On Exploiting Transient Contact Patterns for Data Forwarding in Delay Tolerant Networks

Wei Gao and Guohong Cao
Department of Computer Science and Engineering
The Pennsylvania State University, University Park, PA 16802
{*weigao, gcao*}*@cse.psu.edu*

***Abstract*—Effective data forwarding in Delay Tolerant Networks (DTNs) is challenging, due to the low node density, unpredictable node mobility and lack of global information in such networks. Most of the current data forwarding schemes choose the nodes with the best cumulative capability of contacting others as relays to carry and forward data, but these nodes may not be the best relay choices within a short time period, due to the heterogeneity of the transient node contact patterns. In this paper, we propose a novel approach to improve the performance of data forwarding in DTNs by exploiting the transient node contact patterns. We formulate the transient node contact patterns based on experimental studies of realistic DTN traces, and propose appropriate forwarding metrics based on these patterns to improve the effectiveness of data forwarding decision. When applied to various data forwarding strategies, our proposed forwarding metrics achieve much better performance compared to existing schemes with similar forwarding cost.**

I. Introduction

Delay Tolerant Networks (DTNs) [12] consist of mobile nodes which contact each other opportunistically. Due to the low node density and unpredictable node mobility, only intermittent connectivity among mobile nodes exist in DTNs. To forward data to a destination within a given time constraint, node mobility is exploited to let nodes physically carry the data as relays, and forward data opportunistically upon contacts with others. The key problem for data forwarding in DTNs is therefore how to make effective data forwarding decision, to ensure that the data is carried by the relays with the best chance to contact the destination within the time constraint.

In most of the current data forwarding schemes in DTNs, due to the lack of global information at individual nodes about how to reach the destination, data forwarding decision is made based on the prediction of node capability of contacting others. This capability is indicated by various destination-independent *data forwarding metrics*, which differ in the network information being used, as well as how it is used, for the aforementioned prediction. Some schemes [20], [7], [25] estimate such capability by predicting node mobility and subsequent co-location events. Some others [8], [15], [13] propose data forwarding metrics by exploiting node contact pattern as abstraction of node mobility pattern. Since node contact pattern exhibits the long-term social relation among nodes with better stability, they make data forwarding decision more effective and less susceptible to the node mobility randomness. In these schemes, the data forwarding metrics are calculated based on the nodes' cumulative contact characteristics.

We observe that the transient contact characteristics of mobile nodes during short time periods in DTNs usually differ a lot from their cumulative contact characteristics. The relays selected by existing schemes therefore may not be the best choices for forwarding data with a short time constraint. Based on this observation, in this paper we improve the performance of data forwarding in DTNs, by exploiting the transient node contact patterns from the following two perspectives.

- *Transient contact distribution*, which may be highly skewed during different time periods. For example, a student A may contact his classmate B frequently in the daytime but not at night, while B's roommate C contacts B frequently at night but not in the daytime. Then, if there is data destined to B, it is better to deliver the data to A in the daytime, but to C in the nighttime. However, cumulative contact distribution cannot differentiate these two cases. Thus, we use transient contact distribution to better represent the skewness of contact distribution and improve the data forwarding performance.
- *Transient connectivity*, which indicates that some nodes in DTNs may remain connected with each other during specific time periods to form *transient connected subnets (TCS)*, despite the general absence of end-to-end paths among them. For example, a student remains connected with his classmates during the class and they form a TCS during that time period. Similarly, vehicles also form a TCS when they are waiting for the traffic light at the crossroads [26]. A node has "*indirect*" contacts with all the nodes in a TCS, as long as it directly contacts any node in that TCS. The contact capability of mobile nodes is then increased by exploiting these indirect contacts.

The major contribution of this paper is two-fold. First, we formulate the transient node contact patterns based on experimental studies of realistic DTN traces. Transient contact distribution is formulated as intermittent appearances of short "on-period" where most of the node contacts accumulate, and transient connectivity is formulated as the number of nodes inside a node's TCS during different time periods. These two perspectives are then uniformly represented in the form of Gaussian function. Second, based on the transient contact patterns, we develop data forwarding metrics to analytically

This work was supported in part by the US National Science Foundation (NSF) under grant number CNS-0721479, and by Network Science CTA under grant W911NF-09-2-0053.

predict the contact capability of mobile nodes with better accuracy, so as to improve the effectiveness of data forwarding decision. Through extensive trace-driven simulations, we show that our approach significantly outperforms existing schemes in terms of data delivery ratio, with similar forwarding cost.

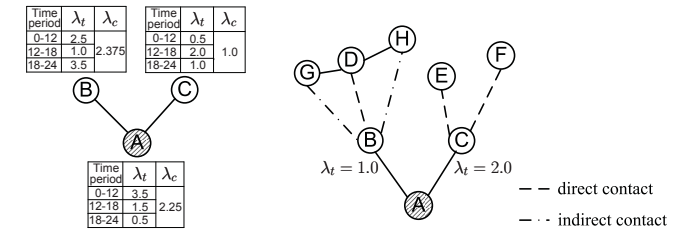
The rest of this paper is organized as follows. Section II briefly reviews the existing work. Section III gives an overview on how to exploit transient node contact patterns for data forwarding. Section IV formulates the transient contact patterns based on experimental observations from realistic DTN traces, and Section V describes our data forwarding approach in detail. Section VI evaluates the performance of our approach by trace-driven simulations. Section VII discusses and Section VIII concludes the paper.

II. Related Work

The research on data forwarding in DTNs originates from Epidemic routing [24] which floods the entire network. Later studies develop data forwarding strategies to approach the performance of Epidemic routing with lower forwarding cost, which is measured by the number of data copies created in the network. While the most conservative approach [23] always keeps a single data copy and Spray-and-Wait [22] holds a fixed number of data copies, most schemes leave such numbers as dynamic and make data forwarding decision by comparing the nodes' data forwarding metrics. In [3], [8], a relay forwards data to another node whose forwarding metric is higher than itself. Delegation forwarding [11] reduces the cost by only forwarding data to the node with the highest metric.

The data forwarding metric, which measures the nodes' capability of contacting others, is generally independent from the data forwarding strategies mentioned above. Various metrics can be applied to the same forwarding strategy for different performance requirements. Some schemes predict node contact capability by estimating their co-location probabilities in different ways, such as the Kalman filter [7] and semi-Markov chains [25]. In some other schemes, node contact pattern is exploited as abstraction of node mobility pattern for better prediction accuracy, based on the experimental [5], [17] and theoretical [4] analysis on the node contact characteristics.

Recent trace-based study on campus wireless networks [14] shows that different nodes have heterogeneity in their contact patterns, and the nodes' capability of contacting others in the future can be predicted based on their cumulative contact records in past. MaxProp [3] estimates the node contact likelihood based on the contact counts in past, and PodNet [18] forwards data to nodes based on their received data enquiries in past. In recent social-based forwarding schemes, SimBet [8] uses ego-centric betweenness as the data forwarding metric, and BUBBLE Rap [15] considers node centrality hierarchically in social community structures. [13] estimate pairwise node contact probabilities in the future based on the previous cumulative node contact rates, by assuming the exponential distribution of pairwise node inter-contact time. However, due to the heterogeneity of transient contact distributions, such estimations may be inaccurate during short time periods.



(a) Transient contact distribution (b) Transient connectivity

Fig. 1. Overview of transient node contact patterns

Most schemes predict node contact capability only based on the direct contacts among mobile nodes. Contact duration has only been analyzed for the properties of temporal network topology [6], or has been considered in case of limited channel bandwidth to determine the appropriate data transmission order [3], [2], [19]. Various weighting methods [3], [2] and the willingness of mobile users [19] are considered to maximize the bandwidth utilization during the limited contact duration. However, contact duration has never been studied for indirect contacts and transient connectivity among nodes, which can significantly increase the node contact capability.

III. Overview

The major focus of this paper is to develop appropriate data forwarding metrics, by exploiting the transient node contact patterns, for more accurate prediction of node contact capability within the given time constraint. Without loss of generality, we consider forwarding one data item to a specific destination. We assume that the data size is small, so that it can be carried by any node, and can be completely transmitted during a contact. The considerations of limited node buffer and bandwidth are orthogonal to the major focus of this paper.

We consider that the time constraint for data forwarding is shorter than one day. In this case, the transient contact characteristics of mobile nodes differ a lot from their cumulative contact characteristics, and data forwarding decisions based on transient contact patterns are more effective. Such advantage is illustrated in Figure 1, where relay A at 12pm carries data with a 6-hour forwarding time constraint, and needs to decide whether to forward data to nodes B and C.

In Figure 1(a), the transient rate (λ_t) of a node contacting others during different hours in a day is listed. The cumulative contact rate during the whole day is calculated as $\lambda_c = \sum_{i=1}^{n} T_i \cdot \lambda_{t_i} / \sum_{i=1}^{n} T_i$, where T_i is the length of the i-th time period. Based on the strategy that a relay forwards data to the node whose forwarding metric is higher than itself, A forwards data to B but not C if λ_c is used as the forwarding metric. However, considering that the transient contact distribution of B is skewed and λ_t of node B during the time period between 12pm and 18pm is essentially low, C is a better relay choice than B due to its higher λ_t during that time period.

Indirect contacts among mobile nodes can also be exploited to improve the performance of data forwarding. As shown in Figure 1(b), although C has a higher λ_t than B, such

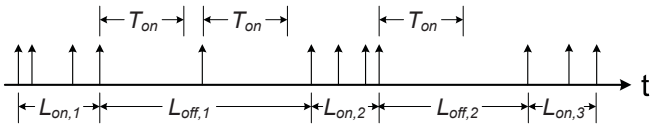


Fig. 2. Alternative appearances of on-period and off-period

λ_t is calculated by counting only the direct contacts among mobile nodes. In contrast, node D has transient connectivity with nodes G and H. By contacting D, node B has indirect contacts with G and H. Node B therefore has higher contact capability than node C, and is a better relay choice.

IV. Trace-based Pattern Formulation

In this section, we formulate transient node contact patterns based on experimental observations from realistic DTN traces. These patterns are formulated with the daily period, which has been shown as the dominant period of node contact patterns in previous trace studies [10], [21].

A. Traces

We study the transient node contact patterns on two sets of DTN traces. These traces record contacts among users carrying hand-held mobile devices on university campus. The devices are equipped with Bluetooth or WiFi interfaces, so as to detect and communicate with each other. In the MIT Reality trace [10], the devices periodically detect their peers via their Bluetooth interfaces, and a contact is recorded when two devices move close to each other. In the UCSD trace [21] which consists of WiFi-enabled devices, the devices search for nearby WiFi Access Points (APs) and associate themselves to the APs with the best signal strength. A contact is recorded when two devices are associated to the same AP. As summarized in Table I, the two traces differ in their scale, detection period, as well as the contact density and duration[1].

TABLE I
Trace summary

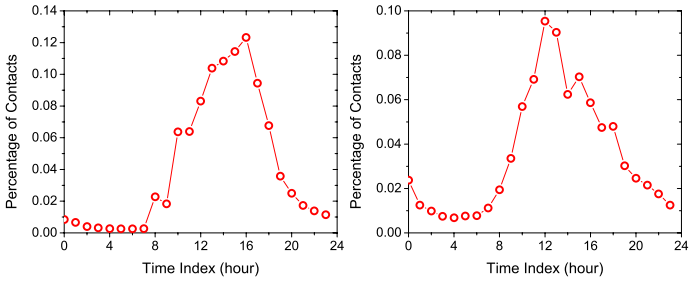
Trace	MIT Reality	UCSD
Network type	Bluetooth	WiFi
Number of devices	97	275
Number of internal contacts	114,046	123,225
Duration (days)	246	77
Granularity (secs)	120	20
Average contact duration (hours)	0.57	10.45
Average inter-contact time (hours)	84.13	47.17

B. Transient Contact Distribution

For each pair of nodes, we formulate their transient contact distribution as alternative appearances of "*on-period*" and "*off-period*". Most contacts happen during the on-periods, and only very few contacts can be found during the off-periods at random.

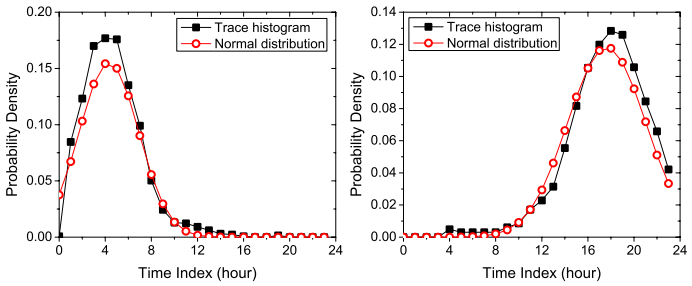
***Definition** 1: An **on-period** $[t_s, t_e]$, which is in units of hours and $0 < t_s - t_e < 24$, is uniquely determined by a set $\mathcal{S}$ of node contacts happened at time $t_1, t_2, ..., t_{|\mathcal{S}|}$, such that for any $1 < k \leq |\mathcal{S}|$, $t_k - t_{k-1} < T_{\text{on}}$, where T_{on} is a pre-defined threshold.*

[1]For any contact with the same starting and ending time, we set its duration to be a half of the detection period.



(a) MIT Reality (b) UCSD

Fig. 3. Skewed distribution of node contacts



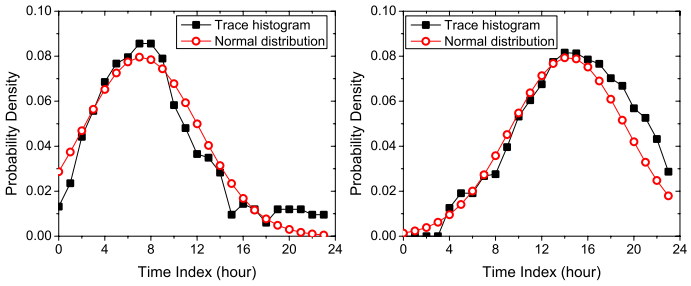
(a) On-period (b) Off-period

Fig. 4. Length distributions of on-period and off-period of the MIT Reality trace

According to Definition 1, an on-period at least contains two contacts, and we fix $T_{\text{on}} = 8$ hours, which is the normal daily working hours. Only the contact process during the on-periods is considered as stable and predictable, and is exploited to predict the node contact capability in the future for data forwarding decision. Individual contacts happened during off-periods are considered as random and unpredictable. This pattern formulation is illustrated in Figure 2, where each vertical arrow indicates a contact. $L_{\text{on},i}$ ($L_{\text{off},i}$) denotes the length of the i-th on-period (off-period). Note that an off-period may last longer than one day; e.g., a student may not contact his classmates during the entire weekend.

This pattern formulation is validated by the skewed distribution of node contacts during different hours in a day, which is shown in Figure 3. In Figure 3(a) of the MIT Reality trace, over 50% of the contacts happen between 12pm to 16pm, while only about 7% of the contacts happen between 22pm to 7am. Similar distribution is shown in Figure 3(b) for the UCSD trace. Such skewed distributions show that most of the contacts happen during the on-periods, which are generally shorter than the off-periods.

We then study the length distributions of on-period and off-period, which may vary as illustrated in Figure 2. The length of off-period is restricted in the modulus of 24 hours, and the results over all pairs of nodes are shown in Figures 4 and 5.



(a) On-period (b) Off-period

Fig. 5. Length distributions of on-period and off-period of the UCSD trace

TABLE II
NUMERIC PARAMETERS OF ON-PERIOD AND OFF-PERIOD

Trace		MIT Reality	UCSD
On-period	Mean (μ_{on})	5.3239	8.1611
	Variance (σ^2_{on})	6.5889	25.109
	Percentage of contacts (%)	89.312	85.624
Off-period	Mean (μ_{off})	18.631	15.334
	Variance (σ^2_{off})	11.387	25.191
	Percentage of contacts (%)	11.687	14.376

We have two observations. First, the length distributions of on-period and off-period are accurately approximated by normal distribution, using the mean and variance of the recorded lengths of on/off-periods as parameters. These parameters are summarized in Table II. Second, in Table II we have $\mu_{\text{on}} + \mu_{\text{off}} \approx 24$ hours for both traces, and over 85% of contacts happen during on-periods which only occupy 25% (MIT Reality) to 33% (UCSD) of the total time. These results validate that transient contact distribution can be formulated as alternative appearances of on-period and off-period.

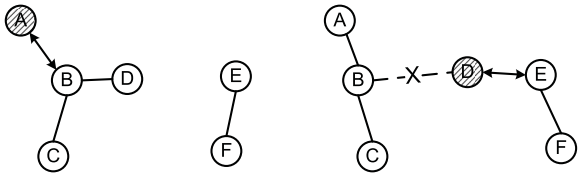
C. Transient Connectivity

The transient connectivity of a node is represented by its TCS size during different time periods.

***Definition** 2: The **Transient Connected Subnet (TCS)** of a node i during the time period $[t_1, t_2]$ is the connected graph $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$, where $i \in \mathcal{V}_i$, and for $\forall (j,k) \in \mathcal{E}_i$, the nodes j and k remain directly contacted with each other during $[t_1, t_2]$. The **size of a TCS** is then defined as $|\mathcal{V}_i|$.*

The TCS size change during different time periods is shown in Figure 6, where the lines between nodes indicate contacts. At time t_1 (Figure 6(a)), node A has indirect contacts with nodes C and D by directly contacting node B, and A's TCS therefore has 4 nodes. Later at time $t_2 > t_1$ (Figure 6(b)), the TCS size of node A reduces to 3 because node D moves away and joins another TCS containing nodes E and F.

The transient connectivity depends on the distribution of contact duration, which is shown in Figure 7. In both traces, there are a large portion of contacts with non-negligible durations. In the MIT Reality trace, there are over 20% of the contacts with durations longer than 1 hour, and this percentage in the UCSD trace is around 30%. Particularly, there are 7%



(a) Time t_1 (b) Time t_2

Fig. 6. Illustration of TCS size change

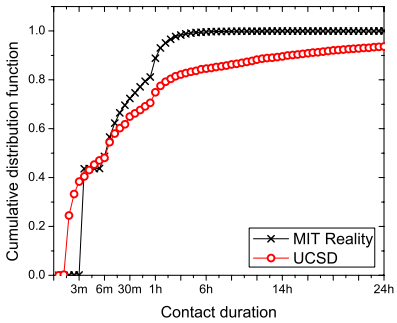


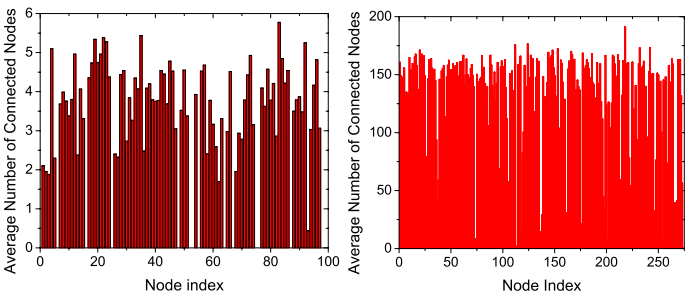
Fig. 7. Cumulative Distribution Function (CDF) of contact durations

of the contacts in the UCSD trace with durations longer than 24 hours, and these long contacts make the average contact duration to be longer than 10 hours.

Nodes that have contacts with long durations have good chances to maintain transient connectivity with others. To study transient connectivity, we replay the traces and calculate the TCS size of each node once every hour to take the average. The results in Figure 8 show that the TCS sizes of over 50% of nodes are larger than 3 in the MIT Reality trace. The transient connectivity is more prevalent in the UCSD trace, where the contacts are recorded based on AP associations.

Moreover, Figure 9 shows that the average TCS size over all the nodes during different time periods in a day can be accurately approximated by Gaussian function in the form of

$$G(t) = A \cdot e^{-\frac{(t-\mu)^2}{\sigma^2}}, \tag{1}$$

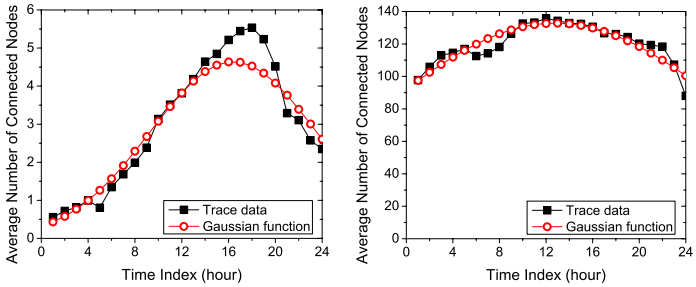


(a) MIT Reality (b) UCSD

Fig. 8. The average TCS sizes

TABLE III
NUMERIC PARAMETERS ON THE TCS SIZE DISTRIBUTION

Trace	MIT Reality	UCSD
A	4.6413	132.85
μ	16.406	12.795
σ^2	49.954	224.97



(a) MIT Reality (b) UCSD

Fig. 9. The temporal distributions of the TCS sizes

and the numerical parameters in Eq. (1) are listed in Table III. By comparing Figure 9 with Figure 3, it is easy to see that the TCS size during a specific time period is proportional to the amount of contacts happened during that time period.

In general, transient contact distribution and transient connectivity can be represented by uniform formulations based on Gaussian approximation. This uniformity facilitates us to combine the two perspectives of transient contact patterns to predict node contact capability in the future for data forwarding decision.

V. DATA FORWARDING APPROACH

A. Overview

In our approach, the data forwarding metric of a mobile node is its capability of contacting others in the future, measured by the expected number of nodes that it can contact within the given time constraint for data forwarding. This constraint is $(t_c, t_e]$ with respective to the notations in Table IV, which are used throughout this section.

It is easy to see that such metric is time-dependent. Since the metric of node i is calculated every time when a relay decides whether to forward data to i, we use C_i to indicate the metric of node i at time t_c without any loss of generality.

C_i is calculated in an accumulative manner, such that

$$C_i = \sum_{j=1, j\neq i}^{N} c_{ij}, \tag{2}$$

where N is the total number of nodes in the network, and c_{ij} indicates the expected number of nodes that i can contact within the time constraint by contacting node j. In the rest of this section, we will focus on how to calculate c_{ij} based on the transient node contact patterns.

If we only consider direct contacts among mobile nodes, c_{ij} is equivalent to the probability for node i to directly contact node j within the time constraint. We evaluate this pairwise node contact probability by exploiting the transient contact distribution of nodes i and j. The basic idea is to calculate this probability only based on the contact process during on-periods.

The node contact capability is further improved by considering the indirect contacts among nodes. c_{ij} is then defined as the expected number of nodes that i can contact in j's TCS, by having contacted with j. We predict this expected number of nodes based on the transient connectivity of node j. Particularly, by taking indirect contacts into account, a relay can decide whether to forward data to a node as long as they are within the same TCS. If a relay and the destination are within the same TCS, the data can be delivered to the destination immediately via multi-hop transmission within that TCS, and the data delivery ratio therefore can be improved.

TABLE IV
NOTATION SUMMARY

Notation	Explanation
t_c	Current time
t_e	The data expiration time
t_l	The time of the last contact happened before t_c
t_n	The time of the first contact to happen after t_c
t_{ls}	The starting time of the last on-period before t_c
t_{le}	The ending time of the last on-period starting before t_c
t_{ns}	The starting time of the first on-period after t_c
N_{on}	The total number of on-periods before t_c
N_{off}	The total number of off-periods before t_c
L_{on}	The total length of on-periods before t_c
C_{on}	The number of contacts happened during the past on-periods
λ	The average contact rate during the past on-periods

B. Characterization of Transient Contact Patterns

Each node characterizes its transient contact patterns at real-time according to the pattern formulations in Section IV. First, for each pair of nodes i and j, the parameters of their on-period and off-period are updated every time they directly contact each other. The update process is described in Algorithm 1 for the case that i contacts j at time t_c, with respective to the notations in Table IV. The quantities N_{on}, N_{off}, C_{on} and L_{on} are initialized as 0 when the network starts. t_{ls} is initialized as 0 and t_{le} is initialized as a small positive real number to ensure that the first contact between i and j starts an on-period.

Second, each node detects its TCS whenever it directly contacts another node, by broadcasting a detecting beacon message. In order to detect the TCS in a multi-hop range, such message is broadcasted among the nodes within the TCS, and each node having received the message acknowledges to the original sender. Transient connectivity is then updated by Gaussian curve fitting based on the recorded TCS sizes during different hours in past. The time needed for transmitting a beacon message is generally much shorter than the contact duration, and the transient connectivity hence can be accurately characterized. Since the broadcasting of beacon messages is only triggered by node contact events, and the sizes of beacon messages and acknowledgements are very small, such TCS detection only produces little data transmission overhead.

Algorithm 1: Updating On/Off-Periods (i, j, t_c)

if $t_c - t_l \leq T_{\text{on}}$ **then** // In an on-period
  $C_{\text{on}} \leftarrow C_{\text{on}} + 1$
  $\lambda \leftarrow C_{\text{on}}/(L_{\text{on}} + t_c - t_{ls})$
  if $t_{ls} < t_{le}$ **then** // New on-period
    $\mu_{\text{old}} \leftarrow \mu_{\text{off}}$
    $\mu_{\text{off}} \leftarrow \frac{\mu_{\text{off}} \cdot N_{\text{off}} + t_l - t_{le}}{N_{\text{off}}+1}$
    $\sigma^2_{\text{off}} \leftarrow \frac{N_{\text{off}}(\sigma^2_{\text{off}} + \mu^2_{\text{old}}) + (t_l - t_{le})^2}{N_{\text{off}}+1} - \mu^2_{\text{off}}$
    $N_{\text{off}} \leftarrow N_{\text{off}} + 1$
    $t_{ls} \leftarrow t_l$
if $t_c - t_l > T_{\text{on}}$ **then** // In an off-period
  if $t_{ls} > t_{le}$ **then** // New off-period
    $\mu_{\text{old}} \leftarrow \mu_{\text{on}}$
    $\mu_{\text{on}} \leftarrow \frac{\mu_{\text{on}} \cdot N_{\text{on}} + t_l - t_{ls}}{N_{\text{on}}+1}$
    $\sigma^2_{\text{on}} \leftarrow \frac{N_{\text{off}}(\sigma^2_{\text{on}} + \mu^2_{\text{old}}) + (t_l - t_{ls})^2}{N_{\text{on}}+1} - \mu^2_{\text{on}}$
    $N_{\text{on}} \leftarrow N_{\text{on}} + 1$
    $t_{le} \leftarrow t_l$
    $L_{\text{on}} \leftarrow L_{\text{on}} + t_{le} - t_{ls}$
$t_l \leftarrow t_c$

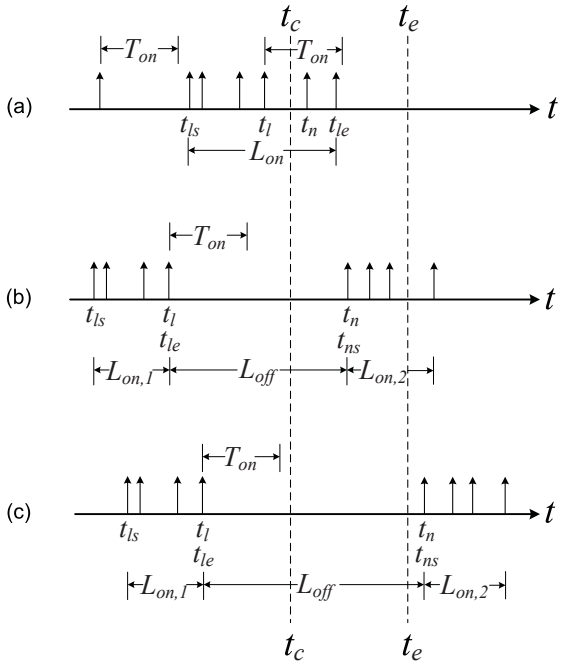


Fig. 10. Cases of predictable node contacts

C. Pairwise Contact Probability

When only direct contacts among mobile nodes are considered for data forwarding, c_{ij} at time t_c is equivalent to the probability p_{ij} for node i to contact node j during $(t_c, t_e]$. We consider that only the contact process during on-periods are stable and predictable, and p_{ij} is calculated based on the past contact records during the on-periods. Since the time period $(t_c, t_e]$ in this paper is shorter than one day, node j can be contacted by node i during $(t_c, t_e]$ in the following two cases:

1) The current time t_c is within an on-going on-period which continues after t_c, as illustrated in Figure 10(a).
2) The current time t_c is within an on-going off-period, but the next on-period will start before t_e, as illustrated in Figure 10(b).

In general, a predictable contact only occurs if at least one on-period overlaps with $(t_c, t_e]$, and the occurrence probability is proportional to the overlapping length. In the first case, we estimate the remaining duration $(t_c, t_{le}]$ of the current on-period after t_c, and in the second case we predict the starting time t_{ns} of the next on-period after t_c. In contrast, node i and j will not contact during $(t_c, t_e]$ in the case illustrated by Figure 10(c), where $(t_c, t_e]$ is totally included in an off-period.

The three cases illustrated in Figure 10 are complete and mutually exclusive for predictable contacts. Therefore, we have $p_{ij} = p^{(1)}_{ij} + p^{(2)}_{ij}$, where $p^{(1)}_{ij}$ and $p^{(2)}_{ij}$ are the contact occurrence probabilities for the cases 1) and 2), respectively. As a prerequisite, we define $p_c(t_1, t_2)$ as the probability that node i and node j contact during the time period $[t_1, t_2]$ within an on-period. Since the contact process between node i and j during on-periods is considered as stable and predictable, we assume that this contact process is a homogeneous Poisson process with the parameter λ, which is updated at real-time according to Algorithm 1. As a result, we have

$$p_c(t_1, t_2) = 1 - e^{-\lambda(t_2 - t_1)}. \tag{3}$$

*1) **Case 1**:* For $\forall t_{le} \in (t_c, t_e]$, node i has the probability $p_c(t_c, t_{le})$ to contact node j. Therefore, due to the memoryless nature of Poisson process we have

$$p^{(1)}_{ij} = \int_0^T p_c(0, T) \cdot f_{\text{on}}(t) dt, \tag{4}$$

where $T = t_e - t_c$, and $f_{\text{on}}(t)$ is the Probability Density Function (PDF) of the length distribution of the on-periods between node i and j, such that

$$f_{\text{on}}(t) = \frac{1}{\sqrt{2\pi}\sigma_{\text{on}}} \cdot e^{-\frac{(t + t_c - t_{ls} - \mu_{\text{on}})^2}{2\sigma^2_{\text{on}}}}. \tag{5}$$

By substituting Eqs. (3) and (5) into Eq. (4), we have

$$\begin{aligned} p^{(1)}_{ij} &= \frac{1}{\sigma\sqrt{\pi}} \cdot \int_0^T (1 - e^{-\lambda t}) \cdot e^{-\frac{(t-\mu)^2}{\sigma^2}} dt \\ &= \frac{1}{2}\Bigg(\text{erf}(\frac{T-\mu}{\sigma}) + \text{erf}(\frac{\mu}{\sigma}) - \\ &\quad e^{\frac{1}{4}\lambda^2\sigma^2 - \lambda\mu} \cdot \Bigg(\text{erf}(\frac{\lambda\sigma^2 + 2T - 2\mu}{2\sigma}) - \text{erf}(\frac{\lambda\sigma^2 - 2\mu}{2\sigma})\Bigg)\Bigg), \end{aligned} \tag{6}$$

where $\mu = \mu_{\text{on}} + t_{ls} - t_c$, $\sigma = \sqrt{2}\sigma_{\text{on}}$, and $\text{erf}(x)$ is the Gaussian error function [1].

*2) **Case 2**:* Similarly we have

$$p^{(2)}_{ij} = \mathbb{P}(t_{le} < t_c) \cdot \int_0^T (1 - e^{-\lambda(T-t)}) \cdot f_{\text{off}}(t) dt,$$

where

$$f_{\text{off}}(t) = \frac{1}{\sqrt{2\pi}\sigma_{\text{off}}} e^{-\frac{(t+t_c-t_{le}-\mu_{\text{off}})^2}{2\sigma_{\text{off}}^2}},$$

and

$$\mathbb{P}(t_{le} < t_c) = \Phi(\frac{t_c - t_{ls} - \mu_{\text{on}}}{\sigma_{\text{on}}})$$

with $\Phi(x)$ indicating the Cumulative Distribution Function (CDF) of standard normal distribution. Therefore,

$$p_{ij}^{(2)} = \frac{\Phi(\frac{t_c-t_{ls}-\mu_{\text{on}}}{\sigma_{\text{on}}})}{\sigma\sqrt{\pi}} \cdot \int_0^T (1 - e^{-\lambda(T-t)}) \cdot e^{-\frac{(t-\mu)^2}{\sigma^2}} dt$$
$$= \frac{1}{2} \cdot \Phi(\frac{t_c - t_{ls} - \mu_{\text{on}}}{\sigma_{\text{on}}}) \cdot \left(\text{erf}(\frac{T-\mu}{\sigma}) + \text{erf}(\frac{\mu}{\sigma}) - \right.$$
$$\left. e^{\frac{1}{4}\lambda^2\sigma^2 - \lambda(T-\mu)} \cdot \left(\text{erf}(\frac{\lambda\sigma^2 + 2\mu}{2\sigma}) - \text{erf}(\frac{\lambda\sigma^2 + 2\mu - 2T}{2\sigma}) \right) \right), \quad (7)$$

where $\mu = \mu_{\text{off}} + t_{le} - t_c$, $\sigma = \sqrt{2}\sigma_{\text{off}}$.

D. Exploiting Transient Connectivity

We improve the node contact capability by considering the indirect contacts among nodes within the same TCS, and we estimate c_{ij} by applying the transient connectivity of node j to the two cases analyzed in Section V-C.

In Section IV-C, the change of the TCS size of node j during different hours in a day is described by the Gaussian function

$$N_G(t) = A_G \cdot e^{-\frac{(t-\mu_G)^2}{\sigma_G^2}}.$$

Hence, for Case 1 in Section V-C, Eq. (4) is rewritten as

$$c_{ij}^{(1)} = \int_0^T p_c(0,T) \cdot N_G(t) \cdot f_{\text{on}}(t) dt.$$

Since transient contact distribution and transient connectivity are represented in uniform forms of Gaussian function, and the set of Gaussian functions is closed under multiplication, we have $c_{ij}^{(1)} = A \cdot p_{ij}^{(1)}$, where $p_{ij}^{(1)}$ takes the similar form as in Eq. (6), with the only difference that

$$\sigma = \sqrt{\frac{\bar{\sigma}_{\text{on}}^2 \cdot \sigma_G^2}{\bar{\sigma}_{\text{on}}^2 + \sigma_G^2}}, \quad (8)$$

$$\mu = \frac{\sigma_G^2 \bar{\mu}_{\text{on}} + \bar{\sigma}_{\text{on}}^2 \mu_G}{\sigma_G^2 + \bar{\sigma}_{\text{on}}^2}, \quad (9)$$

and

$$A = A_G \cdot e^{-\frac{(\mu_G - \bar{\mu}_{\text{on}})^2}{\sigma_G^2 + \bar{\sigma}_{\text{on}}^2}}, \quad (10)$$

where $\bar{\mu}_{\text{on}} = \mu_{\text{on}} + t_{ls} - t_c$, $\bar{\sigma}_{\text{on}} = \sqrt{2}\sigma_{\text{on}}$.

Similarly for Case 2 in V-C, $c_{ij}^{(2)} = A \cdot p_{ij}^{(2)}$. Note that $p_{ij}^{(2)}$ takes the similar form as in Eq. (7), and the quantities $\bar{\mu}_{\text{off}} = \mu_{\text{off}} + t_{le} - t_c$ and $\bar{\sigma}_{\text{off}} = \sqrt{2}\sigma_{\text{off}}$ are used to substitute $\bar{\mu}_{\text{on}}$ and $\bar{\sigma}_{\text{on}}$ in Eqs. (8)-(10). Finally we have $c_{ij} = c_{ij}^{(1)} + c_{ij}^{(2)}$.

E. Analysis on Prediction Error

The prediction error on node contact capability mainly comes from the node contact randomness in the following two perspectives, which may not follow the node contact patterns.

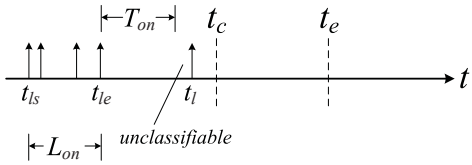


Fig. 11. Illustration of an unclassifiable contact

1) Unclassifiable Contacts: We cannot classify a contact into one of the two cases shown in Figure 10, if it happens in the case shown in Figure 11. Such a contact is the last one happened before t_c, but does not belong to the last recorded on-period. Hence, we cannot determine whether this contact starts a new on-period, or it just happens at random during the off-period starting at t_{le}. Especially when $t_c - t_{le}$ approaches μ_{off}, the calculation of contact probability for either case in Figure 10 may produce non-negligible prediction error.

2) Long Off-periods: In practice, the lengths of some off-periods may exceed 24 hours. For example, two classmates may have no contact during the weekend. The long off-periods affect the prediction accuracy of Case 2 in Figure 10, by postponing the next on-period to happen after t_e.

Nevertheless, the results in Section IV-B show that the alternative appearances of on-period and off-period accurately fit the daily period, and over 85% of node contacts happen during on-periods. Therefore, the unclassifiable contacts and long off-periods are only occasionally found among the nodes with low contact frequency. Since in our approach data is mainly forwarded among the nodes with high contact frequency, the aforementioned prediction error can be effectively eliminated.

VI. Performance Evaluation

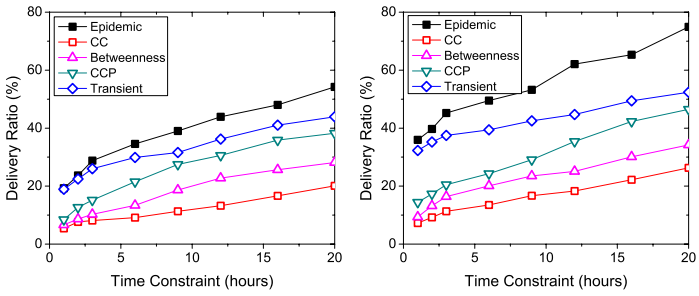
In this section, we compare the performance of our data forwarding approach with existing data forwarding schemes based on cumulative node contact characteristics.

A. Simulation Setup

Our evaluations are performed on the DTN traces described in Section IV-A. We first randomize the data generation time, and then randomly pick the data sources and destinations among the nodes with non-zero contact counts within the time constraint for data forwarding. The transient contact patterns are characterized in real-time as described in Section V-B.

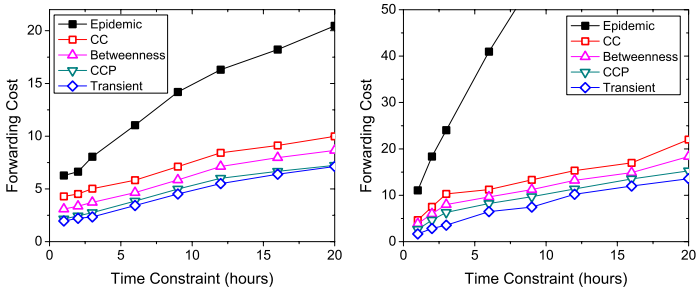
We evaluate the performance of our approach in data delivery ratio and forwarding cost measured by the number of data copies created in the network, and each experiment is repeated 500 times for statistical convergence. The data delivery delay is not considered, as long as the data can be delivered on time. We compare our data forwarding metric with the following existing metrics based on cumulative contact characteristics:

- **Contact Counts (CC)** is calculated cumulatively since the network starts. It was used in [3] for data forwarding decision.
- **Betweenness** evaluates the social importance of a node facilitating the communication among others. It was used in [8], [15] for social-based data forwarding decision.



(a) MIT Reality (b) UCSD

Fig. 12. Delivery ratio with the Compare-and-Forward strategy



(a) MIT Reality (b) UCSD

Fig. 13. Forwarding cost with the Compare-and-Forward strategy

- **Cumulative Contact Probability (CCP)** [13] evaluates the probability of a node contacting others based on its cumulative contact rates.

These data forwarding metrics are applied to the following forwarding strategies. The performance of Epidemic routing is also evaluated as a basic benchmark.

- **Compare-and-Forward**: it has been used in FRESH [9], where a relay always forwards data to the nodes whose forwarding metrics are higher than that of the relay itself.
- **Delegation Forwarding** [11]: each relay records the highest value of data forwarding metric it has ever seen, and only forwards data to the nodes whose forwarding metric is higher than the recorded highest value.
- **Spray-and-Wait** [22]: the maximum number of data copies in the network is fixed as K. The source forwards data to K nodes, such that for $\forall j \in [1, K)$, the $(j+1)$-th node has a higher forwarding metric than the j-th node has. Each relay then follows the Compare-and-Forward strategy, and deletes its local data copy after having forwarded the data to another node.

B. Performance Comparison

We first evaluate the data forwarding performance of our approach with the Compare-and-Forward strategy. For fairness, indirect contacts are exploited in all the schemes for comparison, such that the data is delivered from a relay to the destination as long as they indirectly contact each other.

The results are shown in Figures 12 and 13. Generally speaking, our approach achieves higher data delivery ratio. In cases of short time constraints, the transient contact characteristics of nodes differ more from their cumulative characteristics, and the advantage of our approach is therefore larger. In the MIT Reality trace, as shown in Figure 12(a), when the time constraint is shorter than 3 hours, our approach achieves similar delivery ratio with that of Epidemic, and performs over 100% better than CCP. The performance of other forwarding metrics, including CC and Betweenness, is lower due to their coarse estimation of node contact capability. In the UCSD trace, our approach benefits from the prevalent transient connectivity shown in Figures 8(b) and 9(b). Hence, in Figure 12(b), our approach performs as better as 150%-200% when the time constraint is shorter than 6 hours.

When data forwarding lasts longer, the transient and cumulative node contact characteristics tend to be consistent, and the advantage of our approach decreases accordingly. Nevertheless, during a long time period up to 20 hours, our approach still performs at least 20% better.

Figure 13 shows that the forwarding cost of our approach is similar with that of CCP, and is generally 15%-20% lower than that of CC and Betweenness. When data forwarding lasts longer, the data source and relays have higher chance to contact other nodes, and the forwarding cost increases accordingly in all the schemes.

We also perform the performance evaluation using the Spray-and-Wait forwarding strategy with $K = 5$. The results are shown in Figures 14 and 15. Similar to Figure 12, our approach achieves much higher data delivery ratio, especially in the UCSD trace. The limit on the maximum number of data copies in the network leads to a decrease on delivery ratio, which can be up to 20%-30% compared to that with the Compare-and-Forward strategy. This limit also significantly reduces the data forwarding cost, as shown in Figure 15. In cases of short time constraints, the data source usually cannot contact and forward data to the maximally allowed K nodes, and hence the actual forwarding cost is much lower than K.

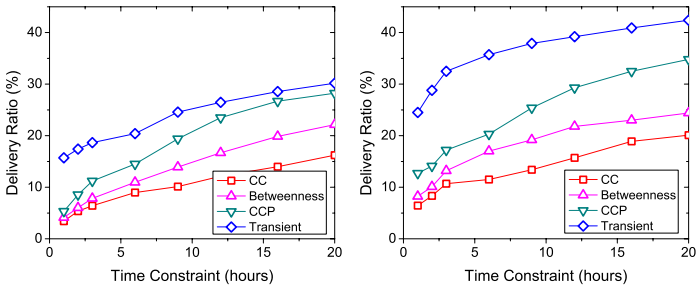
In summary, the Compare-and-Forward strategy leads to higher data delivery ratio by fully exploiting node contacts for data forwarding, and the Spray-and-Wait strategy achieves better data forwarding cost-effectiveness by limiting the data forwarding cost. Our approach significantly outperforms other schemes with both forwarding strategies by exploiting transient node contact patterns.

C. Impact of Transient Contact Distribution

In our approach, transient contact distribution is exploited to estimate the pairwise node contact probability, and therefore has obvious impact on the data forwarding performance.

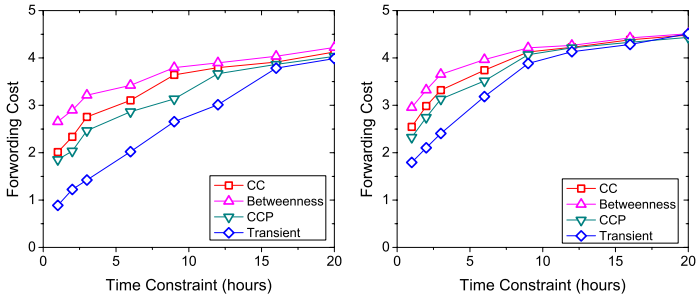
1) Different Values of T_{on}: We first evaluate the impact of different values of the on-period detection threshold T_{on}. In the previous sections T_{on} is fixed as 8 hours, which is validated by Figure 16 as the optimal value to achieve the highest delivery ratio. The reason is that both traces are collected from university campus, where the 8-hour working day is applied.

The delivery ratio decreases when the value of T_{on} deviates from 8 hours, and this decrease is related to the time constraint



(a) MIT Reality (b) UCSD

Fig. 14. Delivery ratio with the Spray-and-Wait strategy



(a) MIT Reality (b) UCSD

Fig. 15. Forwarding cost with the Spray-and-Wait strategy

T_L for data forwarding. The delivery ratio is more sensitive to the change of T_{on} when T_L is small. As shown in Figure 16, in case of $T_L = 3$ hours, the delivery ratio decreases by 50%-75% when the value of T_{on} changes to the smallest (1 hour) or the largest (20 hours). When T_L increases to 12 hours, the delivery ratio only decreases by 15%-25% in similar cases.

2) Excluding Transient Connectivity: To evaluate the standalone impact of transient contact distribution on data forwarding performance, we exclude the consideration of transient connectivity (TC) and indirect contacts from the data forwarding decision of our approach. More specifically, we calculate the data forwarding metric C_i of node i only based on its pairwise contact probabilities p_{ij}, which are calculated in Section V-C. We compare the data forwarding performance of our approach to the CCP forwarding metric with the Delegation forwarding strategy. The other forwarding metrics, including CC and Betweenness, are not considered for comparison due to their low performance shown in Figures 12 and 14.

The results on the MIT Reality trace are shown in Figure 17. When data forwarding is only based on the direct contacts among mobile nodes, the delivery ratio and forwarding cost of both our approach and CCP decrease accordingly, due to the reduction on the nodes' capability of contacting others. Nevertheless, as shown in Figure 17(a), our approach still outperforms CCP by as high as 30%-50% when the time constraint is shorter than 6 hours. This result shows that our approach provides more accurate estimation on the pairwise node contact probabilities, compared to the estimation made by CCP based on cumulative node contact characteristics.

3) Two Cases for Predicting Contacts: In Section V-C, node contacts in the future are predicted in two cases, and we evaluate the impact of each case on the data forwarding performance. To do this, whenever a relay contacts the destination, we classify this contact into one of the two cases.

The results are shown in Figure 18. When the time constraint is short ($<$ 3 hours), over 70% of successful data delivery benefits from contacts of Case 1, which indicates that a currently on-going on-period continues in future. The major reason is that the accuracy of predicting the future on-period occurrences, which corresponds to Case 2, is generally low during such a short time period. When the time constraint increases, the prediction accuracy of Case 2 can be significantly improved, and therefore the importance of the two cases on successful data delivery is gradually balanced.

VII. Discussions

A. Node Buffer Constraint

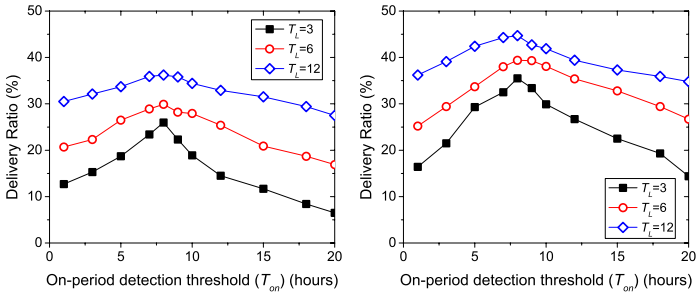
As stated in Section III, the node buffer constraint is not considered in this paper, and the reason is as follows. When only one data item is forwarded in the network, a relay only drops data if it does not have enough buffer to carry the data, and hence considering the buffer constraint is trivial. When multiple data items are forwarded in the network, [13] has shown that data forwarding with the consideration of buffer constraint is formulated as a knapsack problem, where various data forwarding metrics can be applied integrally. In such cases, the consideration of node buffer constraint is orthogonal to the main focus of this paper, which is on developing appropriate data forwarding metrics. The limitation of channel bandwidth is not considered in this paper for similar reason.

B. Daily Pattern Period

In this paper, we study the transient node contact pattern with the daily period, although weekly or monthly period has been used for other trace studies [10], [21]. The major reason is that we focus on forwarding data within a short time constraint, in which the transient contact characteristics of nodes differ more from their cumulative contact characteristics, and the existing data forwarding schemes will not perform well in this case. On the other hand, the transient and cumulative node contact characteristics tend to be consistent during a longer time period, and hence data forwarding will not benefit much by exploiting transient contact patterns with longer periods.

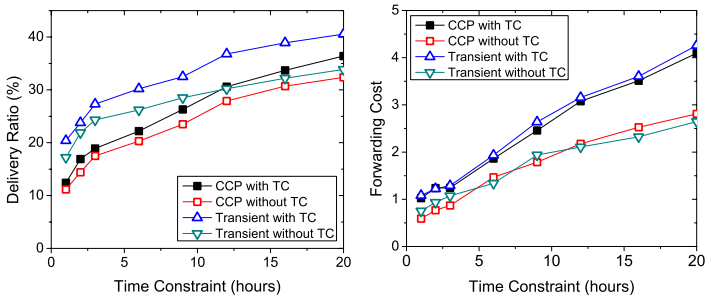
C. Transient Patterns of Social Network Characteristics

Social network characteristics in DTNs, including centrality [8] and social communities [15], are also related to transient node contact patterns. For example, when a student keeps transient connectivity with his classmates during the class, they also form a social community with transient existence. Despite this relation, existing methods for distributed community detection in DTNs [16] suggest that the transient patterns of social network characteristics are much more complicated than the aforementioned simple example. Hence, we leave the study of transient patterns of social network characteristics as



(a) MIT Reality (b) UCSD

Fig. 16. Impact of different values of T_{on}



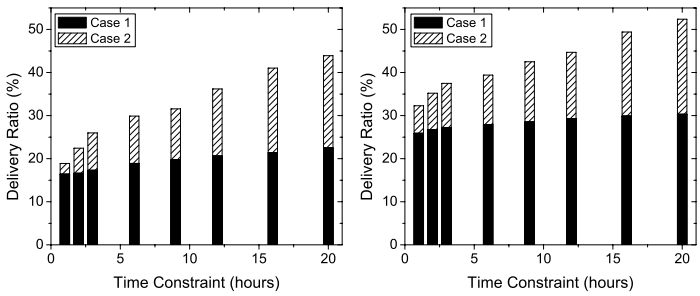
(a) Delivery Ratio (b) Forwarding Cost

Fig. 17. Data forwarding performance on the MIT Reality trace with the Delegation forwarding strategy

future work, and plan to exploit these patterns to improve the performance of social-based data forwarding in DTNs.

VIII. Conclusions

In this paper, we propose effective forwarding metrics to improve the performance of data forwarding in DTNs, by exploiting the transient node contact patterns. We formulate these patterns based on experimental observations from realistic DTN traces, and exploit these patterns for more accurate prediction on the node contact capability. Through extensive trace-driven experiments, we show that our approach significantly improves the data delivery ratio, while keeping similar forwarding cost with the existing schemes based on cumulative node contact characteristics. Future work includes the evaluation on the communication overhead for identifying the transient node contact patterns, as well as the actual impact of prediction error on the data forwarding performance.



(a) MIT Reality (b) UCSD

Fig. 18. Impacts of the two cases for predicting contacts on data delivery ratio

References

[1] M. Abramovitz and I. Stegun. Handbook of mathematical functions, 1972.

[2] A. Balasubramanian, B. Levine, and A. Venkataramani. DTN routing as a resource allocation problem. In *Proceedings of SIGCOMM*, pages 373–384. ACM New York, NY, USA, 2007.

[3] J. Burgess, B. Gallagher, D. Jensen, and B. Levine. Maxprop: Routing for vehicle-based disruption-tolerant networks. *Proc. INFOCOM*, 2006.

[4] H. Cai and D. Y. Eun. Crossing over the bounded domain: from exponential to power-law inter-meeting time in manet. *Proc. MobiCom*, pages 159–170, 2007.

[5] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of Human Mobility on Opportunistic Forwarding Algorithms. *IEEE Trans. on Mobile Computing*, 6(6):606–620, 2007.

[6] A. Chaintreau, A. Mtibaa, L. Massoulie, and C. Diot. The diameter of opportunistic mobile networks. In *Proceedings of the 2007 ACM CoNEXT conference*. ACM New York, NY, USA, 2007.

[7] P. Costa, C. Mascolo, M. Musolesi, and G. Picco. Socially Aware Routing for Publish-Subscribe in Delay-Tolerant Mobile Ad Hoc Networks. *IEEE Journal on Selected Areas in Communications*, 26(5):748–760, 2008.

[8] E. Daly and M. Haahr. Social network analysis for routing in disconnected delay-tolerant MANETs. *Proc. MobiHoc*, 2007.

[9] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli. Age matters: efficient route discovery in mobile ad hoc networks using encounter ages. *Proc. MobiHoc*, pages 257–266, 2003.

[10] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2006.

[11] V. Erramilli, A. Chaintreau, M. Crovella, and C. Diot. Delegation Forwarding. *Proc. MobiHoc*, 2008.

[12] K. Fall. A delay-tolerant network architecture for challenged internets. *Proc. SIGCOMM*, pages 27–34, 2003.

[13] W. Gao, Q. Li, B. Zhao, and G. Cao. Multicasting in delay tolerant networks: a social network perspective. In *Proceedings of MobiHoc*, pages 299–308, 2009.

[14] W. Hsu and A. Helmy. On Nodal Encounter Patterns in Wireless LAN Traces. *Proc. International Workshop On Wireless Network Measurement (WiNMee)*, 2006.

[15] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. *Proc. MobiHoc*, pages 241–250, 2008.

[16] P. Hui, E. Yoneki, S. Chan, and J. Crowcroft. Distributed community detection in delay tolerant networks. *Proc. MobiArch*, 2007.

[17] T. Karagiannis, J.-Y. Boudec, and M. Vojnovic. Power law and exponential decay of inter contact times between mobile devices. *Proc. MobiCom*, pages 183–194, 2007.

[18] V. Lenders, G. Karlsson, and M. May. Wireless ad hoc podcasting. In *Proceedings of SECON*, pages 273–283, 2007.

[19] Q. Li, S. Zhu, and G. Cao. Routing in Socially Selfish Delay Tolerant Networks. *Proc. INFOCOM*, 2010.

[20] A. Lindgren, A. Doria, and O. Schelen. Probabilistic routing in intermittently connected networks. *ACM SIGMOBILE CCR*, 7(3):19–20, 2003.

[21] M. McNett and G. Voelker. Access and mobility of wireless PDA users. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(2):40–55, 2005.

[22] T. Spyropoulos, K. Psounis, and C. Raghavendra. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In *Proceedings of 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 252–259, 2005.

[23] T. Spyropoulos, K. Psounis, and C. Raghavendra. Efficient routing in intermittently connected mobile networks: The single-copy case. *IEEE/ACM Transactions on Networking*, 16(1):63–76, 2008.

[24] A. Vahdat and D. Becker. Epidemic routing for partially connected ad hoc networks. *Technical Report CS-200006, Duke University*, 2000.

[25] Q. Yuan, I. Cardei, and J. Wu. Predict and relay: an efficient routing in disruption-tolerant networks. In *Proc. MobiHoc*, pages 95–104, 2009.

[26] J. Zhao and G. Cao. VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks. *Proc. INFOCOM*, 2006.