

# Optimal Relaying in Heterogeneous Delay Tolerant Networks

Yong Li\*, Zhaocheng Wang\*, Depeng Jin\*, Li Su\*, Lieguang Zeng\*, Sheng Chen†

\*State Key Laboratory on Microwave and Digital Communications

Tsinghua National Laboratory for Information Science and Technology

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

†School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.

Email: jindp@mail.tsinghua.edu.cn

**Abstract**—In Delay Tolerant Networks (DTNs), there exists only intermittent connectivity between communication sources and destinations. In order to provide successful communication services for these challenged networks, a variety of relaying and routing algorithms have been proposed with the assumption that nodes are homogeneous in terms of contact rates and delivery costs. However, various applications of DTN have shown that mobile nodes should be divided into different classes in terms of their energy requirements and communication ability, and real application data have revealed the heterogeneous contact rates between node pairs. In this paper, we design an optimal relaying scheme for DTNs, which takes into account nodes' heterogeneous contact rates and delivery costs when selecting relays to minimise the delivery cost while satisfying the required message delivery probability. Extensive results based on real traces demonstrate that our relaying scheme requires the least delivery cost and achieves the largest maximum delivery probability, compared with the schemes that neglect nodes' heterogeneity.

## I. INTRODUCTION

In Delay Tolerant Networks (DTNs), there exist no end-to-end paths from communication sources to destinations during most of the time due to the factors such as wireless propagation effect, sparse node density and node mobility [1]. Traditional ad hoc routing protocols, which rely on end-to-end paths [2], fail to work in DTNs. Consequently, a new routing mechanism, known as store-carry-and-forward [1], was proposed to provide communications. Due to the randomness of intermittent connectivity between the nodes, the performance of DTN largely depends on how to relay and copy messages to achieve high delivery ratio while maintaining low delivery cost. Researchers have proposed a variety of schemes, such as two-hop relaying [3], epidemic routing [4], spray and wait [3], to tackle this difficult relaying/routing problem. However, these works are based on homogeneous settings of mobile nodes, which are assumed to be identical and to follow the same mobility pattern as well as to have the same contact rate between any two nodes. Recent measurements based on the real traces [5] have revealed the existence of heterogeneity in nodes' contact rates. Moreover, in many envisioned applications of DTN, mobile nodes belong to diverse classes, e.g. vehicles, cellphones, sensors, and so on [1]. These various types of nodes have very different characteristics in terms of their energy consumption, communication ability, mobility patterns and other properties.

The nodes with heterogeneous features will have different message delivery costs associated with them.

The heterogeneous features of mobile nodes should be explicitly exploited in relaying and routing in order to improve the message delivery performance in DTN. In this paper, we consider the problem of optimal relaying for DTN consisting of heterogeneous mobile nodes. We explicitly consider both the heterogeneity in contact dynamics and the heterogeneity in relay costs, caused by different nodes' energy, communication and mobility properties. Our novel contribution is threefold. Firstly, we model the relaying problem by appropriately characterising the nodes' heterogeneity in contact rates and delivery costs, and formulate the optimal relaying in the heterogeneous environment as the corresponding optimisation problem. Secondly, we prove that the hardness of this relaying optimisation problem is NP-hard, and derive a heuristic algorithm based on dynamic programming which is capable of obtaining the optimal relay performance. Thirdly, we evaluate our proposed relaying scheme under the experimental data traces collected from realistic environments, and compare our scheme with some existing schemes that do not consider the node heterogeneity as well as the existing proposals for heterogeneous DTNs. The simulation results obtained demonstrate the effectiveness of our relaying scheme.

Before presenting our optimal relaying scheme and performance evaluation results, we first summarise the related works. Heterogeneity in mobile nodes was considered in the work [6], which proposed a framework to analyse the capacity scaling properties of the mobile heterogeneous DTN. Lee and Eunt [5] focused on how the heterogeneous contact dynamics impact the performance of relaying and routing algorithms. Spyropoulos *et al.* [7] addressed the routing problem in the DTN comprising multiple classes of heterogeneous nodes. However, there are important differences compared with our work. We consider the node heterogeneity in a generic environment where every node is different in terms of the mobility pattern and forwarding cost, while study [7] divided the nodes into some classes of different behaviors but each class containing homogeneous nodes. Moreover, in our extensive simulation under the experimental data traces collected from realistic environments, we demonstrate the performance enhancement of our scheme over the scheme provided in [7].

## II. SYSTEM MODEL AND PROBLEM FORMULATION

### A. System Model

Consider the DTN consisting of a set of wireless mobile nodes, in which the source node is denoted by  $S$ , the destination node is denoted by  $D$ , and  $N + M$  nodes are relay nodes. The set of relay nodes is denoted by  $R = \{R_1, R_2, \dots, R_{N+M}\}$ . Since the density of nodes in DTN is sparse, the relay nodes can communicate with the source or destination node only when they are in the transmission range of each other, which is referred to as a communication contact. The occurrence of the contacts between two nodes is assumed to follow the Poisson distribution, which was validated by the work [8]. Therefore, the contact rate of nodes  $R_i$  and  $R_j$ , where  $i \neq j$ , in our model is assumed to be Poisson distributed with the parameter  $\lambda_{i,j}$ , which characterises the nodes' heterogeneity in terms of contact rate. Let  $T$  be the message life time, and define the message delivery probability as the probability that the message can be delivered to the destination  $D$  within the life time  $T$ . To guarantee a required system performance, we require that the message delivery probability is no less than  $P$ .

In order to deliver a message from the source  $S$  to the destination  $D$ , we select some relays from  $R$ . When the source selects relays to transmit a message, we assume that there are  $M$  nodes in the communication range of the source  $S$ , which are grouped in the set denoted by  $R^I = \{R_1^I, R_2^I, \dots, R_M^I\}$ , while the other  $N$  nodes are not in the range, which are grouped in the set denoted by  $R^O = \{R_1^O, R_2^O, \dots, R_N^O\}$ . Obviously,  $R^O \cup R^I = R$  and  $R^O \cap R^I = \emptyset$ . Selecting relay nodes only from  $R^I$  is referred as the In-Contact Relay (ICR), while selecting relay nodes only from  $R^O$  is known as the Out-of-Contact Relay (OCR). In the ICR, the relays selected by the source  $S$  are in contact with  $S$ . After the relays are selected, the source  $S$  transmits the message to the relays, and then relays transmit the message to the destination  $D$  when they are in next contact with  $D$ . In the OCR, the relays selected are not in contact with the source  $S$ . The source  $S$  has to wait for the relays to come in contact in order to transmit the message to them. Finally the message is successfully transmitted if any one of the relays with the message contacts the destination  $D$ .

We will consider the generic relaying scheme in which the source selects relays from both  $R^I$  and  $R^O$ . We also consider the heterogeneity of relay nodes by assuming that there is a cost factor associated with each relay node. Specifically, let  $C_i^I$  be the cost associated with the relay node  $R_i^I \in R^I$  and  $C_j^O$  be the cost associated the relay node  $R_j^O \in R^O$ . Therefore, the optimal relay selection can be formulated as the optimisation problem of how to choose relay nodes to achieve the message delivery probability  $P$  from the source  $S$  to the destination  $D$  within the time  $T$ , while achieving the minimum total cost.

### B. Problem Formulation

We begin by considering the ICR case. Note that the destination  $D$  is not in contact with the source  $S$ , and any  $R_i^I \in R^I$  is in contact with  $S$  but not in contact with  $D$ . Otherwise,

$S$  can trivially transmit the message to  $D$  immediately. We denote the random variable  $\varphi_i$  as the indicator that  $R_i^I \in R^I$  is contacted by the destination  $D$  within the time  $T$ , which is defined as follows

$$\varphi_i = \begin{cases} 1, & \text{if } R_i^I \text{ can contact } D \text{ within } T, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,  $\varphi_i$  follows the Bernoulli distribution, and its expectation, denoted by  $\bar{\varphi}_i$ , equals to  $1 - e^{-\lambda_{i,d}^I T}$ , where  $\lambda_{i,d}^I$  is the contact rate between node  $R_i^I$  and the destination  $D$ . Define  $p_i = 1 - \bar{\varphi}_i$ , which is the probability that  $R_i^I$  cannot contact the destination  $D$  within the time  $T$ . Further define

$$x_i = \begin{cases} 1, & R_i^I \text{ is selected as a relay,} \\ 0, & \text{otherwise.} \end{cases}$$

as the relay selection indicator for node  $R_i^I \in R^I$ . Under the selection policy  $x_i$ , the message delivery probability that  $R_i^I$  can achieve can be expressed as  $1 - p_i^{x_i}$ , and the delivery cost of  $R_i^I$  can be expressed by  $C_i^I x_i$ .

Next we turn to the OCR case. Since all the relay nodes  $R_j^O \in R^O$  are not in contact with the source  $S$ , we need to consider the time required by  $R_j^O$  to come in contact with the data source. With regarding to the delivery probability that node  $R_j^O$  can achieve, we first consider the distribution of the time needed for  $R_j^O$  to successfully transmit the message to the destination  $D$ , and we denote the associated random variable as  $Y_j$ . From the message transmission process, we can see that  $Y_j$  contains the time the source  $S$  needed to contact  $R_j^O$ , which is denoted as  $T_1$ , and the time  $R_j^O$  needed to contact the destination  $D$ , which is denoted by  $T_2$ . The distribution of  $Y_j$  is given by *Lemma 1*.

*Lemma 1:* The Probability Density Function (PDF) of  $Y_j$ , denoted by  $f_j(y)$ , is given by

$$f_{Y_j}(y) = \frac{\lambda_{j,d}^O \lambda_{j,s}^O (e^{-\lambda_{j,s}^O y} - e^{-\lambda_{j,d}^O y})}{\lambda_{j,d}^O - \lambda_{j,s}^O}, \quad (1)$$

where  $\lambda_{j,s}^O$  is the contact rate between  $R_j^O$  and  $S$ , and  $\lambda_{j,d}^O$  is the contact rate between  $R_j^O$  and  $D$ .

*Proof:* The time for the source  $S$  to deliver a message to the destination  $D$  via relay  $R_j^O$  include  $T_1$  and  $T_2$ , where  $T_1$  is the time needed for  $R_j^O$  to contact  $S$ , and  $T_2$  is the time needed for  $R_j^O$  to contact  $D$  after it has encountered  $S$ . From our system model,  $T_1$  follows the exponential distribution with the parameter  $\lambda_{j,s}^O$ , and its PDF is  $f_{T_1}(y) = \lambda_{j,s}^O e^{-\lambda_{j,s}^O y}$ . According to the memoryless property of the exponential distribution,  $T_2$  also follows the exponential distribution but with the parameter  $\lambda_{j,d}^O$ , and its PDF is  $f_{T_2}(y) = \lambda_{j,d}^O e^{-\lambda_{j,d}^O y}$ . Because  $Y_j = T_1 + T_2$ , the PDF of  $Y_j$  is the convolution of  $f_{T_1}(y)$  and  $f_{T_2}(y)$ , namely  $f_{Y_j}(y) = f_{T_1}(y) \otimes f_{T_2}(y)$ . Therefore, we have

$$\begin{aligned} f_{Y_j}(y) &= \lambda_{j,s}^O \lambda_{j,d}^O \int_0^y e^{-(\lambda_{j,s}^O - \lambda_{j,d}^O)t} e^{-\lambda_{j,d}^O t} dt \\ &= \frac{\lambda_{j,d}^O \lambda_{j,s}^O (e^{-\lambda_{j,s}^O y} - e^{-\lambda_{j,d}^O y})}{\lambda_{j,d}^O - \lambda_{j,s}^O}, \end{aligned}$$

which proves the Lemma.  $\blacksquare$

Let  $q_j = \text{Prob}(Y_j \geq T)$  denote the probability that the message cannot be transmitted to the destination  $D$  via relay node  $R_j^O$  within the time  $T$ . According to the PDF of  $Y_j$  given in Lemma 1,  $q_j$  can be calculated as follows

$$q_j = \int_T^\infty f_{Y_j}(y) dy = \frac{\lambda_{j,d}^O e^{-\lambda_{j,s}^O T} - \lambda_{j,s}^O e^{-\lambda_{j,d}^O T}}{\lambda_{j,d}^O - \lambda_{j,s}^O}. \quad (2)$$

Similarly, we also define  $y_j \in \{0, 1\}$  as the indicator whether node  $R_j^O \in R^O$  is selected as a relay. The delivery cost associated with  $R_j^O$  is obviously  $C_j^O y_j$ .

Recall that our goal is to select appropriate relays from  $R = R^I \cup R^O$  to deliver the message within the time  $T$  to  $D$  with the probability  $P$  at the minimum cost. Combining the both cases of ICR and OCR, we can formulate the optimal relay selection problem as following optimization problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^M C_i^I x_i + \sum_{j=1}^N C_j^O y_j \\ \text{s.t.} \quad & \begin{cases} \prod_{i=1}^M p_i^{x_i} \prod_{j=1}^N q_j^{y_j} \leq 1 - P, \\ x_i \in \{0, 1\}, i = 1, 2, \dots, M, \\ y_j \in \{0, 1\}, j = 1, 2, \dots, N. \end{cases} \end{aligned} \quad (3)$$

By defining

$$\begin{aligned} u_k &= \begin{cases} C_k^I, & 1 \leq k \leq M, \\ C_{k-M}^O, & M+1 \leq k \leq M+N, \end{cases} \\ z_k &= \begin{cases} x_k, & 1 \leq k \leq M, \\ y_{k-M}, & M+1 \leq k \leq M+N, \end{cases} \\ \varpi_k &= \begin{cases} p_k, & 1 \leq k \leq M, \\ q_{k-M}, & M+1 \leq k \leq M+N, \end{cases} \end{aligned} \quad (4)$$

the optimal relaying-node selection problem (3) can be rewritten as

$$\begin{aligned} \min \quad & \sum_{k=1}^{N+M} u_k z_k, \\ \text{s.t.} \quad & \begin{cases} \prod_{k=1}^{N+M} \varpi_k^{z_k} \leq 1 - P, \\ z_k \in \{0, 1\}, k = 1, 2, \dots, N+M. \end{cases} \end{aligned} \quad (5)$$

### III. OPTIMAL RELAYING SCHEME

We refer to the optimal relaying scheme obtained by solving the optimisation problem (5) as the Optimal Heterogeneous Relaying Scheme (OHRS).

#### A. Hardness of the Problem

The hardness of the optimisation problem of (5) is expressed by the following theorem.

*Theorem 1:* The optimal relaying node selection problem (5) is NP-hard.

*Proof:* We begin by considering the optimisation problem

$$\begin{aligned} \max \quad & \sum_{k=1}^{N+M} u_k \hat{z}_k, \\ \text{s.t.} \quad & \begin{cases} \sum_{k=1}^{N+M} w_k \hat{z}_k \leq W, \\ \hat{z}_k \in \{0, 1\}, k = 1, 2, \dots, N+M. \end{cases} \end{aligned} \quad (6)$$

Obviously, the above problem is a knapsack problem, which is an NP-hard problem [9]. Now by defining  $\hat{z}_k = 1 - z_k$ ,

$$w_k = -\log(\varpi_k) \quad (7)$$

and

$$W = \log(1 - P) - \sum_{k=1}^{N+M} \log(\varpi_k), \quad (8)$$

the optimisation problem (5) is transformed exactly into the knapsack problem (6).  $\blacksquare$

Theorem 1 shows that the optimal relaying node selection problem is an NP-hard problem. Thus, we cannot obtain optimal solutions by polynomial-time algorithms, and can at most hope for approximation algorithms of higher complexity. In the following subsection, we design a pseudo-polynomial-time algorithm to obtain the optimal solution of this problem.

#### B. Heuristic Algorithm for the Scheme

We design a heuristic algorithm based on dynamic programming to choose the relay nodes that solve the optimal relaying-node selection problem (5). For notational convenience, we will drop the super-indexes  $I$  and  $O$  from relay nodes, and simply denote all the relay nodes by  $R_k$  for  $1 \leq k \leq N+M$ . We refer to  $w_k$  defined in (7) as the delivery probability gain of relay node  $R_k$  and  $W$  defined in (8) as the total weight. Given the contact rates and the message delivery ratio  $P$ ,  $w_k$  for  $1 \leq k \leq N+M$  and  $W$  can readily be computed. In the relay node selection, we divide the selection process into the  $N+M$  steps, and in the  $k$ th step we decide whether  $R_k$  is selected as a relay. Let  $s_m$  denote the state variable of the quantised delivery probability gain in the  $k$ th step, and  $h_k(s_m)$  denote the maximum weight that we can obtain by selecting the  $k, k+1, \dots$ , and  $M+N$  relay nodes when the quantised delivery probability gain is  $s_m$ . The detailed selection process is shown in Algorithm 1.

The time complexity and memory requirements of this algorithm are in the order of  $(N+M)M_Q$ ,  $\mathcal{O}((N+M)M_Q)$ , which is acceptable because the number of the relay nodes is often limited and most of the source nodes have such computational and memory capacity in real-world networks.

### IV. PERFORMANCE EVALUATION

#### A. Experimental Environment and Settings

We used two experimental traces, *Infocom06* which was collected from MIT Reality Mining Project [10] and *Reality* which was gathered by the Hagggle Project [11]. These two data traces covered two very different DTN environments, namely, a disperse university campus with the experiment period of

---

**Algorithm 1** Heuristic algorithm for the source  $S$  to select relay nodes, where  $\lceil \bullet \rceil$  denotes the integer ceiling operator.

---

- 1: Compute the delivery probability gains  $w_k$  of relay nodes  $R_k$  by their contact rates;
  - 2: Compute the weight  $W$  by the required system delivery probability  $P$  and deliver probability gains  $w_k$ ;
  - 3: Compute the quantisation precision  $M_Q$  by  $M_Q = \left\lceil \frac{W}{\min_{i,j} |w_i - w_j|} \right\rceil$ ;
  - 4: Define  $\bar{w}_k = \frac{\lceil w_k M_Q \rceil}{M_Q}$ ,  $1 \leq k \leq M + N$ , and quantise the set  $[0, W]$  as  $\mathbb{W} = \{0, \frac{W}{M_Q}, \frac{2W}{M_Q}, \dots, W\}$ ;
  - 5: Initialise  $h_k(0) = 0$ ,  $1 \leq k \leq N + M$ ,  $h_1(s) = 0$ ,  $s \in \mathbb{W}$ ;
  - 6: **for** Each relay node  $R_k$ ,  $1 \leq k \leq N + M$ , **do**
  - 7:   **for**  $m \in \{0, 1, \dots, M_Q\}$ , set  $s_m = \frac{mW}{M_Q} \in \mathbb{W}$  **do**
  - 8:     **if**  $s_m < w_k$  **then**
  - 9:       Mark  $R_k$  as Neglected;  $h_k(s_m) = h_{k-1}(s_m)$ ;
  - 10:    **else**
  - 11:     **if**  $h_{k-1}(s_m) > h_{k-1}(s_m - \bar{w}_k) + u_k$  **then**
  - 12:       Mark  $R_k$  as Neglected;  $h_k(s_m) = h_{k-1}(s_m)$ ;
  - 13:     **else**
  - 14:       Mark  $R_k$  as Selected;
  - 15:        $h_k(s_m) = h_{k-1}(s_m - \bar{w}_k) + u_k$ ;
  - 16:     **end if**
  - 17:    **end if**
  - 18: **end for**
  - 19: **end for**
- 

several months (*Reality*) and a concentrated conference site with the experiment period of a few days (*Infocom06*).

As mentioned in the introduction section, most of the existing works only considered the case of homogeneous mobile nodes, and a few exceptions did evaluate how the heterogeneous contact rates influence the performance of relaying algorithms. Therefore, we specifically designed two relay node selection schemes as two benchmarks to reflect this reality. The first relaying benchmark scheme did not consider nodes' heterogeneous features at all and, therefore, was denoted as the Non-Heterogeneous Relay Scheme (NHRS). The second relaying benchmark scheme did consider heterogeneous contact rates but assumed the other nodes' features that influenced the delivery cost being homogeneous, and we referred to this scheme as the Heterogeneous Contact Relay Scheme (HCRS). Also from the introduction section, we noted that [7] was an up to date study on heterogeneous nodes populated DTNs. Since the work [7] only studied several spraying schemes, we needed to change these spraying schemes to the corresponding relaying schemes. The two benchmark relaying schemes so adopted from [7] were the Most-Mobile-First Relaying (MMFR) and Most-Social-First Relaying (MSFR). The MMFR selected the most "mobile" nodes as the relays first. In the work [7], it was assumed that each node carried a label that stated the level of its "mobility". In our simulation, we used the average inter-contact time of each node as its mobility measure. In the MSFR, each node had a metric of "sociability", which was

defined as the number of users that the node had encountered in a given duration of time.

Since there existed many heterogeneous factors that influenced the delivery costs of nodes, the delivery costs might follow some random distribution among the nodes. On the other hand, the traces did not have the relevant information in this aspect and, therefore, we examined several cost distributions among relay nodes, including the fixed constant cost as well as the uniform and normal cost distributions. The expectation of delivery cost distribution of  $C_i^I$  ( $C_j^O$ ) was set to 10 (12). In the fixed constant-cost setting, the delivery costs of all the nodes were set to the same value of 10 (12). In the uniformly distributed cost setting, the delivery costs of nodes were uniformly distributed between 1 (1) and 19 (23), while in the normally distributed cost setting, the delivery costs of nodes followed the normal distribution with the expectation 10 (12) and a variance of 10. Because the values of normal distribution were from negative infinity to positive infinity, we only used the values in the 10% to 90% range of its PDF.

In the simulation, we selected the two most popular nodes that had the largest numbers of contacts with other nodes as the source  $S$  and destination  $D$ , and then divided the one-third of the remaining nodes into  $R^I$  and the other two-third of the nodes into  $R^O$  randomly. In the investigation, we changed the required system delivery probability  $P$ , ran different relaying schemes, and compared their total relay costs.

## B. Evaluation Results

We used the *Infocom06* trace to compare our OHRS with the NHRS and HCRS benchmarks, and the results obtained are plotted in Fig. 1. Specifically, Fig. 1 (a) depicts the case with the fixed constant node cost. As the message delivery probability increased, all the three relay schemes spent higher cost to achieve the desired delivery probability. As expected, our OHRS and the HCRS benchmark achieved almost the identical delivery cost, and they outperformed the NHRS by the 66.4% reduction in cost on average. This was because in this case the nodes' heterogeneity was only featured by the different contact rates of nodes, and both the OHRS and HCRS considered the heterogeneous contact rates while the NHRS did not take into account this heterogeneous feature in choosing relay nodes. Fig. 1 (b) shows the total costs achieved by the three schemes under the uniformly distributed node cost, where we can observe that our OHRS reduced the total cost on average by 64.3% and 88.5%, respectively, over the HCRS and NHRS. The results for the normal distribution of node cost, illustrated in Fig. 1 (c), also demonstrate that our OHRS on average outperformed the HCRS and NHRS by the 67.7% and 87.9% cost reductions, respectively.

We used both the *Infocom06* and *Reality* traces to compare our OHRS with the MMFR and MSFR benchmarks adopted/modified from [7], under the normal distribution of node costs. The results obtained are depicted in Fig. 2 for the two traces. The results of Fig. 2 clearly show that our OHRS achieved a superior performance over the MMFR and MSFR benchmarks. Specifically, for the *Infocom06* trace under the

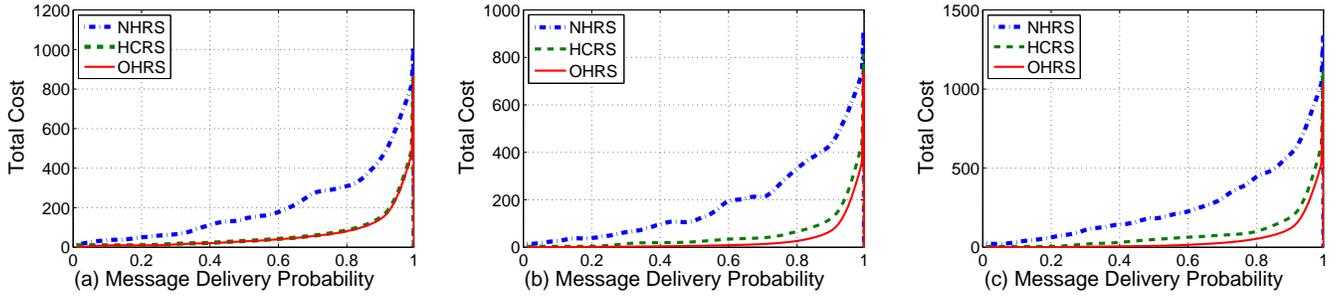


Fig. 1. Performance comparison of our OHRS with the NHRS and HCRS benchmarks with the *Infocom06* trace under the different cost distributions: (a) with the fixed node cost, (b) with the uniformly distributed node cost, and (c) with the normally distributed node cost.

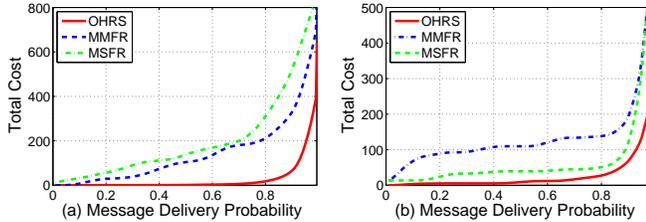


Fig. 2. Performance comparison of our OHRS with the two benchmark schemes adopted from [7] trace with the normally distributed node costs: (a) for the *Infocom06* trace, and (b) for the *Reality* trace.

normal distribution of node costs, the OHRS attained the least cost given the same delivery probability. This was because, even though the MMFR and MSFR did consider the heterogeneous features in terms of nodes' mobility and sociability, respectively, they did not take into account the heterogeneous node costs in selecting relay nodes. It can also be seen that the MMFR achieved a better performance than the MSFR in this case. The reason was that in the *Infocom06* trace, the nodes that moved often encountered the source and destination nodes more frequently. Therefore, the MMFR was able to achieve a smaller delivery cost than the MSFR, given the same delivery probability. For the *Reality* trace under the normal distribution of node costs, the OHRS reduced the average total costs by 23.0% and 92.1%, respectively, in comparison with the MSFR and MMFR. Unlike the *Infocom06* trace, interestingly, the MSFR scheme achieved a much better performance than the MMFR scheme for the *Reality* trace. It appears that the nodes with high social metric could achieve higher delivery probability than the nodes that moved often in the *Reality* trace. The performance of the MMFR and MSFR schemes seemed to depend on specific network environments. By contrast, our OHRS performed equally well in various network environments.

## V. CONCLUSIONS

We have investigated the problem of optimal relaying for DTNs consisting of heterogeneous mobile nodes. Based on the experimentally collected real traces, we have conducted extensive simulations. The results obtained in the experimental

study have demonstrated that our optimal relaying scheme achieves superior performance over these existing benchmark relaying schemes.

## ACKNOWLEDGMENT

This work was supported by National Major Scientific and Technological Specialized Project (No. 2010ZX03004-002-02), National Basic Research Program (No. 2007CB310701) and National High Technology Research and Development Program (No. 2008AA01Z107, No. 2008AA01A331 and No. 2009AA011205), PCSIRT and TNLIST.

## REFERENCES

- [1] K. Fall, "A delay-tolerant network architecture for challenged Internets," in *Proc. ACM SIGCOMM 2003 Conf. Applications, Technologies, Architectures, and Protocols for Computer Communication* (Karlsruhe, Germany), Aug. 25-29, 2003, pp. 27-34.
- [2] M. S. Blumenthal and D. D. Clark, "Rethinking the design of the Internet: The end-to-end arguments vs. the brave new world," *ACM Trans. Internet Technology*, vol. 1, no. 1, pp. 70-109, 2001.
- [3] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case," *IEEE/ACM Trans. Networking*, vol. 16, no. 1, pp. 77-90, 2008.
- [4] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," *Technical Report CS-200006*, Department of Computer Science, Duke University, 2000.
- [5] C.-H. Lee and D. Y. Eunt, "Heterogeneity in contact dynamics: Helpful or harmful to forwarding algorithms in DTNs?" in *Proc. 7th Int. Symp. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks* (Seoul, Korea), June 23-27, 2009, pp. 1-10.
- [6] M. Garetto, P. Giaccone, and E. Leonardi, "Capacity scaling in delay tolerant networks with heterogeneous mobile nodes," in *Proc. ACM MOBIHOC'07* (Montreal, Canada), Sept. 9-14, 2007, pp. 41-50.
- [7] T. Spyropoulos, T. Turletti, and K. Obraczka, "Routing in delay-tolerant networks comprising heterogeneous node populations," *IEEE Trans. Mobile Computing*, vol. 8, no. 8, pp. 1132-1147, 2009.
- [8] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: A social network perspective," in *Proc. 10th ACM Int. Symp. Mobile Ad Hoc Networking and Computing* (New Orleans, LA), May 18-21, 2009, pp. 299-308.
- [9] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms* (2nd Edition). MIT Press, 2001.
- [10] N. Eagle and A. Pentland, "Reality mining: Sensing complex social systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255-268, 2006.
- [11] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay tolerant networks," in *Proc. 9th ACM Int. Symp. Mobile Ad Hoc Networking and Computing* (Hong Kong, China), May 26-30, 2008, pp. 241-250.