ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Computer Networks 00 (2019) 1–23

Computer Networks

www.elsevier.com/locate/procedia

Opportunistic Protocol based on Social Probability And Resources Efficiency for the Intelligent and Connected Transportation System

Tong Wang [a], MengBo Tang [a], Houbing Song[b], Y. Cao[c,*], Zaheer Khan[d]

[a]College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China.
[b]Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA.
[c]Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK
[d]Department of Computer Science and Creative Technologies, University of the West of England, Bristol BS16 1QY, UK.

Abstract

In recent years, Intelligent and Connected Transportation Systems (ICTS) have become a practical and valuable alternative for wide variety of novel applications in road traffic safety. It can be utilized to guarantee road safety and create new forms of inter-vehicle communications. However, due to the high speed of vehicles, the topology of the network is highly dynamic and the network may be disconnected frequently, which will lead to a decline in communication performance. Delay Tolerant Networks (DTNs) follow the approach to store and forward the message. DTNs can adapt to the highly dynamic scenario, envisioned for communication in ICTS suffering from intermittent connection. In this paper, we propose the Social Probability And Resource Effective (SPARE) protocol to improve delivery ratio and minimize the consumption of network resources. In SPARE, we focus on considering four factors that include the nodal resources effective consumption, encounter probability, nodal historical encounter information and the number of messages carried by nodes. We use the nodes' resources efficiency and encounter probability similarity to improve the delivery ratio of SPARE algorithm. In addition, SPARE applies the mechanism of dynamically managing messages to reduce network overhead. Finally, the simulation results show that SPARE achieves a higher delivery ratio and lower overhead ratio, compared to other protocols within resource constrained network situations.

Keywords: Intelligent and Connected Transportation Systems; Delay Tolerant Networks; Resources Efficiency; Encounter Probability; Social Similarity

1. Introduction

With the development of modern transportation, people enjoy more convenient modes of transportation and traffic experience. However, the rapid development of modern traffic has also caused traffic congestion and increased fuel consumption and various potential traffic safety problems. In the past ten years, advances in the Intelligent and Connected Transportation Systems (ICTS) [1] collectively aim to reduce the fuel expenditure by avoiding congested traffic, enhancement of traffic safety and so on. There are a series of

☆Fully documented templates are available in the elsarticle package on CTAN.

*Corresponding author

Email address: yue.cao@northumbria.ac.uk (Y. Cao)

[1]Since 1880.

individual requirements of both safety and non-safety applications in the vehicular communication technology, so it is necessary to build up a new communication technology for integrated solutions of ICTS. It is on this basis that several communication technologies have been developed, such as Vehicle-to-Vehicle (V2V) [2] and Vehicle-to-Infrastructure (V2I) [3] communications. V2I and V2V use a Dedicated Short Range Communication (DSRC) [4] method between either nearby vehicle or roadside equipment facilities. In particular, Vehicular Ad-hoc Networks (VANETs) [5] based on V2V and V2I communications technology have aroused wide attention and gained rapid development. VANETs are expected to support a large spectrum of mobile distributed applications, which range from traffic alert dissemination and dynamic route planning to context-aware advertisement and file sharing. Considering the large number of nodes that participate in these networks and their high mobility, debates still exist about the feasibility of applications that use end-to-end multiceps communication. The main concern is whether the performance of VANETs routing protocols can satisfy the throughput and delay requirements of such applications. Many routing techniques have been designed in VANETs to tackle the limitations of the transmission packet delivery delay, packets being dropped, wasting bandwidth, mobility and security. However, the network communication is easily interrupted when the vehicle nodes are sparse due to the high speed mobility of vehicles. This situation can easily result in the decline of communication performance. To solve the problem, Delay/Interrupt Tolerant Networks (DTNs) [6, 7] came into being. These networks have the following characteristics: poor network stability, nodes in the network are moving at high speed, sparse distribution of cache and limited energy resources. Because these kinds of networks are not stable and reliable channel links, the message transmission between nodes may need to rely on the opportunity to meet.

At present, DTNs protocols are the most mature and fruitful part of DTNs research field. Protocols in DTNs are intended to achieve high delivery rates and low overhead ratios, get along with delivery latency due to the Store-Carry-Forward (SCF) routing mechanism [8, 9]. SCF concerns that contemporaneous end-to-end paths towards destination are unavailable. As shown in [10], although the single message replica forwarding protocol [11, 12] ensure low redundancy and reduces network costs, it also causes low transmission rates and high latency. These exiting protocols [13, 14, 15] aim at improving the delivery ratio as much as possible, which is usually achieved by copying packets of the message. Although a large number of replicated copies can increase the delivery rate of the message, it also increases energy and cache consumption. Based on the algorithm of replica replication strategy, multiple messages are propagated in the network, which improves the transmission rate of the algorithm. However, it also increases the routing overhead and even causes network congestion, which in turn affects the transmission performance of the message. In order to improve the performance of routing protocols, nodes usually select a better next hop according to some utility factors. Spray and forward Routing Protocol [16] based on Spray and Wait routing [17] improves the wait phase and calculates the meeting probability using the Markov location model. However, Spray and forward only considers the meeting possibility at a specific place without considering the possible meeting on the way. Zhenxi Sun and Yuebin Bai devise an algorithm based on Correlated Contact And Message Scheduling Policy (COMSP) [18] with the utility of messages and contacts taken into consideration. All the above protocols are not the mainly concerned about the resource efficiency and social attributes of network. Many mobile nodes such as smart-phones, tablets, EV, PCs and so on have limited energy resources. They use a large amount of resources to transmit and receive messages.

Routing protocols that take consideration of energy consumption of mobile nodes are also necessary in DTNs. Based on the situation of energy consumption, the researchers implement a series of protocols. The energy of the node will be consumed by related process [19], which results in performance degradation. Blindly sending packets without having utility selection strategy such as Epidemic routing protocol [20] leads to high routing overhead using a straightforward flooding method. Routing protocols [21, 22, 23] are inclined to restrict the number of copies to select relay nodes according to the historical information of node encounter. Energy Efficient Routing Protocol for Vehicular Ad hoc Networks (ADTNEER) [24] uses geographical angular region based routing to choose most suitable next hop thereby achieving proper network connectivity among nodes with minimum energy consumption. [25] considers nodes' remaining energy and available free buffer for receiving copies of messages to reduce network overhead and at the same time improve delivery ratio. Only a node with higher energy value than the sending node will receive a copy of the message and store it to send to other nodes or the destination node.

All the above existing DTNs routing protocols fail to have an integrative consideration of resource and social attributes. To solve these problems, we propose the Social Probability And Resource Effective (SPARE) protocol that considers the consumption efficiency of node resources, the Poisson distribution properties of the meeting frequency between nodes and the Irish distribution properties of the meeting waiting time between nodes. The key contributions of our study are summarized as follows:

- We develop a model to capture the resource consumption utilization, and use a mathematical formula to quantify residual resource effectiveness. The formula represents the ratio of the recent energy consumption to the total energy consumption, and the ratio of the recent cache consumption to the total cache consumption in each node.

- SPARE protocol improves the messages' delivery ratio based on the encounter probability and resource utility. We present that the meeting frequency between nodes obeys the Poisson process and give a method for calculating the Poisson parameter. The Resources-Probability Social Similarity Forwarding Strategy ($StiPes$) selects the candidate node according to the nodes' resource consumption utility, nodal historical encounter information and encounter probability among nodes.

- Considering that the frequency of encounter between nodes may be low, we propose the Connection Weight-Waiting Time Similarity Encounter Probability Utility ($SimPro$) based on the Connection Weight Similarity Utility ($SimUtil$) and the Waiting Time Encounter Probability Similarity Utility ($ProUtil$) to determine the message forwarding policy. We present that the meeting waiting time between nodes obeys the Irish distribution, and the Irish distribution parameter is as same as the Poisson parameter. In order to adapt to the dynamic changes of the network, we introduce the idea of graph theory to propose the $SimUtil$.

- We dynamically update the number of replicas generated by the source nodes according to the percentage of residual energy to the initial energy of nodes and the centrality of nodes, which avoids the high network overhead caused by the low energy and redundant copies. The method of dynamically controlling the message copies generated by the source nodes is different from the traditional algorithm for deleting messages.

The rest of this article is organized as follows: Section 2 describes the probability forwarding routings, the social networking protocol and the energy utility protocol. In Section 3, we propose the basic framework of SPARE, which considers the nodal resources effective consumption, encounter probability, nodal historical encounter information and the number of messages carried by nodes. In Section 4, we present the theoretical analysis and some important proofs of the algorithm. In Section 5, SPARE is simulated and analyzed based on the ONE simulation platform, and the simulation results are compared with the existing algorithms. At last, conclusion is made in Section 6.

2. Related Work

Epidemic Protocol [20] based on flooding strategy forwards message to any encountering nodes. Each host maintains a buffer containing messages. Upon meeting, the two nodes exchange summary vectors to determine which messages held by the other have not been seen before. This flooding-based method can guarantee the best delivery ratio, but with possible huge message overheads. In order to solve the problem, other works relay messages either based on probability or social utility metrics. Even if they can achieve a higher delivery than Epidemic, their performance is dramatically degraded in case of sparse network density. Instead, using redundant message copies has been widely investigated, with the following three main branches, based on probability, utility and energy strategy.

2.1. Probability Forwarding Protocol

Prophet [21] is based on Epidemic aiming at using knowledge obtained from past encounters with other nodes to optimize the packet delivery. When the node forwards the message, the node will select the node with the higher forwarding probability as the next hop. Therefore, the Prophet avoids the high overhead caused by the bulk copy of the message. However, Prophet does not control the copy of the message, and the probability calculation is simple, so Prophet performance still needs to be improved. There are a number of successful attempts of improving the performance of the Prophet algorithm. Delivery Probability Routing (DPR) [22] includes spray phase and wait phase. DPR updates delivery probability vector of nodes which can be used to decide message copies assignation and tactics of message transmission. In [26], the Predicted And Forward (PAF) based on Markov meeting time span prediction model is provided. The main idea of PAF is as follows. Each node stores locally a multidimensional array which records the meeting time spans between the node and every other node. Then the general range of the next meeting time span is predicted from the previous time span using Markov model, and the highest utility value is endowed to the node holding the shortest meeting time span with the destination node. During both spray phase and wait phase, the message is forwarded to the node of the highest utility value. Priority-enhanced PRoPHET (Pen-ProPHET) [27] performs a content based message filtering using Natural Language Processing (NLP) and a first level prioritization based on the outcomes of this filtering. A second level on-the-fly prioritization is done by using delivery predictability of forwarder nodes using the Prophet routing protocol.

2.2. Utility Forwarding Protocol

SimBet [28] adopts the 'betweenness' and similarity for routing decision. Compared to other centrality calculation methods, 'betweenness' is better to control the spread of the message. The drawback of Simbet is that it just considers the counts of encounters instead of their social relations between neighbors. Enhanced Spray and Wait (ESW) [29] method that sets the maximum replication number dynamically in accordance with the state of the network and selects appropriate relay destinations, thus facilitating increase in the replication of messages for which the replicas are not sufficiently disseminated in the network and constraining replication of messages having sufficient replicas. Encounter-Based Routing (EBR) [30] is a quota-based routing protocol which limits the number of replicas of any message in the system to minimize network resource usage. EBR makes routing decisions based on nodal encounter rates. Encounter-based Replication Routing (EBRR) [31] defines utility factors based on node history encounter information. It is divided into three phases: in Utility Replication Phase, EBRR balances replication redundancy according to the utility metric; in Conditional Replication, the algorithm takes the utility and message remaining lifetime into account to improve delivery performance; in Probabilistic Replication Phase, considering the worst case, the historical encounter information in relation to destination is unavailable due to rare encounter. But when the messages lifetime is insufficient, the message carrier still forwards the copy based on utility. This kind of routing decision cannot raise the delivery ratio effectively.

2.3. Energy Forwarding Protocol

Although the above two types of algorithms have shown some advantages in different aspects, they did not take into account the node energy limited situation. Next, we will introduce some protocols based on energy utility forwarding strategy. Social Energy Based Routing (SEBAR) [32] belongs to this category and uses the novel concept of social energy to quantify the social ability of a node to forward messages to others. The calculation of social energy considers social energy of encountering nodes and is in favor of the node with a higher social energy in its or the destinations social community. The drawback of SEBAR is that it only considers the centrality of the node by the connection frequency between nodes instead of the global social information. Energy-Efficient Copy Limit-Optimized protocol [33] based on the Box's complex method for epidemic routing is proposed, which is designed to determine the optimal copy limit in multiple communities. Community-Based Energy-Aware Routing Protocol (CBEAR) [34] based on the MSN network is dynamically divided into several different communities, transferring and sharing news in the intra-community and inter-community, considering node energy consumption rate and the encounter probability between nodes and each community to make routing decisions, avoiding some nodes' energy consumptives too fast to realize load balance between nodes. However, CBEAR has a high transmission delay. Resource-Efficient Routing Protocol Based on Historical Encounter Time Interval (RRPHETI) [35] exploits historical encounter time interval to measure social relations between nodes, as the more intimate the nodes are, the smaller the encounter time interval becomes. RRPHETI creates a model to capture the resource consumption behaviors in DTN, and utilizes the maximum likelihood method to estimate the parameters of delivery probability. Because the RRPHETI protocol does not limit the copies, the network overhead performance will become worse when the node density increases.

In this paper, we use the utilization efficiency of nodal energy and historical information to design the SPARE protocol. We tackle resource constraints issue in DTNs comprising the mobility of nodes.

3. Social Probability And Resource Effective Utilization Algorithm

Firstly, we give the encounter factors between nodes N_i and N_J, where N_i, N_j, encounter count $C_{i,j}$, encounter duration $D_{i,j}$ and Encounter period time $P_{i,j}$ are addressed. The lists of commonly used variables are defined in TABLE 1.

3.1. System Model

This paper assumes that the energy of all nodes is limited and the initial energy values are equal. Therefore, we first give the concept of the kinds of energy used in this paper. It includes the initialize energy E_{init}, the transmitting energy $E_{transmit}$ and the scanning energy $E_{scanning}$.

$$\begin{cases} E_{init} = a \\ E_{transmit} = e_t \times \sum_{i=1}^{m} S_i \\ E_{scanning} = e_s \times t_{scanning} \end{cases} \tag{1}$$

When transmitting one byte, e_t is the energy consumed by transmitting and when listening to the neighbor nodes for a second, e_s is the energy consumed by scanning. S_i is the packet size, $t_{scanning}$ is the time for scanning, and m is the number of packets transmitted.

In usual DTNs networks, nodes can be in four states: sleeping, listening, transmitting and receiving packets. If the current node does not handle the event, it will turn into sleeping state to save energy. If the node is in an active state and the message is forwarded, it consumes energy. After forwarding messages, the node has to maintain its remaining energy to judge whether there is enough residual energy for the next message transmission. Here, we analyze energy consumption status, the total energy consumption E_{consum} and the current remaining energy $E_{current}$ of nodes,

$$\begin{cases} E_{consum} = N \times \ E_{transmit} + \frac{T_{Past}}{t_{scanning}} \times \ E_{scanning} \\ E_{current} = E_{init} - E_{consum} \end{cases} \tag{2}$$

where N represents the number of messages forwarded by the node, and T_{Past} is the time that nodes experienced in the network.

Energy and buffer capacity are the key factors that affect the performance of DTNs. When energy is exhausted, the node will stop working, which will affect the whole network lifetime. Due to the limited buffer space, the node can not provide service well when buffer space is full, which may even causes network congestion. The model provides that when the energy or buffer space is about to be exhausted, it will no longer encourage the node to continue to participate in forwarding work. The resource of node has an important effect on message forwarding, and excellent nodes consume less resources to transmit more messages. In addition, the rate of decline in energy consumption is increasing. Based on these analysis, we define residual resource effectiveness function as shown in formula (3):

$$\varphi(E_i, E_c, S_i, S_c) = \frac{E_{current} - E_{last}}{E_{init} - E_{current}} + \frac{S_{current} - S_{last}}{S_{init} - S_{current}} \tag{3}$$

Table 1. List of Notations

$E_{current}$	Node currently available energy space
S_{last}	Available energy value for nodes before message exchange
S_{init}	Initialization buffer space for nodes
$S_{current}$	Node currently available buffer space
S_{last}	Available buffer space for nodes before message exchange
N_i	Message carrying node with destination information
N_j	Encountered node
N_d	Destination of message
M	A message carried by N_i
$D_{j,d}$	Encounter duration between N_j and N_d
$P_{j,d}$	Encounter period time between N_j and N_d
$StiUtil(j,d)$	Separation Time Interval Similarity Utility for node N_d, calculated by the historical meeting separation time interval utility information which is recorded in N_j
$PesUtil(i,d)$	Poisson Encounter Probability Similarity Utility for node N_d, calculated by the historical meeting Poisson probability utility information which is recorded in N_j
$StiPes(j,d)$	Resources-Probability Social Similarity Utility value for N_d, calculated by the historically encounter information and nodal resource consumption behaviors which is recorded in N_j
$SimUtil(i,d)$	Connection Weight Similarity Utility estimated for N_d, calculated by the common neighbors information and weight connection information which is recorded in N_j
$ProUtil(j,d)$	Waiting Time Encounter Probability Similarity Utility for node N_d, calculated by the historically encounter probability similarity during a period of time which is recorded in N_j
$SimPro(j,d)$	Connection Weight-Waiting Time Encounter Probability Similarity Utility value for node N_d, calculated by encounter calculated by waiting time encounter probability similarity and optimized weight betweenness which is recorded in N_j
$C_{j,d}$	Encounter count between N_j and N_d
m_t	Total number of M replicas which stored at N_i
H	Current encounter count
L_{init}	Initialized copy ticket of M
E_i	Current energy of node N_i
K	Total number of nodes in network

where $\frac{E_{current}-E_{last}}{E_{init}-E_{current}}$ denotes the ratio of the latest energy consumption to the total consumption, $\frac{S_{current}-S_{last}}{S_{init}-S_{current}}$ denotes the ratio of the latest cache consumption to the total consumption. The lower ratio of resources consumption are, the more available resources of nodes are.

3.2. Overview of Social Probability And Resource Effective Utilization Protocol

SPARE models DTN routing as a utility-driven resources utilization problem. A packet is routed by replicating it until a copy reaches the destination. The key question is: Given limited resources, how should packets be replicated in the network so as to effectively use a specified routing metric? SPARE derives a per-packet utility function from the routing metric. SPARE protocol mainly considers the four factors; the node's resources, historical meeting information of nodes, nodal encounter probability and the number of messages carried by nodes. The framework of SPARE is shown in Algorithm 1. SPARE has three core components: a resources probability similarity forwarding algorithm, a connection weight-waiting time similarity forwarding algorithm and a message management method. The resources probability similarity forwarding algorithm is used to determine which packets to replicate at a transfer opportunity given their utilities. The connection weight-waiting time similarity forwarding algorithm is used to forwarding messages to nodes with higher network importance. The message management method is used to dynamically manages the number of messages generated by the source node, which guaranteed delivery ratio while reducing network overhead.

Algorithm 1 Routing Strategy of SPARE

initialize the value of $E_{current}$ and L_{init}
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 $EC(i, j)$ **if** $EC(i, j)$ has a copy of M
        update $P(X(t + \triangle t) - X(t))$ **if** $(P(X(t + \triangle t) - X(t)) > 0)$
        if $EC(j, d)$ contains the information of N_j **then**
            update Encounter Probability Similarity Utility and Separation Time Interval Similarity Utility
            if $StiPes(i, d) > StiPes(j, d)$ and $L > 1$ **then**
                replicate $m_i \cdot \frac{EC_j \times E_j}{EC_j \times E_j + EC_i \times E_i}$ to N_j
            else if $P(X(t + \triangle t) - X(t)) > 0$ and $L = 1$ **then**
                forward M to N_j using single copy
                delete the copy from the buffer
            end if
        end if
        if $EC(j, d)$ does not contain the information of N_j **then**
            update Connected Weight Similarity Utility and Waiting Time Encounter Probability Similarity Utility
            if $Pro(X(t + \triangle t) - X(t)) > 0$ and $L > 1$ **then**
                if $SimPro(j, d) > SimPro(i, d)$ **then**
                    replicate $m_i \cdot \frac{EC_j \times E_j}{EC_j \times E_j + EC_i \times E_i}$ to N_j according to $EC(i, j)$
                    keep the rest of M in N_i
                else
                    replicate M to N_j anyway
                    keep the rest for M in N_i
                    delete the copy from the buffer
                end if
            else
                disconnect the connection
            end if
        end if
    end for
end for

SPARE protocol maintains an encounter 'probability' $EC(i, j)$ similars to the thought of literature [30].

$$EC(i, j) = \alpha \cdot CWC(i, j) + (1 - \alpha) \cdot EC(i, j)_{old} \tag{4}$$

$EC(i, j)$ represents the past rate that the node N_i encountered the node N_j, $CWC(i, j)$ represents the number of encounters between node N_i and node N_j. The encounter smoothing factor $\alpha \in [0,1]$ and the value of $EC(i, j) \in [0,1]$.

$EC(j, d) > EC(i, d)$ represents N_j has a higher probability to encounter node N_d than N_i. In SPARE, if $EC(j, d) > EC(i, d)$, N_j will have a higher potential to encounter N_d, then N_i will forward messages to N_j. Furthermore, we define an identifier S that is measured the importance of the node compared to the other nodes in the network. The value of S is determined according to the $EC(i, j)$ defined as shown in formula (5).

$$S = \begin{cases} 1 & \text{if the information of } EC(j, d) \text{ about the node } N_j \text{ exit and } EC(j, d) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The identifier $S = 1$ reflects that the node has contact with other nodes for the past period of time, and the message can be forwarded according to the social probability utility of the system. However, the identifier $S = 0$ indicates that the node has no contact with other nodes in the past time, that is, messages carried by the node are not likely to be forwarded.

3.3. Resources-Probability Social Similarity Forwarding Strategy

Next, we describe how SPARE Protocol can support specific metrics using an algorithm to infer utilities.

Metric 1: ***Separation Time Interval Similarity Utility***: We present SPARE based on efficiently spread limited copies of the same packet in the network and optimize the resource efficiency. We utilize historical encounter time interval to measure social relations between nodes. The smaller the encounter time interval is, the higher the encounter frequency is, that is, the more intimate the nodes are. We should consider $P_{i,j} - D_{i,j}$ instead of $P_{i,j}$. For $P_{i,j} - D_{i,j}$ is influenced by $P_{i,j}$ and $D_{i,j}$ at the same time, $P_{i,j} - D_{i,j}$ will have a low value when $P_{i,j}$ is low or $D_{i,j}$ is long. It means that 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. The number of encounters $C_{i,j}$ is decided by the average value of $P_{i,j} - D_{i,j}$, so different combinations of encounter durations and inter meeting times may result in the same encounter gap. Thus, we define the separation time interval factor ($Sti(i, j)$) as:

$$Sti(i, j) = \frac{T_{i,j}^{(C_{i,j}=1)} + \sum_{(C_{i,j}=2)}^{H} \left((P_{i,j} - D_{i,j})^{(C_{i,j})} \right)}{H} \tag{6}$$

The node N_i has a high willingness to transmit messages to the node N_j, which means that the value of $Sti(i, j)$ should be as small as possible. We define the Separation Time Interval Similarity Utility ($StiUtil(i, d)$) between node N_i and node N_d as shown in formula (7).

$$StiUtil(i, d) = \frac{\frac{1}{Sti(i,d)}}{\frac{1}{Sti(i,d)} + \frac{1}{Sti(j,d)}} = \frac{Sti(j, d)}{Sti(j, d) + Sti(i, d)} \tag{7}$$

Metric 2: ***Poisson Encounter Probability Similarity Utility***: If $Sti(i, d) > Sti(j, d)$, N_j will have a higher potential to encounter N_d, Due to the dynamic changes of the network topology, node N_i and N_j may meet frequently in recent time. After a period of time, the frequency of encounters between node N_i and node N_j maybe change. In order to better explore this dynamic change of the connection frequency between nodes. We make the same assumption as [36] that most nodes' degrees are subject to power-law distributions. Further, we consider the number of encounters within a period of time T between nodes obeys the Poisson process, that is, $P(X_{i,j}(t + T) - X_{i,j}(t) = n) = \frac{e^{-\overline{X_{i,j}}T}}{n!}(\overline{X_{i,j}}T)^n$. The specific analysis is found in section 4.1.

We define the Poisson Encounter Probability Similarity Utility ($PesUtil(i, d)$) between node N_i and node N_d in formula (8).

$$PesUtil(i, d) = \frac{P(i, d)}{P(i, d) + P(j, d)} \tag{8}$$

According to Metric 1, Metric 2 and the resource consumption model, we consider to use the historical separation time interval similarity, the resource consumption behaviors and Poisson process encounter similarity between nodes to determine the message forwarding policy. The Resources-Probability Social Similarity Utility value ($StiPes$) is between 0 and 1. Selecting which node as the best carrier for the message becomes a multiple attribute decision problem, where we wish to select the node that provides the maximum utility for carrying the message. This is achieved using a pairwise comparison matrix on the normalized relative weights of the attributes. Therefore, we define the $StiPes(i, d)$ between node N_i and N_d as shown in formula (9).

$$StiPes(i, d) = \frac{\gamma StiUtil(i, d) + (1 - \gamma)PesUtil(i, d)}{\varphi(E_i, E_c, S_i, S_c)} \tag{9}$$

Where γ is an adjustable parameter and the value of $\gamma \in [0, 1]$. The value of γ is adjusted according to the importance of $StiUtil$ and $PesUtil$.

Based on these routing mechanisms, this algorithm is recorded as previously encountered node utility metric and compared with the node upcoming. In view of this, the condition (5) focused more on practical indicators, only the nodes between the current node and previously encountered node, rather than comparison between the current node and the message carrier. When the value of S is 1 in (5), we determine the forwarding order of the message according to the utility value in formula (9). When the node N_i encounters the node N_j, the packet forwarding policy can be expressed as follows:

- Step 1: Update the value of $EC(j, d)$ and $StiPes(i, d)$. Each node has a utility value to every other node which is enlarged by their frequent meetings and more residual resource, and vice versa.
- Step 2: When two nodes encounter, they record sequence value of encounter time interval, then exchange their utility values and delivery probabilities.
- Step 3: All data packets queued in one node are ordered on a first in first out (FIFO) basis to be sent to others.
- Step 4: If the queue in node N_i is not empty, for a packet in the queue which encounters node N_j exactly, it should be sent to node N_j. Afterwards, it should be deleted from the queue and buffer. If $StiPes(i, d) < StiPes(j, d)$, then node N_i forwards the packet to the node N_j. That is, sender node only forwards message copies to nodes which have the greater utility values.
- Step 5: Update Encounter probability between nodes, if $EC(j, d)$ doesn't contain the information of N_j or the $P(i, j) = 0$, then jump into the Connection Weight-Waiting Time Similarity Forwarding Utility Strategy; otherwise, jump into the Step 6.
- Step 6: Repeat the previous steps until the queue is empty, or the connection is disconnected.

In order to reduce the redundancy of the network and utilize the energy of nodes effectively, we spread the copies according to the $EC(i, j)$. For two nodes N_i and N_j, for every message M_i, node N_i sends $m_i \cdot \frac{\sum_{i=1}^{K} EC(i,j)\times E_j}{\sum_{i=1}^{K} EC(i,j)\times E_j + \sum_{j=1}^{K} EC(i,j)\times E_i}$ message copies of M_i to node N_j.

3.4. Connection Weight-Waiting Time Similarity Forwarding Strategy

When the value of S is 0 in formula (5), that is, $EC(j, d)$ does not contain information about node N_j, which means the node N_j may not encounter node N_d for a period of past time. It is necessary to transfer the message to the destination node before the end of the message life. Furthermore, it results in a longer delay in the delivery, even reduces the probability of delivery due to the short lifetime of the message. To solve these problems, we integrate the waiting time similarity between two nodes and connected weighted similarity to determine the forwarding order of the message, which is discussed as follows:

Metric 1: ***Connection Weight Similarity Utility***: Researchers have found that nodes clustered together tend to exhibit similarities in the social networks, that is, the two nodes have more common points, which means that the two nodes may have more common neighbors. The similarity between nodes has a positive effect on the transmission performance of the algorithm. Liben-Nowell and Kleinberg explored the common neighbors metric between two nodes in order to predict future collaborations on an author database [37]. The probability of a future collaboration $Sim(i, j)$ between authors N_i and N_j was calculated as $Sim(i, j) = |N(i) \bigcap N(j)|$, where $N(i)$ and $N(j)$ respectively represent the number of neighbors of node N_i and N_j. The quantity $Sim(i, j)$ can be described as a basic measure of similarity between N_i and N_j. As long as there is a connection between nodes, there is an edge between the corresponding nodes in the contact graph. The construction method of this connection diagram can not reflect the dynamic changes of the number of connections between the nodes with time. So, only $Sim(i, j)$ is inadequate. In this paper, we introduce a weighted connection method between nodes to further modify similarity. The weighted connection diagram of the node is shown in Fig. 1.

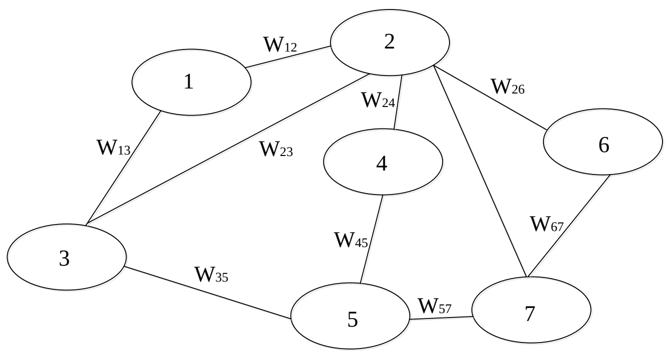


Fig. 1. Node Weight Connection Diagram

In Fig 1, the edge weight W_{13} represents the number of encounters between node 1 and node 3, that is, the connection intensity between node 1 and node 3. The weight constantly changes over time, for this reason, a new method of calculating the connection weight between nodes is defined in formula (10).

$$\overline{Sim(i, j)} = \frac{|N(i) \bigcap N(j)|}{\sqrt{\sum_{j=1}^{K} W_{ij} \times \sum_{i=1}^{K} W_{ij}}} \tag{10}$$

where W_{ij} is the number of connections between N_i and N_j. W_{ij} can better reflect the dynamic characteristics of the social connections among nodes, we define the Connection Weight Similarity Utility ($SimUtil(i, d)$) as shown in formula (11).

$$SimUtil(i, d) = \frac{\overline{Sim(i, d)}}{\overline{Sim(i, d)} + \overline{Sim(j, d)}} \tag{11}$$

Metric 2: ***Waiting Time Encounter Probability Similarity Utility***: Suppose that $X_{i,j}$ represents the number of event A (node i meets with node j) occurs. Denote the time of the first contact for any node in DTNs by T_1. $W_i = \sum_{i=1}^{n} T_i, i = 1, 2, 3, \ldots,$ W_i represents the ith time of event A occurred. The waiting time for the nth meeting between any two nodes obeys the Irish distribution with the parameter λt. The specific analysis is found in section 4.2.

Similarly, we define the $ProUtil(i, d)$ between N_i and N_d as shown in formula (12).

$$ProUtil(i, d) = \frac{Pro(i, d)}{Pro(i, d) + Pro(j, d)} \tag{12}$$

Based on the above analysis, we integrate the connection weight similarity utility and waiting time similarity utility to define the Connection Weight-Waiting Time Encounter Probability Similarity Utility

($SimPro$): Similar to formula (9), we define the $SimPro(i,d)$ between node N_i and N_d as shown in formula (13).

$$SimPro(i,d) = \delta SimUtil(i,d) + (1-\delta)ProUtil(i,d) \quad (13)$$

where δ is an adjustable parameter and the value of $\delta \in [0,1]$. The value of δ is adjusted according to the importance of $SimUtil$ and $ProUtil$.

Next, this portion describes the messages forwarding process based on the Connection Weight-Waiting Time Similarity Strategy. The utility forwarding policy shows the message exchange process between node N_i and N_j.

- Step 1: Upon reception of a Hello message node N_j verifies that node N_i is a new neighbour. Under these circumstances, any messages destined for node N_i are delivered and an encounter request is sent.
- Step 2: Then, node N_i replies with a list of nodes it has encountered. This list of contacts is then used to update the arrival waiting time similarity value on node N_j and the connected weight similarity value.
- Step 3: Then, the two nodes exchange respective summary vector, which contains a list of destination nodes they are currently carrying messages for along with their own $SimUtil$ value and $ProUtil$ value for each destination.
- Step 4: For each destination in the summary vector, node N_j calculates the $SimPro$ value between node N_i and node N_j.
- Step 5: If node N_j has a higher utility for a given destination, the destination is added to a vector of destinations for which messages are requested.
- Step 6: Node N_j sends the message request list to node N_i when all destinations in the summary vector has been compared. Then, node N_i removes all messages destined for the destination node from its queue and forwards them to node N_j. Upon receiving a transfer message from node N_i the message is added to the message queue of node N_j.
- Step 7: Repeat the previous steps until the queue is empty, or the connection is disconnected.

3.5. Message Management Method

The buffer has a significant effect on the number of messages carried by the node. Node' occupied cache is too large can easily lead to the overflow of the messages. Under these circumstances, the copies carried by the node are likely to be discarded and can not be successfully sent to the destination node, which will result in a lower delivery ratio. In addition, the newly generated message also needs to take up some space, in which case network would have to delete or discard some of the messages to avoid the congestion. However, the high delivery ratio is often achieved by sacrificing a large number of network resources, which easily leads to a high network overhead. Therefore, it is particularly important to improve the performance of network delivery while avoiding high network overhead. Several works have been done in the context of buffer management and message scheduling. [38] defines a weight of node according to the message's properties instance message size, remaining time-to-live, hop count, and replication count. It splits messages into High weight message list (HWML) and low weight message list (LWML) according to the weight criterion, then calls the drop event handler to drop the messages from the HWML. [39] by counting the drop time and the number of drop times for each message arriving at their destination node in the message history list, to decide the order of dropping the message. These exiting works mainly delete messages or limit the number of message copies based on some strategy to avoid congestion. These protocols do not take into account the characteristics of the source node itself.

In this paper, we consider that the source nodes dynamically generate messages by current energy of nodes, encounter count H with other nodes and available cache of nodes to balance network performance rather than just considering to delete messages. A Node can forward more messages, if it connects more frequently with other nodes. A node need to consume at least one unit of energy per message transmitted in the network. The more energy the node has, the more messages the node can forward. In addition, we rely on the calculation results of formula (14) to dynamically adjust the the number of message replicas due to the limited buffer of nodes.

$$L = (\frac{E_{current}}{E_{initt}} \times L_{init} + H \times \frac{S_{current}}{S_{init}}) \tag{14}$$

If the node has no contact with other nodes in the past period of time, the value of formula (14) will be zero, which means nodes don't generate messages. However, because of the dynamic changes of the network and the continuous movement of the nodes, the frequency of connection between nodes is constantly changing. Zero message easily causes no message to be forwarded when the source node meets with other nodes next time. To avoid this situation, we define a minimum number $\frac{L_{init}}{2}$ of message generation replicas. Finally, we dynamically update the number of messages generated by the source node according to formula (15).

$$L_{new} = Max[L, \frac{L_{init}}{2}] = Max[(\frac{E_{current}}{E_{initt}} \times L_{init} + H \times \frac{S_{current}}{S_{init}}), \frac{L_{init}}{2}] \tag{15}$$

As the residual energy of the node decreases, the value of L_{new} decreases gradually, which avoids the high network overhead thanks to the fact that the extra messages can not be sent out due to low energy.

4. Theoretical Analysis

4.1. Analysis of Deliver Probability

Although a large number of copies of the message can improve delivery rate, the blind transmission of messages also causes unnecessary resource consumption, even causes congestion. In order to make better use of network resources, we propose the encounter predictability between the nodes as a criterion to determine the next hop of message forwarding.

Power-law distribution [36] refers to the fact that most nodes have a low degree and only a small number of nodes have a high degree. The node degree is the simplest measure to reflect the centrality of the node. In this paper, self-centered network analysis is used to study node degree, which considers the number of nodes that each node encounters in the past period of time T as the centrality of the node. For example, the centrality of node N_i is shown $C_i = \sum_{j=1}^{N} e_T(i, j)$, where $e_T(i, j)$ indicates whether the node N_i and the node N_j meet in the time period T or not. If encountered, $e_T(i, j) = 1$, otherwise $e_T(i, j) = 0$. The larger the C value is, the higher the node' degree is. The node with higher degree has the tendency to connect with the node with higher degrees, and the node with lower degrees has the tendency to connect with the node with lower degrees. In DTNs, the basic event of interest (i.e. pairwise node contact) is a rare one due to sparsity, nevertheless it occurs.

Let the random variable $X_{i,j}(t)$ be the cumulative number of contacts of a node pair N_i and N_j at time t. And we assume that any two contacts between N_i and N_j are independent from each other. Hence, $X_{i,j}(t)$ is a stochastic process with independent and stationary increments, i.e., for any $0 \le t_1 < t_2 < \cdots < t_n$, $X_{i,j}(t_2) - X_{i,j}(t_1)$, $X_{i,j}(t_3) - X_{i,j}(t_2)$, $\ldots$, $X_{i,j}(t_n) - X_{i,j}(t_{n-1})$ are all independent random variables. $X_{i,j}(t)$ is therefore modelled as a homogeneous Poisson process. For any $t > 0$ and interval of length t_0, the number of contacts $X_{i,j}(t + t_0) - X_{i,j}(t)$ between node pair (N_i, N_j) during time t_0 follows Poisson distribution $P(\lambda t_0)$, i.e., Its density function can be described as formula (16):

$$P(X_{i,j}(t + t_0) - X_{i,j}(t) = n) = \frac{e^{-\lambda t_0}}{n!}(\lambda t_0)^n \tag{16}$$

λ is the Poisson parameter. [40] shows that it's reasonable to model DTNs as Poisson processes from the social network perspective, and [41, 42, 43] also design the DTN routing protocols based on the same assumptions. However, above literatures do not determine the evaluation of Poisson parameter λ. Here, we derive the λ by using maximum likelihood estimation method.

Deduction: The likelihood for a sample x_1, x_2, x_3..., x_n in T can be expressed as:

$$\begin{aligned} L(X_1, X_2, X_3, \ldots, X_n; \lambda) &= \prod_{i=1}^{n} \frac{\lambda^{X_i}}{X_i!} e^{-\lambda} \\ &= e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{X_i}}{X_i!} \end{aligned} \tag{17}$$

The log-likelihood is written as:

$$\ln L = -n\lambda + \sum_{i=1}^{n} (X_i \ln \lambda - \ln X_i!) \tag{18}$$

Then derivation on the left and right sides of the formula (18) at the same time, we can get

$$\frac{d \ln L}{d\lambda} = -n + \sum_{i=1}^{n} \frac{X_i}{\lambda} \tag{19}$$

Let

$$\frac{d \ln L}{d\lambda} = 0 \tag{20}$$

According to formula (19) and (20), we can calculate

$$\lambda = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X} \tag{21}$$

The maximum likelihood estimator of λ is:

$$\widetilde{\lambda} = \overline{X} \tag{22}$$

Let the formula (22) substitute for the formula (16), and the probability of the encounter number n between the node N_i and the node N_j for a period of time T can be expressed as formula (23):

$$P(X_{i,j}(t+T) - X_{i,j}(t) = n) = \frac{e^{-\overline{X_{i,j}}T}}{n!} (\overline{X_{i,j}}T)^n \tag{23}$$

Lemma 1: Node N_i and node N_j maybe encounter once, two times, or more for the past period of T. We calculate the probability of meeting at least one time between the two nodes is an exponential growth function with the increase of $\overline{X_{i,j}}$.

Proof:

$$P(X_{i,j}(t+T) - X_{i,j}(t) \geq 1) = \sum_{k=1}^{n} \frac{e^{-\overline{X_{i,j}}T}}{n!} (\overline{X_{i,j}}T)^n = 1 - P(X_{i,j}(t+T) - X_{i,j}(t) = 0) = 1 - e^{-\overline{X_{i,j}}T} \tag{24}$$

It is easy to see that with the increase of the independent variable $\overline{X_{i,j}}$, the value of $P(X_{i,j}(t+T) - X_{i,j}(t) \geq 1)$ increases. The greater the average encounter count $\overline{X_{i,j}}$ between node N_i and node N_j is in a period of T, the greater the $P(X_{i,j}(t+T) - X_{i,j}(t))$ value is. The conclusion coincides with the Poisson process we assume, which shows that it is reasonable that we use the Poisson process to fit the probability of the meeting frequency between the nodes.

Lemma 2: Given a length of time interval T, the probability of meeting between two nodes is similar to the two function distribution of the number of meeting count n. Now, since a Poisson process possesses stationary and independent increments, it seems reasonable that each interval in $[0, T]$ of equal length should have the same probability of containing the event. In other words, the time of the event should be uniformly distributed over $[0, T]$. For $0 \le s \le T$, the probability of meeting n times between node N_i and node N_j within time length s is proportional to $(\frac{s}{T})^n$.

Proof:

$$\begin{aligned}
P(T_1 < s | X_{i,j}(t+T) - X_{i,j}(t) = n) &= \frac{P(T_1 < s, X_{i,j}(t+T) - X_{i,j}(t) = n)}{P(X_{i,j}(t+T) - X_{i,j}(t) = n)} \\
&= \frac{P(X_{i,j}(t+s) - X_{i,j}(t) = n, X_{i,j}(t+T) - X_{i,j}(t+s) = 0)}{P(X_{i,j}(t+T) - X_{i,j}(t) = n)} \\
&= \frac{P(X_{i,j}(t+s) - X_{i,j}(t) = n) P(X_{i,j}(t+T) - X_{i,j}(t+s) = 0)}{P(X_{i,j}(t+T) - X_{i,j}(t) = n)} \\
&= \frac{\frac{e^{-\overline{X_{i,j}}s}}{n!}(\overline{X_{i,j}}s)^n e^{-\overline{X_{i,j}}(T-s)}}{\frac{e^{-\overline{X_{i,j}}T}}{n!}(\overline{X_{i,j}}T)^n} \\
&= (\frac{s}{T})^n
\end{aligned} \tag{25}$$

We can see that the probability of meeting n times between nodes for a period of time T is more reflected the dynamic changes of the network over time than the probability of meeting at least once through formula (24). Thus, we choose that nodes meet n times instead of at least one time to evaluate the possibility of the meeting between the nodes within a period of time T. The detailed theoretical analysis as ***Lemma 3***.

Lemma 3: We use formula (8) to evaluate the Poisson encounter probability similarity between node N_i and node N_j, the probability similarity between the nodes is related to the parameter λ and the number of encounter count of nodes in time T.

Deduction:

$$\begin{aligned}
PesUtil(i,d) &= \frac{P(i,d)}{P(i,d) + P(j,d)} \\
&= \frac{P(X_{i,d}(t+T) - X_{i,d}(t) = n)}{P(X_{i,d}(t+T) - X_{i,d}(t) = n) + P(X_{j,d}(t+T) - X_{j,d}(t) = k)} \\
&= \frac{\frac{e^{-\overline{X_{i,d}}T}}{n!}(\overline{X_{i,d}}T)^n}{\frac{e^{-\overline{X_{i,d}}T}}{n!}(\overline{X_{i,d}}T)^n + \frac{e^{-\overline{X_{j,d}}T}}{k!}(\overline{X_{j,d}}T)^k} \\
&= \frac{e^{-\overline{X_{i,d}}T}\overline{X_{i,d}}^n}{e^{-\overline{X_{i,d}}T}\overline{X_{i,d}}^n + \frac{n!}{k!}e^{-\overline{X_{j,d}}T}\overline{X_{j,d}}^k} \\
&= \frac{1}{1 + \frac{n!}{k!}e^{(\overline{X_{i,d}} - \overline{X_{j,d}})T}\frac{\overline{X_{j,d}}^n}{\overline{X_{i,d}}^k}}
\end{aligned} \tag{26}$$

Under the assumption of $n = k$ in formula (26), we get $PesUtil(i,d) = \frac{1}{1+e^{(\overline{X_{i,d}}-\overline{X_{j,d}})T}(\frac{\overline{X_{j,d}}}{\overline{X_{i,d}}})^n}$. Similarly, we can get $PesUtil(j,d) = \frac{1}{1+e^{(\overline{X_{j,d}}-\overline{X_{i,d}})T}(\frac{\overline{X_{i,d}}}{\overline{X_{j,d}}})^n}$.

4.2. Analysis of Waiting Time Probability

The sequence $T_i, i = 1, 2, ..$ is called the sequence of inter-encounter arrival times, the schematic diagram of sequence as shown in the Fig. 2.

In Fig. 2, $W_n = \sum_{i=1}^{n} T_i, n \ge 1$, W_n denotes the time distribution of the nth event A arrival. W_n is a waiting time sequence corresponding to the Poisson process $X_t, t \ge 0$.

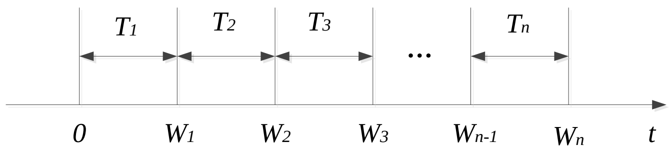


Fig. 2. Schematic diagram of time sequence

Theorem: The meeting frequency between nodes obeys the Poisson process with the parameter of λ, so the waiting time for the nth meeting of the two nodes obeys the Irish distribution with the parameter λ.

Proof: Note that the nth event occurs before the time t, and the event has occurred at least n times before the time t, that is, $X_t \geq n \Leftrightarrow W_n \leqslant t$.

So,

$$P(W_n \leqslant t) = P(X_t \geq n) = 1 - P(X_t < n) = [1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)k}{k!}]u(t) \tag{27}$$

Formula (27) shows that W_n obeys the Irish distribution with parameters λ and n (also known as Γ distribution). The probability of two nodes meeting at least once between before the time t can be expressed as $P(W_1 \leqslant t) = P(X_t \geq 1) = 1 - P(X_t < 1) = 1 - e^{-\lambda t}$ calculated by formula (27). The result is the same as that calculated by formula (24).

According to formula (27), the probability function of the Irish distribution is shown in formula (28):

$$F_{W_n}(t) = P(W_n \leqslant t) = [1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)k}{k!}]u(t) \tag{28}$$

Then derivation on both sides of formula (28) at the same time, we get the Irish probability density function as shown in formula (29).

$$f_{W_n}(t) = \frac{dF_{W_n}(t)}{dt} = \lambda e^{-\lambda t} \frac{\lambda t}{(n-1)!} u(t) \tag{29}$$

Suppose there is no connection for the destination with the node N_j before the time t, the Irish probability $Pro(X) = P(Wn < t|X_s = 0)$ is stated as when the event X does not happen before the time s with event X happening n times before the time t.

$$\begin{aligned} Pro(X) =& P(W_n(X) \leqslant t|X_s = 0) \\ &= P(W_n \leqslant t|X_s = 0) \\ &= \frac{P(X_t \geq n, X_s = 0)}{P(X_s = 0)} \\ &= \frac{P(X(t) - X(s) = n, X_s = 0)}{P(X_s = 0)} \\ &= \frac{P((X(t) - X(s) = n)P(X_s = 0)}{P(X_s = 0)} \\ &= P((X(t) - X(s) = n) \\ &= \frac{e^{-\lambda(t-s)}}{n!}[\lambda(t-s)]^n \end{aligned} \tag{30}$$

We find that the result of formula (30) has an important relationship with the value of time difference

$(t-s)$. If $EC(j,d)$ does not contain information about the node N_j, that is, node N_j has not been encountered the node N_d before the time s, in this case, we select the Ireland distributes probability to determine the message forwarding policy.

Lemma 4: When $\lambda = \frac{n}{t-s}$, the value of probability $P(W_n \leqslant t|X_s = 0)$ is maximal by formula (30). And when n takes 1, the probability maximum value of $P(W_n \leqslant t|X_s = 0)$ is the constant $\frac{1}{e}$.

Proof:

$$\begin{aligned}\frac{dP(W_n \leqslant t|X_s = 0)}{d\lambda} &= \frac{d\frac{e^{-\lambda(t-s)}}{n!}[\lambda(t-s)]^n}{d\lambda} \\ &= \frac{n\lambda^{n-1}(t-s)^n e^{\lambda(t-s)} - \lambda^n(t-s)^{n+1}e^{\lambda(t-s)}}{n!e^{2\lambda(t-s)}} \\ &= \frac{\lambda^{n-1}(t-s)^n[n-\lambda(t-s)]}{n!e^{\lambda(t-s)}}\end{aligned} \tag{31}$$

Let

$$\frac{dP(W_n \leqslant t|X_s = 0)}{d\lambda} = 0 \Longrightarrow \lambda = \frac{n}{t-s} \tag{32}$$

If $\lambda \in (0, \frac{n}{t-s})$, $\frac{dP(W_n \leqslant t|X_s=0)}{d\lambda} > 0$, the value of $Pro(W_n \leqslant t|X_s = 0)$ increases; if $\lambda \in (\frac{n}{t-s}, \infty)$, $\frac{dP(W_n \leqslant t|X_s=0)}{d\lambda} < 0$, the value of $P(W_n \leqslant t|X_s = 0)$ reduces. This shows the $P(W_n \leqslant t|X_s = 0)$ takes the maximum when $\lambda = \frac{n}{t-s}$. We easily calculate the maximum value is $P(W_1 \leqslant t|X_s = 0, \lambda = \frac{n}{t-s}) = \frac{1}{e}$ when n takes 1.

Lemma 5: We use formula (12) to evaluate the arrival waiting time similarity between node N_i and node N_j, the probability similarity between the nodes is related to the parameter λ, the encounter count between the two nodes and the time difference $(t-s)$.

Deduction:

$$\begin{aligned}ProUtil(i,d) &= \frac{Pro(i,d)}{Pro(i,d) + Pro(j,d)} \\ &= \frac{P(W_n(i,d) \leqslant t|X_s = 0)}{P(W_n(i,d) \leqslant t|X_s = 0) + P(W_k(j,d) \leqslant t|X_s = 0)} \\ &= \frac{\frac{e^{-\lambda_{i,d}(t-s)}}{n!}[\lambda_{i,d}(t-s)]^n}{\frac{e^{-\lambda_{i,d}(t-s)}}{n!}[\lambda_{i,d}(t-s)]^n + \frac{e^{-\lambda_{j,d}(t-s)}}{k!}[\lambda_{j,d}(t-s)]^k} \\ &= \frac{[\lambda_{i,d}(t-s)]^n}{[\lambda_{i,d}(t-s)]^n + e^{(\lambda_{i,d}-\lambda_{j,d})(t-s)}\frac{n!}{k!}[\lambda_{j,d}(t-s)]^k} \\ &= \frac{1}{1 + e^{(\lambda_{i,d}-\lambda_{j,d})(t-s)}\frac{n!}{k!}\frac{[\lambda_{j,d}(t-s)]^k}{[\lambda_{i,d}(t-s)]^n}}\end{aligned} \tag{33}$$

Under the assumption of $n = k$ in formula (33), we get $ProUtil(i,d) = \frac{1}{1+e^{(\lambda_{i,d}-\lambda_{j,d})(t-s)}(\frac{\lambda_{j,d}}{\lambda_{i,d}})^n}$. Similarly, we can get $ProUtil(j,d) = \frac{1}{1+e^{(\lambda_{j,d}-\lambda_{i,d})(t-s)}(\frac{\lambda_{i,d}}{\lambda_{j,d}})^n}$.

5. PERFORMANCE EVALUATION

5.1. Experimental setup

In order to study the performance of the protocol proposed in the paper and get reliable data, we use the popular DTNs simulator namely Opportunistic Network Environment (ONE) [44] simulator. ONE is a simulation platform developed by the Nokia Research Center in Finland, funded by the SINDTN and CATDTN projects. ONE is an open source simulation platform based on the Java language. ONE simulation platform supports a variety of mobile models and opportunities network routing algorithm, with a graphical user interface, rich report type. The simulation scenario applies the map with $4500 \times 3400m^2$ area as shown in Fig. 3 in ONE.

Fig. 3. Illustration of Scenario

,

The scheme depends on its current location and moving speed. The communication technique is configured as 4Mbit/t bandwidth and $50m$ transmission range. We use the shortest path map based movement model. Referring to [45], we set the initial number of copies of the network is 6. The default buffer space is limited to 40MB. We assume there are light weight vehicles in the network, and the network is sparse and highly dynamic, we set the number of vehicle nodes to 90. The parameters in the simulations are shown in Table 2.

In order to be based on the same energy consumption conditions, we further import energy models for all evaluated route schemes to show their energy consumption. As shown in Table 2, the initial energy of each node is fixed at $20000J$, the energy consumed by the transmission is fixed at $20J$ per packet, the scanning energy of each scan is $5J$, and the interface is scanned every $100s$. The main purpose of our assessment of energy consumption is to show the fairness of the routing scheme. Here, only the energy consumption is related to the number of nodes relay messages. Good fairness means that the resources, such as the energy used by each node to relay messages, are equal.

Table 2. NETWORK AND PROTOCOLS PARAMETERS

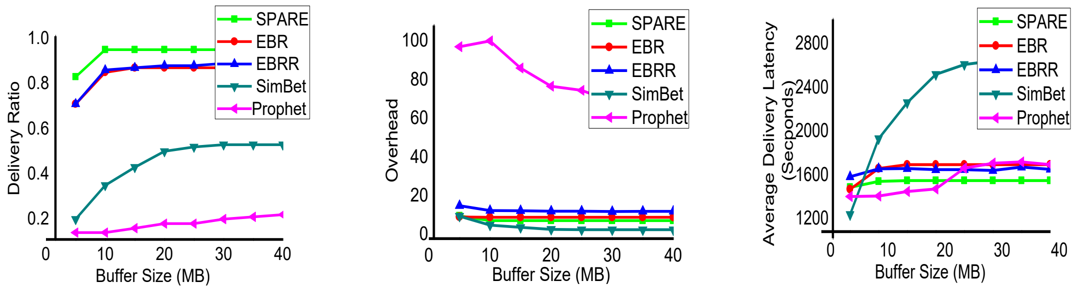
Protocol	Parameter	Value
All	Scenario Time	43200s
	E_{init}	$20000J$
	$E_{transmit}$	$20J$
	$E_{scanning}$	$5J$
	$t_{scanning}$	$100s$
	$Nodes$	90
	L	6
SPARE	α	0.85
	β	0.25
	$\gamma = \delta$	0.5

Each node records all the received message copies to avoid receiving the same message copies. We research the impact of the buffer size, message lifetime, the node density, the message generation interval, the speed of nodes and the initial energy on the performance of these protocols: Prophet [21], EBR [30], EBRR [31], SimBet [28] and SPARE. Prophet selects the node with a higher forwarding probability as the next hop. Both EBR and EBRR protocol use the nodes' historical encounter information to calculate the utility value, according to the utility rank for message forwarding. SimBet exploits the exchange of pre-estimated 'betweenness' centrality metrics and locally determined social 'similarity' to the destinations.

The performance measures considered are: In general, we use three factors to evaluate the performance of the five algorithms: Delivery Ratio, Overhead and Average Delivery Latency.

- Delivery Ratio: The ratio of successful messages to the number of messages generated by the source node.
- Overhead: The ratio of the number of failed messages to the number of successful delivery.
- Average Delivery Latency: The time duration between the messages generation and their delivery.

5.2. *Influence of Buffer Size*



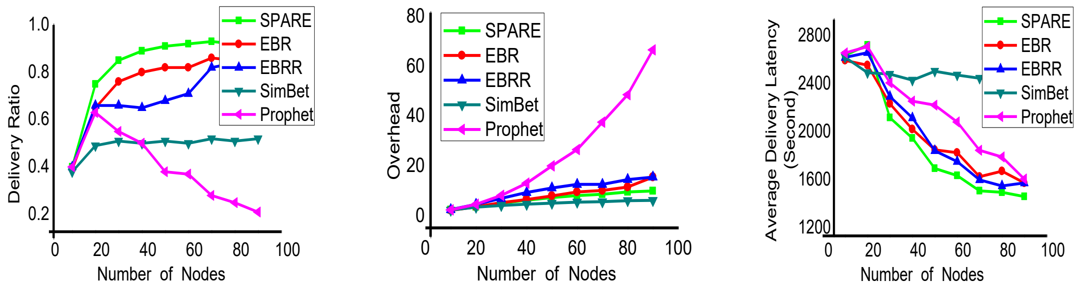
(a) Delivery Ratio (b) Overhead (c) Average Delivery Latency

Fig. 4. Influence of Buffer Size

In Fig. 4(a), 4(b) and 4(c), SPARE is not sensitive to the change of buffer size and keeps a stable performance compared with other algorithms. SPARE performs a little better than the others in most of the time in the same simulation setting. The buffer has a great influence on the Prophet becaucse Prophet doesn't limit the number of copies for replication. As the number of messages generated by the buffer increases, the delivery ratio of the all protocols increase. SimBet in the average delivery latency performance aspect is the worst, and SimBet' delivery ratio is low. SPARE is stable in terms of delivery rate and routing overhead, and due to several other algorithms. When the network is in a harsh condition like the message lifetime maintaining in a low level, most of the algorithms cannot keep a good performance, but SPARE maintains a high level and finishes the work efficiently.

5.3. *Influence of Network Density*

In Fig. 5(a), 5(b) and 5(c), we obverse that the number of nodes have a significant effect on these algorithms. With the number of nodes increase, the message copies generated also increases in the network, which results in all protocols' delivery ratio increasing. However, the Prophet' delivery ratio decreases after the number of nodes more than 20. The reason is that with the increase in the number of nodes, the network will produce a large number of copies of the message. Prophet does not limit the number of message replicas and the cache is limited, so a large number of redundant copies are discarded due to buffer overflows, reducing delivery ratio. SPARE is efficient in a sparse network density by achieving the highest delivery ratio, and the lowest average delivery latency. The reason is that SPARE relays messages to the next hop having a high encounter possibility with each other in sparse networks. All the algorithms benefit from the increased network density by achieving the decreased average delivery latency. In particular, we observe that Prophet does not adequately utilize the increased network density by suffering from the least decrease with respecting to this performance metric. The reason is that the increased network density results in contention for message transmission, which is not properly considered by Prophet.

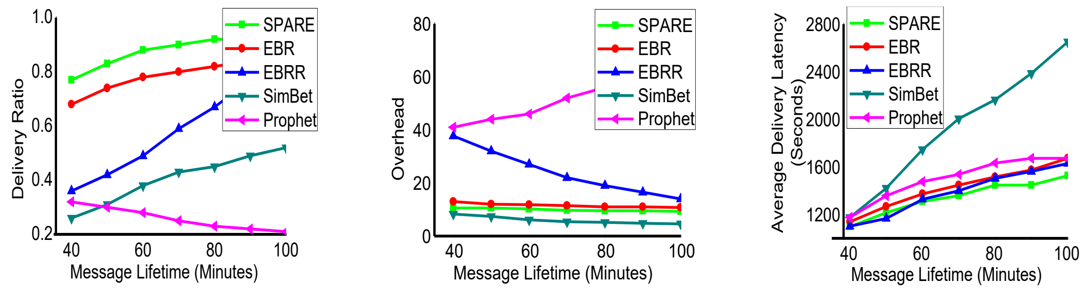


(a) Delivery Ratio (b) Overhead (c) Average Delivery Latency

Fig. 5. Influence of Network Density

5.4. Influence of Message lifetime

In Fig. 6(a), 6(b) and 6(c), we observe that with the increase in message lifetime, a copy of the message in the network has enough time to transmit, which causes more messages to reaching the destination node and improve the delivery ratio of the message. However, Prophet shows a downward trend. Because with the increase of the message lifetime, the total number of messages in the network for a certain period of time increases rapidly, and the buffer is limited to cause network congestion, a large number of message copy loss and retransmission led to a sharp decline in protocol performance. SPARE achieves the highest delivery ratio and the lowest average delivery latency, even in a harsh condition which is limited by message lifetime. Thanks to routing mechanism that based on encountering history information, EBRR is significantly reducing the copies comparing to replication based routing. In particular, the spray based schemes, like EBR and Prophet are with an observable performance improvement. This is because spray based routing schemes rely more on the situation that nodes are sufficiently mobile to encounter each other, as reflected by message lifetime. EBR and EBRR have a low overhead ratio due to controlling the number of copies strictly.

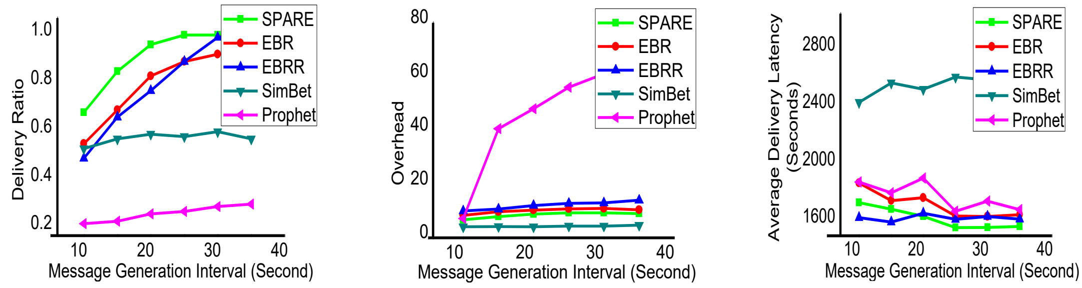


(a) Delivery Ratio (b) Overhead (c) Average Delivery Latency

Fig. 6. Influence of Message Lifetime

5.5. Influence of Generation Interval

We observe that SPARE achieves the highest delivery ratio compared with other algorithms in Fig. 7(a), thanks to only making a limited number of message copies in the network and spraying the copies according



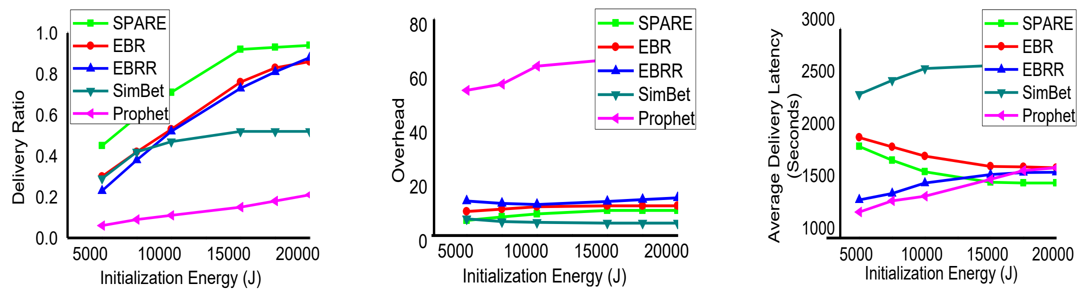
(a) Delivery Ratio (b) Overhead (c) Average Delivery Latency

Fig. 7. Influence of Generation Interval

to the rank of encounter time interval. When the message generation interval becomes longer, the replication based routing like EBRR begins to outperform the spray based routing schemes like EBR, because there are more copies in the network. We obverse that EBR and EBRR have a dramatically increased overhead ratio. To the contrary, SPARE maintains a stable performance due to their forwarding mechanism. It implies that proposed routing is advanced for guaranteeing message delivery when bandwidth is limited. Taking into account the fact that the experiment is carried out under adverse conditions, the results are reasonably reliable.

5.6. *Influence of Initialization Energy*

We observe that no matter what the initial energy of the node is sufficient or limited, SPARE achieves the highest delivery ratio compared with other algorithms in Fig. 8(a). When the initialization energy of the node increases, the rate of delivery of the five agreements increased, because nodes have more energy to transmit messages. The overhead of the SPARE is the lowest except SimBet. When the initialization energy is more than $1000J$, the average delivery latency of SPARE decreases and keeps a low value. In addition, SPARE maintains a stable performance due to their forwarding mechanism. It implies that proposed routing is advanced for guaranteeing message delivery when resource is limited. The reason is that SPRAE takes into account the utilization of resources and dynamically allocates copies of the message according to the node' residual energy, so SPRAE is adapt well to the change of energy compared to other protocols.



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

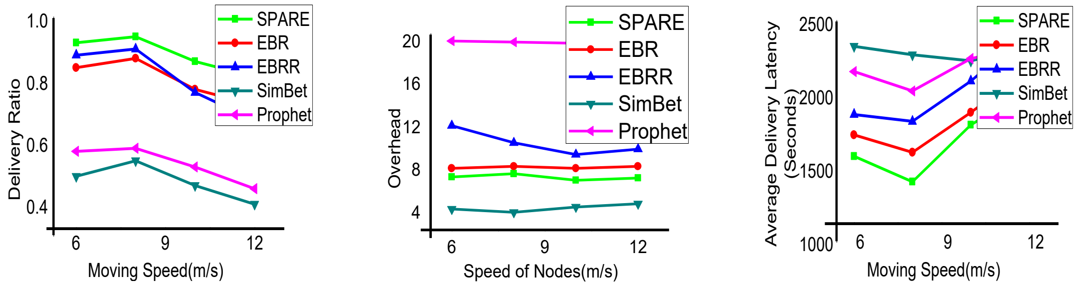
Fig. 8. Influence of Initialization Energy

5.7. *Influence of Speed*

We consider such a scenario that the moving speed of the sparse vehicles is constantly changing. We set the number of vehicles to 70, the range of communication is set to 30m, the messages size varies between

[500k, 1M] and the message lifetime is 90 minutes.

SPARE achieves the highest delivery ratio and the lowest average delivery latency compared with other algorithms as shown in Fig. 9(a), 9(b) and 9(c). The performance of all protocols are good when the speed of nodes is around $8m/s$. The overhead of the SPARE is the lowest except the SimBet. The performance of all protocols present a downward trend, when the speed of nodes is more than $8m/s$. The connection time between nodes is shorter due to the high speed mobility of nodes, which cause the message has not enough time to be completely transmitted. However, the decline trend of SPARE' delivery ratio is the slowest and still shows a better performance compared with other protocols.



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

Fig. 9. Influence of Speed

6. CONCLUSION

In this paper, we address routing resource utilization issue for ICTS by introducing SPARE, SPARE combines nodal resources effective consumption, encounter probability, nodal historical encounter information and the number of messages carried by nodes to make the nodes cooperative. The messages forwarding policy consists of two core components, which are the resources-probability social similarity forwarding algorithm and the weight connection-waiting time similarity forwarding algorithm. In addition, we present a new message management mechanism to balance network performance, which is composed of historical encounter information and resource utilization of nodes. Results under the scenario show that, it is also essential to consider the design of routing framework to reliably and efficiently deliver messages. Compared with the exiting algorithms, SPARE ensures the high delivery ratio while maintaining the low routing overhead. The performance of SPARE is well and stable. When the remaining time of the message is low, the transmission delay of SPARE is higher. In the future work, we will study the node's deeper sociality and the nodal historical information to reduce the transmission delay.

Acknowledgments

This paper was supported by the National Natural Science Foundation (61102105, 51779050), Key Program for International S&T Cooperation Projects of China (No. 2015DFG12150) and the Harbin Science Fund for Technology Innovation Talents (No. 2017RAQXJ036).

References

[1] X. Weng, Y. Liu, H. Song, S. Yao, P. Zhang, Mining urban passengers travel patterns from incomplete data with use cases, Computer Networks 134 (2018) 116–126.

[2] C. Lyu, D. Gu, Y. Zeng, P. Mohapatra, Pba: Prediction-based authentication for vehicle-to-vehicle communications, IEEE Transactions on Dependable & Secure Computing 13 (1) (2016) 71–83.

[3] K. C. Dey, A. Rayamajhi, M. Chowdhury, P. Bhavsar, J. Martin, Vehicle-to-vehicle (v2v) and vehicle-to-infrastructure (v2i) communication in a heterogeneous wireless network c performance evaluation , Transportation Research Part C 68 (2016) 168–184.

[4] A. Balasundram, Poster: A comparison study of wi-fi direct and dedicated short range communication for intelligent transport services, in: International Conference on Mobile Systems, Applications, and Services Companion, 2016, pp. 10–10.

[5] R. Anggoro, M. I. Royyana, H. A. A. M., A survey on non-dtn and dtn geographic based protocols for vanets, in: The Iiae International Conference on Intelligent Systems and Image Processing, 2015, pp. 299–304.

[6] T. Wang, Y. Cao, Y. Zhou, P. Li, A survey on geographic routing protocols in delay/disruption tolerant networks (dtns), International Journal of Distributed Sensor Networks,2016,(2016-2-23) 2016 (6).

[7] T. Wang, Y. Wang, B. Liu, X. Wang, J. Zhang, A. Hussain, P. Wang, Y. Cao, A novel cross-layer communication protocol for vehicular sensor networks, International Journal of Communication Systems (4).

[8] G. Araniti, N. Bezirgiannidis, E. Birrane, I. Bisio, S. Burleigh, C. Caini, M. Feldmann, M. Marchese, J. Segui, K. Suzuki, Contact graph routing in dtn space networks: overview, enhancements and performance, IEEE Communications Magazine 53 (3) (2015) 38–46.

[9] Y. Cao, O. Kaiwartya, C. Han, K. Wang, H. Song, A. Nauman, Toward distributed battery switch based electro-mobility using publish/subscribe system, IEEE Transactions on Vehicular Technology 67 (11) (2018) 10204–10217.

[10] P. Das, K. Dubey, T. De, Credit based routing in delay tolerant networks, in: IEEE International Conference on Parallel Distributed and Grid Computing, 2013, pp. 158–163.

[11] Q. Ayub, S. Rashid, M. S. M. Zahid, A. H. Abdullah, The optimization of spray and wait routing protocol by prioritizing the message forwarding order, International Journal of Innovation & Applied Studies 3 (3) (2013) 794–801.

[12] H. Chen, W. Lou, Gar: Group aware cooperative routing protocol for resource-constraint opportunistic networks, Computer Communications 48 (18) (2014) 20–29.

[13] Y. Mao, Y. Xia, Z. Bong, K. Y. Chai, A. Y. P. Chua, A. J. L. Tan, A. Sivagami, T. K. Ng, K. S. Lai, S. H. Toh, Multi-policy link state routing for disruption tolerant networks, in: Wireless and Mobile NETWORKING Conference, 2013, pp. 1–7.

[14] S. D. Han, W. C. Yun, An improved prophet routing protocol in delay tolerant network, The Scientific World Journal,2015,(2015-1-12) 2015 (2015) 1–7.

[15] Y. Cao, T. Wang, O. Kaiwartya, G. Min, N. Ahmad, A. H. Abdullah, An ev charging management system concerning drivers' trip duration and mobility uncertainty, IEEE Transactions on Systems Man & Cybernetics Systems 48 (4) (2018) 596–607.

[16] X. Liu, G. Jin, X. Jiang, J. Niu, N. University, Meeting delay expect value in spray and forward routing algorithm, Telecommunications Science.

[17] Shally, Performance evaluation of rapid and spray-and-wait dtn routing protocols under black hole attack 03 (1) (2014) 35–40.

[18] Z. Sun, Y. Bai, R. Wang, W. Wang, Comsp: Correlated contact and message scheduling policy in dtn, in: IEEE International Conference on High PERFORMANCE Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing, 2014, pp. 595–602.

[19] S. E. Loudari, M. Benamar, N. Benamar, A. Habbal, The impact of energy consumption on the performance of dtn routing protocols, in: International Conference on Internet Applications, Protocols and Services, 2015.

[20] X. Guan, M. Chen, T. Ohtsuki, Epidemic theory based h + 1 hop forwarding for intermittently connected mobile ad hoc networks, in: IEEE International Conference on Communications, 2012, pp. 2114–2118.

[21] M. Ababou, R. Elkouch, M. Bellafkih, N. Ababou, Antprophet: A new routing protocol for delay tolerant networks, in: Microwave Symposium, 2015, pp. 1–6.

[22] S. Xia, Z. J. Cheng, C. Wang, Y. F. Peng, A deliver probability routing for delay tolerant networks (dtn), in: International Conference on Wireless Communication and Sensor Network, 2015, pp. 407–410.

[23] B. Zhou, V. Lau, Q. Chen, Y. Cao, Simultaneous positioning and orientating (spao) for visible light communications: Algorithm design and performance analysis, IEEE Transactions on Vehicular Technology.

[24] B. Paramasivan, M. Bhuvaneswari, K. M. Pitchai, Augmented dtn based energy efficient routing protocol for vehicular ad hoc networks, in: Sensors, 2016, pp. 1–4.

[25] B. B. Bista, D. B. Rawat, A robust energy efficient epidemic routing protocol for delay tolerant networks, in: IEEE International Conference on Data Science and Data Intensive Systems, 2016, pp. 290–296.

[26] E. Wang, Y. Yang, B. Jia, T. Guo, The dtn routing algorithm based on markov meeting time span prediction model, International Journal of Distributed Sensor Networks,2013,(2013-9-18) 2013 (8) (2013) 286–291.

[27] S. Bhattacharjee, S. Basu, S. Roy, S. D. Bit, Best-effort delivery of emergency messages in post-disaster scenario with content-based filtering and priority-enhanced prophet over dtn, in: International Conference on Communication Systems and Networks, 2016, pp. 1–7.

[28] E. M. Daly, M. Haahr, Social network analysis for routing in disconnected delay-tolerant manets, in: ACM International Symposium on Mobile Ad Hoc NETWORKING and Computing, 2007, pp. 32–40.

[29] K. Henmi, A. Koyama, An enhanced spray and wait dtn routing protocol based on node mobility and maximum number of replications, International Journal of Adaptive & Innovative Systems 2 (1) (2014) 3–14.

[30] S. C. Nelson, M. Bakht, R. Kravets, Encounter: based routing in dtns, Acm Sigmobile Mobile Computing & Communications Review 13 (1) (2009) 56–59.

[31] Y. Cao, N. Wang, Z. Sun, H. Cruickshank, A reliable and efficient encounter-based routing framework for delay/disruption tolerant networks, IEEE Sensors Journal 15 (7) (2015) 4004–4018.

[32] F. Li, H. Jiang, Y. Wang, X. Li, M. Wang, T. Abdeldjalil, Sebar: Social energy based routing scheme for mobile social delay tolerant networks, IEEE Transactions on Vehicular Technology PP (99) (2013) 1–1.

[33] J. Wu, Y. Zhu, L. Liu, B. Yu, J. Pan, Energy-efficient routing in multi-community dtn with social selfishness considerations, in: Global Communications Conference, 2017, pp. 1–7.

[34] D. Bin, Y. Peng, G. Wang, Community-based energy-aware routing protocol in mobile social networks, in: International Conference on Algorithms and Architectures for Parallel Processing, 2015, pp. 311–321.

[35] J. He, C. Xu, Y. Wu, Resource-efficient routing protocol based on historical encounter time interval in dtn, in: IEEE International Conference on Computer and Communications, 2017.

[36] J. Mu, H. Sun, J. Wang, S. Fan, J. Pan, A weighted network model with simultaneous change of node and edge, in: Control Conference, 2014, pp. 2805–2809.

[37] C. A. Bliss, M. R. Frank, C. M. Danforth, P. S. Dodds, An evolutionary algorithm approach to link prediction in dynamic social networks, Journal of Computational Science 5 (5) (2014) 750–764.

[38] S. Rashid, Q. Ayub, A. H. Abdullah, Reactive weight based buffer management policy for dtn routing protocols, Wireless Personal Communications 80 (3) (2015) 993–1010.

[39] L. Yang, S. Chen, D. Wei, A buffer management strategy based on message drop history in dtn satellite network, in: International Conference on Mechatronics Engineering and Information Technology, 2017.

[40] C. Boldrini, M. Conti, A. Passarella, Performance modelling of opportunistic forwarding under heterogenous mobility, Computer Communications 48 (8) (2014) 56–70.

[41] N. Privault, Laplace transform identities for the volume of stopping sets based on poisson point processes, Advances in Applied Probability 47 (4) (2015) 919–933.

[42] Y. Chen, Y. Shen, J. Zhu, X. Jiang, Capacity and delay-throughput tradeoff in icmns with poisson contact process, Wireless Networks 21 (8) (2015) 2453–2466.

[43] D. Matsui, R. Hagihara, Y. Yamasaki, H. Ohsaki, Analysis of geographic dtn routing under random walk mobility model, in: IEEE Computer Software and Applications Conference, 2017, pp. 538–547.

[44] I. Shinko, T. Oda, E. Spaho, V. Kolici, A simulation system based on one and sumo simulators: Performance evaluation of first contact, prophet and spray-and-wait dtn protocols, in: International Conference on Broadband and Wireless Computing, Communication and Applications, 2015, pp. 137–142.

[45] W. Huang, S. Zhang, W. Zhou, Spray and wait routing based on position prediction in opportunistic networks, in: International Conference on Computer Research and Development, 2011, pp. 232–236.