An Optimal Relaying Scheme for Delay-Tolerant Networks With Heterogeneous Mobile Nodes

Yong Li, *Member, IEEE*, Zhaocheng Wang, *Senior Member, IEEE*, Li Su,
Depeng Jin, *Member, IEEE*, and Sheng Chen, *Fellow, IEEE*

Abstract—To provide communication services in delay-tolerant networks (DTNs) where there may exist no end-to-end paths between mobile node pairs, a variety of relaying and routing algorithms have been proposed under the assumption that the mobile nodes are homogeneously distributed in the network with the same contact rate and delivery cost. However, experimental data have revealed the heterogeneous contact rates between node pairs, and various applications of DTNs have shown that the mobile nodes often belong to different types in terms of energy consumption, communication ability, and other properties. Following the philosophy of exploiting the heterogeneous features of nodes to enhance the routing performance, we design an optimal relaying scheme for DTNs, which takes into account the nodes' heterogeneous contact rates and delivery costs when selecting relays to minimize the delivery cost while satisfying the required message delivery probability. We use the trace-driven simulations to demonstrate the effectiveness of our optimal relaying scheme in various distributions of nodes' delivery costs and mobility environments. Simulation results show that our proposed optimal relaying scheme requires the least delivery cost and achieves the largest maximum delivery probability, compared with the schemes that neglect or do not fully take into account nodes' heterogeneity.

Index Terms—Delay-tolerant networks (DTNs), heterogeneous networks, message delivery cost, message delivery probability, optimal relaying scheme.

I. INTRODUCTION

THE traditional view of wireless networks as connected graphs over which end-to-end paths have to be established [1] is carried over from conventional computer communication

networks. This classic concept might not be applicable to some existing and emerging wireless networks, where there may be no contemporaneous end-to-end paths between the source and destination pairs for a variety of reasons, including node mobility, wireless propagation phenomena, power limitations, and malicious attacks [2]. To provide communication services in such highly challenging wireless networks, delay/disruption-tolerant networks (DTNs) are proposed [2]–[5], which exploit the opportunistic connectivity and node mobility to relay and carry messages around the network. When the next hop is not immediately available for the current node to deliver a message, some relay nodes will store the message in their buffers, carry the message along their movements, and forward the message to other nodes when a communication opportunity is occurring, which helps to transmit the message further [3]. Since this store–carry–forward mechanism can transmit data even in the challenging environment of no end-to-end path between the source and destination, many related interesting applications are experimented and conducted, including mobile social networks based on human mobility [6], sensor networks for wildlife tracking and habitat monitoring [2], vehicular ad hoc networks for road safety and commercial applications [7], [8], and deep-space interplanetary networks [9]. Due to the randomness of intermittent connectivity between the nodes, the performance of a 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 to tackle this difficult relaying/routing problem, such as epidemic routing [10], which is a flooding-based protocol, and two-hop relaying [11], spray and wait [11], gossip relaying [12], and MaxProp [7], which aim to reduce the overhead of epidemic routing. Specifically, Spyropoulos *et al.* [11] described a family of spray-and-focus routing algorithms and show that spray and focus cannot only significantly improve the transmission probability per message but reduce the average delivery delay as well, whereas Burgess *et al.* [7] proposed the MaxProp based on prioritizing the schedule of packets to be dropped. In the gossip-based relaying scheme [12], on the other hand, a "gossiping" probability is used to determine the message transmissions. Furthermore, some research aims to achieve optimal system performance [13]–[19]. For example, Picu and Spyropoulos [13], [15] discussed the problem of optimal relay selection, whereas Krifa *et al.* [17], [18] proposed an optimal message dropping policy and a forwarding schedule algorithm to minimize the average delivery delay or to maximize the average

delivery rate. Li *et al.* [19] provided an optimal adaptive buffer management scheme to decide the message transmission and dropping to maximize the average delivery rate or minimize the delivery delay. All the aforementioned works, however, are based on homogeneous settings of mobile nodes, which are assumed to be identical, follow the same mobility pattern, and have the same contact rate between any two nodes. Recent measurements based on the real traces [20] have revealed the existence of heterogeneity in nodes' contact rates. Moreover, in many envisioned applications of DTNs, mobile nodes belong to diverse classes, e.g., vehicles, cellphones, sensors, etc. [3]. These various types of nodes have very different characteristics in terms of their energy consumption, communication ability, mobility patterns, and other properties. These heterogeneity features will result in different message delivery costs of nodes. Some of the aforementioned relay schemes have an additional drawback that they select the relays randomly by considering only a subset of nodes. For example, in the two-hop relaying, the nodes in contact with the source will always be selected as relays first [21]. However, some nodes that are currently out of contact with the source may in fact be able to achieve better relaying and routing performance.

The heterogeneous features of mobile nodes should be explicitly exploited in the relaying and routing schemes to improve the message delivery performance in DTNs. Heterogeneity in mobile nodes was considered in [22], which, however, only analyzes the capacity scaling properties of the mobile DTN with heterogeneous nodes and spatial inhomogeneities but does not consider the design of an optimal relaying scheme. Lee and Eunt [20] focused on how the heterogeneous contact dynamics impact the performance of relaying and routing algorithms by introducing a theoretic analysis framework. However, their work only concerns with the performance evaluation, and it does not consider either how to design the relaying schemes to efficiently utilize the heterogeneity of mobile nodes. Recent works [6], [23]–[26] considered only the heterogeneous contact rates between nodes in DTN routing but did not take into account other heterogeneous features existing in realistic DTNs that may influence the routing performance. Spyropoulos *et al.* [27] addressed the routing problem in the DTN comprising multiple classes of nodes, which are heterogeneous in diverse characteristics and mobility patterns. To our best knowledge, this is the only published work to date to study the DTN routing by considering not only the heterogeneous contact rates but other heterogeneous features as well. However, there are important differences between [27] and this paper, which will be emphasized after we have described our novel contributions.

Against this background, we consider the problem of optimal relaying for DTNs consisting of heterogeneous mobile nodes. Here, heterogeneity refers to the fact that contact dynamics in terms of contact time are different for node pairs, and relay costs are not the same for different nodes. We propose an optimal relaying scheme, which minimizes the delivery cost while satisfying the required message delivery probability. Our scheme explicitly considers both the heterogeneity in contact dynamics and the heterogeneity in relay costs, caused by different nodes' energy, communication, and mobility properties. Our novel contributions are summarized as follows.

1) We model the relaying problem by appropriately characterizing the nodes' heterogeneity in contact rates and delivery costs, and formulate the optimal relaying in the generic heterogeneous environment as the corresponding optimization problem.
2) We prove that the hardness of the formulated relay optimization problem is NP-hard and derive a pseudo-polynomial-time exact algorithm based on dynamic programming, which is capable of achieving the optimal relaying performance.
3) We evaluate our proposed optimal relaying scheme under the experimental data traces collected from realistic DTN environments, and compare our schemes with some existing schemes that do not consider the node heterogeneity and the existing proposals for heterogeneous DTNs. The simulation results obtained demonstrate the effectiveness of our optimal relaying scheme.

We now emphasize that [27] is rather restricted in comparison with our proposed optimal relaying scheme for the generic DTN. First, we consider the node heterogeneity in a generic environment in which every node is different from each other in terms of mobility pattern and forwarding cost, whereas [27] divides the nodes into some classes of different behaviors but each class containing homogeneous nodes. Therefore, our results are more general and realistic. In particular, the heterogeneous metric considered in [27] only includes the contact rates of different groups, whereas this paper considers not only the heterogeneous contact rate but also another important factor for routing, i.e., the heterogeneous message delivery cost. Second, in terms of routing, [27] focuses on the spray-and-focus scheme to select the nodes with higher utility to spray the messages. In this paper, by contrast, we derive the optimal relaying scheme by solving the associated optimization design problem. Our proposed relaying scheme therefore achieves the optimal system performance. More specifically, [27] uses the forwarding utility maximization principle to select the relay nodes, without considering how much the delivery cost will be, whereas this paper takes the cost into consideration to achieve the system required message delivery ratio. Finally, in our extensive simulation under the experimental data traces collected from realistic environments, we demonstrate the performance enhancement of our scheme over the schemes provided in [27].

Since mobility patterns are important to system design, many works [6], [29]–[31], [33]–[37] have studied the mobility modeling and analyzed how mobility patterns influence the performance of routing schemes. Different from previous works [6], [30], [31], [33]–[35], [37], which focus on mobility modeling and its influence on the achievable system performance, in this paper, we study how to efficiently exploit the heterogeneous features of mobile nodes, including the heterogeneity of intercontact times (ICTs) and the heterogeneity of relaying costs, in the relaying scheme design. Specifically, [34] studies the ICT distributions based on the real mobility traces and then evaluates the system performance to show that the heterogeneity of ICTs can be exploited to improve the routing

performance. However, [34] does not deal with the design of routing and relaying schemes; furthermore, it only considers the heterogeneity of ICTs. In the simulation study, we will compare our optimal relaying design with a relaying scheme that only takes into account the heterogeneity of ICTs.

The rest of this paper is organized as follows. We describe the system model and formulate the associated optimization problem in Section II, whereas in Section III, we derive an algorithm for solving the optimal relaying scheme. In Section IV, we introduce the experimental environment for performance evaluation and provide the simulation results. Finally, we conclude the paper in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Here, we first describe our system model and then derive the associated optimization problem for obtaining the optimal relaying scheme.

A. System Model and Definitions

We consider a DTN containing a set of the $N + M + 2$ wireless mobile nodes, which are labeled as $i \in \{1, 2, \ldots, N + M + 2\}$, in which the source node is also denoted by S and the destination node by D, whereas the $N + M$ relay nodes are denoted by $R = \{R_1, R_2, \ldots, R_{N+M}\}$. Because the density of nodes is usually sparse in a DTN, we assume that the relay nodes can communicate with source S or destination D only when they move to within the transmission range of each other, which is referred to as a communication contact. We now define the pairwise ICT between nodes i and j for i, $j \in \{1, 2, \ldots, M + M + 2\}$ and $i \neq j$.

Definition 1—ICT: The ICT of nodes i and j, which is denoted as $\text{ICT}_{i,j}$, is defined as the time that takes them to come within the range of each other again from the last time, denoted as t_0, when they were moving out of the range of each other, i.e.,

$$\text{ICT}_{i,j} = \min_t \left\{ (t - t_0) : \|L_i(t) - L_j(t)\| \le R_{\text{Tx}}^{i,j}, t > t_0 \right\}$$

where $L_i(t)$ and $L_j(t)$ are the positions of nodes i and j at time t, respectively; $R_{\text{Tx}}^{i,j}$ denotes the transmission range between nodes i and j; and $\| \bullet \|$ denotes the distance measure. Note that $\text{ICT}_{i,j}$ is a time duration, which is a random variable and is independent of time t.

In Definition 1, we introduce the pairwise ICTs of every node pairs, which will be used to formulate the optimal relay selection problem. Note that it is reasonable to model an individual pairwise ICT by an exponential distribution [6], [28], [29], [32]. The recent works focusing on the well-known mobility models, such as random waypoint and random direction [29], and the real-life mobility traces [6] have confirmed that the individual ICT follows an exponential distribution. Although some studies [30], [31], [33] suggest a truncated power law for ICTs, these works actually reveal that the aggregated ICT follows a truncated power law distribution. However, the individual ICT with constraints can be better modeled by the exponential distribution with heterogeneous coefficients [6], [34], [35]. Specifically, [6] models the individual contact rates as the Poisson process

and validate this model experimentally by using the χ^2 test on the real traces. In [34], it is revealed that the individual ICT with the constraints of ICT less than 1 week and at least 20 contacts can be modeled better by an exponential distribution with a heterogeneous coefficient. In [35] it is confirmed that, through controlling the shape of the individual intermeeting time distribution by choosing an appropriate domain size with respect to a given time scale, the distribution is exponential. Moreover, recent works [36]–[38] modeling the aggregated ICT of vehicles, which are typical nodes in vehicle DTNs, reveal the exponential distribution of ICT between nodes by analyzing a large amount of car/taxi mobility traces. We emphasize that the model of exponential individual ICT does not render the homogeneous contact time since different ICTs are described by the exponential parameters, i.e., the average contact times, which are different. We can now formally define the contact rate.

Definition 2—Contact Rate: The contact rate of nodes i and j, which is denoted as $\lambda_{i,j}$, is simply defined as $\lambda_{i,j} = (1/E[\text{ICT}_{i,j}])$, where $E[\bullet]$ denotes the expectation operator.

Thus, the contact rate of nodes i and j is Poisson distributed with the mean arrival rate $\lambda_{i,j}$. The nodes' heterogeneous contact dynamics in our system are mainly characterized by the different contact rates $\lambda_{i,j}$, where $i, j \in \{1, 2, \ldots, M + N + 2\}$, and $i \neq j$. Next, we have the following definition.

Definition 3—Message Delivery Ratio: Message delivery ratio P is defined as the average ratio that messages are delivered to destination D within message lifetime T. For a given message, its delivery ratio is equivalent to the average probability that destination D can receive it.

To deliver a message from source S to destination D, the source selects some relays from R. However, we have proven in this paper that how to select appropriate relays to optimize the system performance is an NP-complete problem. Consequently, we need to design an efficient algorithm, which is run each time when source S has a message to send and is able to obtain a pseudo-polynomial-time solution. To formulate this problem, we first categorize the relay nodes by following the terminology used in [6]. Specifically, when the source selects relays to transmit a message, some relaying nodes are in the communication range of the source node, which are referred to as *relays in contact*, whereas the other relays are out of the range, which are referred to as *relays not in contact*. Without loss of generality, we assume that there are M relay nodes in the communication range of the source S, which are grouped in the *relays in contact* subset denoted by $R^I = \{R_1^I, R_2^I, \ldots, R_M^I\}$, whereas the other N relay nodes are not in the range, which are grouped in the *relays not in contact* subset denoted by $R^O = \{R_1^O, R_2^O, \ldots, R_N^O\}$, as shown in Fig. 1. Obviously, $R^O \bigcup R^I = R$, and $R^O \bigcap R^I = \emptyset$. Selecting relay nodes only from R^I is then referred to as the in-contact relay (ICR), whereas selecting relay nodes only from R^O is known as the out-of-contact relay (OCR). In the ICR, the relays selected by source S are in contact with S. After the relays are selected, source S transmits the message to the relays, and then, relays transmit the message to the destination D when they next contact D. In the OCR, the relays selected are not in contact with source S. Source S has to wait for the relays to come in contact to transmit the message to them. Finally, the message
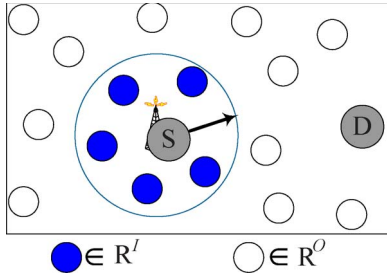
Fig. 1. Generic relaying scheme for message transmission, where subset R^I denotes the relays in contact with the source, whereas subset R^O denotes the relays not in contact with the source.

is successfully transmitted if any one of the relays with the message contacts destination D.

As emphasized previously, apart from the heterogeneous contact rates, there exist other heterogeneous features in realistic DTNs. To take into account these other heterogeneous features, we impose a cost factor associated with each relay node. In the realistic networks, the individual relay cost depends on the node's type, in terms of its energy consumption, communication capability, and mobility features. For example, the node with a larger energy budget will have a lower cost to act as a relay since its energy is relatively sufficient. The relay cost factors depends on the specified system, and how to decide them is beyond the scope of this paper, which is focusing on designing the optimal relay selection scheme. Without loss of generality, however, let C_i^I be the cost associated with relay node $R_i^I \in R^I$ and C_j^O be the cost associated with relay node $R_j^O \in R^O$.

Definition 4—Optimal Relay Selection: The problem of optimal relay selection is defined as follows. Choose relay nodes from R to achieve the specified message delivery probability P by delivering the message from source S to destination D within time T while imposing the minimum total cost.

B. Problem Formulation

We consider the generic relaying scheme in which the source selects relays from both R^I and R^O. The ICR and OCR then can be considered as the special cases of this generic relay scheme.

1) ICR Case: We begin by considering the ICR case. Note that destination D is not in contact with 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 random variable φ_i as the indicator that $R_i^I \in R^I$ is contacted by destination D within 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, φ_i follows the Bernoulli distribution, and its expectation, which is denoted by $\bar{\varphi}_i$, is equal to $1 - e^{\lambda_{i,d}^I T}$, where $\lambda_{i,d}^I$ is the contact rate between node R_i^I and destination D. Define $p_i = 1 - \bar{\varphi}_i$, which is the probability that R_i^I cannot contact destination D within 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 $C_i^I x_i$. Consequently, under $\{x_i\}_{i=1}^M$, the in-contact nodes can achieve the delivery probability of $(1 - \prod_{i=1}^M p_i^{x_i})$ at the delivery cost of $\sum_{i=1}^M C_i^I x_i$.

2) OCR Case: Now, we turn to consider the OCR case. Since none of the relay nodes $R_j^O \in R^O$ are 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 regard 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 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 and source S needed to contact R_j^O, which is denoted as T_1, and the time R_j^O needed to contact destination D, which is denoted by T_2. The distribution of Y_j is given in the following lemma.

Lemma 1: The probability density function (pdf) of Y_j, denoted by $f_j(y)$, is given by

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

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

Proof: The time for source S to deliver a message to the destination D via relay R_j^O includes 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 $f_{T_2}(y)$, i.e.,

$$f_{Y_j}(y) = f_{T_1}(y) \otimes f_{T_2}(y)$$

where $\otimes$ denotes the convolution operator. Let the generating functions of Y_j, T_1, and T_2 be

$$F_{Y_j}(w) = \int_0^\infty e^{-\varsigma w y} f_{Y_j}(y)\, dy$$

$$F_{T_1}(w) = \int_0^\infty e^{-\varsigma w y} f_{T_1}(y)\, dy$$

$$F_{T_2}(w) = \int_0^\infty e^{-\varsigma w y} f_{T_2}(y)\, dy$$

respectively, where $\varsigma = \sqrt{-1}$ is the imaginary unit. We have

$$F_{Y_j}(w) = F_{T_1}(w) F_{T_2}(w).$$

Noting $F_{T_1}(w) = (\lambda_{j,s}^O / \lambda_{j,s}^O + \varsigma w)$ and $F_{T_2}(w) = (\lambda_{j,d}^O / \lambda_{j,d}^O + \varsigma w)$ leads to

$$F_{Y_j}(w) = \frac{\lambda_{j,s}^O \lambda_{j,d}^O}{\left(\lambda_{j,s}^O + \varsigma w\right)\left(\lambda_{j,d}^O + \varsigma w\right)} = \sum_{i=1}^{2} \Lambda_i F_{T_i}(w) \quad (2)$$

where $\Lambda_1 = (\lambda_{j,d}^O / \lambda_{j,d}^O - \lambda_{j,s}^O)$, and $\Lambda_2 = -(\lambda_{j,s}^O / \lambda_{j,d}^O - \lambda_{j,s}^O)$, assuming that $\lambda_{j,d}^O \neq \lambda_{j,s}^O$. Applying the reverse transformation of generating function to (2) yields

$$f_{Y_j}(y) = \int_0^\infty e^{\varsigma w y} F_{Y_j}(w)\, dw = \sum_{i=1}^{2} \Lambda_i \int_0^\infty e^{\varsigma w y} F_{T_i}(w)\, dw$$

$$= \frac{\lambda_{j,d}^O \lambda_{j,s}^O \left(e^{-\lambda_{j,s}^O y} - e^{-\lambda_{j,d}^O y}\right)}{\lambda_{j,d}^O - \lambda_{j,s}^O}.$$

This proves Lemma 1. ∎

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

$$q_j = \int_T^\infty f_{Y_j}(y)\, dy = \frac{\lambda_{j,d}^O e^{-\lambda_{j,s}^O T} - \lambda_{j,s}^O e^{-\lambda_{j,d}^O T}}{\lambda_{j,d}^O - \lambda_{j,s}^O}. \quad (3)$$

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. Under policy y_j, the message delivery probability and cost associated with R_j^O are obviously $1 - q_j^{y_i}$ and $C_j^O y_j$, respectively. Therefore, the out-of-contact nodes can achieve the delivery probability of $(1 - \prod_{j=1}^{N} q_j^{y_j})$ at the total delivery cost of $\sum_{j=1}^{N} C_j^O y_j$.

3) Formulation of Optimal Relay Selection Problem: Recall that our goal is to select appropriate relays from $R = R^I \bigcup R^O$ to deliver the message within time T to destination D with probability P at the minimum cost, as is defined in Definition 4. Combining the previous discussions for both the ICR and OCR cases, this optimal relay selection can be formulated as the following optimization problem:

$$\min \quad \sum_{i=1}^{M} C_i^I x_i + \sum_{j=1}^{N} C_j^O y_j$$

$$\text{s.t.} \quad \begin{cases} \prod_{i=1}^{M} p_i^{x_i} \prod_{j=1}^{N} q_j^{y_j} \leq 1 - P \\ x_i \in \{0,1\}, i = 1, 2, \ldots, M \\ y_j \in \{0,1\}, j = 1, 2, \ldots, N. \end{cases} \quad (4)$$

In the given optimal relay selection problem (4), the node heterogeneity manifests by the fact that each relay node has a different relay cost (C_i^I or C_j^O) and a different delivery ratio ($p_i^{x_i}$ or $q_j^{y_j}$), which is influenced by the node contact rate. As expected, the node heterogeneity enters the problem naturally and has an inherent influence on the relay selection solution. By defining

$$u_k = \begin{cases} C_k^I, & 1 \leq k \leq M \\ C_{k-M}^O, & M+1 \leq k \leq M+N \end{cases}$$

$$z_k = \begin{cases} x_k, & 1 \leq k \leq M \\ y_{k-M}, & M+1 \leq k \leq M+N \end{cases}$$

$$\varpi_k = \begin{cases} p_k, & 1 \leq k \leq M \\ q_{k-M}, & M+1 \leq k \leq M+N \end{cases} \quad (5)$$

the optimal relaying-node selection problem (4) can be rewritten as

$$\min \quad \sum_{k=1}^{N+M} u_k z_k$$

$$\text{s.t.} \quad \begin{cases} \prod_{k=1}^{N+M} \varpi_k^{z_k} \leq 1 - P \\ z_k \in \{0,1\} k = 1, 2, \ldots, N+M. \end{cases} \quad (6)$$

We refer to the optimal relaying scheme obtained by solving the optimization problem (6) as the optimal heterogeneous relaying scheme (OHRS).

Remark 1: Clearly, the OHRS will minimize the system's relaying cost while requiring that the message delivery ratio P is met, and the message is delivered to the destination before its message lifetime T. That is, the average message delay is guaranteed to be upper bounded by T. An alternative optimal relaying scheme is to minimize the average message delay while guaranteeing a given level of relaying cost and a given value of delivery ratio P. The most general optimal relaying scheme would be to simultaneously minimize the system's relaying cost and the average message delay at the given value of delivery ratio P. However, the optimization formulated becomes a challenging multiobjective optimization problem. For such a multiobjective optimization, the optimization algorithm must generate a set of Pareto-optimal solutions, which should be ideally well-distributed across the Pareto frontier [39]. Moreover, even when such a set of the Pareto-optimal solutions can be obtained, a human decision or a preference structure has to be imposed to choose which particular solution to use. This is usually achieved by first specifying what level of average message delay is acceptable and then choosing the corresponding system's relaying cost from the Pareto solutions or, alternatively, by first specifying what level of system's relaying cost is allowed and then choosing the corresponding average message delay from the Pareto solutions. In practice, therefore, the problem is often solved in the approach adopted in this paper.

III. OPTIMAL HETEROGENEOUS RELAYING SCHEME

Before deriving an algorithm to solve the optimization problem (6) for obtaining the OHRS, we first look into the nature of this optimization problem.

A. Hardness of the Problem

The hardness of the optimization problem (6) is expressed in the following theorem.

Theorem 1: The optimal relaying-node selection problem (6) is NP-hard.

Proof: We begin by considering the following optimization problem:

$$\max \quad \sum_{k=1}^{\bar{N}} u_k \hat{z}_k$$

$$\text{s.t.} \quad \begin{cases} \sum_{k=1}^{\bar{N}} w_k \hat{z}_k \leq W \\ \hat{z}_k \in \{0,1\}, k = 1, 2, \ldots, \bar{N}. \end{cases} \quad (7)$$

This is a 0–1 knapsack problem with real-valued weights w_k, which is an NP-hard problem [41]. Now, by defining

$$\hat{z}_k = 1 - z_k \quad (8)$$

$$\bar{N} = N + M \quad (9)$$

$$w_k = -\log(\varpi_k) \quad (10)$$

$$W = \log(1 - P) - \sum_{k=1}^{\bar{N}} \log(\varpi_k) \quad (11)$$

the 0–1 knapsack problem (7) is transformed exactly into the optimization problem (6). This proves that the optimal relaying-node selection problem (6) is NP-hard. ∎

Note that weights w_k, $1 \leq k \leq \bar{N}$, and W are all positive real-valued numbers. Theorem 1 shows that the optimal relaying-node selection problem is an NP-hard problem. Thus, we cannot obtain its optimal solution in general by polynomial-time algorithms, but can only obtain a suboptimal solution by polynomial-time algorithms. Alternatively, we may look for algorithms of higher complexity to achieve the optimal result. In terms of finding polynomial-time suboptimal solutions, the well-known fully polynomial-time approximation scheme (FPTAS) [40] is suitable for the transformed 0–1 knapsack problem of (7), which aims to obtain a $1 - \varepsilon$ approximation to the optimal solution in polynomial time with respect to both the size of the problem $\bar{N}$ and the approximation accuracy $1/\varepsilon$. The FPTAS first scales the values u_k down sufficiently so that the costs u_k of all the relaying nodes R_k are polynomially bounded in $\bar{N}$. It then uses a pseudo-polynomial-time algorithm on the transformed problem to obtain an approximate solution based on the fact that the upper bound of the achievable total profit is not more than $\bar{N}$ times the profit of the most profitable object among all the $\bar{N}$ objects. Inspired by this FPTAS, we note that, in our optimal heterogeneous relay selection problem, the delivery probabilities of the relay nodes are heterogeneous distributed since the contact dynamics in terms of contact time are very different for the node pairs in our problem. Consequently, we can quantize the real-valued weights of all the items without loss of ability to distinguish any two nodes with different delivery probability gains. This allows us to use dynamic programming to obtain the optimal solution. With this motivation, in the following, we design a pseudo-polynomial-time algorithm to obtain the optimal solution of the relaying-node selection problem.

B. Algorithm for the OHRS

We design an algorithm based on dynamic programming to choose the relay nodes to solve the optimal relaying-node

selection problem (6), namely, the 0–1 knapsack problem with real-valued weights (7). For notational convenience, we will drop the superscript indexes ^{I} and ^{O} from relay nodes and simply denote all the relay nodes by R_k for $1 \leq k \leq \bar{N}$. We refer to w_k defined in (10) as the delivery probability gain of relay node R_k and W defined in (11) as the total weight. Given the contact rates and the message delivery ratio P, w_k for $1 \leq k \leq \bar{N}$, and W can readily be computed. Since w_k's are continuous-valued real numbers, we need to quantize these delivery probability gains to run the dynamic optimization procedure. The required quantization precision is defined by

$$\Delta w = \min_{\forall i,j: i \neq j} |w_i - w_j| \quad (12)$$

and the corresponding quantization step is given by

$$M_Q = \left\lceil \frac{W}{\Delta w} \right\rceil \quad (13)$$

where $\lceil \bullet \rceil$ denotes the integer ceiling operator. In Algorithm 1, we divide the relay node selection process into $\bar{N}$ steps, and in the kth step, we decide whether R_k is selected as a relay. Furthermore, s_m denotes the state variable of the quantized delivery probability gain in the kth step, and $h_k(s_m)$ is the maximum weight that we can obtain by selecting the $k, k+1, \ldots,$ and $\bar{N}$ relay nodes when the quantized delivery probability gain is s_m.

Algorithm 1 Algorithm for source S to select relay nodes.

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 quantization step M_Q as given in (13);
4: Define $\bar{w}_k = \lceil (w_k M_Q / W) \rceil \cdot (W/M_Q)$, $1 \leq k \leq \bar{N}$, and quantize the set $[0, W]$ as $\mathbb{W} = \{0, 1 \cdot (W/M_Q), 2 \cdot (W/M_Q), \ldots, M_Q \cdot (W/M_Q)\}$;
5: Initialize $h_k(0) = 0$, $1 \leq k \leq \bar{N}$, $h_1(s) = 0$, $s \in \mathbb{W}$;
6: **for** Each relay node R_k, $1 \leq k \leq \bar{N}$, **do**
7: **for** $m \in \{0, 1, 2, \ldots, M_Q\}$, set $s_m = m \cdot (W/M_Q) \in \mathbb{W}$ **do**
8: **if** $s_m < \bar{w}_k$ **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; $h_k(s_m) = h_{k-1}(s_m - \bar{w}_k) + u_k$;
15: **end if**
16: **end if**
17: **end for**
18: **end for**

We now show that Algorithm 1 achieves the optimal solution of the optimal relaying-node selection problem (6).

Theorem 2: By using Algorithm 1, we can obtain the optimal solution for the optimization problem (7).

Proof: We note that, in Algorithm 1, the quantization precision Δw is used to quantize the delivery probability gain w_k. Since $\Delta w = \min_{\forall i, j : i \neq j} |w_i - w_j|$ and the "quantized" weights are

$$\bar{w}_k = \left\lceil \frac{w_k M_Q}{W} \right\rceil \cdot \frac{W}{M_Q} = I_k \cdot \frac{W}{M_Q}, 1 \leq k \leq \bar{N} \tag{14}$$

we have

$$\bar{w}_i \neq \bar{w}_j \quad \forall i \quad \forall j : i \neq j \text{ and } w_i \neq w_j.$$

That is, we have quantized w_k to $\bar{w}_k$ without the loss of ability to distinguish any two nodes with different delivery probability gains. Formally, this quantization process yields the following 0–1 knapsack problem with quantized real-valued weights $\bar{w}_k$, i.e.,

$$\max \quad \sum_{k=1}^{\bar{N}} u_k \hat{z}_k$$
$$\text{s.t.} \quad \begin{cases} \sum_{k=1}^{\bar{N}} \bar{w}_k \hat{z}_k \leq W \\ \hat{z}_k \in \{0, 1\}, k = 1, 2, \ldots, \bar{N}. \end{cases} \tag{15}$$

This 0–1 knapsack problem is exactly equivalent to the following 0–1 knapsack problem with integer weights I_k:

$$\max \quad \sum_{k=1}^{\bar{N}} u_k \hat{z}_k$$
$$\text{s.t.} \quad \begin{cases} \sum_{k=1}^{\bar{N}} I_k \hat{z}_k \leq M_Q \\ \hat{z}_k \in \{0, 1\}, k = 1, 2, \ldots, \bar{N}. \end{cases} \tag{16}$$

Since the dynamic programming within Algorithm 1 can find the optimal solution for the 0–1 knapsack problem (16), as proven in [41], Algorithm 1 achieves the optimal solution of the optimization problem (15).

Therefore, we only need to show that the optimal solution of the 0–1 knapsack problem (15) with quantized real-valued weights is the optimal solution of the 0–1 knapsack problem (7) with real-valued weights, In other words, let $\{\hat{z}_k^*\}_{k=1}^{\bar{N}}$ be the optimal solution of the 0–1 knapsack problem (15), i.e., $\{\hat{z}_k^*\}_{k=1}^{\bar{N}}$ maximize the utility function $\sum_{k=1}^{\bar{N}} u_k \hat{z}_k$ while satisfying the following constraint:

$$\sum_{k=1}^{\bar{N}} \bar{w}_k \hat{z}_k^* \leq W. \tag{17}$$

We need to show that $\{\hat{z}_k^*\}_{k=1}^{\bar{N}}$ maximize the utility function $\sum_{k=1}^{\bar{N}} u_k \hat{z}_k$ while satisfying the following constraint:

$$\sum_{k=1}^{\bar{N}} w_k \hat{z}_k^* \leq W. \tag{18}$$

Note that (17) automatically implies (18) because $\bar{w}_k \geq w_k$, $1 \leq k \leq \bar{N}$.

In fact, assume that $\{\hat{z}_k^*\}_{k=1}^{\bar{N}}$ is not the optimal solution of the 0–1 knapsack problem (7). Instead, the optimal solution of

the 0–1 knapsack problem (7) is $\{\hat{z}_k^o\}_{k=1}^N$. Consider the generic case, where $\hat{z}_i^* = 1$ and $\hat{z}_j^* = 0$; $\hat{z}_i^o = 0$ and $\hat{z}_j^o = 1$; and $\hat{z}_k^o = \hat{z}_k^*$, $k \neq i$ and $k \neq j$. Since $\{\hat{z}_k^*\}_{k=1}^{\bar{N}}$ is the optimal solution of the 0–1 knapsack problem (15), we must have either

$$\sum_{k=1}^{\bar{N}} u_k \hat{z}_k^o = \sum_{k=1}^{\bar{N}} u_k \hat{z}_k^* + u_j - u_i < \sum_{k=1}^{\bar{N}} u_k \hat{z}_k^* \tag{19}$$

or

$$\sum_{k=1}^{\bar{N}} \bar{w}_k \hat{z}_k^o > W. \tag{20}$$

The former is

$$u_i > u_j \tag{21}$$

whereas the latter is

$$\sum_{k=1}^{\bar{N}} \bar{w}_k \hat{z}_k^* + \bar{w}_j - \bar{w}_i > W. \tag{22}$$

On the other hand, since $\{\hat{z}_k^o\}_{k=1}^N$ is the optimal solution of the 0–1 knapsack problem (7), we have

$$\sum_{k=1}^{\bar{N}} u_k \hat{z}_k^o = \sum_{k=1}^{\bar{N}} u_k \hat{z}_k^* + u_j - u_i > \sum_{k=1}^{\bar{N}} u_k \hat{z}_k^*. \tag{23}$$

This excludes (21). Furthermore, we must have

$$\sum_{k=1}^{\bar{N}} w_k \hat{z}_k^o = \sum_{k=1}^{\bar{N}} w_k \hat{z}_k^* + w_j - w_i \leq W. \tag{24}$$

Noting (18), (24) implies that $w_j - w_i \leq 0$, which leads to

$$\bar{w}_j - \bar{w}_i \leq 0. \tag{25}$$

Considering (17) and (25), we must have

$$\sum_{k=1}^{\bar{N}} \bar{w}_k \hat{z}_k^* + \bar{w}_j - \bar{w}_i \leq W \tag{26}$$

which contradicts (22). ∎

Remark 2: The complexity and memory requirements of this algorithm are on the order of $\bar{N} M_Q$, which is denoted as $\mathcal{O}(\bar{N} M_Q)$. Obviously, Algorithm 1 is most effective when the minimum difference Δw between the delivery probability gains w_k, $1 \leq k \leq \bar{N}$, is large. In other words, if the network is very heterogeneous, as manifested by a small M_Q, our algorithm finds the optimal solution with linear complexity. This is much faster than the FPTAS [40], which can only find a suboptimal solution with polynomial complexity. As the network becomes less heterogeneous or more homogeneous, M_Q increases, and our algorithm finds the optimal solution with an increasing complexity. At the certain M_Q value, the complexity will become more than polynomial time, and our optimal scheme becomes less efficient than the polynomial-time suboptimal FPTAS. For the extreme case of $\Delta w = 0$, our optimal algorithm cannot be used.

The "quantization precision" of our algorithm M_Q is a "physical" quantity specified by the underlying network that tells the designer what to use for the quantization step. Note that W depends on the required delivery ratio P and all the w_i, whereas the w_is are specified by the contact dynamics of the network. If a network is difficult, as manifested by the required level of P, the values of w_i and the value of Δw, such that M_Q is too large, then the designer may have to compromise, e.g., by reducing the desired level of delivery ratio P to reduce M_Q. Alternatively, the suboptimal design may be adopted by using for example the polynomial-time FPTAS. It is worth emphasizing that, since we consider the realistic heterogeneous network, where the heterogeneity is reflected by the fact that different nodes have different relay costs and contact times are very different for different node pairs, there exist a wide range of network conditions to enable the efficient application of our optimal relaying-node selection algorithm.

Remark 3: Using Algorithm 1 to select relay nodes, the source node needs to know the contact rates and delivery costs of the nodes in the network. The delivery costs and contact rates between nodes are often quasi-statistic information. In the infrastructure-based wireless network, such as wireless local area network or cellular network, all the nodes of the network can transmit their contact rates and associated delivery costs to the source node, as they can access the service provided by the network. In other networks, where such an infrastructure-based service is not provided, the nodes can still transmit the contact rates and costs by the so-called in-band metachannel provided by the opportunistic contacts, whose practicality and efficiency have been demonstrated in [25] and [42]. For example, in the mobile social network, nodes can use the mobile-to-mobile DTN main channel to transmit large-size contents at a relatively high data rate, while using the in-band channel of limited rate to transmit the required system information. Therefore, although the required information for the source to choose the relay nodes are distributed in the network, this information can be obtained in a centralized or distributed way by the source node. In this paper, we assume that the nodes' contact rates and delivery costs collected by the source are accurate, and they do not change during the decision process of selecting relay nodes. However, it is worth emphasizing that the collected information may be inaccurate, particularly if they are obtained distributively with the in-band metachannel of very limited rate. It is also worth emphasizing that if the nodes' contact rates and delivery costs change during the actual message delivery process, the relay decision made will no longer be optimal. Studying the robustness of the OHRS under inaccurate system information will be our future investigation.

IV. Performance Evaluations

We performed an extensive simulation study to evaluate our proposed OHRS by using some existing relay schemes for DTNs as the benchmarks.

A. Evaluation Environment and Experimental Settings

We used the four human mobility traces, i.e., *Infocom05, Infocom06, Reality,* and *Cambridge*, and one vehicular mobility trace, i.e., *Shanghai*. Among them, *Reality* was collected from MIT Reality Mining Project [43], and the other three human traces were gathered by the Haggle Project [26]. These traces recorded contacts among users carrying Bluetooth devices, which periodically discover their contacts with peers in the communication range and record them. The vehicular trace of *Shanghai* [37] records the positions of vehicles carrying GPS devices. Specifically, about 2100 operational taxis were running for the whole month of February 2007 in Shanghai city to collect the trace. The five data traces are summarized in Table I, where we observe that the chosen traces covered diverse DTN environments, from concentrated conference sites (*Infocom05* and *Infocom06*) to dispersed university campus (*Reality* and *Cambridge*), and vehicular scenario covering a large city (*Shanghai*), with experiment period from a few days (*Infocom05, Infocom06* and *Cambridge*), to a month (*Shanghai*), and several months (*Reality*).

We evaluated the performance of our proposed optimal relaying scheme, which were obtained using Algorithm 1 presented in Section III-B. The simulation was carried out in Matlab, which processes the experimental trace data to obtain the information required by the algorithm. As aforementioned, most of the existing works only considered the case of homogeneous mobile nodes, whereas some studies did evaluate how the heterogeneous contact rates influence the performance of relaying algorithms. Moreover, these existing schemes are not relaying-node selection schemes. For a meaningful comparison, therefore, we specifically designed two relaying-node selection schemes as two benchmarks to reflect this reality. The first relaying benchmark scheme designed does not consider nodes' heterogeneous features at all and, therefore, is denoted as the nonheterogeneous relay scheme (NHRS). The second relaying benchmark scheme that we designed does consider heterogeneous contact rates but assume the nodes' other features that influence the delivery cost being homogeneous, and we refer to this scheme as the heterogeneous contact relay scheme (HCRS). From the discussion in Section III-A, we note that the optimal relay selection problem is equivalent to an NP-hard 0–1 knapsack problem, and the well-known FPTAS algorithm [40] is a pseudo-polynomial-time algorithm that finds an $1 - \varepsilon$ approximation to this 0–1 knapsack problem. We therefore used the FPTAS algorithm with $\varepsilon = 0.3$ as the third benchmark algorithm. In addition, from the introduction, we note that [27] is an up-to-date study on heterogeneous-node-populated DTNs. Since [27] only studied several spraying schemes, we needed to change these spraying schemes to the corresponding relaying schemes for a meaningful comparison. The two benchmark relaying schemes adopted from [27] were the most-mobile-first relaying (MMFR) and most-social-first relaying (MSFR). The MMFR selects the most "mobile" nodes as the relays first owing to the consideration of nodes' heterogeneity in terms of different moving areas and speeds. In [27], it was assumed that each node carries a label that states the level of its "mobility." In our simulation, we used the average ICT 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.

TABLE I
TRACE SUMMARY

Trace	Infocom05	Infocom06	Reality	Cambridge	Shanghai
Mobility			Human		Vehicular
Device	iMote	iMote	Cell	iMote	GPS
Number of devices	41	98	97	54	2100
Duration (days)	3	3	246	11	30
Granularity (secs)	120	120	300	600	800
Number of contacts	22,459	191,336	54,667	10,873	1,317,060
Contact frequency (per day)	4.6	6.7	0.024	0.345	0.373
Settings for T (secs)	5,600	1,900	60,000	36,800	30,000

The delivery costs of nodes are influenced by many heterogeneous factors, including nodes' energy variance, communication capacity, mobility patterns, and spatial positions. Consequently, the delivery costs may follow some random distribution among the nodes in reality. On the other hand, the traces did not have the relevant information in this aspect; therefore, we examined several cost distributions among relay nodes, including the equal constant cost, the discrete V class cost distribution, and the uniform and normal cost distributions. The expectation of delivery cost distribution of C_i^I (C_j^O) was set to 10 [see (12)]. In the fixed equal-cost setting, the delivery costs of all the nodes were set to the same value of 10 [see (12)], which means that the nodes' delivery costs are homogeneous. Although the homogeneous cost case is unrealistic in practice, many of the existing approaches, including the two benchmarks NHRS and HCRS, assume this property. The physical interpretation of the discrete V cost classes is that the nodes or devices in the network may be grouped into the V types, in terms of their energy budgets for relaying. In the simulation, we set $V = 4$ with the cost factors of 1 [see (3)], 7 [see (9)], 13 [see (15)], and 19 [see (21)], respectively. Furthermore, each group contained approximately one quarter of the nodes. The limiting case of very large V may be viewed to have the uniform distribution. In the normal cost distribution, the nodes with extremely large cost and extremely small cost are rare, and the costs of the most network nodes are within the two standard deviations of the mean delivery cost. Therefore, the Gaussian distribution may be considered to be a good representation or approximation of the cost distribution in practical networks. In the uniformly distributed cost setting, the delivery costs of nodes were uniformly distributed between 1 [see (1)] and 19 [see (23)], whereas in the Gaussian distributed cost setting, the delivery costs of nodes followed the normal distribution with the expectation of 10 [see (12)] and a variance of 10. Because the values of normal distribution are from negative infinity to positive infinity, we only used the values in the 10%–90% range of its pdf.

In the simulation, any nodes could be selected as the source or destination to send or receive messages. From the consideration of achieving high message delivery probability, we selected the two most popular nodes, which had the largest numbers of contacts with other nodes, as source S and destination D. Since the traces did not have related information about the source coverage, we divided the one third of the remaining nodes into R^I and the other two thirds of the nodes into R^O randomly. In the simulation investigation, all the contact rates and delivery costs were available to the source to make relay selection decision, and they were assumed to be accurate. Furthermore,

during the subsequent message relaying process, the nodes' contact rates and delivery costs remained unchanged. We measured the relay costs of all the selected relays as the system achieved total cost for each algorithm tested. In the simulation, we changed the required system delivery probability P, ran different relay schemes, and compared their total delivery costs. The simulation results were obtained by simulating the delivery of 500 messages, and for each message, the relay nodes' costs were randomly generated according to the chosen distribution. For the convenience of comparison, we allowed the delivery probability of the relaying schemes to approach 100% and set the different T values for the different traces, as shown in Table I. We used the traces of *Infocom05* and *Cambridge* to evaluate our OHRS in comparison with the NHRS, HCRS, and FPTAS benchmarks. We compared our OHRS with the other two benchmarks, i.e., the MMFR and MSFR, using the traces of *Infocom06, Reality,* and *Shanghai.*

To quantitatively analyze the performance enhancement of our proposed OHRS over the other benchmark schemes, we calculated the average cost reductions achieved by the OHRS over the other schemes, according to the average cost reduction (ACR) defined by

$$\text{ACR} = \frac{1}{99} \sum_{i=1}^{99} \frac{C^{(\text{other})}(P_i) - C^{(\text{OHRS})}(P_i)}{C^{(\text{other})}(P_i)} \qquad (27)$$

where $C^{(\text{other})}(P_i)$ denoted the achieved cost of the compared benchmark scheme while $C^{(\text{OHRS})}(P_i)$ was the achieved cost of our OHRS when the message delivery probability was set to P_i, and $P_i = 0.01 \cdot i$ for $1 \leq i \leq 99$.

B. Comparison With NHRS, HCRS, and FPTAS

The results for the trace of *Infocom05* are shown in Figs. 2–5. As expected, when the required message delivery probability increased, all the relay schemes spent more to achieve the desired delivery probability. Specifically, Fig. 2 depicts the case of equal node cost. In this case, the nodes' heterogeneity is only featured by the different contact rates of nodes, and both the OHRS and HCRS consider the heterogeneous contact rates, whereas the NHRS does not take into account this heterogeneous feature in choosing relay nodes. Therefore, we can predict that the performance of our optimal OHRS and the HCRS benchmark will be the same in this case. As confirmed in Fig. 2, our OHRS and the HCRS benchmark achieved almost the identical delivery cost, and they outperformed the NHRS by the 56.2% reduction in cost on average. In the case of equal node cost, it can be observed that the FPTAS achieved almost
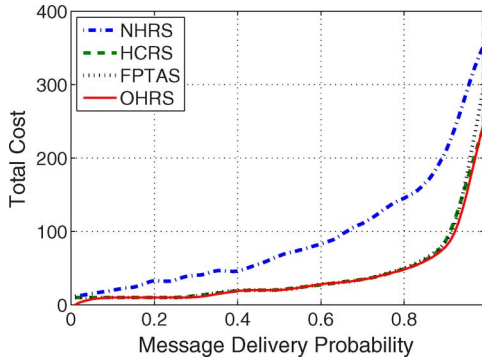
Fig. 2. Simulation results of different relaying schemes for the trace of *Infocom05* with equal constant node cost.
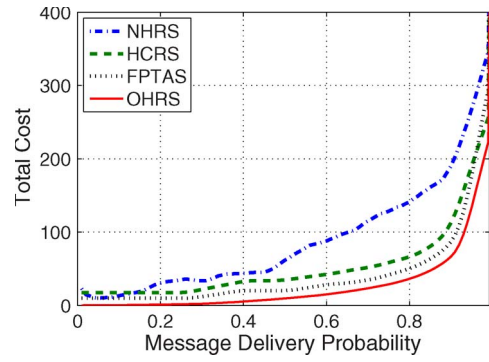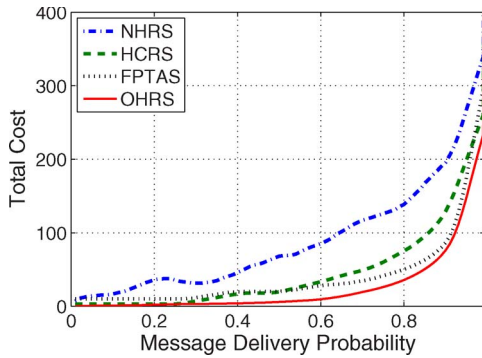

Fig. 3. Simulation results of different relaying schemes for the trace of *Infocom05* with discrete $V = 4$ class-distributed node cost.
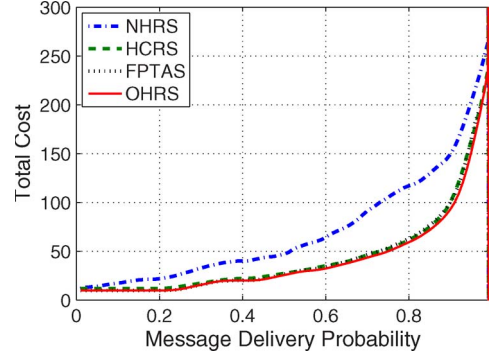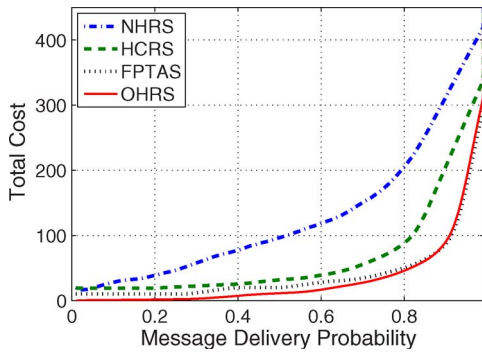

Fig. 4. Simulation results of different relaying schemes for the trace of *Infocom05* with uniformly distributed node cost.
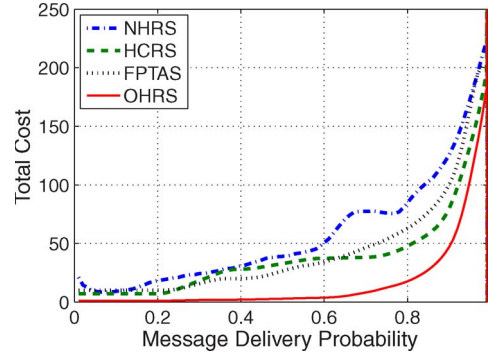

Fig. 5. Simulation results of different relaying schemes for the trace of *Infocom05* with normally distributed node cost.

Fig. 6. Simulation results of different relaying schemes for the trace of *Cambridge* with equal constant node cost.

Fig. 7. Simulation results of different relaying schemes for the trace of *Cambridge* with discrete $V = 4$ class-distributed node cost.

the same performance as our OHRS. The results for the discrete $V = 4$ class distribution of cost are shown in Fig. 3, which demonstrate that our OHRS on average outperformed the FP-TAS, HCRS, and NHRS by the 27.4%, 55.6%, and 70.8% cost reductions, respectively. Fig. 4 shows the total costs achieved by the four schemes under the uniformly distributed node cost, where we observe that our optimal OHRS reduced the total cost on average by 11.2%, 67.5%, and 83.2%, respectively, over the FPTAS, HCRS, and NHRS benchmarks. The results for the normal distribution of node cost, shown in Fig. 5, also demonstrate that our OHRS on average outperformed the FPTAS, HCRS, and NHRS by the 26.6%, 48.2% and 87.5% cost reductions, respectively. As expected, the solution found by the FPTAS algorithm is generally inferior to that obtained by our OHRS simply because the FPTAS is suboptimal, whereas

our OHRS attains the optimal solution. Most significantly, our OHRS imposes considerably lower computational complexity than the FPTAS. This will be demonstrated later together with the simulation results for the *Cambridge* trace.

Figs. 6–9 depict the results obtained by the four relaying schemes for the trace of *Cambridge* under the four different node cost distributions. In terms of delivery cost, our OHRS achieved better performance than the three relaying benchmarks, i.e., the NHRS, HCRS, and FPTAS. More specifically, under the equal node cost setting, our OHRS outperformed the NHRS significantly while achieving the same performance as the HCRS and FPTAS. The average cost reductions of 49.4%, 73.3% and 80.8% were achieved by our OHRS over the HCRS, FPTAS, and NHRS benchmarks, respectively, under the discrete $V = 4$ class distribution of node cost. With the
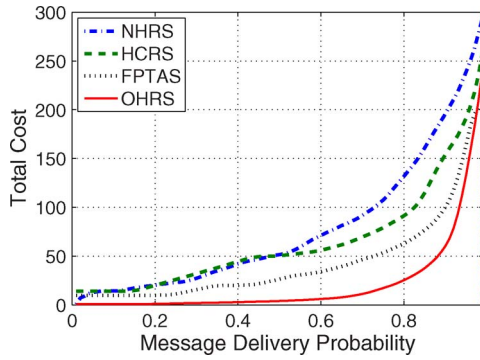
Fig. 8. Simulation results of different relaying schemes for the trace of *Cambridge* with uniformly distributed node cost.
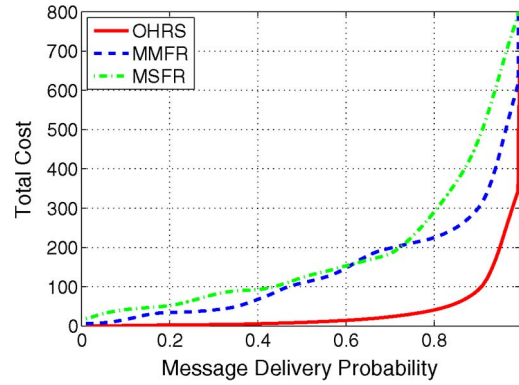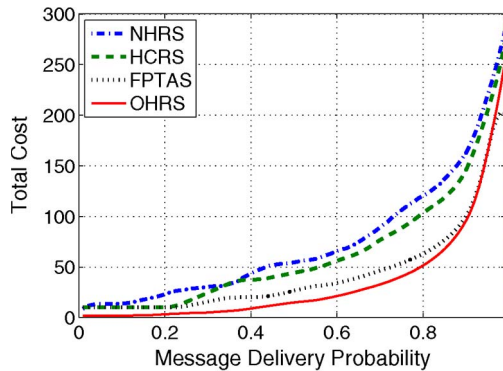

Fig. 9. Simulation results of different relaying schemes for the trace of *Cambridge* with normally distributed node cost.

TABLE II
RUN TIMES OF THE FPTAS AND OHRS ALGORITHMS IN THE
SIMULATION OF THE CAMBRIDGE AND INFOCOM05 TRACES

Cost Distribution	Equal Cost	$V = 4$ Classes	Uniform	Normal
FPTAS with *Infocom05*	44.82s	45.00s	45.01s	44.97s
OHRS with *Infocom05*	13.41s	13.43s	13.45s	13.44s
FPTAS with *Cambridge*	13.88s	13.95s	13.98s	13.93s
OHRS with *Cambridge*	9.61s	9.82s	9.54s	9.54s

uniform distribution of node cost, the OHRS reduced the total costs on average by 29.7%, 78.7%, and 80.6%, respectively, in comparison with the FPTAS, HCRS, and NHRS. The average cost reductions of 12.4%, 76.6%, and 81.8% were achieved by our OHRS over the FPTAS, HCRS, and NHRS benchmarks, respectively, under the normal distribution of node cost.

Combining the results for the traces of *Cambridge* and *Infocom05*, we can observe that our optimal OHRS outperformed the three benchmark schemes, i.e., the NHRS, HCRS, and FPTAS, considerably. Furthermore, our optimal OHRS attains the solution with significantly lower computational complexity than the FPTAS. To demonstrate the efficiency of our proposed optimal OHRS, in comparison with the well-known pseudo-polynomial-time suboptimal FPTAS, we collected the run times of the FPTAS and OHRS algorithms in the simulation of the *Cambridge* and *Infocom05* traces. The run-time results are shown in Table II, where it can be seen that our OHRS reduced the running time by about 70% and 31%, respectively, for the traces of *Infocom05* and *Cambridge*, compared with the run time of the FPTAS.

Fig. 10. Performance comparison between our OHRS and the two benchmark schemes adopted from [27] for the trace of *Infocom06* with discrete $V = 4$ class-distributed node cost.
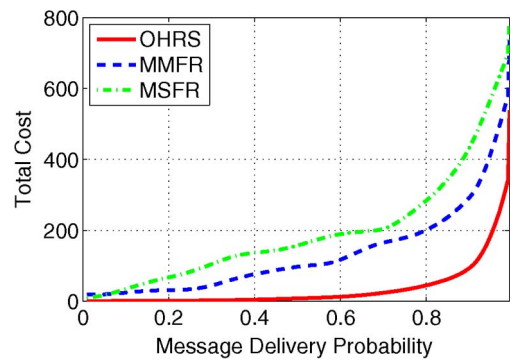

Fig. 11. Performance comparison between our OHRS and the two benchmark schemes adopted from [27] for the trace of *Infocom06* with uniformly distributed node cost.

C. Comparison With MMFR and MSFR

We then evaluated our OHRS and the two benchmarks, i.e., the MMFR and MSFR, using the traces of *Infocom06, Reality*, and *Shanghai*. The MMFR and MSFR, which are adopted/modified from [27] and briefly described in Section IV-A, represented the existing schemes for heterogeneous DTNs. To compare these schemes in the truly heterogeneous environment, we used the discrete $V = 4$ class, uniform and normal distributions of node cost for the *Infocom06* and *Reality* traces, whereas we use the discrete $V = 4$ class and normal distributions of node cost for the *Shanghai* trace.

Figs. 10–12 compare the total costs achieved by the three schemes for the *Infocom06* trace under the discrete $V = 4$ class, uniform, and normal distributions of node cost, respectively. The results of Figs. 10–12 clearly show that our proposed OHRS achieved superior performance over the MMFR and MSFR benchmarks. For example, with the discrete $V = 4$ class distribution of node cost, the OHRS reduced the total costs on average by 65.9% and 70.1%, respectively, in comparison with the MMFR and MSFR. For the cases of uniform and normal cost distributions, the OHRS also attained the least cost given the same delivery probability, compared with the MMFR and MSFR. This is because, although the MMFR and MSFR do consider the heterogeneous features in terms of nodes' mobility and sociability, respectively, they do not explicitly take into account the heterogeneous node costs in selecting relay nodes.
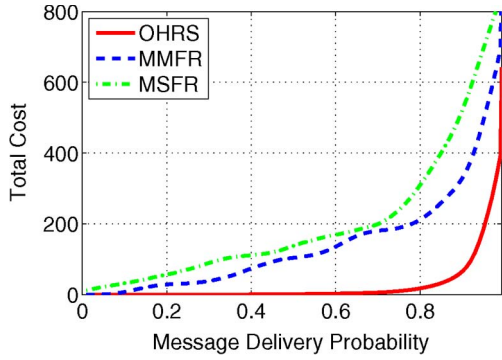
Fig. 12. Performance comparison between our OHRS and the two benchmark schemes adopted from [27] for the trace of *Infocom06* with normally distributed node cost.
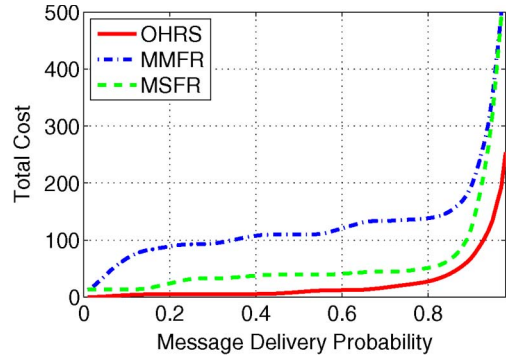

Fig. 15. Performance comparison between our OHRS and the two benchmark schemes adopted from [27] for the trace of *Reality* with normally distributed node cost.
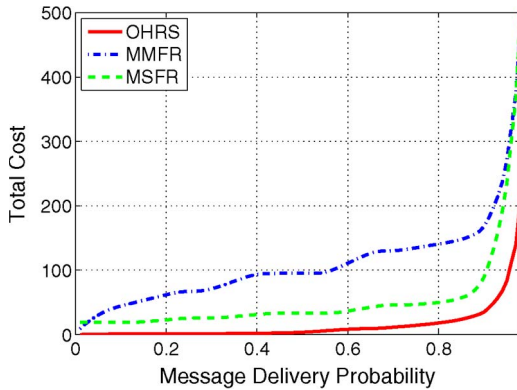

Fig. 13. Performance comparison between our OHRS and the two benchmark schemes adopted from [27] for the trace of *Reality* with discrete $V = 4$ class-distributed node cost.
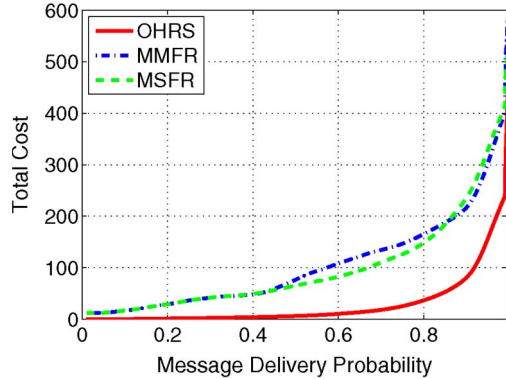

Fig. 16. Performance comparison between our OHRS and the two benchmark schemes adopted from [27] for the trace of *Shanghai* with discrete $V = 4$ class-distributed node cost.
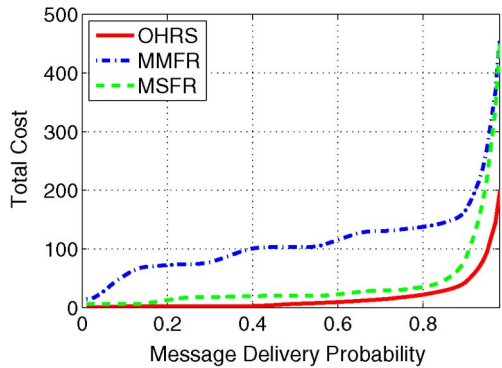
the discrete $V = 4$ class distribution of node cost, the OHRS reduced the average total costs by 82.6% and 91.3% over the MSFR and MMFR, respectively. Under the uniform distribution of node cost, the OHRS reduced the average total costs by 13.8% and 90.2% over the MSFR and MMFR, respectively. The average reductions of 23% and 92.1% in the total delivery cost were obtained by the OHRS over the MSFR and MMFR, respectively, under the normal distribution of node cost. Unlike the *Infocom06* trace, interestingly, the MSFR scheme achieved a much better performance than the MMFR scheme. It appears that the nodes with a high social metric can achieve higher delivery probability than the nodes that move often in the *Reality* trace. The performance of the MMFR and MSFR schemes seem to depend on specific network environments.

For the *Shanghai* trace, Figs. 16 and 17 show the results obtained by the three schemes under the discrete $V = 4$ class and normal distributions of node cost, respectively. Again, our OHRS achieved the best performance. For example, with the discrete $V = 4$ class distribution of node cost, the OHRS reduced the average total costs by 55.3% and 56.3% over the MSFR and MMFR, respectively. Unlike the cases of the *Infocom06* and *Reality* traces, in the vehicular environment of *Shanghai* trace, the MSFR and MMFR schemes achieved similar performance. This was because, in this vehicular environment, "mobility" and "sociability" were similar metrics. This further demonstrated that the performance of the MMFR and

Fig. 14. Performance comparison between our OHRS and the two benchmark schemes adopted from [27] for the trace of *Reality* with uniformly distributed node cost.

It can also be seen that the MMFR achieved better performance than the MSFR. The reason is that in the *Infocom06* trace, the nodes that move often encounter 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, Figs. 13–15 show the results obtained by the three schemes under the discrete $V = 4$ class, uniform, and normal distributions of node cost, respectively. Again, our OHRS achieved the best performance. Specifically, with
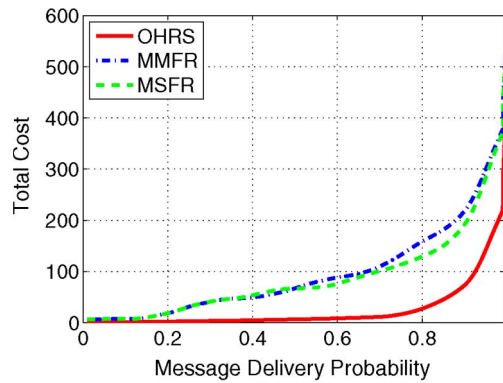
Fig. 17.   Performance comparison between our OHRS and the two benchmark schemes adopted from [27] for the trace of *Shanghai* with normally distributed node cost.

MSFR schemes depended on specific network environments. By contrast, our optimal OHRS performed equally well in various network environments.
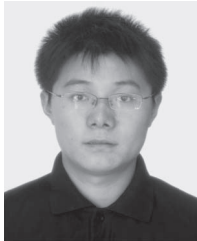
V. Conclusion

We have investigated the problem of optimal relaying for DTNs consisting of heterogeneous mobile nodes. In our proposed relaying scheme, we have used the heterogeneous delivery costs and contact rates of nodes to enhance system performance. We have formulated the optimal relaying problem as a knapsack optimization problem and have derived an algorithm based on dynamic programming to obtain the optimal relaying-node selection solution. Based on the experimentally collected real traces of *Reality, Cambridge, Infocom05, Infocom06,* and *Shanghai*, we have conducted extensive simulations and have adopted from the literature some existing relaying schemes, which assume homogeneous nodes, and the existing schemes that take into account the nodes' heterogeneous features as the benchmarks. The results obtained in the experimental study have demonstrated that our optimal heterogeneous relaying scheme achieves superior performance over these existing benchmark relaying schemes.

References

[1] 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 Technol.*, vol. 1, no. 1, pp. 70–109, Aug. 2001.

[2] K. Fall, "A delay-tolerant network architecture for challenged Internets," in *Proc. ACM SIGCOMM Conf. Appl., Technol., Architect., Protocols Comput. Commun.*, Karlsruhe, Germany, Aug. 25–29, 2003, pp. 27–34.

[3] K. Fall and S. Farrell, "DTN: An architectural retrospective," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 5, pp. 828–836, Jun. 2008.

[4] H. Wu, Y. Wang, H. Dang, and F. Lin, "Analytic, simulation, and empirical evaluation of delay/fault-tolerant mobile sensor networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 9, pp. 3287–3296, Sep. 2007.

[5] H. Dang and H. Wu, "Clustering and cluster-based routing protocol for delay-tolerant mobile networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 1874–1881, Jun. 2010.

[6] 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 Netw. Comput.*, New Orleans, LA, May 18–21, 2009, pp. 299–308.

[7] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, "Maxprop: Routing for vehicle-based disruption-tolerant networks," in *Proc. 25th IEEE INFOCOM*, Barcelona, Spain, Apr. 23–29, 2006, pp. 1–11.

[8] Y. Li, Z. Wang, D. Jin, L. Zeng, and S. Chen, "Collaborative vehicular content dissemination with directional antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1301–1306, Apr. 2012.

[9] S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, and H. Weiss, "Delay-tolerant networking: An approach to interplanetary Internet," *IEEE Commun. Mag.*, vol. 41, no. 6, pp. 128–136, Jun. 2003.

[10] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Dept. Comput. Sci., Duke Univ., Durham, NC, Tech. Rep. CS-200006, Apr. 2000.

[11] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case," *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, pp. 77–90, Feb. 2008.

[12] M. Abdulla and R. Simon, "A simulation analysis of multicasting in delay tolerant networks," in *Proc. 38th Conf. Winter Simul.*, Monterey, CA, Dec. 3–6, 2006, pp. 2234–2241.

[13] A. Picu and T. Spyropoulos, "Distributed stochastic optimization in opportunistic networks: The case of optimal relay selection," in *Proc. 5th ACM Workshop Challenged Netw.*, Chicago, IL, Sep. 20–24, 2010, pp. 21–28.

[14] Y. Li, Y. Jiang, D. Jin, L. Su, L. Zeng, and D. Wu, "Energy-efficient optimal opportunistic forwarding for delay-tolerant networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 9, pp. 4500–4512, Nov. 2010.

[15] A. Picu and T. Spyropoulos, "Minimum expected *-cast time in DTNs," in *Bioinspired Models of Network, Information, and Computing Systems*, Y. Hayel, E. Hart, R. El-Azouzi, I. Carrera, and E. Altman, Eds. New York: Springer-Verlag, 2010, pp. 103–116.

[16] Y. Li, P. Hui, D. Jin, L. Su, and L. Zeng, "Optimal distributed malware defense in mobile networks with heterogeneous devices," *IEEE Trans. Mobile Comput.*, to be published.

[17] A. Krifa, C. Barakat, and T. Spyropoulos, "Optimal buffer management policies for delay tolerant networks," in *Proc. 5th IEEE SECON*, San Francisco, CA, Jun. 16–20, 2008, pp. 260–268.

[18] A. Krifa, C. Barakat, and T. Spyropoulos, "An optimal joint scheduling and drop policy for delay tolerant networks," in *Proc. 2nd WoWMoM*, Newport Beach, CA, Jun. 23–26, 2008, pp. 1–6.

[19] Y. Li, M. Qian, D. Jin, L. Su, and L. Zeng, "Adaptive optimal buffer management policies for realistic DTN," in *Proc. IEEE GLOBECOM*, Honolulu, HI, Nov. 3–Dec. 4, 2009, pp. 1–5.

[20] C.-H. Lee and D. Y. Eunt, "Heterogeneity in contact dynamics: Helpful or harmful to forwarding algorithms in DTNs?" in *Proc. 7th Int. Symp. Model. Optim. Mobile, Ad Hoc, Wirel. Netw.*, Seoul, Korea, Jun. 23–27, 2009, pp. 1–10.

[21] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Perform. Eval.*, vol. 62, no. 1–4, pp. 210–228, Oct. 2005.

[22] M. Garetto, P. Giaccone, and E. Leonardi, "Capacity scaling in delay tolerant networks with heterogeneous mobile nodes," in *Proc. ACM MOBIHOC*, Montréal, QC, Canada, Sep. 9–14, 2007, pp. 41–50.

[23] C. Liu and J. Wu, "An optimal probabilistic forwarding protocol in delay tolerant networks," in *Proc. 10th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, New Orleans, LA, May 18–21, 2009, pp. 105–114.

[24] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant MANETs," in *Proc. 8th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Montréal, QC, Canada, Sep. 9–14, 2007, pp. 32–40.

[25] A. Balasubramanian, B. N. Levine, and A. Venkataramani, "DTN routing as a resource allocation problem," in *Proc. ACM SIGCOMM Conf. Appl., Technol., Architect., Protocols Comput. Commun.*, Kyoto, Japan, Aug. 27–31, 2007, pp. 373–384.

[26] 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 Netw. Comput.*, Hong Kong, China, May 26–30, 2008, pp. 241–250.

[27] T. Spyropoulos, T. Turletti, and K. Obraczka, "Routing in delay-tolerant networks comprising heterogeneous node populations," *IEEE Trans. Mobile Comput.*, vol. 8, no. 8, pp. 1132–1147, Aug. 2009.

[28] Y. Li, P. Hui, D. Jin, L. Su, and L. Zeng, "Evaluating the impact of social selfishness on the epidemic routing in delay tolerant networks," *IEEE Commun. Lett.*, vol. 14, no. 11, pp. 1026–1028, Nov. 2010.

[29] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Performance analysis of mobility-assisted routing," in *Proc. 7th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Florence, Italy, May 22–25, 2006, pp. 49–60.

[30] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. Mobile Comput.*, vol. 6, no. 6, pp. 606–620, Jun. 2007.

[31] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovi, "Power law and exponential decay of inter contact times between mobile devices," *IEEE Trans. Mobile Comput.*, vol. 9, no. 10, pp. 1377–1390, Oct. 2010.

[32] Y. Li, G. Su, D. Wu, D. Jin, L. Su, and L. Zeng, "The impact of node selfishness on multicasting in delay tolerant networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2224–2238, Jun. 2011.

[33] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *Proc. ACM SIGCOMM Workshop Delay Tolerant Netw.*, Philadelphia, PA, Aug. 22–26, 2005, pp. 244–251.

[34] V. Conan, J. Leguay, and T. Friedman, "The heterogeneity of inter-contact time distributions: Its importance for routing in delay tolerant networks," *arXiv:cs/0609068 [cs.NI]*, Sep. 2006.

[35] 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 Trans. Netw.*, vol. 17, no. 5, pp. 1578–1591, Oct. 2009.

[36] 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*, San Diego, CA, Mar. 14–19, 2010, pp. 1–9.

[37] H. Zhu, L. Fu, G. Xue, Y. Zhu, M. Li, and L. Ni, "Recognizing exponential inter-contact time in VANETs," in *Proc. IEEE INFOCOM*, San Diego, CA, Mar. 14–19, 2010, pp. 1–5.

[38] Y. Li, D. Jin, P. Hui, L. Su, and L. Zeng, "Revealing contact interval patterns in large scale urban vehicular ad hoc networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 42, no. 4, pp. 299–300, Oct. 2012.

[39] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multiobjective optimization," *Evol. Comput.*, vol. 3, no. 1, pp. 1–16, Spring 1995.

[40] V. V. Vazirani, *Approximation Algorithms*. Berlin, Germany: Springer-Verlag, 2004.

[41] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA: MIT Press, 2001.

[42] Y. Wang, H. Wu, and N.-F. Tzeng, "Cross-layer protocol design and optimization for delay/fault-tolerant mobile sensor networks (DFT-MSN's)," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 5, pp. 809–819, Jun. 2008.

[43] N. Eagle and A. Pentland, "Reality mining: Sensing complex social systems," *Pers. Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, Mar. 2006.

Li Su received the B.S. degree from Nankai University, Tianjin, China, in 1999 and the Ph.D. degree from Tsinghua University, Beijing, China, in 2007, both in electronics engineering.

He is currently a Research Associate with the Department of Electronic Engineering, Tsinghua University. His research interests include telecommunications, future Internet architecture, and on-chip networks.

Depeng Jin (M'09) received the B.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 1995 and 1999, respectively, both in electronics engineering.

He is currently an Associate Professor and the Vice Chair with the Department of Electronic Engineering, Tsinghua University. His research interests include telecommunications, high-speed networks, application-specific integrated circuit design, and future Internet architecture.

Dr. Jin received the National Scientific and Technological Innovation Prize (Second Class) in 2002.

Sheng Chen (M'90–SM'97–F'08) received the B.Eng. degree in control engineering from the East China Petroleum Institute, Dongying, China, in 1982; the Ph.D. degree in control engineering from the City University, London, London, U.K., in 1986; and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005.

From 1986 to 1999, he held research and academic appointments with the University of Sheffield, University of Edinburgh, and University of Portsmouth. Since 1999, he has been with the Department of Electronics and Computer Science, University of Southampton, where he is currently a Professor of intelligent systems and signal processing. He is also a Distinguished Adjunct Professor with King Abdulaziz University, Jeddah, Saudi Arabia. He is an ISI highly cited researcher in the engineering category (March 2004). He is the author of over 456 research papers. His recent research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, and evolutionary computation methods and optimization.

Dr. Chen is a Chartered Engineer and a Fellow of The Institution of Engineering and Technology.

Yong Li (M'09) received the B.S. degree from Huazhong University of Science and Technology, Wuhan, China, in 2007 and the Ph.D. degree from Tsinghua University, Beijing, China, in 2012.

He is currently a Postdoctoral Researcher with Tsinghua University. His research interests include mobile delay-tolerant networks, including forwarding policy design, buffer management design, and performance evaluation; mobility modeling; and mobility management in next-generation wireless internet protocol (IP) networks, including Mobile IP, Session Initiation Protocol, Proxy mobile IP, cross-layer design, multicast mobility, modeling for mobility performance evaluation, enhancing handoff performance, and proposing mobility management architecture.

Dr. Li serves as a Reviewer for several international conferences, including IEEE International Conference on Communications, the IEEE Vehicular Technology Conference, the International Conference on Information Networking, and the IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications.

Zhaocheng Wang (M'09–SM'11) received the B.S., M.S., and Ph.D. degrees from Tsinghua University, Beijing, China, in 1991, 1993, and 1996, respectively.

From 1996 to 1997, he was with Nanyang Technological University (NTU), Singapore, as a Post Doctoral Fellow. From 1997 to 1999, he was with OKI Techno Centre Pte. Ltd., Singapore, first as a Research Engineer and then as a Senior Engineer. From 1999 to 2009, he worked with SONY Deutschland GmbH, first as a Senior Engineer and then as a Principal Engineer. He is currently a Professor with the Department of Electronic Engineering, Tsinghua University. He is the author of over 60 technical papers and the holder of 22 U.S./EU patents. His research interests include wireless communications, digital broadcasting, and millimeter-wave communications.

Dr. Wang is a Fellow of The Institution of Engineering and Technology. He has served as a Cochair and a member of technical program committees of many international conferences.