A Reliable and Efficient Encounter-Based Routing Framework for Delay/Disruption Tolerant Networks

Yue Cao, Ning Wang, *Member, IEEE*, Zhili Sun, *Member, IEEE* and Haitham Cruickshank, *Member, IEEE*

***Abstract*—This article addresses Delay/Disruption Tolerant Networking (DTN) routing under a highly dynamic scenario, envisioned for communication in Vehicular Sensor Networks (VSNs) suffering from intermittent connection. Here, we focus on the design of a high level routing framework, rather than the dedicated encounter prediction. Based on an analyzed utility metric to predict nodal encounter, our proposed routing framework considers the following three cases: 1) Messages are efficiently replicated to a better qualified candidate node, based on the analysed utility metric related to destination. 2) Messages are conditionally replicated if the node with a better utility metric has not been met. 3) Messages are probabilistically replicated if the information in relation to destination is unavailable in the worst case. With this framework in mind, we propose two routing schemes covering two major technique branches in literature, namely Encounter-Based Replication Routing (EBRR) and Encounter-Based Spraying Routing (EBSR). Results under the scenario applicable to VSNs show that, in addition to achieving high delivery ratio for reliability, our schemes are more efficient in terms of a lower overhead ratio. Our core investigation indicates that apart from what information to use for encounter prediction, how to deliver messages based on the given utility metric is also important.**

***Index Terms*—VSNs, VANETs, DTN, Routing Framework, Message Replication, Efficiency.**

I. Introduction

In recent years, a new infrastructure for monitoring the physical world is emerging, named Vehicular Sensor Networks (VSNs) [1]. VSNs consist of highly dynamic mobile vehicles equipped with on-board sensors to relay data messages via wireless communication, and are envisioned to support a variety of urban monitoring and safety applications such as cooperative traffic monitoring, prevention of collisions and road surface monitoring.

Different from traditional wireless sensor networks, vehicles in VSNs are not limited by energy, and can process high complexity mobile computing based on the powerful processing units. Furthermore, since vehicles move at relevant high speeds, the topology of network is highly dynamic and the network topology may be even frequently disconnected. Such network properties, which are concerned by the research related to Delay/Disruption Tolerant Networks (DTNs) [2], make routing in VSNs a challenge due to the intermittent connection. Here, DTNs were originated from interplanetary communication, and have been widely envisioned for a range of terrestrial applications including Vehicular Ad hoc NETworks (VANETs) [3] and sensor networks [4].

Routing in DTNs aims to achieve high message delivery ratio, along with a low redundancy and delivery delay. Here, messages are relayed at each encounter opportunity and eventually delivered by destination via the Store-Carry-Forward (SCF) routing mechanism, concerning a contemporaneous end-to-end path towards destination is unavailable. As reviewed in [5], although forwarding a message using single copy [6]–[8] (meaning no message copies will be generated) guarantees low redundancy, the routing performance dramatically suffers from low delivery ratio and high delay. In contrast, it is more effective to generate multiple message copies [9], which promotes fast diffusion and increases the possibility that one of them would be delivered before a given message expiration deadline. It is worth noting that although the latter operation increases message delivery ratio, the suffered replication redundancy inevitably deteriorates the routing efficiency.

Concerning how to balance these objectives, mainly two branches of researches have made contributions to solving this problem in literature, where:

1) The former branch replicates messages to any better candidate node based on the utility metric [10]–[14], where the utility metric can be defined in various ways based on the historical encounter information. Note that the number of copies that a message can be replicated is unlimited in this branch.
2) The latter branch however limits the number of replicated copies of a message up to L, together with selecting the candidate node via utility metric [15]–[18]. Here, L is decided in relation to underlying scenario.

In comparison, the latter branch benefits more from the scenario where nodes are highly mobile to deliver messages in an efficient way, whereas the former branch guarantees the delivery reliability with a trade off additional replication redundancy in a relatively low dynamic network. Despite of these contributions, previous works do not adequately consider the following two limitations:

- Due to an infrequent encounter in sparse VANETs [19]

The authors are with the Institute for Communication Systems (ICS), University of Surrey, Guildford, UK. email: y.cao; n.wang; z.sun; h.cruickshank@surrey.ac.uk.

This work is supported by EU FP7 C-DAX project and EU FP7 Mission project.

Manuscript received 23 September, 2014.
Manuscript revised 2 December, 2014.
Manuscript revised 3 February, 2015.
Manuscript accepted 22 February, 2015.

DOI: 10.1109/JSEN.2015.2410297
Version of Record at http://ieeexplore.ieee.org/document/7055213/

or VSNs[1], the candidate node with a better quality might not be available. In this case, to still keep messages until this given candidate node is met, would however result in a longer delivery delay. This situation is critical and often happens, when the nodal utility metric is unstable due to highly dynamic characteristic.
- Due to the same reason, the candidate node selection operations might not be performed, if the information in relation to destination is unavailable. Although using flooding scheme such as Epidemic routing [9] is straightforward, it is however inefficient considering the resource constraint in the network. Here, the resource constraint normally refers to limited bandwidth for transmitting messages given short encounter duration, or limited buffer space to persistently carry messages.

Concerning the nodal mobility for prediction, as pointed out in [20], our interest focuses on the routing schemes in the branch[2] exploring the nodal encounter history as stochastic process for prediction [13], different from those in another branch making prediction based on their mobility patterns [21]. It is known that, apart from the number of encounters [10] and encounter duration [22] adopted by previous works, the inter-meeting time [13], [23] as another important factor[3] measures how regular pairwise nodes will encounter. In literature, it has been shown that using more encounter information helps to improve the routing performance.

However, even though a short inter-meeting time indicates a frequent encounter, the short encounter duration may deteriorate message delivery considering the limited bandwidth for transmitting large size messages. Given the nature of SCF routing behavior that messages are stored for a long time in nodal buffer until a new encounter happens, the limited buffer space may be insufficient for storage purpose, even if they have been successfully transmitted thanks to a long enough encounter duration. In light of this, apart from a well designed routing framework and a useful utility metric for candidate node selection, it is also essential to manage message transmission and storage.

Instead of focusing on predicting nodal encounter, we address routing in DTNs from another aspect, by designing a reliable and efficient routing framework that can be generalized for other routing schemes in literature. Therefore, any prediction for opportunistic encounter based on our proposed routing framework, can facilitate the communication in VSNs where the intermittent connectivity is concerned. Below are our contributions in this article:

We analyze the encounter factors affecting the message delivery, then define a utility metric based on jointly considering the nodal encounter history information. Further to this, we propose a replication based routing scheme, namely Encounter-Based Replication Routing (EBRR), consisting of three phases:

- The Utility Replication Phase controls replication redundancy, by selecting the candidate node via the utility metric. Considering the nodal mobility of possible next hop, only the node with a higher delivery potential reserves the bandwidth and buffer space for receiving message copies.
- The Conditional Replication Phase enhances the delivery reliability, given that the message delivery will be degraded if the better candidate node is not met. Here, the condition depends on the utility metric and message remaining lifetime.
- The Probabilistic Replication Phase considers the worst case, that the historical encounter information in relation to destination is unavailable due to rare encounter. Here, the replication probability depends on message remaining lifetime and its number of copies.

By generalizing the properties of EBRR, we further propose a spray[4] based routing scheme, namely Encounter-Based Spray Routing (EBSR) scheme, in which the number of replications for each message is limited by an initialized value. It is worth noting that compared to EBRR replicating message copies without a limitation of L, EBSR which distributes and relays $(L-1)$ message copies, is proposed for achieving a lower overhead ratio and therefore performs better in resource constrained network. However, EBSR suffers more from delivery deterioration in a relatively low dynamic network due to replicating a limited number of message copies.

The rest of this paper is organized as follows. Following a summary of related work in Section II, Section III presents the assumption and discussion for the encounter factors affecting the message delivery. We then propose the EBRR and EBSR in Sections IV and V respectively, together with a discussion on their properties. In Section VI, their performance are evaluated and compared with other well known routing schemes, followed by the conclusion made in Section VII.

II. Related Work

First, we introduce two benchmark routing schemes. Direct Delivery (DD) [24] only relays each message if the destination is in proximity. Although this scheme performs only one transmission, it suffers from dramatically low delivery ratio and high delivery delay. Epidemic [9] blindly floods each message to achieve the high delivery ratio. However, this scheme only performs well if there is no contention exists in relation to bandwidth and buffer space, because it results in huge replication redundancy. In order to reduce the replication redundancy, the following two branches have been investigated in literature.

[1]In sparse VANETs or VSNs, network nodes are normally connected in a sparse mode and also opportunistically. As such, this type of networks is expected to experience frequent network fragmentation in rural locations with sparsely populated roads, and also during non-busy hours such as late night.

[2]The schemes in former branch are normally applied for highly dynamic and sparse scenario, such as VSNs with a light number of vehicles.

[3]For example, assuming the encounter duration between any two nodes is same, two nodes may encounter 2 times in 1 hour, as compared to those with the same movement interest may encounter 2 times with 10 minutes. Obviously, a shorter inter-meeting time implies a more frequent encounter opportunity.

[4]The term "spray" means given predefined value L, each message can only be replicated for $(L-1)$ times. Note that the initialization of L is application dependent, e.g., based on number of nodes in network to support delay tolerant requirement, since using more message copies expedites delivery. In this article, we set L equals to the 10% number of nodes in network, by referring to [15].

A. Not Limiting the Number of Copies for Replication

A variety of history encounter information could be used to define a utility metric for encounter prediction. The value of utility metric is updated when pairwise nodes encounter, as a criterion to qualify the delivery potential of each node. In literature, the forwarding based routing schemes [6]–[8] have paid extensive attention on the definition of utility metric, in case that there is no additional message copies replicated through communication. Upon these contributions, the replication based routing schemes focus more on enhancing the message delivery in sparse networks, if any encountered node to which messages are replicated, is qualified with a higher delivery potential [10]–[14]. Targeting to VANETs scenario with the same network properties of VSNs, Packet Oriented Routing (POR) [25], Resource Allocation Protocol for Intentional DTN (RAPID) [13] have showed their good performance. Further to these, Approach-and-Roam (AaR) [12] which estimates a movement range of destination and geographically replicates messages for delivery, shows both its delivery reliability and efficiency under such highly dynamic scenario. To enhance routing efficiency, Delegation Forwarding (DF) [26] enables message to cache an updated threshold value equals to the topology based utility metric for message destination, enabling message replication to the candidate node with better utility metric than this cached threshold.

B. Limiting the Number of Copies for Replication

As compared to those in the former branch, the routing schemes in this branch assume that when enough nodes in the network are sufficiently mobile, a limited number of replications of each message can guarantee the message delivery. In [15], a scheme called Spray-and-Wait (SaW) is proposed, where a copy ticket for each message is initialized as L. Then, only $(L-1)$ message copies are allowed for replication. Previous works [16]–[18] in this branch also consider using the utility metric for encounter prediction. In particular, instead of relying on the direct encounter between destination and one of the $(L-1)$ message copies, relaying them via intermediate nodes [15], [18] can speed up the message delivery. For example, Spray-and-Focus (SaF) [15] further adopts a forwarding scheme to decrease the delivery delay via a utility forwarding scheme.

It is worth nothing the binary spray mechanism has been proved to be effective and widely used by existing works, where the way to distribute message copies is independent of any encounter prediction. Contrary to above works, [27]–[29] relies on a specific encounter prediction to heuristically distribute message copies. The focus in [27] determines a multiple periods for message copies spraying, where some number of additional copies are sprayed in each period. A similar work in [28] considers to dynamically reduce the number of copies that a message can be replicated. The work in [29] proposes a heuristic scheme to proportionally distribute message copies.

C. Contributions

Following above previous works, we observe that even with years of investigation, most previous works investigate "*what historical encounter information to maintain for predicting delivery potential*", rather than "*how to reliably and efficiently deliver messages based on the delivery potential deducted from historical encounter information*".

For instance, assuming that node i learns that j is a better candidate node to carry message, whereas they rarely encounter most likely. Following the routing strategy adopted by previous works, node i has to always keep on carrying this given message, until node j is in proximity, or until this message is inevitably deleted due to expiration. As such, the routing reliability in terms of both delivery delay and delivery ratio are degraded. In addition to this, instead of always flooding message, what to do when node i has no information about other useful candidate node is also important.

In this article, we are mainly concerned with a well designed routing framework that plays an important role on driving the routing performance, instead of relying on a utility metric for encounter prediction. Therefore, the contributions from existing works on encounter prediction could be integrated with our proposed EBRR and EBSR for further improvement, in spite that integrating those encounter prediction based heuristics [27]–[29] to spray message copies will need further consideration.

III. Preliminary

The encounter happens when pairwise nodes come into the transmission range of each other. Based on the routing decision, a copy of message is replicated to encountered node, while the current carrier still keeps its message. A message is either replicated via intermediate nodes, or directly delivered when destination is in proximity. Mainly, the delivery ratio and overhead ratio are more concerned, whereas the delivery delay is less important due to the delay tolerant nature of DTNs.

Here, the routing information is exchanged simultaneously when pairwise nodes encounter. Given a simple MAC layer protocol, only one connection can be established among a set of nodes. In other words, the routing decision/transmission will not be operated in relation to an encountered node if it is busy with transmitting messages to other nodes. Also, in networks that are quite sparse, we expect that only a few nodes would be close enough each time to compete for the same bandwidth. Considering DTNs are generally with sparse network density, referring to a light weight VSNs, the overhead in relation to exchanging routing information can be ignored by referring to [12]. Meanwhile, messages are with large size compared to the size of routing table, thus we are mainly concerned with the buffer occupation for data messages.

A. Single Message Case

Based on a pure flooding scheme Epidemic [9], the Delivery Probability (DP) that considers only one message for transmission, is calculated as:

$$DP = 1 - (1 - P_R)^R \tag{1}$$

where R as the replication redundancy, is the number of message copies including the original message in the network. Meanwhile, P_R is the probability that each copy is delivered along an independent routing path, before message expiration deadline. The equation (1) calculates the probability that at least one of the R message copies is delivered. Here, we observe that a larger value of P_R and R increase the message delivery probability.

Since a node normally has sufficient buffer space to receive single message, P_R can be converted as equation (2), where D is the encounter duration and b is the transmission bandwidth.

$$P_R = \prod_{x=1}^{Z} \left(\frac{D \times b}{S} \right) \tag{2}$$

Where:

$$P_R = \begin{cases} 0 & \text{if } \frac{D \times b}{S} < 1 \\ 1 & \text{if } \frac{D \times b}{S} \geq 1 \end{cases} \tag{3}$$

Equations (2) and (3) imply that the encounter duration D must be long enough at each hop, for successfully transmitting each message with the size of S. Here, we denote Z as the maximum hop of delivered message.

Based on [30], under Random WayPoint (RWP) mobility model, the replication redundancy R when the network size grows large, is presented as:

$$R = \frac{K}{1 + (K-1)e^{-\beta K t}} \tag{4}$$

where t is the current time in the network, and K is the total number of nodes. Note that the condition $(t > 0)$ holds the nature that a network is active, meaning no message generation or nodal movement will start given $(t = 0)$. Besides, $(\beta = \frac{1}{T})$ as the encounter rate, is inverse to the inter-meeting time T.

Considering both encounter duration and buffer space are sufficient, equation (1) is converted as:

$$DP = 1 - \left(1 - \left(\frac{\text{Initial Message Lifetime} - t}{\text{Initial Message Lifetime}}\right)\right)^{\frac{K}{1+(K-1)e^{-\frac{Kt}{T}}}} \tag{5}$$

Here, with a large value of K:

$$DP \approx 1 - \left(1 - \left(\frac{\text{Initial Message Lifetime} - t}{\text{Initial Message Lifetime}}\right)\right)^{e^{\frac{Kt}{T}}} \tag{6}$$

It is observed that given an initial message lifetime, the delivery probability is increased by using more message copies as well as diffusing them fast, requiring a small T which is inherently in relation to either a small network size or large transmission range.

Since the remaining message lifetime is reduced following the increased t, the message should be delivered before this expiration deadline for reliability. Here, a smaller value of t implies a less number message copies generated, whereas a larger P_R requires a smaller t. Therefore, awaiting to generate more message copies does not always play a positive effect on message delivery, given a message expiration deadline. With this in mind, if considering the quality of mobile node to relay the message, selecting the one with good delivery potential to carry message copy is essential to achieve a fast and efficient delivery.

B. Multiple Messages Case

If considering there are multiple messages for transmission, then:

$$P_R = \prod_{x=1}^{Z} \left(\frac{(D \times b - \vartheta \times S)}{S} \right) \tag{7}$$

Here, ϑ is the number of other messages have been successfully transmitted/received during an encounter. Since a large R reflects a large number of other messages to be transmitted/received, decreasing T does not always contribute to a large P_R. This is because that if $\left(\frac{(D \times b - \vartheta \times S)}{S}\right) < 1$, the delivery probability of this given message is deteriorated.

Besides, since a large number of generated messages yields a large number of copies, the nodal buffer space B should be considered at receiver side, where:

$$P_R = \prod_{x=1}^{Z} \left(\frac{(B - \vartheta \times S)}{S} \right) \tag{8}$$

If too many messages are replicated at this encounter opportunity, the available buffer space may be insufficient for receiving the subsequent messages. Here, we consider the general behavior of routing scheme, that receiver will allocate buffer space for any incoming message, by deleting a message from its local buffer space. In light of this, some of the already carried messages on receiver side have to be deleted, as such the delivery of these deleted messages is deteriorated due to a less number of their copies existing in network.

Generally, although preventing message replication does reduce ϑ, this operation inevitably decreases the delivery probability for other messages, considering T and message lifetime. With this in mind, the research in this article focuses on designing a routing framework that controlling the ϑ via efficient decision, meanwhile guaranteeing message delivery.

C. Definition of Utility Metric

TABLE I
LIST OF NOTATIONS

N_i	Message carrier
N_j	Encountered node
N_d	Destination of message
M	A message carried by N_i
T_M^{ini}	Initial message lifetime
T_M^{ela}	Elapsed time since message generation
$D_{j,d}$	Historically encounter duration between N_j and N_d
$T_{j,d}$	Historically inter-meeting time between N_j and N_d
$C_{j,d}$	Historically encounter count between N_j and N_d
$U_{j,d}$	Utility value estimated for N_d, based on the information recorded in N_j
∇_M	Threshold value cached in M, for recording $U_{j,d}$
C_M	Carrier list of M
RP_M	Replication probability in EBRR
SP_M	Spraying probability in EBSR
L	Initialized copy ticket of M
K	Total number of nodes in network
U_M	Message utility

Based on above analysis, a node with better quality of utility metric implies a higher message delivery potential. Given the encounter factors between nodes N_i and N_j, where $N_i, N_j \in K$, we address encounter count $C_{i,j}$, encounter duration $D_{i,j}$ and inter-meeting time $T_{i,j}$.

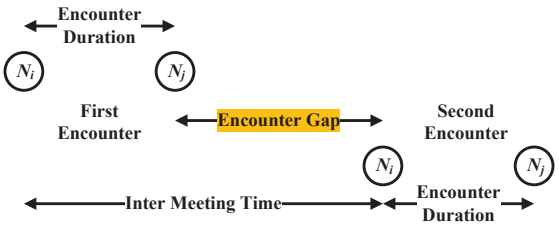


Fig. 1. Illustration of Nodal Encounter

With the aforementioned discussion on the encounter factors that affect delivery probability, we observe that $(T_{i,j} - D_{i,j})$ has advantage than only considering $T_{i,j}$. This is because that both a lower $T_{i,j}$ and longer $D_{i,j}$ reflects a lower value of $(T_{i,j} - D_{i,j})$. In other words, N_i and N_j would have short time to encounter each other, while with a long encounter duration for message transmission at previous encounter opportunity, as the encounter gap highlighted in Fig.1.

It is worth noting that since different combinations of encounter durations and inter meeting times may result in the same encounter gap, their number of encounters $C_{i,j}$ is considered to average the value of $(T_{i,j} - D_{i,j})$. Then, we define the utility $U_{i,j}$ as:

$$U_{i,j} = \frac{T_{i,j}^{(C_{i,j}=1)} + \sum_{(C_{i,j}=2)}^{H} \left(T_{i,j}^{(C_{i,j})} - D_{i,j}^{(C_{i,j}-1)}\right)}{H} \tag{9}$$

where H is the value of current encounter count. For example, assuming $T_{i,j}^{(C_{i,j}=1)} = 20$, $D_{i,j}^{(C_{i,j}=1)} = 2$ at the 1^{st} encounter, while $C_{i,j} = 2$ and $T_{i,j}^{(C_{i,j}=2)} = 10$ are recorded for the second encounter, then $U_{i,j}$ is calculated as:

$$U_{i,j} = \frac{20 + (10 - 2)}{2} = 14 \tag{10}$$

Note that given the 1^{st} time encounter, the encounter duration is 0 as pairwise nodes did not meet in the past. This intention to estimate an average value of $(T_{i,j} - D_{i,j})$ is because the topology based utility metric inevitably becomes to be unstable, particularly under the high dynamic scenario.

D. Update Routing Information

Regarding information update between any pairwise encountered nodes, $T_{i,j}$ is always updated prior to $D_{i,j}$, because the valid encounter duration is only calculated when the link between two nodes is disrupted. Besides, $T_{i,j}$ is updated when $C_{i,j}$ changes, which implies a new value of $T_{i,j}$ should be calculated for a new encounter. Considering their encounter for the 1^{st} time, $U_{i,j}$ is simply updated given $(T_{i,j} - 0)$. Otherwise, it is updated according to equation (9).

TABLE II
AN EXAMPLE OF ROUTING TABLE IN N_1

Node ID	Inter-Meeting Time	Encounter Duration	Encounter Count	Cumulative Utility	Utility
N_4	20	6	1	20	20
N_6	14	8	2	19	9.5

Using an example in TABLE II, $U_{1,4} = (20 - 0)$ is calculated for the 1^{st} encounter. Based on the existing information about N_6, assuming that if N_1 encounters N_6 for the 3^{rd} time with $T_{1,6}^{(C_{1,6}=3)} = 16$, then $U_{1,6} = \frac{(T_{1,6}^{(C_{1,6}=1)} - 0) + (T_{1,6}^{(C_{1,6}=2)} - D_{1,6}^{(C_{1,6}=1)}) + (T_{1,6}^{(C_{1,6}=3)} - D_{1,6}^{(C_{1,6}=2)})}{3} = \frac{19 + (16 - 8)}{3} = 9$. Note that the cumulative utility is the historically recorded cumulative value, which equals to $T_{1,6}^{(C_{1,6}=1)} + \sum_{(C_{1,6}=2)}^{2} \left(T_{1,6}^{(C_{1,6})} - D_{1,6}^{(C_{1,6}-1)}\right) = 19$ illustrated in TABLE II.

Algorithm 1 Update Routing Information

1: **if** A connection between N_i and N_j is established **then**
2: both N_i and N_j update the number of encounter $C_{i,j}$
3: both N_i and N_j update the inter-meeting time $T_{i,j}$
4: both N_i and N_j update the utility $U_{i,j}$
5: **for** each $N_x \in K$ encountered by N_i in the past **do**
6: **if** $N_x = N_j$ **then**
7: skip this N_x
8: **else if** N_x is encountered by N_j in the past **then**
9: **if** $(U_{i,x} > U_{i,j} + U_{j,x})$ **then**
10: replace the value $U_{i,x}$ with $(U_{i,j} + U_{j,x})$
11: **end if**
12: **end if**
13: **end for**
14: **else if** The connection between N_i and N_j is disrupted **then**
15: both N_i and N_j update the encounter duration $D_{i,j}$
16: **end if**

Here, denoting N_d as the destination of message M, N_j is considered as a better candidate node than N_i, given that $(U_{j,d} < U_{i,d})$. In addition to above direct encounter, the transitivity is based on the observation that if N_i has high potential to encounter N_j, and similarly for the encounter between N_j and N_d. Then N_i has a good chance to relay its carried message M to N_d through N_j. Such a relation is implemented as follows:

$$U_{i,d} = \begin{cases} U_{i,j} + U_{j,d} & \text{if } (U_{i,d} > U_{i,j} + U_{j,d}) \\ U_{i,d} & \text{else} \end{cases} \tag{11}$$

Note that this transitivity for N_i is not updated if N_i itself has a better potential for message delivery. This distributed updating is implemented after information exchange between pairwise nodes. In Algorithm 1, the pseudo-codes between lines 9 and 10 present the process of updating the transitivity of $U_{i,j}$. Here, both N_i and N_j update the transitivity for any N_x that encountered in the past, where N_x includes N_d.

IV. DESIGN OF EBRR

A. Routing Framework

Upon previous background on utility metric, EBRR is presented based on the proposed routing framework in this section. As illustrated in Algorithm 2, before making any routing decision, N_i will deliver the message M destined to N_d, only if N_i directly encounters N_d. Besides, if a message copy is already in the buffer of N_j, then the routing framework ignores the process of this message.

1) Utility Replication: This routing phase happens if N_j has obtained the information about $U_{j,d}$. In this case, the condition $(U_{i,d} > U_{j,d})$ is used to make message replication, since N_j has a higher potential to encounter N_d.

Further to this, the concept of DF [26] is utilized to enhance routing efficiency. We define an additional flag ∇_M in message

Algorithm 2 Routing Strategy of EBRR

```
initialize the value of ∇_M
for each encounter between N_i and N_j do
    for each M carried by N_i do
        directly deliver M if it is destined to N_j
        update ∇_M if N_j has a copy of M
        update ∇_M if ((U_{i,d} < ∇_M) and (U_{i,d} ≠ N/A))
        if (U_{j,d} ≠ N/A) then
            if (∇_M > U_{j,d}) then
                update ∇_M towards U_{j,d}
                replicate M to N_j
            else if (∇_M > T_M^{ini} − T_M^{ela}) then
                replicate M to N_j
            end if
        else
            replicate M according to RP_M
        end if
    end for
end for
```

M, initialized as:

$$\nabla_M = \begin{cases} U_{i,d} & \text{if } (N_i \text{ has knowledge about } U_{i,d}) \\ +\infty & \text{else} \end{cases} \tag{12}$$

For each message generation, ∇_M is initialized as an infinitely large value if N_i did not encounter N_d in the past, which implies the least delivery potential. Otherwise, ∇_M is initialized as the value of $U_{i,d}$, considering that N_i has already met N_d since message generation.

Upon this initialization, the message M is replicated to N_j given the condition:

$$(\nabla_M > U_{j,d}) \tag{13}$$

Here, the value of ∇_M will be updated towards $U_{j,d}$, recorded as the utility metric of previous encountered node, to compare with that of an upcoming encountered node. In light of this, the condition (13) focuses on comparing the utility metrics, only between the currently encountered node and previously encountered node, instead of comparing that between the currently encountered node and message carrier.

If considering that N_j already has a copy of M, the values of ∇_M in the messages carried by both N_i and N_j, are updated towards a smaller value between them. This is important to make a converged routing decision, as the highly dynamic mobility results in large variation of nodal utility metric.

We further make a modification that the ∇_M is updated according to $U_{i,d}$, given $(\nabla_M > U_{i,d})$. Considering the condition $(\nabla_M > U_{i,d} > U_{j,d})$, this is different from original DF which does not adopt the utility metric of message carrier for comparison, even if the latter is with a better potential. It is worth noting that since the proposed utility metric is estimated in an accumulative manner, $U_{i,d}$ would be with a small variation. As such, adaptively updating ∇_M to $U_{i,d}$ further promotes the convergence of routing decision.

2) Conditional Replication: Considering that N_j does not meet the condition (13), to always keep the message would result in a longer delivery delay, or even degrade the delivery probability due to a short message lifetime.

Therefore, we present the condition (14) to deliver message before expiration deadline, as given by:

$$(T_M^{ela} + U_{i,d} \leq T_M^{ini}) \tag{14}$$

Here, T_M^{ela} is denoted as the elapsed time since message generation, while T_M^{ini} is denoted as the initialized message lifetime. For message delivery before remaining message lifetime $(T_M^{ini} - T_M^{ela})$, $U_{i,d}$ should not be longer than the value of $(T_M^{ini} - T_M^{ela})$. Thus the intention of using condition $(U_{i,d} > T_M^{ini} - T_M^{ela})$ to make message replication, implies N_j would encounter other nodes meet the condition $(U_{i,d} > U_{j,d})$ in future.

Since replicating additional message copies increases routing overhead, we further utilize DF and convert the condition $(U_{i,d} > T_M^{ini} - T_M^{ela})$ into:

$$(\nabla_M > T_M^{ini} - T_M^{ela}) \tag{15}$$

The condition (15) implies that, even ∇_M as the smallest value of $U_{j,d}$ recorded in the network, is still longer than the remaining message lifetime. Therefore, the redundancy is reduced, by using this routing phase to compensate message delivery if the better candidate node is not met.

3) Probabilistic Replication: This operation is performed if the value of $U_{j,d}$ is "N/A", meaning that N_j did not meet N_d in the past. Here, instead of blindly flooding messages, a replication probability RP_M is defined:

$$RP_M = \left(1 - \frac{T_M^{ini} - T_M^{ela}}{T_M^{ini}}\right)^{C_M} \tag{16}$$

Based on equation (16), we observe a shorter remaining message lifetime $(T_M^{ini} - T_M^{ela})$ contributes to a larger RP_M, meaning that M is not with sufficient period to survive in the network. Particularly, RP_M is also increased based on a small C_M which is as the estimated number of copies for M. In this context, both a long message remaining lifetime as well as a large number of copies contribute to a high delivery probability.

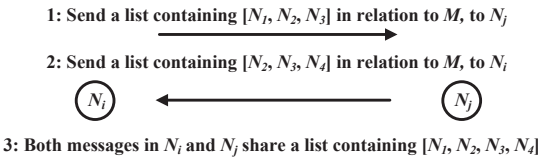


Fig. 2. Updating The Knowledge About Number of Nodes Have Carried M

Since all the nodes in the networks are differentiated by their IDs, we propose a heuristic method to estimate the number of nodes which have carried the message. Here, an additional flag in each message is used to keep a list of these IDs, and its initialization is performed by recording the ID of node that generates the message. Therefore, the initial value of C_M equals to 1.

Upon a successful transmission from N_i to N_j, the message including its replicated copy in both N_i and N_j, will record the ID of N_j. Moreover, the nodal IDs will be exchanged when pairwise encountered nodes both carrying the message or its copy. As an example in Fig.2 where N_i and N_j encounter each other and both of them already carried the same M, then the knowledge about the number of nodes which carried M in the past, is extended by sending the handshake request to exchange nodal IDs. Note that this information exchange is

operated together with the exchange of routing information, when N_i and N_j encounter.

B. Message Management

1) Message Transmission: As aforementioned, $(T_M^{ini} - T_M^{ela})$ is calculated as the remaining message lifetime. Then, a positive value of $((T_M^{ini} - T_M^{ela}) - \nabla_M)$ implies this message can be delivered within the remaining lifetime, given their encounter potential $U_{j,d}$ as recorded in ∇_M. Thus, considering the mobility of the candidate node, $\frac{(T_M^{ini}-T_M^{ela})-\nabla_M}{(T_M^{ini}-T_M^{ela})}$ implies the possibility that each message copy is successfully delivered before the expiration deadline.

Next, by considering the number of message copies, a larger $(K - C_M)$ implies this message has not been extensively replicated, in the network consisting of total K nodes. In other word, this given message can still be replicated with $(K-C_M)$ times, as C_M copies have already existed. Since increasing the number of message copies enhances the message delivery ratio, the message with a larger value of $(K - C_M)$ is more important.

Based on the above, we thus define the utility U_M to qualify each message, where:

$$U_M = 1 - \left(1 - \frac{(T_M^{ini} - T_M^{ela}) - \nabla_M}{(T_M^{ini} - T_M^{ela})}\right)^{(K-C_M)} \tag{17}$$

The equation (17) estimates the possibility of a message, that can be delivered via at most $(K - C_M)$ copies within the following $(T_M^{ini} - T_M^{ela})$ remaining lifetime. As such, the message with the largest U_M is with the highest priority for transmission.

In another case if the information about $U_{j,d}$ is unavailable, U_M is then defined as:

$$U_M = 1 - \left(1 - \frac{T_M^{ini} - T_M^{ela}}{T_M^{ini}}\right)^{(K-C_M)} \tag{18}$$

Here, since the delivery potential of encountered node is not qualified in this case, messages are only qualified according to $(T_M^{ini} - T_M^{ela})$. The equation (18) estimates the possibility that a message can be delivered before its maximum expiration deadline T_M^{ini}, similar to the discussion on equation (5).

Based on the above discussion, U_M is defined as:

$$U_M = \begin{cases} 1 - \left(1 - \frac{(T_M^{ini}-T_M^{ela})-\nabla_M}{(T_M^{ini}-T_M^{ela})}\right)^{(K-C_M)} & \text{if } U_{j,d} \neq N/A \text{ (Case 1)} \\ 1 - \left(1 - \frac{T_M^{ini}-T_M^{ela}}{T_M^{ini}}\right)^{(K-C_M)} & \text{else (Case 2)} \end{cases} \tag{19}$$

Comparing with those replicated via Utility Replication Phase, messages replicated via Probabilistic Replication Phase are considered with a lower priority for transmission. This is because that it is desirable to transmit the message with a higher priority if the utility metric in relation to destination is available, rather than naively transmitting those if not knowing their destinations.

2) Buffer Management: If N_j does not have sufficient buffer space to receive the incoming M, N_j then deletes its stored messages from the one with the lowest U_M. Following the same rule for message transmission, messages with the U_M estimated in Case 2, are deleted prior to those with the U_M estimated in Case 1.

If a message copy is successfully delivered, it is essential to delete other copies of this message in the network, in order to free the buffer space for the undelivered messages. In this case, each node maintains a list to record the IDs of delivered messages in the network, then exchanges and updates the information in this list. In the worst case that a node does not have this knowledge, may constantly carry the delivered message copy until the destination node is in proximity, the destination will delete this copy since it has been already received.

Here, when pairwise nodes encounter, the knowledge on IDs of delivered messages will be exchanged together with nodal maintained encounter information, where fast disseminating this knowledge benefits from a highly mobile network characteristic. Since the ID is with string based format, the bandwidth consumption regarding exchanging those IDs information could be ignored comparing with transmitting data messages.

C. Properties of EBRR

The analysis is based on the Random WayPoint (RWP) mobility model, where the meeting times of pairwise nodes are Independent and Identically Distributed (IID). The number of nodes in the network is denoted as K. Here, we assume that bandwidth and buffer space are infinite, to avoid the factor of resource contention.

Property IV.1 *The Utility Replication Phase enhances the routing efficiency.*

Proof: Referring to [26], based on a given utility metric for making routing decision, using original DF reduces the redundancy dramatically from $O(K)$ to $O(\sqrt{K})$, where the replication redundancy when using original DF in Utility Replication Phase is derived as:

$$R_{DF}(g) \approx (1 + \sqrt{g}) \times \sqrt{K} = (1 + \sqrt{g}) \times 2^x \tag{20}$$

Here, g as the gap between the maximum delivery potential and initial value, is calculated as $g = 1 - \frac{1}{U_{i,d}}$ given that the delivery potential is inverse to $U_{i,d}$. For example, the delivery potential is considered as 0 if N_i did not meet N_d in the past. Meanwhile, the assumption $2^x = \sqrt{K}$ made in [26] implies that the routing decision goes above the threshold by updating ∇_M for x times. Therefore, in the worst case that N_i never met N_d, we have $U_{i,d} = +\infty$. Then we approximate $g = 1 - \frac{1}{U_{i,d}} \approx 1 - 0 = 1$, as such $R_{DF}(g) \approx 2\sqrt{K}$.

In our Utility Replication Phase:

$$R'_{DF}(g) \approx (1 + \sqrt{g}) \times (1 + \lambda)^x \leq R_{DF}(g) \tag{21}$$

Here, the replication process in Utility Replication Phase is controlled by λ, where $\lambda \in [0, 1]$ is in relation to dynamically updating ∇_M towards $U_{i,d}$ given the condition $(\nabla_M > U_{i,d})$. Note that since only $(\lambda = 1)$ is held if $(\nabla_M \leq U_{i,d})$, then the above discussion implies the further improved efficiency of Utility Replication Phase. Although the above analysis is based

on the ideal case that the node encounter rate is independent of their utility metric, the improvement in replication redundancy when using DF is even greater in case that utility metric and encounter rate are identical, as discussed in [26]. In this context, this property is also applicable to the scenario with more complicated mobility pattern, which will be shown in section VI. ■

Property IV.2 *The Conditional Replication Phase enhances the performance regarding delivery ratio and delivery delay.*

Proof: According to [30], the scale of delivery delay in Direct Delivery, $Delay(DD)$ follows EMT, as the Expected Meeting Time for any pairwise encountered nodes. Also, the scale of delivery delay in Epidemic, $Delay(Epidemic)$ follows $\left(\frac{EMT \times \ln K}{K-1}\right)$, as an optimal redundancy based multi-hop routing. Then:

$$Delay(DD) > Delay(Utility) > Delay(Epidemic) \tag{22}$$

In the worst case that the encountered node has a lower value $U_{j,d}$ is always unavailable, the Utility Replication Phase degrades to be Direct Delivery, because the message is only delivered only when the destination is met. In contrast, it behaves as Epidemic, because messages are always replicated to encountered nodes, if any encountered node has a lower value of $U_{j,d}$ than ∇_M.

Recall that the condition $(\nabla_M > T_M^{ini} - T_M^{ela})$ makes message replication in Conditional Replication Phase, only when the condition $(\nabla_M > U_{j,d})$ in Utility Replication Phase is not met. In light of this, a larger probability that messages are replicated from N_i to N_j reduces the delivery delay, as such increasing the possibility to deliver messages before the expiration deadline. ■

Property IV.3 *By using DF in Conditional Replication Phase, the replication redundancy is reduced as compared to the case if not using DF.*

Proof: Recall that the condition $(U_{i,d} > T_M^{ini} - T_M^{ela})$ is used if not using DF. While in Conditional Replication Phase, the condition $(\nabla_M > T_M^{ini} - T_M^{ela})$ is adopted instead. Based on the discussion for Property IV.1, we observe that the condition $(\nabla_M \leq U_{i,d})$ is always met, given the dynamic updating process in Utility Replication Phase. In light of this, by using the smallest value of $U_{i,d}$, implied as the best delivery potential recorded in the past to compare with $(T_M^{ini} - T_M^{ela})$, the replication redundancy is further reduced. ■

The performance is evaluated using the Opportunistic Network Environment (ONE) [31] version 1.41, a well known java based simulator used for the research on routing in DTNs. We validate the properties of EBRR, under the RWP scenario with 1000×1000 m^2 area, consisting of $K = 100$ mobile nodes with the constant 3 m/s moving speed. The communication technique is set with 10m transmission range. Messages are randomly generated from mobile nodes for every 30s with 1KB size and default 90 minutes lifetime, while the buffer space for all nodes in the network is configured to be infinite. Therefore, there is no contention from bandwidth and buffer space.

Here, we specify different versions of EBRR to examine the discussed properties:

- EBRR (U): The version only the Utility Replication Phase is performed.
- EBRR (U&C): The version that Utility Replication and Conditional Replication Phases are performed.
- EBRR (U&P): The version that Utility Replication and Probabilistic Replication Phases are performed.

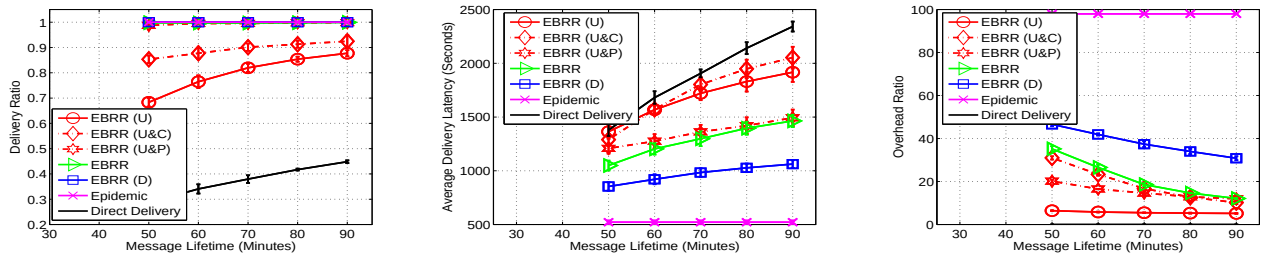
Algorithm 3 Strategy of EBRR (D)

```
for each encounter between N_i and N_j do
    for each M carried by N_i do
        directly deliver M if it is destined to N_j
        if (U_{j,d} ≠ N/A) then
            if (U_{i,d} is unavailable) then
                replicate M
            else if (U_{i,d} > U_{j,d}) then
                replicate M to N_j
            else if (U_{i,d} > T_M^{ini} − T_M^{ela}) then
                replicate M to N_j
            end if
        else
            replicate M according to RP_M
        end if
    end for
end for
```

The intentions for this scenario configuration are because: Firstly, the exponential distribution of EMT is a widely adopted assumption in literature [32]. Although the distribution in realistic motilities may be more complicated, we assume an exponential distribution of delivery time[5] in this article. Moreover, further simulation results in Section VI show that this tractable simplification performs well in a realistic Helsinki city scenario. Secondly, we guarantee a sparse network density following the assumption in Section III, by using a small device transmission range where end-to-end connectivities are difficult to create, partially referring to [15]. Thirdly, an appropriately set moving speed is required, as for 3 m/s in the 1000×1000 m^2. This is because that either a slow moving speed can not enable EBRR to achieve a close performance of Epidemic, or a fast moving speed can not reflect the performance gaps between different versions of EBRR.

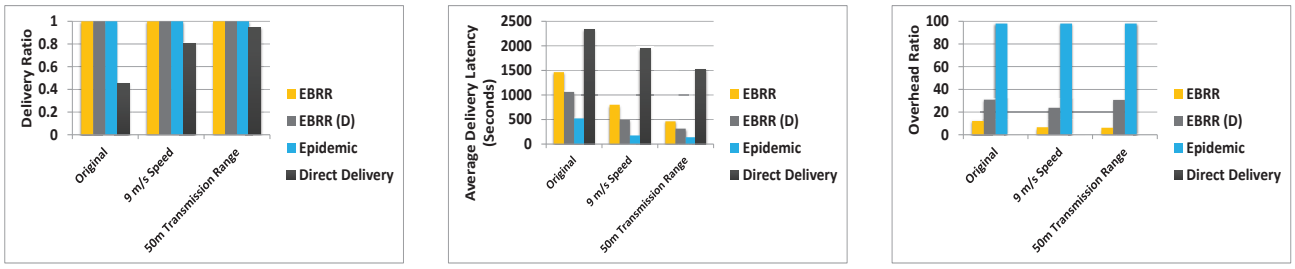
For well examining the properties, we further present a degraded version of EBRR that not using DF, namely EBRR (D) as presented in Algorithm 3. To measure the full activity of the network, the message generation ends up before 16200s with additional 5400s to consume the unexpired messages while results are plotted with 95 $\%$ confidence interval. The delivery ratio is given by the ratio between the number of messages delivered and the total number of messages generated. The overhead ratio is given by the ratio between the number of relayed messages (excluding the delivered messages) and the number of delivered messages. The average delivery latency is given by the end-to-end average time spent for delivering messages from the source to destination.

[5]The encounter meeting times between destination and different nodes carrying a message copy are exponentially distributed with means $\frac{1}{\beta_1}$, $\frac{1}{\beta_2}$.... $\frac{1}{\beta_R}$ where $R \leq K$. Then the minimum delivery delay that the destination meets any one of these nodes is $\frac{1}{\beta_1+\beta_2+.....\beta_R}$.



(a) Delivery Ratio (b) Average Delivery Latency (c) Overhead Ratio

Fig. 3. Results Regarding the Discussion of EBRR



(a) Delivery Ratio (b) Average Delivery Latency (c) Overhead Ratio

Fig. 4. Performance Given Increased Speed and Transmission Range

Here, Epidemic and Direct Delivery are the upper bound and low bound of the performance. In Fig.3(a), the improvement of Conditional Replication Phase regarding delivery ratio, as discussed in Property IV.2, is observed given the comparison between EBRR (U) and EBRR (U&C). Meanwhile, the improvement of this regarding delivery delay is shown given comparing EBRR with EBRR (U&P) in Fig.3(b).

In Fig.3(c), when comparing EBRR with EBRR (U&C), we observe Probabilistic Replication Phase does not yield too much redundancy. This is because a long message lifetime results in a small RP_M, meaning that it is unnecessary to replicate message if it can still survive for certain time. Here, the Conditional Replication Phase makes more message replication if with short message lifetime. In contrast, the replication redundancy is dramatically reduced, due to using ∇_M to compare with remaining long message lifetime. Further concerning EBRR (D), Properties IV.1 and IV.3 enable EBRR to achieve a lower overhead ratio. This observation shows that Conditional Replication Phase overcomes the limitation of Utility Replication Phase, in which the encounter gap is defined based on inter-meeting time, number of encounters as well as encounter duration.

In Fig.4(a), Fig.4(b) and Fig.4(c), we further increase the nodal speed and transmission range, where EBRR achieves the best improvement in terms of decreased delivery latency while with a high delivery ratio. This is because that a faster speed and large transmission range increases the possibility that destination is met. In particular, compared with EBRR (D), EBRR achieves a decreased overhead ratio given large transmission range. This shows the efficiency of EBRR if being applied in realistic VANETs even the network is relatively not sparse.

V. Design of EBSR

Given the advantage of proposed routing framework, we then extend EBRR by limiting the number of message copies and propose a spray based routing scheme, namely Encounter Based Spray Routing (EBSR). Note that both EBRR and EBSR share the same operation for updating the routing information.

Algorithm 4 Strategy of EBSR

initialize the value of ∇_M
for each encounter between N_i and N_j **do**
  for each M carried by N_i **do**
    directly deliver M if it is destined to N_j
    update ∇_M if N_j has a copy of M
    update ∇_M if $((U_{i,d} < \nabla_M)$ and $(U_{i,d} \neq N/A))$
    if $(U_{j,d} \neq N/A)$ **then**
      if $((\nabla_M > U_{j,d})$ and $(L > 1))$ **then**
        update ∇_M towards $U_{j,d}$
        replicate M to N_j with $\frac{L}{2}$
        keep $(L - \frac{L}{2})$ for M in N_i
      else if $((\nabla_M > U_{j,d})$ and $(L = 1))$ **then**
        update ∇_M towards $U_{j,d}$
        forward M to N_j using single copy
      else if $((\nabla_M > T_M^{ini} - T_M^{ela})$ and $(L > 1))$ **then**
        replicate M to N_j with $\frac{L}{2}$
        keep $(L - \frac{L}{2})$ for M in N_i
      end if
    else if $((U_{j,d} = N/A)$ and $(L > 1))$ **then**
      replicate M to N_j with $\frac{L}{2}$, according to SP_M
      keep $(L - \frac{L}{2})$ copy tickets for M in N_i
    end if
  end for
end for

In EBSR, each message is only replicated up to $(L-1)$ times, where L is a parameter defined according to the scenario. Given the nature of spray based routing scheme, the value of L will be distributed upon each message replication.

Here, we adopt the binary tree mechanism[6] to distribute the value of L.

A. Routing Framework

1) Utility Spraying: Similar to EBRR, the condition $(\nabla_M > U_{j,d})$ is adopted to select a better relay node, whereas the routing decision depends on the current value of L.

- The $(L > 1)$ case: Since this case implies that message can still be sprayed, the message with $(L > 1)$ copy tickets is replicated to N_j given the condition $(\nabla_M > U_{j,d})$. Meanwhile, the value of L will be equally distributed, for the replicated message in N_j and the original carried message in N_i. That is, the replicated message is allocated with $\frac{L}{2}$ copy tickets, while the original message keeps the rest $(L - \frac{L}{2})$ copy tickets, as illustrated between lines 8 and 11 in Algorithm 4.
- The $(L = 1)$ case: Here, the message is only forwarded using single copy, towards a better relay node meets the condition $(\nabla_M > U_{j,d})$, as illustrated between lines 12 and 14 in Algorithm 4. This operation further expedites the message process by utilizing the mobility of other intermediate nodes, rather than only waiting for the direct encounter between N_d.

2) Conditional Spraying: Considering that the Utility Spraying Phase does not work given $(\nabla_M \leq U_{j,d})$, the message spraying is determined by the condition $(\nabla_M > T_M^{ini} - T_M^{ela})$, with the operations presented at lines 16 and 17. Referring to the motivation of Conditional Replication Phase in EBRR, here a message copy allocated with $\frac{L}{2}$ copy tickets is replicated to N_j.

3) Probabilistic Spraying: Recall that this routing phase is performed only if $(U_{j,d} = N/A)$ while the number of message copies is limited up to $(L - 1)$ in EBSR, then the message spraying probability SP_M is defined as:

$$SP_M = \begin{cases} \left(1 - \frac{T_M^{ini} - T_M^{ela}}{T_M^{ini}}\right)^{C_M} & \text{if } C_M < L \\ \left(1 - \frac{T_M^{ini} - T_M^{ela}}{T_M^{ini}}\right)^{L} & \text{else} \end{cases} \tag{23}$$

Since only the message with $(L > 1)$ copy tickets is processed and M is normally sprayed with $\frac{L}{2}$ copy tickets in EBSR, then the given $(L-1)$ message copies requires $O(\log_2(L))$ epochs to be fully sprayed.

Similar to that in EBRR, the value of C_M is initialized with 1 since message generation and increased by 1 upon each successful transmission. In EBSR, we still consider the message with $(L > 1)$ copy tickets as one message copy, although it can be further sprayed. For example, assuming L is initialized as 8 and equally distributed for the 1^{st} time spraying, then if a message with 4 remaining copy tickets is requested for the 2^{nd} time spraying, SP_M is calculated as $\left(1 - \frac{T_M^{ini} - T_M^{ela}}{T_M^{ini}}\right)^2$ because the value of C_M is updated to 2 after 1^{st} time spraying.

[6]Given that the binary spray mechanism has been rigorously proven to be effective [15] and widely used by previous works, thus EBSR adopts this mechanism. Since our focus in this article is the design of the routing framework consisting of the proposed three phases, further concerning about how to distribute such $(L - 1)$ copies depending on encounter prediction is out of discussion herein.

B. Message Management

The U_M in EBSR is defined as follows:

$$U_M = \begin{cases} 1 - \left(1 - \frac{(T_M^{ini} - T_M^{ela}) - \nabla_M}{(T_M^{ini} - T_M^{ela})}\right)^{L} & \text{if } U_{j,d} \neq N/A \text{ (Case 1)} \\ 1 - \left(1 - \frac{T_M^{ini} - T_M^{ela}}{T_M^{ini}}\right)^{L} & \text{else (Case 2)} \end{cases} \tag{24}$$

Given that the value of L follows the binary distribution, the message entitled with L copy tickets implies it can still be replicated with $(L - 1)$ times. Therefore, a large value of L increases the possibility that one of $(L - 1)$ copies could be delivered. Similar to that in EBRR, the message transmission and storage management in EBSR follow the same rule as discussed for EBRR in Section 4.

C. Properties of EBSR

Property V.1 *The Utility Spraying Phase reduces the transmission overhead for the message with* $(L = 1)$ *copy ticket.*

Proof: This is because this given message is relayed to the node with the smallest value of $U_{j,d}$ recorded in the past, rather than being relayed to the node only with a smaller value than $U_{i,d}$. Therefore, a less number of nodes not significantly contributing to delivery will participate intermediate relay, as such the number of transmissions is further reduced based on the converged routing decision. ■

Property V.2 *The Conditional Spray Phase enhances the performance regarding delivery ratio the delivery delay.*

Proof: This proof can be referred to Property IV.2, as discussed for EBRR in Section 4. ■

For validation, we initialize $(L = 10)$, which is chosen following 10% number of nodes in the network, as suggested by [15]. In particular, referring to the design of original EBSR and EBRR (D), EBSR (D) as the degraded version of EBSR is also evaluated.

Here, the characteristics of three routing phases of EBSR as well as its properties, are observed from the results in TABLE III. Note that the reason that EBSR is with a slightly higher average delivery latency, is because EBSR only selects the nodes with the historically best $U_{j,d}$, different from EBSR (D) only selecting that with a currently better potential. Therefore, some of the intermediate nodes are not selected in EBSR, which reduces the overhead ratio. Meanwhile, since the number of message copies is limited up to $(L-1)$ herein, a slightly higher overhead ratio achieved by EBSR is considered to be less critical.

VI. Performance Evaluation

The nature of EBRR and EBSR have been discussed and validated in previous sections. Here, the main evaluation is based on the medium Helsinki city scenario in ONE shown in Fig.5. Considering as a community mobility model, we

TABLE III
RESULTS REGARDING THE DISCUSSION OF EBSR

Routing Algorithm	Delivery Ratio	Average Delivery Latency	Overhead Ratio
50 Minutes Message Lifetime			
EBSR (U)	0.6673 (±0.022)	1368s (±55)	6.54 (±0.21)
EBSR (U&C)	0.7642 (±0.020)	1372s (±62)	11.09 (±0.35)
EBSR (U&P)	0.9080 (±0.011)	1278s (±49)	11.51 (±0.28)
EBSR	0.9148 (±0.015)	1237s (±42)	12.94 (±0.39)
EBSR (D)	0.9210 (±0.008)	1195s (±76)	16.04 (±0.46)
90 Minutes Message Lifetime			
EBSR (U)	0.8691 (±0.017)	1955s (±103)	5.21 (±0.23)
EBSR (U&C)	0.8920 (±0.021)	2027s (±110)	5.84 (±0.29)
EBSR (U&P)	0.9839 (±0.006)	1623s (±64)	9.10 (±0.15)
EBSR	0.9852 (±0.003)	1609s (±101)	9.03 (±0.11)
EBSR (D)	0.9944 (±0.004)	1471s (±65)	12.37 (±0.27)

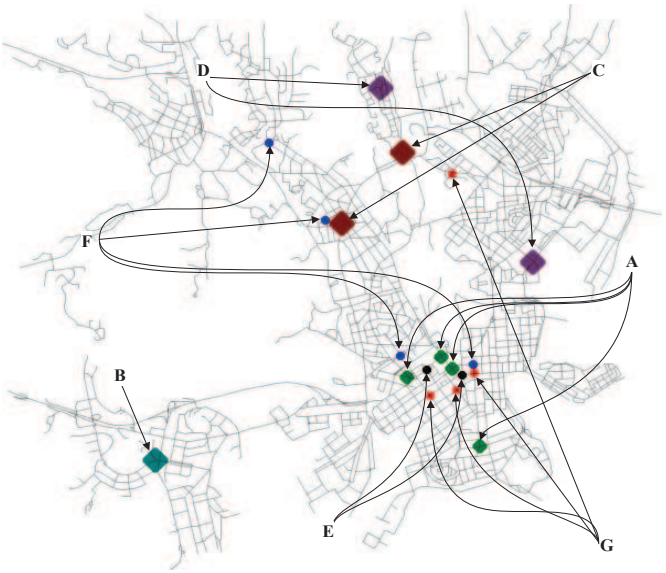


Fig. 5. Illustration of Scenario

deploy 7 types of interest points on the map. 10 mobile nodes of each group are allocated with individual interest points, as highlighted from A-G on the map, with 80% probability moving around these points and 20% probability just roaming in the entire network. Note that mobile nodes will encounter more likely and frequently due to a high moving probability around interest points.

Each node chooses the shortest path to an interested point via the Dijkstra's shortest path scheme, depending on their current location and moving speed. Referring to [33], the communication technique is configured as 4 Mbit/s bandwidth and 30m transmission range, considering as a low power WiFi technique. The default buffer space is limited to 40MB. We assume there are light weight vehicles in the network, with moving speed varies between [3∼10] m/s. Note that the network is sparse and highly dynamic, since the number of connectivities is small as compared to underlying 8300×7400 m^2 geographic area.

Envisioning for large file transmission like multimedia content, the message size is set with 1MB such that a transmission contention would exist. Messages are generated for a randomly selected source-destination pair, with 30s generation interval, 90 minutes lifetime. In order to fully examine the delivery reliability, the message generation starts from 0s and ends at 37800s, with additional 5400s to consume the unexpired messages.

It is highlighted that the ONE simulator has a simple MAC implementation. Before making any routing decision/transmission, there is a check function to determine whether the encountered node is in the status of transferring, meaning that node is communicating with others at the moment. If that happens, no additional operation in relation to that node will be done. This makes our simulation valid for transmitting large file messages as configured herein.

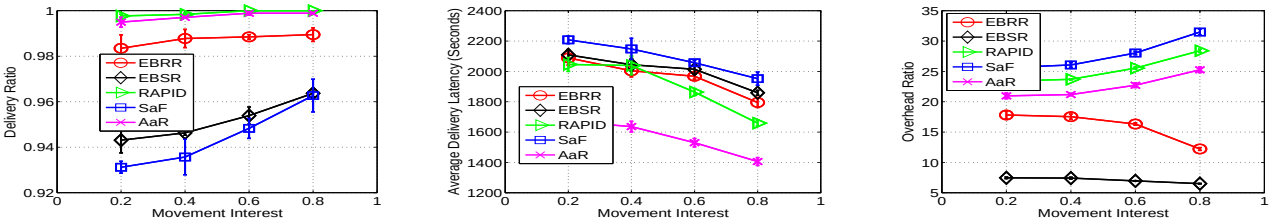
We select the following previous works for comparison:

- **RAPID** [13]: It adopts a random variable to represent the encounter between pairwise encountered nodes, and replicates messages in the descending order according to a marginal utility. In detail, the marginal utility is calculated based on the ratio between the decreased delivery delay and message size. The message estimated with a positive value of the marginal utility is then replicated for bandwidth usage. Its application scenario is applicable for VSNs.
- **SaF** [15]: A spray based routing scheme selected to compare with EBSR. SaF initially sprays a number of L copies of each message via the binary spray mechanism. Once L message copies have been distributed to $(L-1)$ intermediate nodes, these L copies are further relayed towards a better relay node based on recent encounter time.
- **AaR** [12]: A replication based routing scheme designed for VSNs scenario, using historical geographic information of mobile destination. AaR is selected to compare against the topology based routing schemes (e.g., RAPID, SaF, EBRR and EBSR). In AaR, messages are replicated via two phases, namely approach and roam phases, to efficiently and fast make message replication towards an estimated movement range of destination and effectively guaranteeing message replication within this range for delivery.

The default value L for EBSR and SaF is configured as 7, referring to [15] that choosing L equals to about 10% of the total number of nodes in the network. Meanwhile, the number of nodes K is 70 under scenario.

A. Influence of Movement Interest

In Fig.6(a), we observe that increasing the movement interest enhances the delivery ratio for all routing schemes. This is because that a higher movement interest yields more encounter opportunities at such interest points. In particular, the spray based schemes, like EBSR and SaF are with an observable performance improvement. This is because spray based routing schemes rely more on the situation where nodes are sufficient mobile to encounter each other, as reflected by a high movement interest. Of course, if mobile nodes encounter more likely thanks to an increased movement interest, the average delivery latency is therefore reduced in Fig.6(b). This is because the possibility to encounter message destination is increased around interest points. Regarding the overhead



(a) Delivery Ratio (b) Average Delivery Latency (c) Overhead Ratio

Fig. 6. Influence of Movement Interest

ratio in Fig.6(c), EBRR is with a dramatically decreased level compared to previous works. It is worth noting that since EBSR initially limits the number of message copies, it is with a smoothly decreased overhead ratio.

B. Influence of Message Generation Interval

In Fig.7(a), EBSR achieves the highest delivery ratio even with an aggressive traffic contention, thanks to only making a limited number of message copies in the network for efficient delivery. Given a low traffic contention, the replication based routing schemes, like AaR, EBRR and RAPID begin to outperform the spray based routing schemes like EBSR and SaF, because of using more message copies. Furthermore, EBSR is less sensitive to the traffic contention, by achieving the smallest fluctuation regarding average delivery latency in Fig.7(b). Interestingly, we observe that AaR, RAPID and SaF suffer from a dramatically increased overhead ratio, in contrast to EBRR and EBSR maintaining a relatively stable performance in Fig.7(c). This implies the proposed routing framework is advanced for guaranteeing message delivery given limited bandwidth, particularly when all schemes achieve a close delivery ratio in Fig.7(a).

C. Influence of Buffer Space

Although EBRR and EBSR share the same design of routing framework, the latter only yields a limited number of message copies in the network, as such it performs better given small buffer space in Fig.8(a). Thanks to the advantage of geographic utility metric discussed in [12], AaR outperforms RAPID. Although SaF initially sprays a limited number of message copies, it has limitation due to not jointly considering the mobility pattern. Therefore, the contention from limited buffer space becomes to be critical at some key nodes which bridge communication. In Fig.8(b), RAPID and SaF suffer from a higher increase regarding average delivery latency. In contrast, if with large buffer space, the average delivery latency is increased since more messages will exist in network for delivery. Apart from AaR, EBSR also maintains a stable performance. Upon the discussion of EBRR given small buffer space, it achieves a dramatically decreased overhead ratio in Fig.8(c). Meanwhile, thanks to the nature of EBSR that limiting the number of message copies, it maintains the lowest and stablest performance. Here, although AaR is the best related to the decrease of overhead ratio, its value is much higher than EBRR.

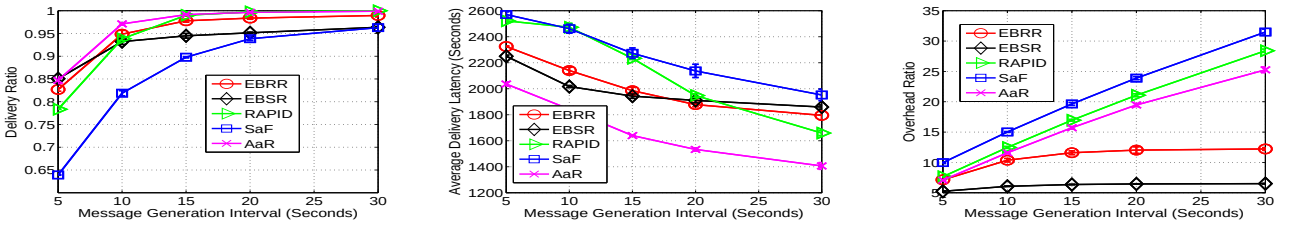
D. Further Discussion

Fig.9(a) shows that EBSR is more advanced than SaF in case that $(L = 1)$. This implies that the definition of $U_{j,d}$ better captures the nodal mobility pattern for encounter prediction. Besides, if increasing the initial value of L, both EBSR and SaF achieve the converged delivery ratio, because the message delivery is improved by using more message copies.

Upon previous results mainly discussing the characteristic of routing framework, we now proceed discussing the definition of utility metric. Here, we select EBRR (U&C) and EBSR (U&C) of which the performance are purely in relation to utility metric. In Fig.9(b), we observe that the case only using Inter-Meeting Time (IMT) as the utility metric achieves a worse performance than the case using our defined metric. Considering that a node with a short inter-meeting time may not have a long enough encounter duration, it may fail to successfully transmit a message in spite of frequent encounter opportunity. In addition to this, pairwise nodes may stay close for a certain time duration due to the same movement interest, whereas only using inter-meeting time does not take this factor into account.

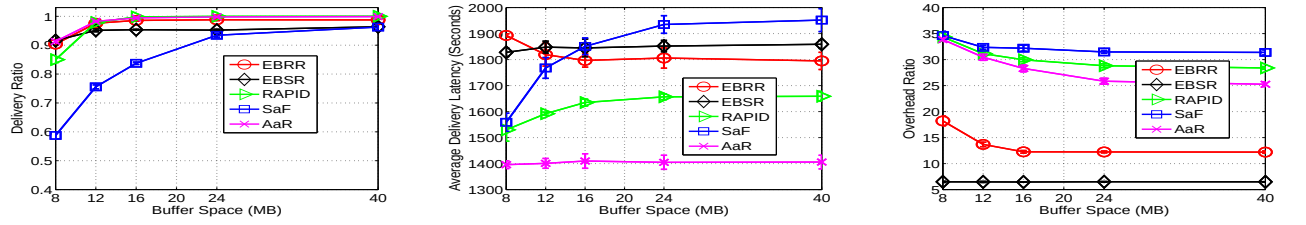
We further reduce the network dynamism, by setting a maximum waiting time for each node. Fig.9(c) shows that AaR performs worst in this case. This is because that the advantage of geographic based utility metric is only applicable under highly dynamic scenario. Concerning the spray based routing schemes like EBSR and SaF, they suffer more from performance degradation than replication based routing schemes EBRR and RAPID. This is due to the nature of spray based routing schemes that relying more on the sufficient nodal mobility for efficient delivery.

In general, with Global Position System (GPS) this assumption is natural in a pure VSNs, such that to support the application of AaR. But more often than not, in a heterogeneous network where only a proportion of vehicles are with GPS, the intermediate communication relying on those not with GPS is also necessary. Such situation could benefit from using history encounter information, as EBRR and EBSR. In both EBRR and EBSR, the storage overhead of maintaining updated routing formation is $O(K^2)$, which is the same as compared routing schemes. In particular, the result in Fig.9(b) indicates that using more historical encounter information could improve routing performance in DTNs, which is the same as the conclusion of previous work. Recall that since the proposed routing framework is somehow independent of the utility metric for encounter prediction, a practical concern for routing overhead in relation to information updating is out



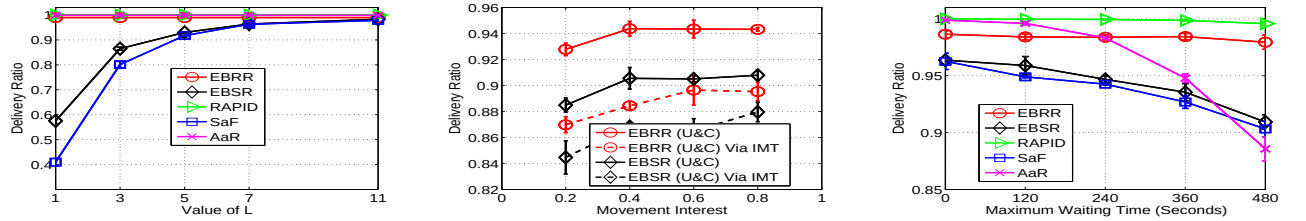
(a) Delivery Ratio (b) Average Delivery Latency (c) Overhead Ratio

Fig. 7. Influence of Message Generation Interval



(a) Delivery Ratio (b) Average Delivery Latency (c) Overhead Ratio

Fig. 8. Influence of Buffer Space



(a) Delivery Ratio vs Initial L (b) Delivery Ratio vs Utility Metric (c) Delivery Ratio vs Maximum Waiting Time

Fig. 9. Influence of Other Conditions

of the scope in this article.

In VANETs, if there are some other vehicles (e.g. proceeding and following vehicles in each direction), then the whole encounter duration (for these two vehicles) might not be used for transmission/reception. Certain media access algorithm [34] will determine and limit the accessing the media and its duration (allocation time) within this encounter duration. For practical implementation purpose, algorithms should be tested with IEEE 802.11p + IEEE 1604 in VANETs applications.

Further to this, IPv6 uses the hop count in the Hop Limit field instead of using time seconds in the Time-To-Live (TTL) field, as in IPv4. Although IPv4 packets are designed to carry the packet lifetime in seconds in the TTL field, this field is used for the hop count in practice. In this case, there is no way to determine how long vehicles will keep carrying messages when delivering messages in DTN nature. The pioneer work in [35] proposing approaches for the delivery of IPv6 packets in a DTN nature in partially connected networks could be referred for this concern.

VII. Conclusion

In this article, we addressed routing issue in DTNs from another aspect, by focusing on the design of routing framework. Upon the discussion on encounter factors affecting message delivery, we proposed a novel routing framework consisting of three phases. With this design in mind, we generalized EBRR and EBSR, as the guideline to design routing schemes following two major branches in literature. Results under the scenario envisioned for VSNs showed that, apart from which utility metric adopted for candidate node selection (e.g., geographic based metric is advanced than topology based encounter metric under highly dynamic scenario), it is also essential to consider the design of routing framework to reliably and efficiently deliver messages. Of course, since the main focus in this article is the proposal of routing framework that can be applied generally, the performance of EBRR and EBSR could be improved if further effort mainly focuses on dedicated encounter prediction.

References

[1] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, and A. Corradi, "Dissemination and Harvesting of Urban Data Using Vehicular Sensing Platforms," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 2, pp. 882–901, February, 2009.

[2] C. Caini, H. Cruickshank, S. Farrell, and M. Marchese, "Delay and Disruption-Tolerant Networking (DTN): An Alternative Solution for Future Satellite Networking Applications," *Proceedings of IEEE*, vol. 99, no. 11, pp. 1980–1997, July, 2011.

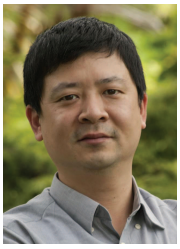
[3] P. Pereira, A. Casaca, J. Rodrigues, V. Soares, J. Triay, and C. Cervello-Pastor, "From Delay-Tolerant Networks to Vehicular Delay-Tolerant Networks," *IEEE Communications Surveys Tutorials*, vol. 14, no. 4, pp. 1166–1182, Fourth Quarter, 2011.

[4] H. Ochiai, H. Ishizuka, Y. Kawakami, and H. Esaki, "A DTN-Based Sensor Data Gathering for Agricultural Applications," *IEEE Sensors Journal*, vol. 11, no. 11, pp. 2861–2868, November, 2011.

[5] Y. Cao and Z. Sun, "Routing in Delay/Disruption Tolerant Networks: A Taxonomy, Survey and Challenges," *IEEE Communications Surveys Tutorials*, vol. 15, no. 2, pp. 654–677, Second Quarter, 2013.

[6] S. Medjiah, T. Taleb, and T. Ahmed, "Sailing over Data Mules in Delay-Tolerant Networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 5–13, January, 2014.
[7] H. Dang and H. Wu, "Clustering and Cluster-Based Routing Protocol for Delay-Tolerant Mobile Networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 1874–1881, June, 2010.
[8] M. Musolesi and C. Mascolo, "CAR: Context-Aware Adaptive Routing for Delay-Tolerant Mobile Networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 2, pp. 246–260, February, 2009.
[9] A. Vahdat and D. Becker, "Epidemic Routing for Partially-Connected Ad Hoc Networks," Duke University Technical Report Cs-2000-06, Tech. Rep., 2000.
[10] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic Routing in Intermittently Connected Networks," *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 7, pp. 19–20, July, 2003.
[11] H.-Y. Huang, P.-E. Luo, M. Li, D. Li, X. Li, W. Shu, and M.-Y. Wu, "Performance Evaluation of SUVnet With Real-Time Traffic Data," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 6, pp. 3381–3396, November, 2007.
[12] Y. Cao, Z. Sun, H. Cruickshank, and F. Yao, "Approach-and-Roam: A Geographic Routing Scheme for Delay/Disruption-Tolerant Networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 266–281, January, 2014.
[13] A. Balasubramanian, B. Levine, and A. Venkataramani, "Replication Routing in DTNs: A Resource Allocation Approach," *IEEE/ACM Transactions on Networking*, vol. 18, no. 2, pp. 596–609, April, 2010.
[14] L. Ding, B. Gu, X. Hong, and B. Dixon, "Articulation Node Based Routing in Delay Tolerant Networks," in *IEEE PerCom '09*, Galveston, Texas, USA, March, 2009.
[15] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Multiple-Copy Case," *IEEE/ACM Transactions on Networking*, vol. 16, no. 1, pp. 77 –90, February, 2008.
[16] B. Pasztor, M. Musolesi, and C. Mascolo, "Opportunistic Mobile Sensor Data Collection with SCAR," in *IEEE MASS '07*, Pisa, Italy, October, 2007.
[17] T. Spyropoulos, T. Turletti, and K. Obraczka, "Routing in Delay-Tolerant Networks Comprising Heterogeneous Node Populations," *IEEE Transactions on Mobile Computing*, vol. 8, no. 8, pp. 1132–1147, August, 2009.
[18] C. Boldrini, M. Conti, J. Jacopini, and A. Passarella, "HiBOp: A History Based Routing Protocol for Opportunistic Networks," in *IEEE WoWMoM '07*, Helsinki, Finland, June, 2007.
[19] H. Kang, S. H. Ahmed, D. Kim, and Y.-S. Chung, "Routing Protocols for Vehicular Delay Tolerant Networks: A Survey," *International Journal of Distributed Sensor Networks*, pp. 1–9, Accepted in 2014.
[20] W. Gao, G. Cao, A. Iyengar, and M. Srivatsa, "Supporting Cooperative Caching in Disruption Tolerant Networks," in *IEEE ICDCS '11*, Minneapolis, USA, June, 2011.
[21] Q. Yuan, I. Cardei, and J. Wu, "An Effiicient Prediction-Based Routing in Disruption-Tolerant Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 1, pp. 19–31, January, 2011.
[22] M. Demmer and K. Fall, "DTLSR: Delay Tolerant Routing for Developing Regions," in *ACM NSDR '07*, Kyoto, Japan, August, 2007.
[23] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, and O. Tonguz, "Routing in Sparse Vehicular Ad Hoc Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 8, pp. 1538–1556, October, 2007.
[24] M. Grossglauser and D. Tse, "Mobility Increases the Capacity of Ad Hoc Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477– 486, August, 2002.
[25] X. Li, W. Shu, M. Li, H. Huang, and M.-Y. Wu, "DTN Routing in Vehicular Sensor Networks," in *IEEE GLOBECOM '08*, New Orleans, Louisiana, USA, December, 2008.
[26] V. Erramilli, M. Crovella, A. Chaintreau, and C. Diot, "Delegation Forwarding," in *ACM MobiHoc '08*, Hong Kong, China, May, 2008.
[27] E. Bulut, Z. Wang, and B. Szymanski, "Cost-Effective Multiperiod Spraying for Routing in Delay-Tolerant Networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1530–1543, October, 2010.
[28] X. F. Guo and M. C. Chan, "Plankton: An Efficient DTN Routing Algorithm," in *IEEE SECON' 13*, New Orleans, USA, June, 2013.
[29] S. Nelson, M. Bakht, and R. Kravets, "Encounter-Based Routing in DTNs," in *IEEE INFOCOM '09*, Rio de Janeiro, Brazil, April, 2009.
[30] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance Modeling of Epidemic Routing," *Elsevier Computer Networks*, vol. 51, no. 10, pp. 2867–2891, July, 2007.
[31] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE Simulator for DTN Protocol Evaluation," in *ICST SIMUTools '09*, Rome, Italy, March, 2009.
[32] H. Wen, F. Ren, J. Liu, C. Lin, P. Li, and Y. Fang, "A Storage-Friendly Routing Scheme in Intermittently Connected Mobile Network," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 1138–1149, March, 2011.
[33] A. Kernen and J. Ott, "Increasing Reality for DTN Protocol Simulations," Helsinki University of Technology, Networking Laboratory Technical Report, Tech. Rep., 2007.
[34] A. Mostafa, A. M. Vegni, and D. P. Agrawal, "A Probabilistic Routing by Using Multi-hop Retransmission Forecast with Packet Collision-aware Constraints in Vehicular Networks," *Elsevier Ad Hoc Networks*, vol. 14, no. 12, pp. 118–129, March, 2014.
[35] K. K. Sevimli and M. Soyturk, "Enabling Delay-Tolerant Communications for Partially Connected Vehicular Ad Hoc Networks," *International Journal of Ad Hoc Ubiquitous Computing*, vol. 11, no. 2/3, pp. 157–168, November, 2012.

Yue Cao joined the Institute for Communication Systems (ICS) formerly known as Center for Communication Systems Research (CCSR), at University of Surrey, Guildford, UK in 2009 and obtained his PhD degree in 2013. He is currently a Research Fellow at the ICS. His research interests focus on Delay/Disruption Tolerant Networks, Electric Vehicle (EV) communication, Information Centric Networking (ICN).



Ning Wang received his PhD degree from Institute for Communication Systems (ICS) formerly known as Center for Communication Systems Research (CCSR), at University of Surrey, Guildford, UK in 2004. He is also a Senior Lecturer at the ICS and his research interests mainly include energy-efficient networks, network resource management, Information Centric Networking (ICN) and QoS mechanisms.

Zhili Sun (Chair of Communication Networking) is a Professor at the Institute for Communication Systems (ICS) formerly known as Center for Communication Systems Research (CCSR), at University of Surrey, Guildford, UK. He obtained his PhD degree from Lancaster University, Lancaster, UK, in 1991. His research interests include wireless and sensor networks, satellite communications, mobile operating systems, traffic engineering, Internet protocols and architecture, QoS, multicast and security.

Haitham Cruickshank is a Senior Lecturer at the Institute for Communication Systems (ICS) formerly known as Center for Communication Systems Research (CCSR), at University of Surrey, Guildford, UK. He obtained his PhD degree from Cranfield University, Cranfield, UK, in 1989. His main research interests are network security, satellite network architectures, DTN security, and quality-of-service (QoS) provisioning.