An Application-Oriented Buffer Management Strategy in Opportunistic Networks

Meihua Liu[1], Xinchen Zhang[2, *], Shuangkui Ge[3], Xiaoli Chen[1], Jianbin Wu[2] and Mao Tian[1]

Abstract: In Opportunistic networks (ONs), buffer management is critical to improve the message exchanging efficiency due to the limited storage space and transmission bandwidth at the wireless edge. Current solutions make message scheduling and drop policy based on assumptions that messages can always been forwarded in a single contact, and all node pairs have the same contact rates. However, such ideal assumptions are invalid for realistic mobility traces of hand-held. Recent studies show that the single contact duration is limited and the mobility of nodes is heterogeneous in reality. In this paper, a buffer management strategy based on contact duration and heterogeneous mobility is proposed to improve the efficiency of buffer policy in the practical applications. We mainly focus on the minimization of the total expected delivery delay for all messages in ONs with resource constraints. Using the global network information including existing copies of message in the network, the distribution of pair-wise inter-contact time and contact duration between nodes, we develop a function to compute per-message utility which reflects the contribution of single message to the total expected delivery delay. Messages are scheduled or dropped according to their utilities. Simulation results show that our proposed strategy not only achieves lower delivery delay than mainstream strategies, but also keeps a high delivery ratio and a low network overhead.

Keywords: Opportunistic networks, buffer management, contact duration, heterogeneous mobility.

1 Introduction

With the popularity of smart handheld devices such as mobile phone and laptop, the demand of communication grows rapidly. As a result, the cellular system is overloaded with huge amount of traffic, and undergo quality deterioration. As an effective solution, the opportunistic contacts of the mobile users are utilized to offload part of the wireless traffic, especially for delay-tolerant data, for example, announcements or various video clips [Li, Zhang, Gan et al. (2015); Si, He, Yao et al. (2016)]. Specifically, the mobile devices with short-range interfaces such as Bluetooth and WiFi can form an Opportunistic Network (ON) by exploiting their opportunistic device-to-device communication [Xia, Liu, Li et al.

[1] School of Electronic Information, Wuhan University, Wuhan, China.
[2] School of Physical Science and Technology, Huazhong Normal University, Wuhan, China.
[3] Beijing Institute of Electronic Technology Application, Beijing, China.
[*] Corresponding Author: Xinchen Zhang. Email: zxc9501@hotmail.com.

(2017)]. Data is shared in a store-carry-forward fashion to handle the intermittent connectivity of mobile nodes [Yongxuan, Xing and Minghong (2016)]. That is, a message could be sent over an existing link, get buffered at current node until the next link in the path comes up, until it arrives to its destination. Such exchanges can be used to extend the cellular networks' coverage. And what counts most is, such exchanges can support more users at a lower cost.

However, in the face of huge number of contents, the performance of routing in ON can be reduced seriously with limited storage space and transmission bandwidth at the wireless edge [Shen, Moh, Chung et al. (2014); Erramilli and Crovella (2008)], so buffer management becomes necessary [Moetesum, Hadi, Imran et al. (2016)]. Specifically, there are two issues that need to be considered in the buffer management: First of all, due to the limited contact duration and transmission bandwidth, some messages in buffer cannot be successfully exchanged in a single contact between nodes. Therefore, the scheduling order of messages is very important for delivery efficiency. Secondly, in the process of message forwarding, multi-copy scheme is used by most existing routing protocols for ON to improve message delivery ratio [Wei, Liang and Xu (2014)]. This message redundancy mechanism coupled with long retention time of messages, imposes a great deal of buffer consumption on mobile devices, resulting in buffer overflows. Therefore, the order in which messages are discarded when buffer is full should be determined prudently, in order to release limited storage resources and ensure message transmission efficiency.

Many buffer management strategies in ON have been proposed. Several works introduce drop policies such as drop last, drop front, or drop most forwarded [Liu and Bai (2015)]. These strategies cannot achieve good performance as they do not utilize global network information [Liu and Bai (2015); Silva, Obraczka, Burleigh et al. (2015)]. Krifa et al. [Krifa, Barakat and Spyropoulos (2012); Wang, Wang, Feng et al. (2017); Wang, Yang and Wu (2015)] have made some improvements by exploiting network-wide information such as node mobility model and the number of existing copies of each message in the network. However, these works neglect the limited contact duration and the heterogeneous mobility of nodes. They simply assume that messages can always been forwarded in a single contact, and all node pairs have the same contact rates. Unfortunately, such ideal assumptions are invalid for realistic mobility traces of hand-held devices, therefore, it has become a challenge to further improve the efficiency of the buffer management strategy in the real situation.

In real applications, single pair-wise contact duration is usually short, since the commonly used interfaces of ON mobile devices such as Bluetooth and WiFi have short communication range, and encounters between nodes may happen randomly [Li, Liu, Zhu et al. (2015)]. For instance, for the smart phones carried by pedestrians, the Bluetooth interface can support a wireless range of about 10 m while the average walking speed of the pedestrians are about 1.5 m/s. Thus the contact duration tends to be as short as few seconds. For the moving vehicles which generally have higher speed, the contact duration is still short even though they communicate via WiFi that can support a longer range. Meanwhile, with the development of society, the demand for sharing large multimedia contents is increasing, which requires longer transmission time. Therefore, if the pair-wise contact duration is not taken into account, it is likely that messages with high theoretic delivery probability fail to transmit, which decrease the realistic efficiency of the buffer policy.

In addition, node mobility is heterogeneous in reality. Studies show that the hand-held mobile devices exhibit the characteristics of human society, such as activity and community [Zhang and Cao (2017); Li, Wang, Yang et al. (2014)]. Different nodes have different activity level, and the contact frequency of node pairs which belong to the same community is higher than that of node pairs which belong to different communities [Wei, Zeng, Guo et al. (2014)]. Therefore, uniform distribution cannot reflect the difference of activity degree of node, nor the realistic node contact behavior. And if the heterogeneous mobility is not considered, the subsequent optimized model will not be in accordance with the actual situation. As a result, the buffer policy will lack practicability and accuracy. To make buffer policy more practical than the existing studies, it becomes a mandatory requirement to apply the two factors in the optimization model.

In this paper, an efficient buffer management strategy is proposed for practical ON applications. First, using global information such as existing copies of message in the network, the distribution of pair-wise inter-contact time, and considering practical constraints including heterogeneous mobility, pair-wise contact duration and limited bandwidth, we develop a function to compute per-message utility which reflects the contribution of a single copy to the total expected delivery delay. The limited contact duration and bandwidth are used to obtain a threshold, message could be transmitted successfully in a contact only if its size is beyond the threshold. And using the heterogeneous mobility, the expected delay of a message is estimated more accurately than using the homogeneous mobility. Then an optimization model aiming at minimizing the total expected delivery delay for all messages is established, and messages are dropped and scheduled according to their utilities to achieve optimization. The simulation results show that compared with the mainstream buffer strategies, the proposed strategy has the lowest delivery delay, and achieves better performance in terms of delivery ratio and network overhead.

2 Related work

Generally, mobile devices have limited resource [Jin, Chunwei, Yu et al. (2018)]. Since buffer management can greatly influence the routing performance when nodes have limited buffer, several relevant strategies have been proposed in ON. For example, Drop Front and Drop Last, determine the discard order of messages according to the storage time of messages. DF discards messages with the longest storage time, while DL discards messages with the shortest storage time, and Drop Random is randomly discarded [Liu and Bai (2015)]. All of them do not consider any message properties or node information [Liu and Bai (2015); Silva, Obraczka, Burleigh et al. (2015)], so their performance is not very well. Lindgren [Lindgren (2006)] evaluate the performance of different combinations of buffer strategies for Prophet routing. Results show that the strategy which drops the messages with the most forwarded or replicated times and sends those with the highest delivery probability can achieve the best performance, in terms of the metrics of message delivery ratio and delay. Similarly, Erramilli et al. [Erramilli and Crovella (2008)] firstly estimate the number of replication or forwarding times of messages, and then they discard the message that have been forwarded or replicated the most. Compared with traditional DF, DL and other strategies, they can acquire better transmission performance due to the utilization of partial local information.

However, these strategies cannot reach the optimum since they ignore the global information of the network, such as the distribution of inter-contact time and contact duration between nodes, and the number of existing message copies. Wang et al. [Wang, Wang, Feng et al. (2017)] take the weighted sum of the time, existing message copies, remaining life time and other attributes of messages, so as to evaluate the drop priority of messages. Krifa et al. [Krifa, Barakat and Spyropoulos (2012); Krifa, Barakat and Spyropoulos (2008)] use the inter-contact time between nodes to estimate the expected delivery delay, and a utility function for message is designed to minimize the overall expected delay of all messages. Message with the minimum utility value are dropped when buffer overflow occurs. However, the two strategies ignore the impact of contact duration on message delivery with the assumption that all messages are of the same size or that the bandwidth is infinite. Wang et al. [Wang, Yang and Wu (2015); Liu, Wang, Zhang et al. (2011)] calculate the utility value of each message based on the dissemination state of message copies, and then make the discard decision. However, they assume homogeneous node mobility, that is, all nodes have the same contact rates, the pair-wise inter-contact rates between nodes are subject to a uniform exponential distribution, which is uncommon in practice.

As a recap, existing strategies do not utilize the global information, or not consider both the contact duration and the heterogeneous mobility between nodes in the optimization process. Their relaxed assumptions can degrade the final efficiency of the strategy in real application. The main reasons can be demonstrated as follows. Uniform distribution cannot reflect the difference of activity degree of node, nor can it reflect the real node contact behavior. In addition, due to the short contact duration, finite bandwidth and large application-level data units, not all messages can be forwarded successfully. If these facts are ignored, the optimized model will not be in accordance with the actual situation. In this paper, based on the global information commonly used in existing studies, we consider the two practical elements as additional constraints in the optimization process. And the proposed strategy is validated in the simulation section.

3 System analysis

Some studies have shown that many popular mobility models, such as random waypoint, random walk, and community-based model, have such a characteristic that the pair-wise inter-contact time and contact duration are exponentially distributed or have exponential tails [Spyropoulos, Psounis and Raghavendra (2006); Batabyal and Bhaumik (2015)]. In this paper, we propose a method to estimate utility based on the hypothesis that the pair-wise inter-contact time and contact duration between nodes are independent random variables and follow exponential distribution. Furthermore, we assume that inter-contact time and contact duration of different node pairs have different rate parameters, and single contact duration is short. Data can be exchanged between nodes through a short-range communication interface with the same and limited bandwidth. This implies that not all messages can be transmitted in a single contact duration. Besides, all nodes have the same buffer size and message size varies. Each message has a lifetime and will be discarded by nodes that store copies of it once the lifetime is over. In addition, a routing protocol is also required in the transmission of messages. Since we mainly focus on buffer management strategy, when comparison to existing buffer management strategies in the simulation section, we base our research on the classic Epidemic protocol [Vahdat and Becker (2000)], where two nodes exchange messages that they don’t have in common when they encounter each other.

4 Buffer management strategy

To minimize the total expected delivery delay for all messages in ONs with resource constraints, we have to obtain the expected delay of a single message. The expected delay of message is estimated by the global information such as mobility of nodes and existing copies of message. In this section, we firstly present the global network information that are needed. Then we develop a function to compute per-message expected delay utility. Finally, an optimization model aiming at minimizing the total expected delivery delay for all messages is established, and messages are dropped and scheduled according to their utilities to achieve optimization.

4.1 The global network information that nodes need to collect

To compute per-message expected delay, each node needs to collect and estimate the global network state. We summarize the notations in Tab. 1. The global information to be obtained in this paper is: $n_i(T_i)$-the number of existing copies of message *i* after the elapsed time T_i since its creation. $\{H_{1,i}, H_{2,i}, \cdots, H_{n,i}\}$ – the time at which $n_i(T_i)$ copies are received and stored at their carrier nodes. $\{\lambda_{1,d_i}, \lambda_{2,d_i}, \cdots, \lambda_{n,d_i}\}$-the inter-contact rates between nodes that store copies of message *i* and the destination of message *i*. $\{\mu_{1,d_i}, \mu_{2,d_i}, \cdots, \mu_{n,d_i}\}$-the contact duration rates between nodes that store copies of message *i* and the destination of message *i*. All nodes obtain and update the global information through encounters. Due to the propagation delay in ON, global network information recorded through node encounters may be unable to be updated in time. But according to the results of studies [Krifa, Barakat and Spyropoulos (2012); Wang, Yang and Wu (2015)], although such information may be inaccurate, the optimization algorithms have significantly improved the transmission efficiency, and achieved better performance than existing algorithms that do not utilize any extra network information.

Table 1: Notations

Symbol	Description
T_i	Elapsed time since the creation of message *i*
$n_i(T_i)$	The number of existing copies of message i after the elapsed time T_i since its creation
$\{H_{1,i}, H_{2,i}, \cdots, H_{n,i}\}$	The time when $n_i(T_i)$ copies of message *i* are received and stored separately at carrier nodes (since the creation of message *i*)
$\{\lambda_{1,d_i}, \lambda_{2,d_i}, \cdots, \lambda_{n,d_i}\}$	the inter-contact rates between nodes that store copies of message *i* and the destination of message *i*
$\{\mu_{1,d_i}, \mu_{2,d_i}, \cdots, \mu_{n,d_i}\}$	the contact duration rates between nodes that store copies of message *i* and the destination of message *i*

4.2 Computation of per-message expected delay utility

Since message transmission occur only when two nodes are in contact, the inter-contact time (the time elapsed between nodes encounters) and contact duration are the basic components of delivery delay. They are both determined by the exponential distributed models we adopt, thus the expected delay of message can be calculated. In addition, the limited contact duration and bandwidth are used to obtain a threshold, message could be transmitted successfully in a contact only if its size is beyond the threshold.

4.2.1 Expected delivery delay *of a message*

The expected delivery delay of a single message is computed as the marginal utility of a copy of this message. Let us denote the delivery delay of message *i* with random variable X_i. If we take at instant T_i a snapshot of the network, the expected delay of message *i* is given by,

$$\mathrm{E}[X_i] = \mathrm{P}[X_i \leq T_i] \times \mathrm{E}[X_i|X_i \leq T_i] + \mathrm{P}[X_i > T_i] \times \mathrm{E}[X_i|X_i > T_i] \tag{1}$$

where $\mathrm{P}[X_i \leq T_i]$ and $\mathrm{E}[X_i|X_i \leq T_i]$ respectively represent the probability and expected delay of message *i* that has been delivered successful before T_i. And $\mathrm{P}[X_i > T_i]$ and $\mathrm{E}[X_i|X_i > T_i]$ represent the probability and expected delay of message *i* that is delivered successful after T_i. We assume that after T_i, message *i* will be transmitted only to its destination node and not be replicated or dropped for the rest of its lifetime, to avert the impact of further change of the number of message copies. While variable X_i can be expressed as:

$$X_i = min_{k=1}^{n_i(T_i)}\{H_{k,i} + I_{k,d_i} + C_{k,d_i}\big|C_{k,d_i} \geq C_i\} \tag{2}$$

$$C_i = \frac{size(i)}{bandwidth} \tag{3}$$

where the random variables I_{k,d_i} and C_{k,d_i} represent the pair-wise inter-contact time and contact duration between the carrier node *k* and the destination node *d* of message *i*, respectively. Node *k* receives message *i* at time $H_{k,i}$ and C_i is the time needed to transmit message *i*, it is computed as the ratio of the size of message *i* ($size(i)$) to the bandwidth. Note that X_i, which is the delivery delay of message, exists only if the message is successfully delivered by its carrier *k*. This implies that the contact duration between the carrier node and the destination node should not be less than the time it takes to transmit the message. That is to say, the prerequisite condition that $C_{k,d_i} \geq C_i$ must be considered.

The derivation procedure of $\mathrm{E}[X_i]$ is presented as follows.

1) The computation of $\mathrm{P}[X_i > T_i]$ and $\mathrm{P}[X_i \leq T_i]$

The probability that message *i* has not been delivered by time T_i is:

$$\begin{aligned}\mathrm{P}[X_i > T_i] &= P\left[min_{k=1}^{n_i(T_i)}\{H_{k,i} + I_{k,d_i} + C_{k,d_i} > T_i|C_{k,d_i} \geq C_i\}\right] \\ &= \prod_{k=1}^{n_i(T_i)} P[I_{k,d_i} + C_{k,d_i} > T_i - H_{k,i}|C_{k,d_i} \geq C_i]\,.\end{aligned} \tag{4}$$

The conditional probability is calculated as follows:

$$P\big[I_{k,d_i} + C_{k,d_i} > T_i - H_{k,i}\big|C_{k,d_i} \geq C_i\big] = \frac{P[I_{k,d_i}+C_{k,d_i}>T_i-H_{k,i},C_{k,d_i}\geq C_i]}{P[C_{k,d_i}\geq C_i]}\;. \tag{5}$$

Since I_{k,d_i} and C_{k,d_i} are mutually independent random variables following exponential distribution, for random variables $Z_{k,d_i} = I_{k,d_i} + C_{k,d_i}$ and $C_{k,d_i} \geq C_i$, the joint probability distribution function is:

$$F_{k,d_i}(z) = P\big[I_{k,d_i} + C_{k,d_i} < \mathrm{z}, C_{k,d_i} \geq C_i\big]$$

$$= \int_{C_i}^{z} \mu_{k,d_i} \cdot e^{-\mu_{k,d_i} \cdot x} \cdot \int_0^{z-x} \lambda_{k,d_i} \cdot e^{-\lambda_{k,d_i} \cdot y} \cdot dy \cdot dx. \tag{6}$$

And the denominator $P[C_{k,d_i} \geq C_i]$ in Eq. (5) is given by:

$$P\left[C_{k,d_i} \geq C_i\right] = e^{-\mu_{k,d_i} \cdot x}. \tag{7}$$

Then we obtain:

$$P\left[I_{k,d_i} + C_{k,d_i} < T_i - H_{k,i} \middle| C_{k,d_i} \geq C_i\right]$$
$$= 1 - \frac{\lambda_{k,d_i}}{\lambda_{k,d_i} - \mu_{k,d_i}} \cdot e^{\mu_{k,d_i} \cdot (C_i - T_i + H_{k,i})} + \frac{\mu_{k,d_i}}{\lambda_{k,d_i} - \mu_{k,d_i}} \cdot e^{\lambda_{k,d_i} \cdot (C_i - T_i + H_{k,i})}. \tag{8}$$

Plugging Eq. (8) into Eq. (4), we obtain:

$$\mathrm{P}[X_i > T_i] = \prod_{k=1}^{n_i(T_i)} \left\{1 - P\left[I_{k,d_i} + C_{k,d_i} < T_i - H_{k,i} \middle| C_{k,d_i} \geq C_i\right]\right\}$$
$$= \prod_{k=1}^{n_i(T_i)} \left[\frac{\lambda_{k,d_i}}{\lambda_{k,d_i} - \mu_{k,d_i}} \cdot e^{\mu_{k,d_i} \cdot (C_i - T_i + H_{k,i})} - \frac{\mu_{k,d_i}}{\lambda_{k,d_i} - \mu_{k,d_i}} \cdot e^{\lambda_{k,d_i} \cdot (C_i - T_i + H_{k,i})}\right]$$
$$= A_i \ . \tag{9}$$

Based on Eq. (9), the probability that message *i* has already been delivered before T_i is:

$$\mathrm{P}[X_i \leq T_i] = 1 - \mathrm{P}[X_i > T_i]$$
$$= 1 - \prod_{k=1}^{n_i(T_i)} \left[\frac{\lambda_{k,d_i}}{\lambda_{k,d_i} - \mu_{k,d_i}} \cdot e^{\mu_{k,d_i} \cdot (C_i - T_i + H_{k,i})} - \frac{\mu_{k,d_i}}{\lambda_{k,d_i} - \mu_{k,d_i}} \cdot \quad e^{\lambda_{k,d_i} \cdot (C_i - T_i + H_{k,i})}\right]$$
$$= B_i \tag{10}$$

2) The computation of $\mathrm{E}[X_i | X_i > T_i]$

Intuitively, the expected delay of message *i* conditioned on $X_i > T_i$ can be computed as the sum of the elapsed time and the period from current time to the time when the first copy of *i* reaches the destination.

$$\mathrm{E}[X_i | X_i > T_i] = T_i + E[min_{k=1}^{n_i(T_i)} \{I_{k,d_i} + C_{k,d_i} | C_{k,d_i} \geq C_i\}] \tag{11}$$

Generally, the contact duration C_{k,d_i} is small compared to the inter-contact time I_{k,d_i}. For simplicity, we assume that $C_{k,d_i} \approx C_i$. Then Eq. (11) can be simplified to:

$$\mathrm{E}[X_i | X_i > T_i] = T_i + C_i + E\left[min_{k=1}^{n_i(T_i)} \{I_{k,d_i}\}\right]$$
$$= T_i + C_i + \frac{1}{\sum_{k=1}^{n_i(T_i)} \lambda_{k,d_i}}$$
$$= T_i + C_i + \frac{1}{n_i(T_i) \cdot \Theta_{d_i}} \tag{12}$$

where Θ_{d_i} represents the average encounter rate between the carrier nodes of message *i* and its destination node.

$$\Theta_{d_i} = \frac{\sum_{k=1}^{n_i(T_i)} \lambda_{k,d_i}}{n_i(T_i)} \tag{13}$$

3) The computation of $\mathrm{E}[X_i|X_i \leq T_i]$

Considering that there may be message replication during the time interval $[0, T_i]$, it is difficult to get the exact solution for the expected delay X_i conditioned on $X_i \leq T_i$. We approximate $\mathrm{E}[X_i|X_i \leq T_i]$ as the average value of the expected delivery delay of copies of message *i*. Mentioned here that only those less than T_i are selected. That is:

$$\mathrm{E}[X_i|X_i \leq T_i] = \frac{\sum_{k=1}^{n_i(T_i)} E[M_{k,i}|M_{k,i} \leq T_i]}{n_i(T_i)} \tag{14}$$

where $M_{k,i}$ is delivery delay of a copy of message *i* carried by node k, and $E\left[M_{k,i}\middle|M_{k,i} \leq T_i\right]$ is given by:

$$\begin{aligned}\mathrm{E}\left[M_{k,i}\middle|M_{k,i} \leq T_i\right] &= \int_{H_{k,i}+C_i}^{T_i} P\left[M_{k,i} = x\right]xdx \\ &= \int_{H_{k,i}+C_i}^{T_i} P\left[H_{k,i} + I_{k,d_i} + C_{k,d_i} = x|C_{k,d_i} \geq C_i\right]xdx \\ &= \int_{H_{k,i}+C_i}^{T_i} P\left[I_{k,d_i} + C_{k,d_i} = x - H_{k,i}|C_{k,d_i} \geq C_i\right]xdx\end{aligned} \tag{15}$$

For random variable $Z_{k,d_i} = I_{k,d_i} + C_{k,d_i}$, under the condition of $C_{k,d_i} \geq C_i$, the cumulative distribution function is calculated and proved to be continuous and derivable. Then the probability density function (PDF) is given by:

$$\begin{aligned}P\left[I_{k,d_i} + C_{k,d_i} = x - H_{k,i}|C_{k,d_i} \geq C_i\right] &= \frac{\lambda_{k,d_i} \cdot \mu_{k,d_i}}{\lambda_{k,d_i} - \mu_{k,d_i}} \cdot [e^{\mu_{k,d_i} \cdot (C_i + H_{k,i} - \mathrm{x})} \\ &\quad - e^{\lambda_{k,d_i} \cdot (C_i + H_{k,i} - \mathrm{x})}]\end{aligned} \tag{16}$$

Then we can obtain:

$$\begin{aligned}\mathrm{E}\left[M_{k,i}\middle|M_{k,i} \leq T_i\right] &= \frac{\lambda_{k,d_i} \cdot \mu_{k,d_i} \cdot T_i + \mu_{k,d_i}}{\left(\lambda_{k,d_i} - \mu_{k,d_i}\right) \cdot \lambda_{k,d_i}} \cdot e^{\lambda_{k,d_i} \cdot (C_i + H_{k,i} - T_i)} \\ &\quad - \frac{\lambda_{k,d_i} \cdot \mu_{k,d_i} \cdot T_i + \lambda_{k,d_i}}{\left(\lambda_{k,d_i} - \mu_{k,d_i}\right) \cdot \mu_{k,d_i}} \cdot e^{\mu_{k,d_i} \cdot (C_i + H_{k,i} - T_i)} \\ &\quad + \left(C_i + H_{k,i} + \frac{1}{\lambda_{k,d_i}} + \frac{1}{\mu_{k,d_i}}\right) \\ &= D_{k,i}\end{aligned} \tag{17}$$

Pugging Eq. (17) into Eq. (14), $\mathrm{E}[X_i|X_i \leq T_i]$ can be approximated as:

$$\mathrm{E}[X_i|X_i \leq T_i] = \frac{\sum_{k=1}^{n_i(T_i)} D_{k,i}}{n_i(T_i)} \tag{18}$$

Based on the analysis above, the $\mathrm{E}[X_i]$ is obtained.

$$\mathrm{E}[X_i] = A_i \cdot \left(T_i + C_i + \frac{1}{n_i(T_i) \cdot \Theta_{d_i}}\right) + B_i \cdot \frac{\sum_{k=1}^{n_i(T_i)} D_{k,i}}{n_i(T_i)} \tag{19}$$

4.2.2 Utility function

To investigate the effect of drop and receiving a copy on the expected delivery delay of the message, we differentiate $\mathrm{E}[X_i]$ with respect to $n_i(T_i)$.

$$\frac{\partial E[X_i]}{\partial n_i(T_i)} = -\frac{1}{n_i(T_i)^2} \cdot \left(\frac{A_i}{\Theta_{d_i}} + B_i \cdot \sum_{k=1}^{n_i(T_i)} D_{k,i}\right) \tag{20}$$

Then, considering the fact that message is unsegmented, we discretize and replace the $\partial n_i(T_i)$ by $\Delta n_i(T_i)$ to obtain:

$$\Delta E[X_i] = -\frac{1}{n_i(T_i)^2} \cdot \left(\frac{A_i}{\Theta_{d_i}} + B_i \cdot \sum_{k=1}^{n_i(T_i)} D_{k,i}\right) \cdot \Delta n_i(T_i) \tag{21}$$

$$\text{where } \Delta n_i(T_i) = \begin{cases} -1 & Discarding\ msg\ i \\ 0 & Continue\ storing\ msg\ i \\ 1 & Receiving\ msg\ i \end{cases} \tag{22}$$

To better reflect the contribution of a single copy to the delivery delay of the message, the utility value of message *i* is given by:

$$U_i = -\Delta E[X_i]/\Delta n_i(T_i) = \frac{1}{n_i(T_i)^2} \cdot \left(\frac{A_i}{\Theta_{d_i}} + B_i \cdot \sum_{k=1}^{n_i(T_i)} D_{k,i}\right) \tag{23}$$

Next, we will present the proposed buffer management strategy and illustrate why the U_i can represent the per-message utility with respect to minimizing the overall expected delivery delay for all messages.

4.3 Drop and scheduling policy

The purpose of buffer management in this paper is to minimize the overall expected delay of all messages stored in the network by selecting messages for dropping or scheduling. Let $E[N_S]$ denote the overall expected delay for all messages, and $N(t)$ denote the number of unique messages in the network at time instant *t*. Then the improvement in $E[N_S]$ is:

$$\Delta E[N_S] = \sum_{i=1}^{N(t)} \Delta E[X_i] = \sum_{i=1}^{N(t)} -U_i \cdot \Delta n_i(T_i) \tag{24}$$

If a node discards an already existing copy of message *i* from its buffer, then $\Delta n_i(T_i) = -1$ and $\Delta E[N_S] = \Delta E[X_i] = U_i > 0$, which means the overall expected delay will increase. Thus, to minimize the increase of $E[N_S]$, the message copy with the smallest utility value should be dropped. Here we denote it as $i_{U_{min}}$:

$$i_{U_{min}} = \mathrm{argmin}_i \sum_{i=1}^{N} U_i \cdot \Delta n_i(T_i) \tag{25}$$

If a node receives and stores an new copy of message *i* from its encounter node, then $\Delta n_i(T_i) = 1$ and $\Delta E[N_S] = \Delta E[X_i] = -U_i < 0$, which means the overall expected delay will decrease. Thus, to maximize the decrease of $E[N_S]$, the message copy with the largest utility value should be forwarded. Here we denote it as $i_{U_{max}}$:

$$i_{U_{max}} = \mathrm{argmax}_i \left|\sum_{i=1}^{N} U_i \cdot \Delta n_i(T_i)\right| \tag{26}$$

To sum up, when two nodes meet, the sender replicates messages to the receiver in decreasing order of their utility values. On the other hand, if the receiver's buffer overflows, it drops messages (including the newly-received message) in its buffer in increasing order of their utility values, subject to the constraint that the receiver node never drop its own source messages. An example of the proposed strategy is shown in Fig. 1, messages in buffer has already sorted in ascending order $m1 < m2 < \cdots < m7$. When two node s and v encounter each other, s replicates $m5$, $m3$, $m1$ in descending order to v, and meanwhile it receives $m7$, $m4$ from v. Since the buffer of s is full, the $m1$ and $m3$ are dropped due to their smaller utilities. Similarly, v will only store $m5$, in order to achieve optimization.

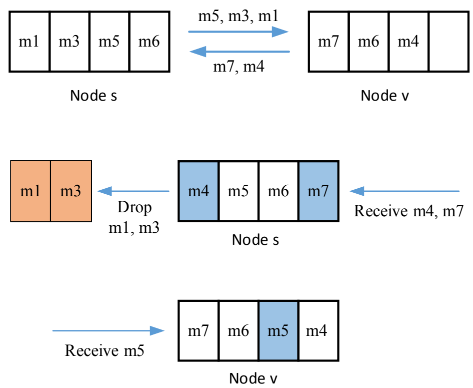


Figure 1: Message drop and scheduling policy

4.4 The estimation of exponential parameters

The pair-wise inter-contact rate $\lambda_{i,j}$ and contact duration rates $\mu_{i,j}$ between node i and j can be computed by exploiting their encounter history:

$$\lambda_{i,j} = \frac{n}{\sum_{k=1}^{n} T_k^{ICT}} \tag{27}$$

$$\mu_{i,j} = \frac{n}{\sum_{k=1}^{n} T_k^{CT}} \tag{28}$$

where $\{T_1^{ICT}, T_2^{ICT}, \cdots, T_n^{ICT}\}$ are the pair-wise inter-contact time samples, and $\{T_1^{CT}, T_2^{CT}, \cdots, T_n^{CT}\}$ are the pair-wise contact duration samples. For scenarios such as mobile social network or vehicular network, the node movement is usually regular [15], so this estimation method is reliable. Furthermore, the $\lambda_{i,j}$ and $\mu_{i,j}$ will be updated when the next encounter between the two node occurs.

5 Simulation and analysis

5.1 Simulation setup

We evaluated the proposed buffer management strategy, named as Utility, on the ONE simulator. First we compared the Utility strategy with existing buffer strategies based on Epidemic routing protocol. Then we applied the Utility to some classic routing algorithms to investigate its improvement on routing performance. Specific environment parameters are shown in Tab. 2. The following metrics are used in the simulations.

a. Delivery ratio, which is defined as the ratio of the number of delivered messages to the total number of unique messages.

b. Delivery delay, which is defined as the average delivery delay of all delivered messages.

c. Overhead ratio, which is defined as the ratio of the number of messages that are not successfully delivered to their destination node and the number of messages that are successfully delivered to their destination node.

Table 2: Simulation parameters

Parameters	Setting
Simulation time	800000 s
Scenario size	4500 m*3500 m
Types of nodes	Pedestrians, taxis trams
Number of nodes	124
Buffer size	15 M
Communication range	10m
Transmit speed of nodes	250 Mbps
Routing protocol	Epidemic
Lifetime of message	8 hours
Message generation interval	Varying between 30 s-120 s
Message size	Varying between 500 kB-2 M

5.2 Comparison to existing buffer management strategies

In this section, we compared the proposed Utility strategy with existing strategies such as Drop Front (DF), Drop Last (DL) [Liu and Bai (2015)], Global Knowledge based Scheduling and Drop (GBSD) [Krifa, Barakat and Spyropoulos (2012)]. DF discards the message that are stored first, while DL discards the message that stored last. GBSD derives per-message utility using global information, and schedules or drops messages according to their utility value to minimize the expected delivery delay. However, it assumes homogeneous node mobility, and ignores the duration of contact between nodes. Figs. 2, 3, 4 illustrate the performance of the four buffer strategies with different buffer size. The range of buffer size which varies from 5 M to 50 M.

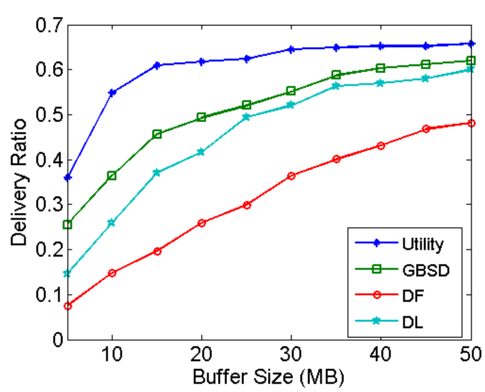


Figure 2: Delivery ratio by varying buffer sizes

As is shown in Fig. 2, although the delivery rate of all four strategies gradually increase with the buffer size and then become stable, the Utility strategy has the highest delivery ratio in the four strategies at the same buffer size. For example, when buffer size is 20 M, the Utility strategy has a delivery ratio 25% higher than GBSD, 137% higher than DF, and 48% higher than DL. This is because DF and DL only make use of a small amount of local information (the amount of time a carrier node stores the message), while GBSD and Utility make message drop decision from a global perspective. Thus, the latter two approaches can optimize the forwarding process and acquire higher delivery ratio. Compared with GBSD, the Utility considers additional constraints for realistic ONs such as heterogeneous mobility, pair-wise contact duration, limited bandwidth and varied message size, which makes the optimization more precise than GBSD. Therefore, more messages can be delivered successfully, in this case the Utility outperforms GBSD in terms of delivery ratio. In addition, the delivery ratio gap between Utility and other strategies is bigger at low buffer sizes, where a larger number of drop decisions is made. As a recap, the advantage of our proposed buffer management strategy is significant in networks with high congestion.

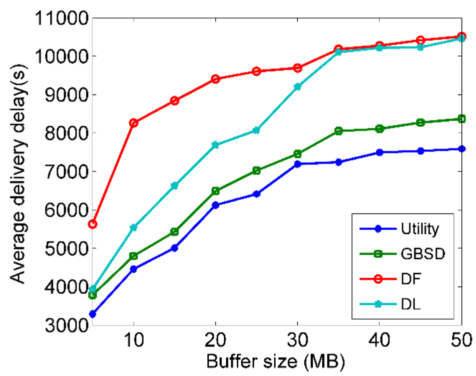


Figure 3: Delivery delay by varying buffer sizes

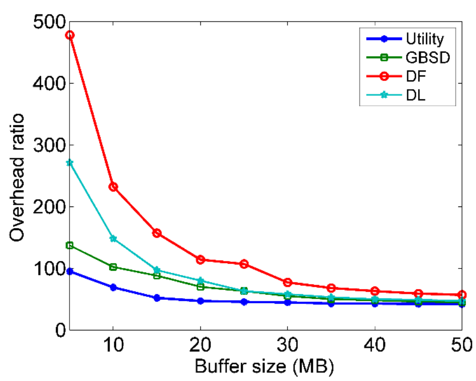


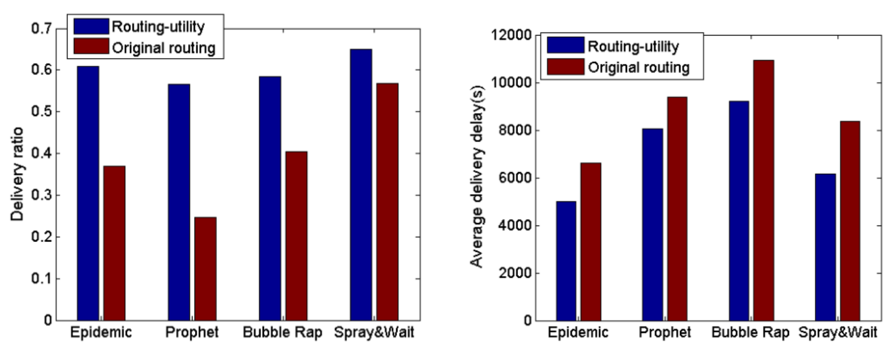
Figure 4: Overhead ratio by varying buffer sizes

Fig. 3 illustrates the performances of the four buffer strategies in terms of average delivery delay. The delivery delay of the four strategies all increase with the cache, which is reasonable because when the cache is large, some messages that are normally discarded under a small cache will wait for a long time until they are delivered, thus increasing the average delivery delay. More importantly, Utility outperforms other strategies, followed by GBSD, and the average delay of the Utility and GBSD are significantly less than those of DF and DL. For example, when buffer size is 20 M, the average delay of Utility is 25% of delay of GBSD, 137% of DF, and 48% of DL. Thanks to the utilization of the mobility model and several global information, GBSD and Utility estimate the utility of message more accurately, so they can conduct buffer management with the goal of minimizing the overall expected message delivery delay of the network. Furthermore, compared to GBSD, Utility calculates the expected delivery delay with higher accuracy by taking the heterogeneous mobility of nodes into account. In addition, message forwarding failure resulting from the ignorance of time required for message transmission is avoided since we consider the pair-wise contact duration overlooked by GBSD. Therefore, the Utility can schedule and drop messages more efficiently, thus achieving a lowest average delay.

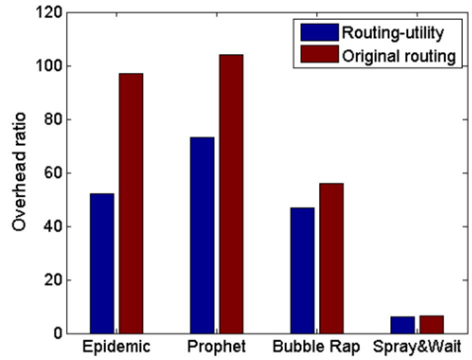
Fig. 4 illustrates the performances of the four buffer strategies in terms of overhead ratio. Since a big buffer size can reduce the retransmission times resulting from message drop, overhead of the four strategies all decrease as buffer size increases. Similar to the other two delivery metrics, Utility has advantages over the other three strategies in terms of overhead ratio, followed by GBSD. For example, when buffer size is 15 M, the overhead of Utility is 59% of overhead of GBSD, 33% of DF, and 54% of DL. Since the Utility and GBSD optimize the message forwarding process from global perspective, the useless forwarding times can be significantly reduced. And the Utility considers extra information to derive per-message utility which is more precise than the utility in GBSD, thus it can avoid discarding messages with high delivery probability, leading to the improvement of the performance.

5.3 Improvement on routing performance

To further investigate the efficiency of the Utility, in this paper we use Utility for some mainstream routing protocols in ONs such as Prophet, Bubble Rap and Spray and Wait [Wei, Liang and Xu (2014)], and then compare the performance of the routing with and without Utility. Meanwhile, we use Epidemic as reference. Figs. 5(a), 5(b), 5(c) illustrate the impact of Utility on the performance of routing protocols.



(a) (b)



(c)

Figure 5: (a) Performance in terms of delivery ratio; (b) Performance in terms of average delay; (c) Performance in terms of overhead ratio

From Figs. 5(a), 5(b), 5(c), it can be seen that all the routing protocols with Utility strategy acquire improved performances in terms of delivery ratio, average delay and overhead ratio. For Epidemic which don't have any optimizing process, Utility can significantly improve its performance. Besides, for optimized routing protocols like Prophet, Bubble Rap and Spray and Wait, the application of Utility still improves the efficiency of message delivery. These results also verify the effectiveness and reliability of proposed Utility strategy.

6 Conclusion

In this paper, an efficient application-oriented buffer management strategy for opportunistic networks is proposed to cope with the limited storage space and transmission bandwidth of the wireless devices while with huge number of contents. Specifically, we develop a message drop and scheduling policy named Utility in the message forwarding process. As two non-negligible factors in practical applications, the limited contact duration and heterogeneous mobility are utilized in Utility to calculate the per-message expected delay utility more practically and accurately than the existing algorithms. Messages are dropped and scheduled according to their utilities to minimize the overall expected delivery delay for all messages. Several comparisons are conducted in the simulation. The results show that our proposed strategy outperforms existing buffer management strategies in terms of average delivery delay, while still keep a high delivery ratio and a low network overhead. And the proposed strategy can significantly improve the efficiency of various routing protocols in ONs.

Acknowledgement: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was financially supported by National Natural Science Foundation of China (Grant Nos. U1736121 and U1536104).

References

Batabyal, S.; Bhaumik, P. (2015): *Analysing Social Behaviour and Message Dissemination in Human Based Delay Tolerant Network*. Springer-Verlag New York, Inc.

Erramilli, V.; Crovella, M. (2008): *Forwarding in Opportunistic Networks with Resource Constraints*. ACM Workshop on Challenged Networks.

Jin W.; Chunwei J.; Yu G.; Sangaiah, A. K.; Kim, G. J. (2018): A PSO based energy efficient coverage control algorithm for wireless sensor networks. *Computers, Materials & Continua*, vol. 56, no. 3, pp. 433-446.

Krifa, A.; Barakat, C.; Spyropoulos, T. (2008): Optimal buffer management policies for delay tolerant networks. *IEEE Communications Society Conference on Sensor, Mesh & Ad Hoc Communications & Networks.*

Krifa, A.; Barakat, C.; Spyropoulos, T. (2012): Message drop and scheduling in dtns: theory and practice. *IEEE Transactions on Mobile Computing,* vol. 11, no. 9, pp. 1470-1483.

Li, Y.; Zhang, J.; Gan, X.; Yang, F. (2015): Contract-based traffic offloading over delay tolerant networks. *IEEE Global Communications Conference*.

Li, Z.; Liu, Y.; Zhu, H.; Sun, L. (2015): Coff: contact-duration-aware cellular traffic offloading over delay tolerant networks. *IEEE Transactions on Vehicular Technology*, vol. 64, no. 11, pp. 5257-5268.

Li, Z.; Wang, C.; Yang, S.; Jiang, C.; Li, X. (2014): Lass: local-activity and social-similarity based data forwarding in mobile social networks. *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 1, pp. 174-184.

Lindgren, A. (2006): Evaluation of queueing policies and forwarding strategies for routing in intermittently connected networks. *International Conference on Communication System Software & Middleware.*

Liu, F.; Bai, X. (2015): Research on the buffer management algorithm in DTN. *International Conference on Information Science and Control Engineering*, vol. 50, pp. 442-446.

Liu, Y.; Wang, J.; Zhang, S.; Zhou, H. (2011): A buffer management scheme based on message transmission status in delay tolerant networks. *Global Telecommunications Conference.*

Moetesum, M.; Hadi, F.; Imran, M.; Minhas, A. A.; Vasilakos, A. V. (2016): An adaptive and efficient buffer management scheme for resource-constrained delay tolerant networks. *Wireless Networks*, vol. 22, no. 7, pp. 1-13.

Shen, J.; Moh, S.; Chung, I., Sun, X. (2014): Buffer scheme optimization of epidemic routing in delay tolerant networks. *Journal of Communications and Networks*, vol. 16, no. 6, pp. 656-666.

Si, P.; He, Y.; Yao, H.; Yang, R.; Zhang, Y. (2016): Dave: offloading delay-tolerant data traffic to connected vehicle networks. *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3941-3953.

Silva, A. P.; Obraczka, K.; Burleigh, S.; Hirata, C. M. (2015): Smart congestion control for delay- and disruption tolerant networks. *Ad Hoc Networks*, vol. 25, pp. 480-494.

Spyropoulos, T.; Psounis, K.; Raghavendra, C. S. (2006): Performance analysis of mobility-assisted routing. *Proceedings of MOBIHOC'06*, pp. 49-60.

Vahdat, A.; Becker, D. (2000): Epidemic routing for partially connected ad hoc networks. *Technical Report CS-200006.* Duke University, USA.

Wang, E.; Yang, Y.; Wu, J. (2015): A Knapsack-based message scheduling and drop strategy for delay-tolerant networks. *Wireless Sensor Networks*.

Wang, H.; Wang, H.; Feng, G.; Lv, H. (2017): Nwbbmp: a novel weight-based buffer management policy for DTN routing protocols. *Peer-to-Peer Networking and Applications*, vol. 2, pp. 1-7.

Wei, K.; Liang, X.; Xu, K. (2014): A survey of social-aware routing protocols in delay tolerant networks: applications, taxonomy and design-related issues. *IEEE Communications Surveys & Tutorials,* vol. 16, no. 1, pp. 556-578.

Wei, K.; Zeng, D.; Guo, S.; Xu, K. (2014): On social delay-tolerant networking: aggregation, tie detection, and routing. *IEEE Transactions on Parallel & Distributed Systems*, vol. 25, no. 6, pp. 1563-1573.

Xia, F.; Liu, L.; Li, J.; Ma, J.; Vasilakos, A. V. (2017): Socially-aware networking: a survey. *IEEE Systems Journal*, vol. 9, no. 3, pp. 904-921.

Lai, Y. X.; Gao, X.; Liao, M. H.; Xie, J. S.; Lin, Z. Y. (2016): Data gathering and offloading in delay tolerant mobile networks. *Wireless Networks*, vol. 22, no. 3, pp. 959-973.

Zhang, X.; Cao, G. (2017): Transient community detection and its application to data forwarding in delay tolerant networks. *IEEE/ACM Transaction on Networking*, vol. 25, no. 5, pp. 2829-2843.