M_2(t), \dots M_K(t))$. Once a new node gets message, we can say that one-step transition occurs. We assume that from time *t*, one-step transition finished at time $t+\Delta t$, and then the state of the network at time $t+\Delta t$ must be one element of the set $X(t+\Delta t)$, which is shown as follows,

$$X(t+\Delta t) = \{X_1(t+\Delta t), X_2(t+\Delta t), \dots, X_K(t+\Delta t)\}$$
(2)

Through one-step transition, only one node can get message, and the node may be in any class. If the node belongs to *i* class, we can say that the network is in state $X_i(t+\Delta t)$ at time $t+\Delta t$, and the state can be described as follows,

$$X_{i}(t + \Delta t) = \{M_{1}(t), M_{2}(t), ..., M_{i}(t) + 1, M_{i+1}(t), ...M_{K}(t)\}$$
(3)

But if all the nodes in *i* class have got message at time *t*, that is $M_i(t)=N_i$, there is no node that can get message again at time $t+\Delta t$. Therefore, the system cannot go into state $X_i(t+\Delta t)$ through one-step transition in this situation. For simplicity, when $M_i(t)=N_i$, we set,

$$X_i(t + \Delta t) = \emptyset \tag{4}$$

Symbol \emptyset means the invalid state. So we can say that from time *t*, the network may come into any valid state in set

 $X(t+\Delta t)$.

Obviously, one node in i class may get message from nodes in any class. If this node gets message from the one belonging to the same class, the rate can be described as follows,

$$\lambda p_i M_i(t) (N_i - M_i(t)) \tag{5}$$

That is, at time *t*, there are $M_i(t)$ nodes that have message in *i* class, so the last $N_i - M_i(t)$ nodes do not have message. Because two nodes encounter with each other following an exponential distribution with parameter λ , there are $M_i(t)$ (N_i - $M_i(t)$) possible Poisson process which may induce one node get message. Further, due to nodes in *i* class forward toward each other with probability p_i , we can get (5). Using the same method, we can get the rate that nodes in *i* class get message from nodes in *j* class,

$$\lambda M_{i}(t)(N_{i} - M_{i}(t))p_{ii}$$
(6)

From analysis above, we know that the total rate that one node in *i* class gets message from time *t* can be shown as follows $(X_i(t+\Delta t) \text{ is a valid state}),$

$$M(t) \to X_i(t + \Delta t), \text{ with rate}$$

$$\sum_{j=1 \& j \neq i}^{K} \lambda M_j(t) (N_i - M_i(t)) p_{ji} + \lambda M_i(t) (N_i - M_i(t)) p_i$$
(7)

It is easily to see that the number of nodes in *i* class that have message may be any integer from 0 to N_i . Therefore, the total number of the states is,

$$\prod_{i=1}^{K} (1+N_i) \tag{8}$$

If every node in the network has message, the state cannot change again, so the network goes into the absorption state which is denoted by $Dst=(N_1, N_2, ..., N_K)$. In addition, states other than the absorption state can be seen as transient state.

Now, we define the generate matrix Q as follows,

$$Q = \begin{pmatrix} T & R \\ 0 & 0 \end{pmatrix} \tag{9}$$

This matrix presents the transition rate from one state to another. T is a sub-matrix and it denotes the rate of the transition from one transient state to another, and the state include Dst. does not Obviously, T is $((1+N_1)(1+N_2)...(1+N_K)-1) \times ((1+N_1) (1+N_2)...(1+N_K)-1)$ matrix. R is a column vector with $(1+N_1)(1+N_2)\dots(1+N_K)-1$ elements and it denotes the rate of the transition from one transient state to the absorbing state Dst. The left 0 is a row vector with $(1+N_1)(1+N_2)\dots(1+N_K)$ -1 elements and it denotes the rate of the transition from Dst to any transient state. The right 0 is a vector with only one element and it can denote the rate of the transition from Dst to Dst. Combining with (7), we can get every element of the matrix T easily. For example, the rate of the transition from state M(t) to other transient states is shown as follows,

$$\begin{cases} T(M(t), X_{i}(t + \Delta t)) = \sum_{j=1 \& j \neq i}^{K} \lambda M_{j}(t)(N_{i} - M_{i}(t))p_{ji} \\ +\lambda M_{i}(t)(N_{i} - M_{i}(t))p_{i}, 1 \le i \le K \& M_{i}(t) < N_{i} \\ T(M(t), others) = 0 \end{cases}$$
(10)

Symbol *others* presents the states which are not in the set $X(t+\Delta t)$. If the system reaches to the absorption state *Dst* from M(t) through one-hop transition, M(t) must satisfy the following condition, that is,

$$\exists l \le i \le K, satisfy M_i(t) = N_i - 1 \& M_i(t) = N_i, for \forall j \ne i$$
(11)

Similar to (10), we can get the rate of the transition from state M(t) to *Dst* easily when M(t) satisfies the condition shown in (11),

$$R(M(t), Dst) = \sum_{j=1\& j\neq i}^{K} \lambda M_j(t) p_{ji}$$

$$+\lambda M_i(t) p_i, M_i(t) = N_i - 1\& M_i(t) = N_i, j \neq i$$
(12)

Now, we begin to explore the one-hop transition probability matrix M according to Q, which is shown as follows,

$$M = \begin{pmatrix} P & PR \\ 0 & 0 \end{pmatrix}$$
(13)

P is a sub-matrix and it denotes the probability of the transition from one transient state to another when leaving the current state. The number of the elements in *P* equals to that in *T*. Similarly, *PR* denotes the probability of the transition from one transient state to *Dst* when leaving the current state. We can know the meaning of the two 0 elements easily through above analysis.

According to (10), the total rate of leaving state M(t) is,

$$Trate(M(t)) = \sum_{i=1}^{K} \left(\sum_{j=1 \& j \neq i}^{K} \lambda M_{j}(t) (N_{i} - M_{i}(t)) p_{ji} + \lambda M_{i}(t) (N_{i} - M_{i}(t)) p_{i} \right)$$
(14)

So the probability of the one-step transition from state M(t) to another state LS is,

$$\begin{cases} P(M(t), LS) = T(M(t), LS) / Trate(M(t)) \\ PR(M(t), Dst) = R(M(t), Dst) / Trate(M(t)) \end{cases}$$
(15)

C. Optimal Performance Control

Though there are N mobile phones in the network, the content providers may only forward L < N replicas of the same data. To reduce the energy cost and offload the traffic, the content providers want to make the value of L as small as possible. However, because the limited patience of users, the content providers must make sure that all the users can get the data before certain time. Here, we assume that the deadline is denoted by T.

Now, our problem is how to select the smallest value of L under limited time constraint T. First, we will explore the delivery delay (the time needed to make all users get data) according to the data dissemination process in above section.

Let DT(k) denote the expected delivery delay till all the

users get data (reaches to the absorption state Dst), starting from state k. Obviously, DT(Dst)=0. Similarly, let ST(k)denote the residence time in state k. From (14) we know that the total rate of leaving state k equals to Trate(k). So the residence time in state k can be described as,

$$ST(k) = 1/Trate(k), k \neq Dst$$
 (16)

Because every state other than the absorption state *Dst* is the transient state, the total rate of leaving current state must be greater than 0. Above equation is right for every transient state. For state *Dst*, the transition rate is 0, so the residence time is infinite.

Suppose the set of all states is denoted by *S*. Therefore, by conditioning on the one-hop transition out of the current state, we have,

$$DT(k) = \sum_{j \in S - \{Dst\}} P(k, j) DT(j) + ST(k)$$

+ $PR(k, Dst) DT(Dst)$
= $\sum_{j \in S - \{Dst\}} P(k, j) DT(j) + ST(k)$ (17)

Define DT as a column vector of the expected delivery delay starting from any valid transient state, and ST a column vector of the residence time. We can obtain the following formula easily,

$$DT = P * DT + ST \tag{18}$$

Let *I* denote the identity matrix, we can get,

$$DT = (I - P)^{-1}ST \tag{19}$$

Vector *DT* includes the expected delivery delay from any transient state.

Therefore, under the limited time T, the feasible initial distribution Fstate(T) is shown as follows,

$$Fstate(T) = \{i \mid DT(i) \le T, i \in S\}$$
(20)

In fact, we can obtain the total number TN(k) of users that get data in any state $k=(k_1, k_2, ..., k_k)$ easily which is shown in (21),

$$TN(k) = \sum_{j=1}^{K} k_j \tag{21}$$

Combined with (20) and (21), we can get the smallest value of the number of initial replicas L.

$$L = \min\{TN(i) \mid i \in Fstate(T)\}$$
(22)

When we get the optimal initial number of replicas, how to deploy them optimally is another problem. Because our main object is to make all users get data as soon as possible under limited initial replicas, we think that the optimal distribution method should have the smallest delivery delay. In fact, it is easily to see that one distribution of the initial replicas is corresponding to a state, so the distribution also can be seen as the initial state or initial policy. Suppose the optimal distribution is $k=(k_1, k_2, ..., k_K)$, then it must satisfy the following constraint,

$$\begin{cases} DT(k) \le DT(i), \forall i \in \{1, 2, \dots K\} - \{k\} \\ \sum_{j=1}^{K} k(j) \le L \end{cases}$$

$$(23)$$

Therefore, the optimal distribution policy is a feasible policy, that is, the number of users receiving data in the initial distribution policy is not more than L. In addition, the optimal policy has the smallest delivery delay in all of the feasible polices. Note that there may be more than one optimal distribution policy, and we can adopt any one of them. With above constraint, we can get one of them easily through traverse the set DT.

III. SIMULATION AND NUMERICAL RESULTS

In this section, we will check the accuracy of our model, and we run several simulations using the Opportunistic Network Environment (ONE) simulator [16]. Our first simulation is based on the Random Waypoint (RWP) mobility model, which is commonly used in many mobile wireless networks. Here, we set K=3, $N_1=10$, $N_2=15$, and $N_3=20$. That is, the users are divided into 3 classes. The first class has 10 users, the second class has 20 users and the third class has 30 users. These users move according to the RWP mobility model within a 100m×100m terrain according to a scale speed chosen from a uniform distribution from 4m/s to 10m/s.

In fact, the selfish degree of different users may be different and there are may be infinite possible combinations. Therefore, we cannot carry out the simulation on every situation. For simplicity, in this section, we set $p_1 = p_2 = p_3 = 0.7$, and users in one class forward toward users in another class with probability 0.2.

Both the optimal distribution policy and the optimal initial value *L* can be got based on (19). Therefore, if we can prove the equation is accuracy, we can say that our model is accurate. To check the accuracy of (19), we may carry out the simulation from any initial state. In this section, we randomly select two cases. The first one is from the state s_1 = (2, 5, 8) and the second one is from state s_2 = (3, 2, 1). Let the transmission range *r* increase from 1m to 10m, we can get the simulation and numerical results as follows,



The second simulation is based on a real motion traces from about 2100 operational taxis for about one month in Shanghai city collected by GPS [17]. The location information was recorded at every 40 seconds within an area of 102 km². We divide nodes in the dataset into two classes, and we randomly select 10 nodes from the dataset in the first class. Let the number of nodes in the second class increase from 10 to 100. These nodes are also picked randomly. For the theoretical related parameters, we set $p_1 = p_2 = 0.7$, and users in the first class forward toward users in the other class with probability 0.2. Here, we also randomly select two initial distributions. The first one is from the state $s_1 = (2, 5)$ and the second one is from state $s_2 = (3, 2)$. Simulation result is shown in Fig. 3.



Fig. 3. Simulation with the real motion trace.

From Fig. 2 and Fig. 3 we can see that the theoretical results are very close to that obtained by simulation in both simulations, and the average deviations between the simulation and theoretical results are about 5.62% in RWP model, and about 3.88% in Shanghai city trace. This demonstrates the accuracy of our model.

IV. CONCLUSION

This paper explores the mobile data offloading policy through DTN. In this policy, only part of the users download data through traditional cellular network directly, and other users get data through opportunistic communication from users who have data. By markov process, we propose an optimal policy to select the number of initial replicas and an optimal policy to distribute the limited initial replicas under the influence of selfish behaviors. Simulation results show the accuracy of our model. Numerical results show that our optimal policy can get better performance and the selfish behaviors can greatly lower the performance of the offloading method.

REFERENCES

- S. Bell, A. McDiarmid, and J. Irvine, "Nodobo: mobile phone as a software sensor for social network research," in *Proc. IEEE VTC*, 2011, pp. 1-5.
- [2] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, 2010, pp. 140-150.

- [3] U. Lee, S. Jung, D.-K. Cho, A. Chang, J. Choi, and M. Gerla, "P2P content distribution to mobile Bluetooth users," *IEEE Transactions on Vehicular Technology*, vol. 59, 2010, pp. 344-355.
- [4] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, "Mobile data offloading: how much can wifi deliver?" in *Proc. ACM CoNEXT*, 2010.
- [5] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proc. ACM SIGCOMM*, 2003.
- [6] Q. Li, S. Zhu, and G. Cao, "Routing in socially selfish delay tolerant networks," in *Proc. IEEE INFOCOM*, 2010, pp.1-9.
- [7] W. Gao and G. Cao, "User-centric data dissemination in disruption tolerant networks," in *Proc. IEEE INFOCOM*, 2011.
- [8] P. Hui, J. Crowcorft, and E. Yoneki, "Bubble rap: social-based forwarding in delay tolerant networks," in *Proc. ACM MOBIHOC*, 2008.
- [9] T. Spyropoulos, T. Turletti, and K. Obrazcka, "Routing in delay tolerant networks comprising heterogeneous populations of nodes," *IEEE Transactions on Mobile Computing*, vol. 6, 2009.
- [10] M. Waldir and .M. Paulo, "Survey on opportunistic routing for delay/disruption tolerant networks," *Technical report*, SITI-TR-11-02, University Lus ófona, 2011.
- [11] T. Karagiannis, J.-Y. L. Boudec, and M. Zojnovic, "Power law and exponential decay of inter-contact times between mobile devices," in *Proce. ACM MOBICOM*, 2007.
- [12] H. Cai and D. Y. Eun, "Crossing over the bounded domain: from exponential to power-law intermeeting time in mobile ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 17, 2009, pp.1578-1591.
- [13] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: a social network perspective," in *Proc. ACM MOBIHOC*, 2009, pp.299-308.
- [14] K. Lee, Y. Yi, J. Jeong, H. Won, I. Rhee, and S. Chong, "Max-Contribution: on optimal resource allocation in delay tolerant networks," in *Proc. IEEE INFOCOM*, 2010, pp.1-9.
- [15] H. Zhu, L. Fu, G. Xue, Y. Zhu, M. Li, and L. M. Ni, "Recognizing exponential inter-contact time in VANETs," in *Proc. IEEE INFOCOM*, 2010, pp.1-5.
- [16] A. Keränen, J. Ott, and T. Kärkkänen, "The ONE simulator for DTN protocol evaluation," in *Proc. Simulation Tools and Techniques*, 2009, pp. 1-10.
- [17] SJU Traffic Information Grid Team. Grid Computing Center. *Shanghai taxi trace data*. [Online]. Available: http://wirelesslab.sjtu.edu.cn/



Zhen Liu was born in 1976, PhD. He is an associate professor of the National University of Defense Technology, Changsha, Hunan, China. His main research interests include cyber physical system, internet of things and delay tolerant networks.



Yahui Wu was born in 1984. He is a PhD candidate of the National University of Defense Technology, Changsha, Hunan, China. His main research interests include cyber physical system, delay tolerant networks and wireless sensor networks.



Su Deng was born in 1963. He is a professor and PhD supervisor of the National University of Defense Technology, Changsha, Hunan, China. His main research interests include cyber physical system, internet of things, information integrated processing and decision support.