A packet replication algorithm for delay tolerant networks based on multi-metric

Yongguang Liu
Institute of Information, Guangdong Industry Polytechnic, Guangzhou, Guangdong, 510300, China
Corresponding author’s e-mail: liu.yongguang@gmail.com

Abstract. Utility function of packet is often used to decide how to replicate packet in DTN networks. But this method has more uncertainty and relies on a single performance metric. For this problem, in order to reduce the impact of a single metric uncertainty, multiple utility metrics are introduced in the new algorithm. A packet replication probability calculation method based on entropy weight is designed. By calculating the entropy weight of each metric, the algorithm obtains the replication probability of each packet and takes the probability as the priority of packet replication. Because of considering the two metrics of packet expected transmission delay and node encounter possibility, the algorithm effectively reduces the influence of encounter time distribution problem and direction prediction problem in the original algorithm, and reduces the uncertainty of utility function. Simulation results show that the algorithm reduces the packet replication number and the average delay, improves the successful packet delivery rate. The overall performance of the network is further improved.

1. Introduction

DTN (Delay Tolerant Network) is a special wireless network, which often has some or all of the following characteristics: limited communication bandwidth, bad communication conditions, blocked communication process and interfered and heterogeneous networks of various communication modes[1]. DTN has a wide application in many areas, such as military tactical Internet, vehicle network in the city, sensor networks in various complex environments.

The characteristics of DTN show that the whole network cannot maintain stable TCP/IP communication. Frequent network topology changes induce that the traditional IP routing protocol based on routing discovery can’t complete the topology maintenance and satisfy network requirements. For the characteristics of DTN, to solve the DTN network problem in complex situations, the researchers have proposed some new routing algorithms and mechanisms. Considering that the maintenance routing table is not suitable for DTN application scenarios, the existing studies have adopted storage and forwarding method. According to how one node processes a packet, it can be divided into single-copy routing and multi-copy routing. For single-copy routing[2], a node will removes the stored packet when it forwards it. The advantage of this method is that only one copy of the packet in the network exists, which reduces the storage and forwarding overhead of the network. But the disadvantage is also obvious. Since only one copy spreads in the network, the delivery rate of the packet is reduced and the delay of the packet is increased. Given the shortcomings of single-copy routing, more researchers use multiple-copies as DTN routing strategies and focus on how to reduce the number of multiple copies in the network. [3] is a simple flood routing algorithm. It adopts the propagation method and achieves a high success rate. But large number of copies in the network leads
to the congestion and performance decline. [4] and [5] are two representative algorithms and derive several improvement protocols. The former reduces the copies by introducing the encounter probability of nodes. But the performance deteriorates dramatically when the network scale becomes large. The latter limits the number of copies by setting the number of copies of the packet to a fixed value. The algorithm will stop replicating when the number of copies is 1 and enter a WAIT stage. Although the load is less, the direct transmission mode reduces the data transmission rate in the final stage. Although the above algorithms improve the performance, the algorithms have a large accidental in improving a certain performance. In order to improve the deterministic of the algorithm, [6] constructs a utility function for a metric, preferentially ensuring the transmission and replication within the limited transmission time and transmission resources. However, the algorithm adopts exponential distribution as the encounter time between nodes, only for computational convenience and no theoretical basis. Moreover, the imprecise information adopted in the algorithm can support the experimental needs in its setting scenarios, and it is unknown whether it is sufficient and feasible in other scenarios. [7] improves the algorithm and redesigned the utility function, but also has uncertainty in direction prediction.

Routing algorithms based on location information have been widely used in Ad hoc networks. Some researches apply location information to the DTN routing. The nodes in [8] obtain the positioning and movement information from receiving beacon message from the anchor node periodically and use these to transmit data in the cache. It is a central based topology that one node obtains the position of the other nodes through the center node to improve the delivery success rate. But the network will be paralyzed when the anchor node fails. [9] and [10] have improved the [5] using location information. [11] proposes a location-based urban bus DTN network in which one node obtains accurate destination node direction with the support of Bus Information System (BIS) and forwards data in that direction for good performance. But the dependence on BIS limits its application. [12] takes advantage of the location, speed and movement direction information. If the time elapsed too long after the meeting, the algorithm performance drops sharply.

According to the advantages and disadvantages of the above algorithms, this paper introduces two metrics to reduce the dependency and forwards packets according to a probability. Through the location estimation of target nodes and neighbor nodes, it not only reduces the number of message replication, but also avoids the uncertainty of utility function. The algorithm does not need central node or additional assistance. It increases applicability and achieves good results.

2. Algorithm design

In the current research, there are various strategies in forwarding packets. These methods have the common idea that selects the nodes which are most likely to reach the destination to forward packet. In terms of research idea, one is to predict the transmission delay of each message and determine which message to forward first. The second is to predict the movement of the node and decide which message to copy according to the movement direction of the node carrying the message. The two methods have their own advantages and disadvantages. The purpose of this study is to combine the two decision methods and integrates the two metrics. This paper designs how to forward each message according to the two metrics. The following introduces the comprehensive forward strategy by the two metrics. In this algorithm, two metrics are considered. One is the probability that the current node meets the destination node. The other is the packet expected transmission delay.

2.1. Packet expected transmission delay

The forward strategy is proposed in [6]. The core of the utility function is the design of metric. Five metrics are designed in the literature. but only the expected transmission delay is the basis, which is also used in our research. The calculation of the metric is introduced as following.

Step 1: Each node X maintains a message queue in which the packets’ destination is Z. Each packet i in this queue has a lifetime T(i) (from creation to present). The packets in the queue are sorted
in descend order by $T(i)$. Define $B$ as the average bytes that $X$ and $Z$ transmit at each encounter and $b(i)$ as the packet bytes before $i$ in the queue.

**Step 2:** If there is the opportunity to submit packet $i$, the encounter times between $X$ and $Z$ must be . Define $M_{XZ}$ as the time that $X$ and $Z$ meets each time. $M_{XZ}$ is a random number and obeys some distribution. The delay before $X$ transmitting packet $i$ is:

$$M_{XZ}(i) = \sum^n_{i=1} M_{XZ}$$

(1)

**Step 3:** Define $X_1, X_2, \ldots X_k$ as the collection of the nodes with a copy of the packet $i$, the expected rest transmission time of packet $i$ is:

$$a(i) = \min \{M_{X_1}(i), M_{X_2}(i), \ldots, M_{X_k}(i)\}$$

(2)

When the encounter time is subject to the exponential distribution, define $n_1(i), n_2(i), \ldots, n_k(i)$ as the encounter times between $X_1, X_2, \ldots, X_k$ and $Z$, then:

$$P(a(i) < t) = 1 - e^{-\frac{\lambda_1 n_1(i) + \lambda_2 n_2(i) + \cdots + \lambda_k n_k(i)}{1}}$$

$$E(a(i)) = \frac{\lambda_1 n_1(i) + \lambda_2 n_2(i) + \cdots + \lambda_k n_k(i)}{1}$$

(3)

(4)

**Step 4:** The delay of the packet $i$ is:

$$D(i) = T(i) + E(a(i))$$

(5)

2.2. **Probability of meeting destination nodes**

As discussed in [7], the predicted time of the node $S$ to the destination node $D$ is calculated as follows:

$$\begin{align*}
(y - y_D - v_D \cos \beta_D)^2 + (x - x_D - v_D \cos \alpha_D)^2 &= R^2 \\
y - y_S &= \frac{\sin \beta_S}{\sin \beta_D} (x - x_S) \\
(\sqrt{(y - y_S)^2 + (x - x_S)^2} = v_st)
\end{align*}$$

(6)

From this equations group, parameter $t$ can be gotten. Let’s assume the survival time of each message is deadline. The encounter probability between message $i$ and the destination node is defined as follows:

$$P(i) = \begin{cases} \frac{\text{deadline}}{t} & \text{deadline} < t \\ 1 & \text{deadline} \geq t \end{cases}$$

(7)

2.3. **Message replication and forwarding policy under multi-metric**

For combining the above two metric, this paper proposes a replication strategy based on metric entropy weight. The implementation details are as follows:

2.3.1. **Standardize metric to eliminate the effects of dimensions**

For each message, its single metric is standardized as follows:

$$X_i = \frac{\max(X_i) - X_i}{\max(X_i) - \min(X_i)} \times \alpha_i + (1 - \alpha_i)$$

(8)

For the encounter probability, no standardization is required because its value is between 0 and 1. $\alpha$ is the influence factor of the $i$ metric. In this project, there are two metric and each influence factor is set to 0.5.

2.3.2. **Calculate the information entropy**

The information entropy for the metric $i$ of each message is calculated as follows:

$$E_{ni} = \frac{1}{-\ln n} \sum^n_{j=1} F_i \ln F_i$$

(9)

For metric $i$: 


F_j = \frac{X_{ij}}{\sum_{i=1}^{n_i} X_{ij}} \quad (10)

2.3.3. The weight of the information entropy
The weight of the information entropy for metric i is calculated as follows:

w_i = \frac{1 - EN_i}{\sum (1 - EN_i)} \quad (11)

2.3.4. Replication probability of each message
The replication probability of each message is defined as follows:

p(i) = 1 - \frac{c_i}{\sum c_i} \quad (12)

Where each message i has M metrics, the C value of the message i is defined as:

C_i = \sum_{j=1}^{M} w_j X_{ij} \quad (13)

After the replication probability of each message is calculated, it will be sorted according to the replication probability in descend order. In the limited time of the node encounter, the message with large replication probability will be copied in priority.

3. Algorithm performance evaluation
In this part, the entropy-based replication policy algorithm (EBA) is evaluated by simulation. In the design process of the algorithm, the research results of the [6] and [7] are used. Therefore, our algorithm performance will be evaluated by comparing with the two algorithms. The main simulation parameters used in this algorithm are shown in Table 1.

| Parameters                                | Value                        |
|-------------------------------------------|------------------------------|
| Area size                                 | 10k m \times 10k m           |
| The Mac protocol                          | IEEE 802.11 DCF              |
| Physical layer model                      | PHY802.11b                   |
| Move the model                            | Random Waypoint              |
| Node transfer radius                      | 250m                         |
| Number of most recent location sampling values N | 10                           |
| Sampling time interval T                  | 1min                         |
| Value of k in PRR                         | 2                            |
| Number of nodes                           | 20-50                        |
| Cache size                                | 1M                            |
| Message size                              | 1024Byte                     |
| Flow model                                | CBR                           |
| deadline                                  | 20s                           |

Figure 1-3 are the simulation results in the case of 20 nodes, in which the node movement speed is 10m/s. The transverse axis is the number of messages generated per 50s between per pair of source and destination nodes. Figure 1 is a comparison result of the average delay for each successfully arrived message. It can be seen from this figure that as the number of messages increases, the number of messages waiting in the cache increases and the delays increase while the node density remains
constant. Under the mobile model of this paper, for considering both the two metrics, our algorithm reduced the influence of uncertainty of the encounter time and used the moving characteristics. We can send the message directly to the moving direction to obtain less delay. For the same reason, this algorithm made it easier for messages to reach the target node in some direction at some time before being discarded, but the successful delivery rate decreased as the number of messages increased.

Figure 3 shows that the algorithm obtains smaller message replication. This is mainly based on two reasons: Firstly, the propagation of the algorithm message has the influence of multiple factors and further enhances the adaptability of the algorithm; secondly, the multi-metric-based message replication rules have more detailed restrictions to reduce the number of message replication.

Figure 1. Comparison of average delay.

Figure 2. Comparison of message submission rate.

Figure 3. Comparison of number of message replication.

Figure 4-6 are simulation results at a moving speed of 10m/s and one packets per second generated between per source and destination. The transverse axis is the number of nodes in the network. Compared with Figure 2 and 3, with the increase of nodes, each node obtains more information to the destination node. The efficiency function is more accurate and more forward paths to the destination node are found, which lead to less delay to the destination node. The probability of messages being successfully submitted is higher. Our proposed algorithm combines the advantages of the two algorithms, thus making further improvement in performance, which is shown in Figures 4 and 5.
Figure 4. Comparison of average delay.

Figure 5. Comparison of delivery rates.

Figure 6 describes the change of the number of messages duplicated under current simulation conditions. Due to the increase of number of nodes, the number of nodes participating in the message replication increases. The number of messages duplicated increases. However, when the density of nodes increased to a certain extent, due to more accurate path information to the destination node, some nodes did not need to do useless replication at this time. Therefore, although the number of nodes increased, the number of replication messages decreased, and the simulation results of the three algorithms showed this trend.

4. Conclusion
In this paper, encounter probability and entropy weight were defined. The replication probability of each message was obtained and messages were duplicated in a finite time according to the probability. Simulations showed that the algorithm effectively solved direction prediction problem in the RAPD algorithm and time distribution problem in PRR algorithm by multi-metric, reduced the uncertainty of utility function and achieved good performance.

Acknowledgments
This research was supported by the universities and colleges key program(2020ZDZX1074), which is sponsored by Guangdong Province Education Department of China.
References

[1] Khabbaz M., Assi C. M., and Fawaz W. F.,(2012)Disruption-Tolerant networking: a comprehensive survey on recent developments and persisting challenges. IEEE Commun. Surveys & Tutorials.14: 607-640.

[2] Spyropoulos T., Psounis K., Raghavendra C. S.(2008)Efficient routing in intermittently connected mobile networks: The single-copy case. IEEE Trans. on Net., 16:63–76.

[3] Vahdat A.,Becker D.(2000)Epidemic routing in partially-connected ad hoc network.Technical Report,CS-200006:345-350.

[4] Lindgren A., Doria A. and Schelén O.(2003)Probabilistic routing in intermittently connected networks.In:ACM SIGCOM. New York. pp. 19-20.

[5] Spyropoulos T., Psounis K.(2015)Spray and wait: An efficient routing scheme for intermittently connected mobile networks.In: ACM SIGCOM. New York.pp. 252–259.

[6] Balasubramanian A.,Levine B. N., and Venkataramani A.(2007)DTN routing as a resource allocation problem.In: SIGCOMM, London.373-384.

[7] Liu Y. G.(2015)Position-Based RAPID routing algorithm for delay tolerant networks. Journal of UEST,44:840-844.

[8] Shen J., Moh S., and Chung I.(2010)A priority routing protocol based on location and moving direction in delay tolerant networks. IEICE Trans. on Info. and Sys., 5:2763-2775.

[9] Dang F.,Yang X.(2010)Longping jet forwarding algorithm: a DTN routing algorithm based on Markov position prediction model. Chinese Science: Info. Sci,40 : 1312-1320.

[10] Guo H.,Wang X.(2012)Location information-based jet routing algorithm in DTN. Mini-Micro Comp. Sys., 33 : 2481-2484.

[11] Park H., Jang J.(2012)Position-based DTN routing in metropolitan bus network.In: ICSAI, Washington.1449-1453.

[12] Andrade G. E., Calsavara A.(2013)Routing protocol based on the position,velocity,and direction of nodes,.In: WAINA,Singapore. 971-976.