DSearching: Using Floating Mobility Information for Distributed Node Searching in DTNs

Kang Chen, Haiying Shen, and Li Yan
Department of Electrical and Computer Engineering
Clemson University, Clemson, SC 29631
Email: {kangc, shenh, lyan}@clemson.edu

Abstract*—In delay tolerant networks (DTNs), enabling a node to search and find an interested mobile node is an important function in many applications. However, the movement of nodes in DTNs makes the problem formidable. Current node searching methods in disconnected networks mainly rely on fixed stations in the network and infrastructure-based communication to collect node position information, which is difficult to implement in DTNs. In this paper, we present DSearching, a distributed mobile node searching scheme for DTNs that requires no infrastructure. In DSearching, the entire DTN area is split into sub-areas, and each node summarizes its mobility information as both transient sub-area visiting record and long-term movement pattern. Upon arriving at a sub-areas, a node generates a new visiting record for the sub-area and distributes it to nodes that are likely to stay in the previous sub-area, so that visiting records forms a chain for the locators to trace it. Each node also stores different parts of its long-term mobility pattern to long-staying nodes in different sub-areas for others to trace it when visiting records are absent. Considering that nodes in DTNs usually have limited resources, DSearching constrains the communication and storage cost in the information distribution while enabling efficient node searching. Advanced extensions that can further improve the searching efficiency is also proposed in this paper. Extensive trace-driven experiments with real traces demonstrate the high efficiency and high effectiveness of DSearching.

I. Introduction

In recent years, delay tolerant networks (DTNs) [1] have attracted significant research interests. In DTNs, nodes are sparsely distributed, and no end-to-end connection can be ensured, Therefore, enabling a node to search and find another interested node is an important function for node management and many applications in DTNs. For example, in the DTN that exploits sensors on animals (e.g., ZebraNet [2]) for various purposes (e.g., environment monitoring and animal tracking), we may need to find a specific sensor to upgrade or repair it. For a DTN in battlefield, the system administrator needs to not only isolate a misbehaving device but also find the owner who carries the device. In the DTN formed by mobile device users, the node search function can support allowing a person to find and meet another person distributively. Figure 1 demonstrates the problem of node searching in DTNs, in which the *locator node* searches for the *target node*.

This paper addresses the node searching problem in DTNs embracing social network properties: nodes move with certain patterns and have skewed visiting preferences [3]–[5]. Such properties can be found in many scenarios. For example, in the DTN consisting of wild animals, animals usually move between different places for food, water, and flock gathering with certain patterns. For DTNs in a battlefield, vehicles and soldiers mainly move on specific routes. In a DTN consisting of mobile devices on campus, device holders (i.e., students) often visit certain buildings repeatedly , e.g., library, department building, and dorm.

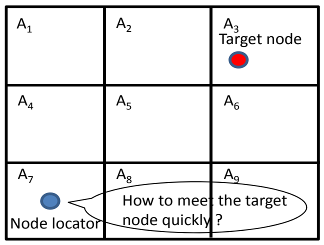


Fig. 1: Demonstration of node searching in DTNs.

There are already some works [2], [6]–[8] for object tracking (i.e., animal and people) under the context of DTNs or mobile networks. In these methods, the position or mobility information of a target object is proactively collected and gathered to search for it. Specifically, these methods require either central stations [2], [7], [8] or infrastructure-based communication (i.e., GPSR [8] and satellite communication [6]) to collect position information. Though these techniques are possible and even common in certain scenarios (e.g., campus), they are costly to use and even impractical in DTNs designed for certain areas, e.g., rural and mountain areas.

Therefore, decentralized node searching method is favorable in DTNs. One intuitive way for node searching is to follow the DTN routing algorithms [4], [9]–[14]. These methods deduce a node's probability of meeting other nodes and forward a message to nodes that have higher probability of meeting their destinations. Then, the node locator can find the target node by following the movement of the holders of the message destined for the target node. However, this scheme may lead to a long delay because it only provides indirect mobility information of the target node (i.e., other nodes' probability of meeting the target node). Without knowing the direct mobility information of the target node, a node locator cannot move actively toward the target node for efficient node searching.

In DTNs, nodes meet opportunistically with limited communication range, leading to frequent network partitions and posing great challenges on the retrieval of node mobility information. Though the target node can be sensed by its neighbors, its position information can hardly be forwarded to the locator quickly. Further, the limited resources on mobile nodes make the broadcasting of mobility information or the storage of all mobility information on every node not applicable in DTNs. As a result, a distributed and lightweight algorithm is needed to distribute node mobility information in the network and make such information easily accessible for locators.

In this paper, we propose DSearching, a distributed and lightweight node searching algorithm in DTNs. The design of DSearching is based on two social network properties in DTNs. First, mobile nodes in DTNs usually exhibit certain movement patterns [5]. Indeed, one previous research reveals that the mobility of mobile devices carried by students on a campus is predictable [15]. Second, nodes always visit certain places and stay there for a relatively long time [4].

In DSearching, the entire DTN area is split into sub-areas, and each node collects its mobility information as both transient sub-area visiting records and long-term movement patterns between sub-areas. Both types of mobility information are distributed to selected nodes to control the overhead and meanwhile support effective node searching. The transient sub-area visiting record indicates where the node just visits. Each node distributes its visiting record for a newly entered sub-area to nodes that are more likely to stay in the previous sub-area, so that the visiting records form a chain to enable locators to know targets' movement paths. The long-term mobility pattern of a node in a sub-area reflects how it transits to next sub-areas in general. Each node distributes its long-term mobility pattern in a sub-area to long-staying nodes in the sub-area, and distributes its mobility pattern in its home-area to all sub-areas. Then, when the visiting records are absent, such information can be used to search for the missing VRs to recover the target node's movement path.

In summary, our contributions are threefold:

(1) We propose a lightweight method to distribute and store node mobility information on mobile nodes, which enables a locator to easily access such information.
(2) We propose a node searching scheme that can efficiently and timely find a target in a decentralized manner.
(3) We have conducted extensive trace-driven experiments to show the efficiency and effectiveness of DSearching.

The remainder of this paper is organized as follows. Related work is introduced in Section II. Section III presents the detailed design of DSearching. In Section IV, the performance of DSearching is evaluated through trace-driven experiments. Section V concludes this paper with future work.

II. Related Work

Tracking Objects in Disconnected Networks. Tracking objects in the network without constant connections has been studied in several previous works [2], [3], [6]–[8]. In SOMA [3], each node utilizes its previous records on encountering nodes and places to predict the places it is going to visit. ZebraNet [2] tracks Zebras in Kenya by configuring tracking collars on them, which record their positions and send the data back to the central station through hop-by-hop broadcasting. Cenwits [7] and SenSearch [8] aim to search people in wilderness areas without a connected network. They both utilize the opportunistic encountering among nodes to forward their position information to stations or access-points in the network. The work in [6] utilizes the flock behavior to reduce costs needed to track sheep in wild areas. It basically lets the flock leader monitor and report the positions of other sheep to the server through GPRS or satellite communication. Different from these methods, DSearching does not require central stations, GPRS or satellite communication to collect position information, but enables a node locator to actively find the target node in a completely distributed manner.

Routing in DTNs. Routing in DTNs has been widely studied in recent years [4], [9]–[14]. These methods can generally be classified into two groups: probabilistic routing methods [9]–[12] and geographical routing methods [4], [13], [14]. In the former group, RAPID [9] and MaxContribution [10] predict a node's future encountering probability with the destination based on previous encountering records and forward packets to nodes with higher probability of meeting their destinations. They also specify the forwarding or storage priorities of each packet based on their delivery probabilities. BUBBLE [11] and SimBet [12] further use social factors (i.e., centrality and similarity with the destination node) to deduce a node's probability of meeting a packet's destination.

In the latter group of methods, GeoDTN [13] predicts node encountering possibility based on previous movement and forwards a packet to nodes that are more likely to meet the destination node. GeoOpps [14] deduces each possible route's minimal estimated time of delivery (METD) to reach the closest point to the destination and forwards packets to nodes that lead to smaller METD. LOOP [4] exploits mobility patterns of mobile nodes to predict their future movement to forward packets to certain areas in the network.

These algorithms only provide indirect information about the target node's position or mobility, i.e., how other nodes meet it. Therefore, they lead to low efficiency on node searching. On the contrary, DSearching uses a novel set of data structures to directly depict a node's mobility information. Therefore, DSearching enables a locator to actively move towards the target node, leading to more efficient node searching.

III. System Design

We assume a DTN with n nodes denoted by N_i ($i = 1, 2, ...n$). We regard the mobility of a node as the transits between sub-areas (different places) in the network. This raises a question on how to decide sub-areas? Clearly, the more sub-areas are partitioned, the more overhead is needed on maintaining node status, but also the more accurate depict of node mobility. Considering that the sub-area division is uniform for all nodes, we decide sub-areas based on popular places that are frequently visited by all nodes.

Specifically, we first select popular places in DTNs, which are common in DTNs with social network properties, e.g., villages in the DTNs on maintain or rural areas. Then, we partition the entire DTN area into sub-areas by below rules:

- Each sub-area contains only one popular place.
- The area between two popular places is evenly split to the two sub-areas containing the two places.

- There is no overlap among sub-areas.

Also, sub-areas with size smaller than a threshold (i.e., minimal sub-area size) are merged with the smallest neighbor sub-area. This is to prevent too many small sub-areas. The popular places and minimal sub-area size are decided by the network/application administrator.

The area partition is completed off-line to generate an area map with sub-areas. Each node is configured with the area map when it initially joins in the DTN with the DSearching enabled. This means that the map does not scale on individual node's visiting preference, and all nodes have the same map. Therefore, DSearching needs the application scenario information, i.e., popular places and general node mobility information, before applying it to a DTN. Though all sub-areas in the illustration in this paper have regular shapes, i.e., squares, they can be any shapes. Each sub-area is stored as the set of positions of its vertexes.

We assume that each node has a GPS in DSearching. Thus, each node can know the sub-areas in which it is located. The GPS is different from the infrastructures since it can easily be installed on mobile devices. Furthermore, a node does not need to query the GPS frequently to save energy, i.e., it can query the position only when finding it may have transited to a new sub-area. Other energy efficient localization techniques that use other signals such as WiFi can also be adopted to reduce energy consumption for this purpose.

A. Representation of Node Mobility

In DSearching, each node records its mobility information for locators to search for it. First, to enable the locator to know its actual movement path, each node leaves transient sub-area visiting records as hints along the movement path. This is like the phenomenon in which an ant leaves a trail of pheromone as it looks for food for others to follow. Second, to enable the locator to find its regular movement pattern, each node creates a mobility pattern table (MPT) to summarize its staying in different sub-areas and its transits between sub-areas. We introduce the two types of mobility information below.

1) Transient Visiting Record: Each node leaves a hint, i.e., visiting record (VR), when it enters a new sub-area:

$$VR :< N_i, A_{new}, A_{prev}, Time, T_s, Seq >$$

where N_i is node ID, A_{new} and A_{prev} are the newly entered sub-area and the previous sub-area, respectively, $Time$ denotes the time when the node finds that it enters A_{new}, T_s is the TTL of the VR, and Seq is the sequence number. Seq increases by 1 when a new VR is created. Considering that these records are time sensitive, we often set T_s to a relative short period of time, i.e., half day or one day. The visiting record is designed to ensure that once the locator arrives at A_{prev}, it knows that the target node goes to A_{new} after moving out of A_{prev}.

2) Long-Term Mobility Pattern: A node's mobility pattern table (MPT) contains two types of information for each of its regularly visited sub-areas: staying probability and transit probabilities. This is because nodes usually present skewed preferences on staying at certain sub-areas or transiting from a sub-area to another. The staying probability of a node, say N_i, at a sub-area indicates how likely the node is in the sub-area and is calculated as

$$Ps_i(A_m) = D_m/D \tag{1}$$

where D_m represents the total time N_i has stayed in sub-area A_m and D is the period of time the node has lived.

The transit probabilities of a node, say N_i, at a sub-area, say A_m, describes its probability to transit to another sub-area, say A_n, in next movement, denoted $Pt_i(A_m \to A_n)$. It is calculated as below:

$$Pt_i(A_m \to A_n) = T_{mn}/T_{ma} \tag{2}$$

where T_{mn} is the number of occurrences that the node has transited from A_m to A_n and T_{ma} denotes the total number of occurrences that the node moves out of sub-area A_m.

TABLE I: Mobility pattern table on a node.

Rank	Sub-area	Staying Prob.	Next sub-areas and probabilities	Seq
1	A_5	0.50	A_8(0.8), A_2(0.2)	1
2	A_2	0.25	A_1(0.7), A_3(0.3)	1
3	A_6	0.15	A_5(1)	1
4	A_8	0.08	A_7(0.6), A_9(0.4)	1
...	...	...	...	...

After accumulating sufficient movement records, each node builds a MPT as shown in Table I. In the table, "Sub-area" records the regularly-visited sub-areas of the node, which are sorted in descending order of the staying probability. In the row for a sub-area, say A_m, the "Next sub-areas and probabilities" records the probabilities that the node moves from A_m to corresponding sub-areas. For example, $P_{ti}(A_2 \to A_1) = 0.7$, which means that the node's probability of transiting from A_2 to A_1 is 0.7. "Seq" is the sequence number of the table. It increases by 1 whenever the table is updated. Nodes update their MPTs periodically in DSearching.

B. Mobility Information Distribution

We assume nodes are non-malicious and are willing to carry the mobility information of others to support the node searching function. We leave the work on how to motivate nodes to follow rules in DSearching to future work.

1) Distribute Visiting Record: When a node, say N_i, moves from sub-area A_m to sub-area A_n, it creates a visiting record as introduced in Section III-A1, denoted VR_{imn}. Recall that the purpose of the visiting record is to enable the locator to know the movement path of the target node. Following this direction, DSearching requires that each node to distribute VRs to nodes to ensure that the discovery probability that the locator can find them in the previous sub-area (e.g., A_m for VR_{imn}) is larger than a threshold Th_d.

Specifically, we use $Pd_i(A_m)$ to denote the discovery probability to find N_i's visiting records in A_m. Then, we copy VR_{imn} to nodes in A_n that is likely to transit to A_m and has a high probability to stay in A_m to satisfy that

$$Pd_i(A_m) = 1 - \prod_{r=1}^{k}(1 - Pt_r(A_n \to A_m) * Ps_r(A_m)) \geq Th_d \tag{3}$$

where k is the number of selected nodes and $Pt_r(A_n \to A_m)$ and $Ps_r(A_m)$ denote the transit probability from A_n to A_m and the staying probability in A_m of the r-th node, respectively.

We adopt two additional strategies to ensure that above requirement is satisfied with a controllable amount of cost. Firstly, considering each visiting record has a small TTL, we can select more nodes to hold VRs. Secondly, we can relay VRs to nodes that have high probability to stay in the previous sub-area. Therefore, we set Th_d to 0.8 in this paper.

Finally, whenever a node moves from one sub-area to another sub-area, visiting records are created to leave hint in the previous sub-area. As shown later in Section III-C, these hints form a linked VR chain to help the locator find the target node along its actual path gradually and effectively

2) Distribute the MPT: DSearching utilizes the storage on mobile nodes to store the MPT in a distributed manner for node searching. Since nodes usually have limited storage, it is not practical to store a node's MPT on every node. Also, a node's MPT should be easily accessed by its locators, which can appear in any sub-aeas. Therefore, a node, say N_i, needs to choose a subset of nodes to store its MPT so that

- The locators for N_i can easily retrieve the MPT of N_i;
- The overhead for distributing N_i's MPT is controlled.

We take advantage of node mobility pattern and the mobility of locators to realize the two goals. The general method in DSearching is to only store the necessary part of a MPT in each sub-area for easy retrieval.

Storage Host List: Recall that DSearching is proposed for DTNs with social network property that each node has several places it stays for a relatively long time [4]. This means a sub-area may have nodes that often stay in it. We then define the nodes with staying probability in sub-area A_i larger than a pre-defined threshold as A_i's hosts.

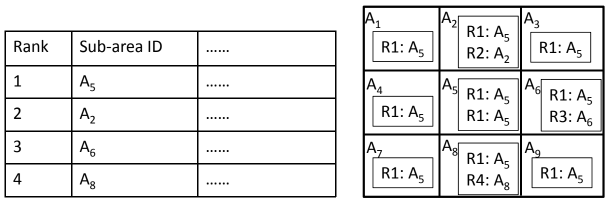
Then, in order to store a node's MPT in a sub-area, a certain number of hosts of the sub-area are selected to store the MPT. These nodes guarantee that even when nodes move continually, the probability that there is at least one copy of MPT stays in the sub-area is higher than a threshold (Th_t). In detail, each node maintains a *host list* for each sub-area containing the hosts in the sub-area that store its MPT, as shown in Table II.

TABLE II: Host list for each sub-area.

Sub-area	Host list	Staying prob.	MPT staying prob.
Sub_1	N_1, N_2, N_4	0.90, 0.8, 0.7	0.994
Sub_2	N_9	0.99	0.99
...			
Sub_M	N_3, N_4	0.6, 0.7	0.88

The "MPT staying prob." means the probability that at least one copy of MPT is in the sub-area and is calculated by $1 - \prod_{r=1}^{k}(1 - Ps_r(A_m))$, where k is the number of hosts and $Ps_r(A_m)$ is the r-th host's staying probability in A_m.

Each node determines its host lists during its movement. Specifically, suppose N_i tries to decide whether N_j can be added to its host lists. For each sub-area in the table, N_j is temporarily added to the host list for the sub-area when it satisfies: 1) N_j is a host of the sub-area with available memory, and 2) is not in the host list of the sub-area. Then, if the MPT staying probability of the sub-area is smaller than Th_t, no further action is needed. Otherwise, the host with the least staying probability is removed until the MPT staying probability reaches the minimum value that is no less than Th_t. The host list of each sub-area can also be determined off-line based on node movement patterns (similar to the sub-area division process). We set $Th_t = 0.7$ in this paper.



Rank	Sub-area ID	......
1	A_5	......
2	A_2	......
3	A_6	......
4	A_8	......

(a) Mobility pattern table (b) Mobility information distribution

Fig. 2: Distribution of MPT entries (R1 denotes Row 1).

MPT Distribution and Update: After creating the host lists from either on-line or off-line method, each node distributes its MPT to nodes in the host list for each sub-area when encountering them. In this step, DSearching does not store the entire MPT on each node. Instead, only a part of the MPT is stored on a node. Since the locator searches in a sub-area by sub-area manner, the MPT in each sub-area only needs to ensure that the locator knows where to search in the next step. For example, based on Table I, a locator in A_5 only needs to know the row for A_5 in the target's MPT in order to know where to search in the next step.

Therefore, when N_i meets a node in its host lists, say N_j, N_i decides the content to be stored in N_j as below.

- N_i copies the first row of its MPT to N_j.
- If the home sub-area of N_j, denoted $hSub(N_j)$, exists in N_i's MPT, N_i copies the corresponding row to N_j.
- When a node receives the row of a MPT that it already has, it only keeps the latest one.

We define a node's home sub-area as the sub-area corresponding to the first row in its MPT. The row for the home sub-area is distributed to all sub-areas because it is the sub-area that the target node is most likely to stay in. Algorithm 1 summarizes the above MPT distribution process. Figure 2 also shows an example of the distribution of a MPT. We see that except the first row of the MPT, the hosts in sub-areas A_5, A_2, A_6 and A_8 only store the corresponding rows in N_i's MPT.

Algorithm 1 Pseudo-code of the MPT distribution when N_i meets N_j.

```
1: procedure CHECKTODISTRIBUTEMPT($N_j$)
2:     if $N_j \in N_i$.HostList then
3:         $N_j \leftarrow$ the first row in $N_i$'s MPT
4:     else if $N_j$.HomeSubarea $\in N_i$.MPT then
5:         $N_j \leftarrow$ the row for $N_j$.HomeSubarea in $N_i$'s MPT
6:     end if
7: end procedure
```

Finally, this strategy can realize the aforementioned two goals. Firstly, wherever a locator starts searching for the target node, it can easily retrieve the needed mobility pattern information from hosts in current sub-area for efficient node searching. Secondly, each node only stores part of its MPT on a selected set of nodes, which saves communication and storage cost on the MPT distribution.

C. Node Searching

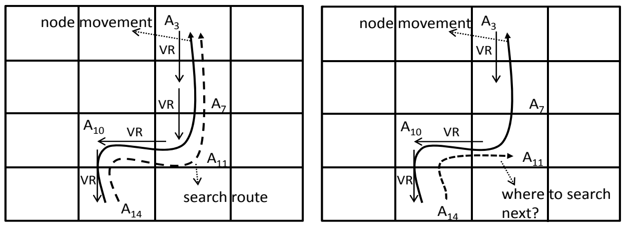
In node searching, we assume that the locator can move much faster than the target node, i.e., can pass more sub-areas in a unit time in average than the target node. This is reasonable because the locator is dedicated for node searching while the target node may sojourn at some places. When a locator moves lower than the target node, the searching efficiency will become very low in DSearching since it is mainly designed for the locators to actively approach the targets. In that case, finding a place to wait for the target node would be more effectively, which is not the focus of this paper. We also assume that when the locator arrives at a sub-area, it can effectively search around to determine whether the target node is in the sub-area with a very high probability.

1) Overview: DSearching utilizes the visiting records and MPT tables distributed in the network to search for the target node. When a locator is initialized, it first searches the home sub-area of the target. Then, considering the chain of visiting records provides information on actual node movement, the locator tries to follow the VR chain to search for the target node along its movement path. When a visiting record on the VR chain cannot be found, i.e., there is a gap on the VR chain, DSearching uses the MPT to search a valid VR that can bridge the gap. During this process, whenever a valid visiting record is obtained, the locator moves to the A_{new} indicated in it. Above process repeats until the target is found.

2) Searching Startup: When a locator is initialized, it knows nothing about the mobility of the target node and can only search randomly. However, as mentioned in Section III-B2, the first row (home sub-area) of each node's MPT table is copied to all sub-areas. Therefore, the locator can easily know the target's home sub-area. Then, the locator moves to the home sub-area of the target to search for it. The rationale behind such a design is that the target stays in the home sub-area longer than in any other sub-areas.

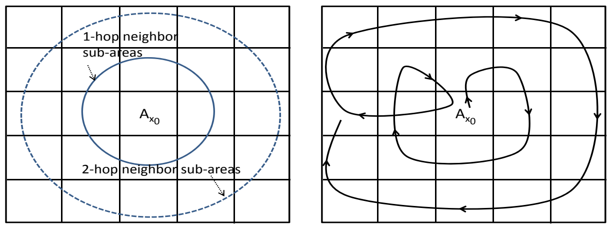
3) Node Searching with VRs: Whenever the locator discovers one or more VRs of the target node that are newer than the previous one it uses, it moves to the A_{new} in the latest VR, i.e., the one with the largest sequence number, to search for the target node. In this sub-area, if the target node cannot be found, the locator is supposed to discover the VR indicating where the target node moves to from the sub-area. When the locator finds such a VR, it again goes to search the A_{new} in the VR. Ideally, following this manner, the locator searches for the target node along its movement path indicated in a chain of VRs. As shown in Figure 3(a), the locator searches along the actual movement path of the target with the help of VRs, i.e., $A_{14} \rightarrow A_{10} \rightarrow A_{11} \rightarrow A_7 \rightarrow A_3$.

4) Node Searching without VRs: Though the design in Section III-B1 requires that a VR should exist in the previous sub-area with a high probability ($\geq Th_d$), the VRs actually are floating in the network due to node mobility. Therefore, it is common that a certain VR cannot be discovered by the locator, leading to a gap on the VR chain. As shown in Figure 3(b), the VR created in sub-area A_7 fails to reach A_{11}. Then, after searching A_{11}, the locator cannot know where to search for the next step. In this case, the MPT table and geographical limitations are jointly considered to provide a effective solution with a low cost.



(a) Ideal situation. (b) A VR is missing.

Fig. 3: Node searching with VRs.



(a) N-hop neighboring sub-areas. (b) A searching route.

Fig. 4: Node searching without VRs.

Specifically, suppose the locator fails to find the expected VR in a sub-area, say A_{x_0}, which indicates where the target moves to from A_{x_0}. Then, the locator moves around to find the missing VR or a VR created after it. We define the cost in this process as the expected number of searching hops, denoted $\mathcal{W}$. One searching hop refers to the movement from one sub-area to another sub-area. Then, the goal is *minimizing the $\mathcal{W}$ needed to find a VR that can bridge the gap on the VR chain.* In below, we first assume that when arrive at a sub-area, the locator can obtain the corresponding row for the sub-area in the target's MPT from nearby nodes, which can easily be realized based on the design in Section III-B2. We then discuss the case when such MPT information cannot be obtained later.

Optimal Solution: We first model the searching process to find the optimal searching route. Suppose there are total M sub-areas and the locator searches sub-areas in the sequence of $A_{x_1}, A_{x_2}, A_{x_3}, \cdots, A_{x_M}$. Then, the probability that a VR that can bridge the gap in the VR chain can be found in A_{x_r} ($r \in [1, M]$), denoted $Pf(A_{x_r})$, can be expressed as

$$Pf(A_{x_r}) = Pv(A_{x_r})Pd(A_{x_r}) \tag{4}$$

where $Pv_(A_{x_r})$ denotes the probability that the target has visited A_{x_r} after moving out of A_{x_0}, and $Pd(A_{x_r})$ denotes the probability that the VR created in the sub-area immediately after A_{x_r} can be found in A_{x_r}, which is introduced in Equation 3. We then have

$$\mathcal{W} = \sum_{r=1}^{M} S(A_{x_{r-1}}, A_{x_r}) * (\prod_{s=1}^{r-1}(1 - Pf(A_{x_s})))Pf(A_{x_r}) \tag{5}$$

where $S(A_{x_{r-1}}, A_{x_r})$ denotes the number of hops needed to move from $A_{x_{r-1}}$ to A_{x_r}. Therefore, the optimal searching route is a set of $A_{x_1^*}, A_{x_2^*}, A_{x_3^*}, \cdots, A_{x_M^*}$ that results in the minimal $\mathcal{W}$. However, such a route can hardly be found.

Firstly, $Pf(A_{x_r})$ cannot be obtained by the locator. For $Pv(A_{x_r})$, it should consider all possible moving paths from A_{x_0} to A_{x_r}. However, the locator can only know the target's transit probabilities in A_{x_0} from the MPT, which only shows the 1-hop path from A_{x_0} to neighboring sub-areas. For $Pd(A_{x_r})$, the information needed to calculate it (e.g., in Equation 3) cannot be known by the locator. Secondly, $S(A_{x_{r-1}}, A_{x_r})$ depends on both the locations of $A_{x_{r-1}}$ and A_{x_r}. Therefore, even when $Pf(A_{x_r})$ is known, there is no efficient way to find the optimal searching route.

A Practical Method: In order to find a practical and effective method, we first investigate the geographical limitations. We define H-hop neighbor sub-area set of a sub-area, say A_{x_0}, as the set of sub-areas that a node needs at least H hops to reach them from A_{x_0}. In Figure 4(a), the 1-hop and 2-hop neighbor sub-areas of A_{x_0} are connected by two circles, respectively. We can easily find that after moving out of A_{x_0}, the target node needs to visit at least one H-hop neighbor ($H = 1$ or 2) sub-area in order to pass through the area covered by these sub-areas. Therefore, it is highly possibility that a VR that can bridge the gap on the VR chain can be found in these sub-areas. We then limit the searching for VRs within 1-hop and 2-hop neighbor sub-areas. Considering the 1-hop sub-area set has smaller size than the 2-hop sub-area set (8 vs.16) and is closer to A_{x_0}, we let the locator first searches all 1-hop neighbor sub-areas and then all 2-hop neighbor areas.

Furthermore, we let the locator only searches sequentially along the circle connecting all H-hop sub-areas, i.e., the circles in Figure 4(a). This is because the movement from one H-hop sub-area to another non-neighboring H-hop sub-area needs at least 2 and at most $2 * H$ hops, while the sequential searching takes only $8H$ hops to search all sub-areas. With this limitation, $S(A_{x_{r-1}}, A_{x_r}) = 1$ in Equation 7, which greatly reduces the complexity of finding the most efficient route.

With above simplification, when searching the H-hop neighbor sub-area set, we only needs to determine the start sub-area and the searching direction, which has only $2 * 8 * H = 16H$ cases since each sub-area has two directions. Specifically, for 1-hop neighbor sub-area set, Equation 4 can be calculated by

$$\begin{aligned} Pf(A_{x_r}) &= Pv(A_{x_r}) Pd(A_{x_r}) \\ &= Pt(A_{x_0} \rightarrow A_{x_r}) * Th_d \end{aligned} \quad (6)$$

where $Pv(A_{x_r})$ and $Pd(A_{x_r})$ are approximated by $Pt(A_{x_0} \rightarrow A_{x_r})$ and Th_d, respectively. For the former, the simplification only considers the 1-hop transit from A_{x_0} to A_{x_r}, i.e., $Pt(A_{x_0} \rightarrow A_{x_r})$, because 1) the 1-hop transit accounts for the majority of all transits and 2) this is the only information the locator can get from the target's MPT so far. For the latter, since DSearching always try to ensure that $Pd(A_{x_r})$ is larger than Th_d, which is set to a large number, e.g., 0.8, it is acceptable to use Th_d to represent $Pd(A_{x_r})$. Then, DSearching calculate the $\mathcal{W}$ for all the 16 cases, i.e., 8 possible start sub-areas and each has two directions, by below.

$$\begin{aligned} \mathcal{W} &= \sum_{r=1}^{M} S(A_{x_{r-1}}, A_{x_r}) * (\prod_{s=1}^{r-1} (1 - Pf(A_{x_s}))) Pf(A_{x_r}) \\ &= (\prod_{s=1}^{r-1} (1 - Pf(A_{x_s}))) Pf(A_{x_r}) \end{aligned} \quad (7)$$

With all $\mathcal{W}$s, the start sub-area and searching direction that lead to the minimal $\mathcal{W}$ is selected.

After searching all 1-hop neighbor sub-areas, if the VR that can bridge the gap on the VR chain is not found, the locator then searches the 2-hop neighbor sub-area set. This process is the same as that for the 1-hop neighbor sub-areas except the calculation of $Pf(A_{x_r})$. In this case,

$$Pv(A_{x_r}) = \sum_{r=1}^{R_r} Pt(A_{x_0} \rightarrow A_{y_r}) * Pt(A_{y_r} \rightarrow A_{x_r}) \quad (8)$$

where A_{y_r} denotes the intermediate 1-hop sub-area through which the target node can move from A_{x_0} to A_{x_r}, and R_r is the number of such sub-areas. Then, the start sub-area and searching direction that lead to the minimal $\mathcal{W}$ is selected to search all 2-hop neighbor sub-areas. Note that $Pt(A_{y_r} \rightarrow A_{x_r})$ can be learn from the MPT rows for 1-hop neighbor sub-areas, which are obtained in the previous step.

Figure 4(b) demonstrates the searching of the 1-hop and 2-hop neighbor sub-areas. During this process, if a VR that can bridge the gap on VR chain can be found, the locator simply follows the VR to continue the search. If not, the locator moves randomly out of the areas covered by the 1-hop and 2-hop neighbor sub-areas to search for VRs.

Without MPT Information: Due to node mobility, it is possible that the MPT information in a 1-hop or 2-hop neighbor sub-area cannot be obtained, which means that the locator cannot know corresponding $Pt(A_{x_0} \rightarrow A_{x_r})$. In this case, we simply regard that each $Pt(A_{x_0} \rightarrow A_{x_r})$ equals to the average value, i.e., all transits have the same possibility.

D. Summary of the Behaviors of Nodes and Locators

We further summarize the behaviors of nodes and locators in DSearching. For mobile nodes, they first collect enough movement records to create the mobility pattern table as introduced in Section III-A2. Meanwhile, when a node enters a new sub-area, it creates a visiting record as introduced in Section III-A1. Both visiting records and mobility pattern tables are distributed to nodes in the network following the methods in Section III-B1 and III-B2, respectively. Algorithm 2 summarizes the behavior of node N_i.

Algorithm 2 Pseudo-code of the behavior of node N_i.

```
1:  while System.Continue() = true do
2:      if $N_i$.CheckArriveNewSubarea() = true then
3:          $N_i$.UpdateMPT()
4:          $N_i$.CreateNewVR()
5:          if meet node $N_j$ then
6:              $N_i$.CheckToDistributeMPT($N_j$) (Algorithm 1)
7:              $N_i$.CopyNewVRTo($N_j$)
8:          end if
9:      end if
10: end while
```

The locator first moves to the home sub-area of the target, which can be known from the collected MPT of the target node, to search for the target, as introduced in Section III-C2. Then, from the home sub-area, the locator follows the VR chain to search for the target along its actual movement path, as introduced in Section III-C3. In case an expected VR cannot be found, i.e., there is a gap on the VR chain, the locator follows the method in Section III-C4 to find such a VR. Once an expected VR is found, the locator again moves along the VR chain to search for the target node. Such a process repeats until the target node is found. Algorithm 3 demonstrates the behavior of a locator.

Algorithm 3 Pseudo-code of the behavior of the Locator $\mathbb{L}$ searching for target node N_i.

```
1:  while $\mathbb{L}$.TargetFind() = false do
2:      if $\mathbb{L}$.CheckArriveNewSubarea() = true then
3:          $\mathbb{L}$.SearchCurrentSubarea();
4:          if $\mathbb{L}$.TargetFind() = false then
5:              if $\mathbb{L}$ hasn't searched $N_i$.HomeSubarea() then
6:                  $\mathbb{L}$.NextSubarea $\leftarrow$ NextSubareaTo($N_i$.HomeSubarea( ))
7:              else if a VR generated in a neighbor subarea is found then
8:                  $\mathbb{L}$.NextSubarea $\leftarrow$ the neighbor subarea
9:              else
10:                 $\mathbb{L}$.NextSubarea $\leftarrow$ $\mathbb{L}$.NextSubareaWithoutVR( )
11:             end if
12:         end if
13:         $\mathbb{L}$ move to $\mathbb{L}$.NextSubarea
14:     end if
15: end while
16: procedure NEXTSUBAREAWITHOUTVR( )
17:     if the neighbor searching start sub-area and direction exist then
18:         return the next sub-area in the determined search direction
19:     else
20:         $minW \leftarrow 0$;
21:         $StartSubarea \leftarrow$ null
22:         $Direction \leftarrow$ null
23:         for each neighbor sub-area $A_k$ do
24:             $W \leftarrow$ result of Equation (7) in clock direction
25:             if $W$ ¡ $minW$ then
26:                 $StartSubarea \leftarrow A_k$
27:                 $Direction \leftarrow$ clock direction
28:             end if
29:             $W \leftarrow$ result of Equation (7) in reverse clock direction
30:             if $W$ ¡ $minW$ then
31:                 $StartSubarea \leftarrow A_k$
32:                 $Direction \leftarrow$ reverse clock direction
33:             end if
34:         end for
35:         Take $StartSubarea$ and $Direction$ as the determined start sub-
    area and direction for neighbor searching when VR is missing
36:     end if
37: end procedure
```

E. Advanced Extensions of DSearching

We further introduce additional mobility information that can be used to enhance the node searching efficiency. First, in DTNs with social network properties, the sub-areas that a mobile node frequently visits often are not independent. For example, in DTNs on campus, students are likely to go to library after finishing lectures in their department buildings; professors often go back to their offices after giving lectures. We denote such a correlation as sub-area visiting correlation. Second, the transit pattern in each day tends to be stable for many nodes. For example, a Ph.D. student often attends classes from 9AM to 10AM and 2PM to 3PM each Monday. We define such a pattern as the daily routine.

The locator can better summarize a node's staying in different sub-areas by considering its sub-area visiting correlation and daily routine. We introduce how to utilize the sub-area visiting correlation and daily routine for more efficient node searching below.

1) MPT Considering Sub-area Visiting Correlation: In the mobility pattern table described in Section III-A2, a node's transit probability from one sub-area to another sub-area is solely determined based on previous frequency of such 1-hop transits, as shown in Equation (2). In other words, it assumes that a node's next transit is independent with its previous transits. However, in some cases, such an assumption may not be accurate and may lead to misleading results.

We take the transit records of Node 27 in the Dartmouth trace (DART) [16] as an example to illustrate this point. Through analyzing the trace, we find that the frequency of transit "$AcadBldg34 \rightarrow ResBldg82$" is larger than that of the transit "$AcadBldg34 \rightarrow LibBldg4$". Thus, according to the design of current MPT, the probability that after staying in $AcadBldg34$, the node will move to $ResBldg82$ is larger than the probability that the node will move to $LibBldg4$. Nevertheless, if we consider the previous transit through which the node moves to $AcadBldg34$, the next hop transit probability would be different. Specifically, the frequency of transit "$ResBldg82 \rightarrow AcadBldg34 \rightarrow LibBldg4$" is larger than that of the transit "$ResBldg82 \rightarrow AcadBldg34 \rightarrow ResBldg82$". This means that if node 27 transits to $LibBldg4$ from sub-area $ResBldg82$, it is more likely to transit to $LibBldg4$ after staying at sub-area $AcadBldg34$. Such a case shows that the previous transit of a node tends to have a certain correlation with its next transit, which can be utilized to enhance the accuracy of node transit prediction. In below, we first introduce the improved MPT based on such conditional transit probabilities and then present how to utilize the improved MPT for more efficient node searching.

Improved MPT: With above findings, we try to integrate the previous sub-areas into the MPT to more accurately deduce the probabilities on which sub-area the node may move to in next transit. Based on our previous investigation [17], considering previous one hop transition can lead to the highest accuracy on the prediction of the next hop transit. Therefore, we mainly consider the transit probabilities in condition of the previous transit in this paper. In detail, we partition the overall transit records into two categories: the transit without previous sub-area, which is referred to as $static\ transit$, and the transit with previous sub-area, which is referred to as $dynamic\ transit$. We then design an improved mobility pattern table that contains the transit probabilities for both static transit and dynamic transit. The probability of a static transit is calculated in the same way as in Equation (2). The probability of a dynamic transit is calculated on condition of the previous transit. For instance, the transit record of a node, say N_i, at a sub-area, say A_m, describes its intention to transit to another sub-area, say A_n, in next movement. The last transit record of the node, say from sub-area A_l to A_m, describes where the node comes from. So the new transit probability can be denoted $P_{t_i}(A_m \rightarrow A_n | A_l \rightarrow A_m)$, which can be calculated as below:

$$\begin{aligned} P_{t_i}\left(A_m \rightarrow A_n | A_l \rightarrow A_m\right) &= \frac{P\left(A_l \rightarrow A_m, A_m \rightarrow A_n\right)}{P\left(A_l \rightarrow A_m\right)} \\ &= \frac{T_{lmn}/T_{N_i}}{T_{lm}/T_{N_i}} = \frac{T_{lmn}}{T_{lm}} \end{aligned} \tag{9}$$

where $P\left(A_m \rightarrow A_n, A_l \rightarrow A_m\right)$ is the node's probability of transiting from sub-area A_l to sub-area A_m and from sub-area A_m to sub-area A_n in a sequence. It is calculated as the number of the two-hop transits, denoted T_{lmn}, over the total number of transits of the node, denoted T_{N_i}; $P\left(A_l \rightarrow A_m\right)$ is the probability of the node moving from sub-area A_l to sub-area A_m, which is represented as the number of the transits from sub-area A_l to sub-area A_m, denoted T_{lm}, over T_{N_i}.

After running for some time, each node can accumulate sufficient transit records to calculate the probabilities of both static transits and dynamic transits, which are included in the improved MPT. For a node, it may come to a sub-area from several different sub-areas, which are all recorded as the previous sub-areas in the mobility pattern table. Table III presents one example of the improved mobility pattern table. In the table, for each present sub-area, the sub-row that takes "-" as the previous sub-area represents the probabilities for static transits, and each sub-row with a specific sub-area as the previous sub-area denotes the probabilities for dynamic transits. Further, the staying probability in a sub-area is calculated following the same method in the original MPT, i.e., based on the portion of staying time in the sub-area.

Utilizing the Improved MPT: The improved MPT is distributed following the same way for the original MPT, as introduced in Section III-B. We then present how to use the improved MPT to further improve the searching efficiency. Specifically, when tracing a node to a sub-area, the locator will consider the previous sub-area that the node transits to this sub-area to predict the next sub-area that the node transits to, i.e., the sub-area to search in the next step. For example, suppose Table III is the improved MPT for a target node, if the locator for the target node arrives at A_5 and finds that the target node previously moves to A_5 from A_1. Then, following the dynamic transit probabilities, the locator will move to A_6 to search for the target, rather than A_8 as indicated by the static transit probabilities.

TABLE III: Improved mobility pattern table with static and dynamic transits.

Rank	Present sub-area	Staying prob.	Previous sub-area	Next sub-areas and prob.
1	A_5	0.5	A_3	$A_8(0.74)$, $A_4(0.13)$, $A_2(0.13)$
			A_1	$A_8(0.4)$, $A_6(0.6)$
			-	$A_8(0.6)$, $A_6(0.2)$, $A_4(0.1)$, $A_2(0.1)$
4	A_2	0.25	A_9	$A_1(1.0)$
			A_4	$A_3(1.0)$
			A_5	$A_1(0.6)$, $A_3(0.4)$
			-	$A_1(0.7)$, $A_3(0.3)$
3	A_6	0.15	-	$A_5(1.0)$

TABLE IV: Routine table.

Node 0			
Entry	Time interval	Landmark ID	Prob.
1	$08:00 \sim 09:00$	$AcadBldg34$	0.43
2	$10:00 \sim 12:00$	$AcadBldg8$	0.21
$\cdots$	$\cdots$	$\cdots$	$\cdots$

2) Exploiting the Daily Routine: In this subsection, we explain how to represent and distribute nodes' daily routine information, as well as the utilization of such information for more efficient node searching.

Motivation: Through the analysis of the Dartmouth trace [16] and DieselNet AP trace [18], we found that a large part of nodes have preferences in visiting sub-areas. Moreover, the intervals of time during which nodes visit their preferreed sub-areas are relatively more stable than other rarely visited sub-areas. Take the transit preference of Node 0 as an example, the most frequently appeared transit record is from $AcadBldg34$ to $AcadBldg8$. Also, quite naturally, $Node\ 0$ often moves from $AcadBldg34$ to $AcadBldg8$ within the time interval between 1:30PM and 2:30PM. This matches with our daily experience that people usually have certain routine on visiting places such as work place, supermarket, and home.

Such routine information can help realize more efficient node searching. After discovering the daily routine of the target node, the locator can compare such information with the current system time and then decide whether to go to the places indicated in the routine information to search for the target node directly. With such an improvement, the locator can quickly move to places where the target node visits regularly, thereby reducing the searching delay.

Such routine information can potentially enable efficient node searching when the locator moves slower than target. In this case, the locator can simply mainly stay in places the target node visits regularly to meet for the target. Since this is not the focus of this paper, we do not elaborate details here.

Summarizing Routine Information: We build a routine table to summarize a node's routine information. The routine table has three types of records: $Time\ interval$, $Landmark\ ID$, and $Probability$. The $Time\ interval$ indicates during what periods of time the node has routines in visiting places. The $Landmark\ ID$ indicates where the node will possibly be in the corresponding interval of time. The $Probability$ describes the probability of staying in $Landmark\ ID$ during the $Time\ interval$. Only the time intervals with probability larger than a threshold, e.g., 0.2, have a corresponding entry in the routine table. This is because only stable routines can help node searching. The detailed structure of the routine table is shown in Table IV.

The key issue with the routine table is how to determine the list of time intervals. This is because even if a large part of nodes reflect obvious skewed preference in visiting certain places, the variance of the stay length may vary significantly. This means that the time intervals for the routinely visited landmarks tend to be not stable. As a tradeoff, the mostly overlapped time interval for a landmark will be chosen as the time interval for the routine table. Then, the probability for each time interval is calculated as

$$P_{r_i}\{Node\ 0\ is\ in\ LM_k\ during\ Interval\ i\} = \frac{T_{i,LM_k}}{T} \tag{10}$$

where the T_{i,LM_k} is the accumulated amount of time that Node 0 stays in LM_k during interval i and T is the total active time of Node 0 during interval i.

Distribution of Routine Table: Recall that we assume that nodes are willing to carry all required mobility information and exchange information as required upon meeting in this paper. However, even when nodes are cooperative, duplicating the whole routine table is exhausting and unnecessary. Then, in order to reduce storage cost and guarantee the effectiveness of the routine table entries collected by the locator, the distribution of the routine table must meet the following goals:

- The overhead for storing and transferring the routine table is acceptable for every node;
- The routine entry collected by the locator can reflect the routine of the target node as accurate as possible.

To control the storage overhead, the size of the distributed routine table should be limited. For a locator, the interval of time covering present system time is much more important than other intervals listed in the routine table. This is because the real-time location of the target is the information that the locator cares for the most. Thus, we partition the routine table into several small-sized units, which are called transient routine records. Each transient routine record consists of two entries of the routine table, which are the entry covering present system time and the entry covering the next time interval. The structure of a transient routine record is as shown in Table V, in which the current system time is 8:30AM.

TABLE V: Transient routine record.

Node ID			
Seq	0	TTL	30min
Entry	Interval of time	Landmark ID	Prob.
1	$08:00 \sim 09:00$	$AcadBldg34$	0.43
2	$10:00 \sim 12:00$	$AcadBldg8$	0.21

To disseminate the routine information of each node efficiently, every node is required to generate a transient routine record from its routine table periodically, e.g., every 1 hour, according to the system time. Then, each node only distributes the current transient routine record to nodes it meets until the generation of the next record. Thus the routine information of each node is floating in the network.

On the other hand, in order to provide reference for locator, each node maintains a first-in-first-out queue to store the transient routine records it receives from other nodes. All records in the queue have identical TTL (time to live) which decreases 1 per time unit. The queue is updated only under four cases. 1) The new record comes from the node that the queue doesn't cover. In this case, the new record will be allowed to enter the queue; 2) The queue already has a record of the node that the new record belongs to but the new record covers a different interval of time. The new record will also be allowed to enter the queue in this case. 3) The TTL of some records has reduced to 0, the record will be removed from the queue. 4) If the new record is a newer version of an existent record in the queue, i.e., comes from the same node and covers the same interval of time but has a more recent version stamp, the new version of the record will replace the old version one in the queue. However, in this case, the TTL of the record is still the old version one, i.e., not refreshed. Therefore, the routine table of each node will be disseminated and updated constantly in the network.

To control the storage overhead, the queue length should have an upper bound. We set the maximum queue length to be the number of nodes subtracting one. This is because the routine table of the node holding the queue is not needed to be stored in the queue. When the number of entries is less than the upper bound of queue length, the arrived transient routine record that is allowed to enter the queue will be pushed into the queue directly. As long as the queue is full, an entry of the node having the most transient routine records will be removed. Such a setting is to ensure that each node can have at least one transient routine record in queue.

Routine Table Aided Node Searching: After running for some time, every node can generate a summary of its daily routines in the form of the routine table based on its historical transit records. Each node also distributes transient routine records to other nodes according to the method mentioned in the previous subsection. The routine table can be used as an auxiliary positioning method for node searching.

Specifically, each time the locator meets a node, it tries to fetch the transient routine record of the target node from the newly encountered node. Then, the locator can find the most possible landmark(s) the target will be based on current system time and the collected routine information. Consequently, the locator can directly move to the indicated landmark(s) to search for the target. Figure 5 shows one case that the search speed is accelerated through utilizing the routine information. Suppose the target node, Node 0, currently is at sub-area $AcadBldg8$, and the system time is 11:30AM. We also assume a received transient routine record is as shown in Table V. Generally, to reach sub-area $AcadBldg8$, the locator has to go through sub-area $LibBldg1$. Now with the second entry in Table V, the locator knows that Node 0 is most likely to stay in sub-area $AcadBldg8$ now. Thus, rather than searching $LibBldg1$ first, it moves towards sub-area $AcadBldg8$ along the geographically closest path directly.

On the other side, though the routine information is useful in indicating the places that the target node is likely to be, VRs and MPT information is the primarily way to decide the searching direction. This is because the routine table may only cover a short period of time in each day, i.e., a node may have a routine only in a part of time daily. Therefore, the routine table will be used to infer the possible position of the target node only when available, i.e., the VR and MPT are the major mechanisms to guide the node searching.

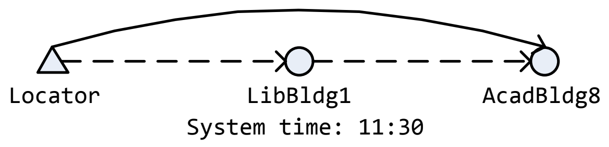


Fig. 5: Demonstration of routine table aided node searching.

IV. Performance Evaluation

In this section, we evaluate the performance of DSearching with two real DTN traces and the trace obtained by 9 students carrying a mobile phone on our campus. We first disable the advanced extensions mentioned in Section III-E to show the advantage of basic DSearching. We then evaluate the improvement of the advanced extensions in Section IV-F.

A. Empirical Datasets

The first trace, denoted Dartmouth trace (DART) [16], records the association of students' digital devices with APs on the Dartmouth campus. We regarded each building as a sub-area and merged neighboring records for to the same mobile device and the same sub-area. We also removed short connections ($< 200s$) and nodes with few records (< 500). Finally, we obtained 320 nodes and 159 sub-areas.

The second trace, namely DieselNet AP trace (DNET) [18], collects the AP association records of 34 buses in the downtown area of a college town. Since there are many APs that do not belong to the experiment in the outdoor environment, APs with few appearances (< 50) were removed from the trace. We mapped APs that are within certain distance ($< 1.5km$) into one sub-area. The trace is pre-processed similarly as the DART trace. Finally, we obtained 34 nodes and 18 sub-areas.

The characteristics of the two traces are shown in Table VI.

TABLE VI: Characteristics of mobility traces.

	DART	DNET
# Nodes	320	34
# Sub-areas	159	18
Duration	119 days	20 days
# Transits	477803	25193

B. Experiment Setup

We set the initial period to 30 days for the DART trace and 2.5 days for the DNET trace, during which nodes collect mobility information to build the MPT. Then, locators were generated with random start sub-area and target node at the rate of R_p per day, which was set to 40 by default. Considering students move less frequently than buses, the default locator TTL was set to 24 hours in the DART trace and 4 hours in the DNET trace. Since both traces do not provide the map information, we assume that the locator needs 10 minutes to move from one sub-area to another sub-area on average.

We compared DSearching with three representative methods: an encountering based method (Cenwits) [7], a routing based method (PROPHET) [19], and a random searching method (Random). In Cenwits, nodes record their meeting locations and times with other nodes and exchange such information with others. The locator collects such information from encountered nodes and moves to the most recent place where the target node appears to search for it. In PROPHET, the locator follows the node that has the highest possibility to meet the target node to search for it. In Random, the locator moves randomly to search for the target node.

We measured four metrics: *Success rate*, *Average delay*, *Average path length*, and *Average node memory usage*. The former three refer to the percentage of locators that successfully find their target nodes and the average delay and average path length of these locators, respectively. The last one denotes the average number of memory units used by each node. For DSearching, we take one row of the MPT and four visiting records as one memory unit. For Cenwits, 4 meeting records with others are regarded as one memory unit. For PROPHET, 8 meeting probabilities are regarded as one memory unit. We set the confidence interval to 95% in the paper.

C. Experiments with Different Locator Rates

In this test, we varied the locator rate R_p from 20 to 70.

1) Success Rate: Figure 6(a) and Figure 7(a) present the success rates of the four methods in the tests with the DART trace and the DNET trace, respectively. We see that the success rates follow: DSearching>Cenwits>Random>PROPHET.

PROPHET has the lowest success rate because a locator does not move actively to search for the target node but only follow the mobile node that has the highest probability to meet the target node, which has its own mobility pattern. Therefore, a lot of locators expire due to TTL, leading to the lowest success rate. For Random, locators move actively but search blindly, resulting in a low success rate. Cenwits explores the witness of the target node's appearances in different sub-areas to actively search for it. Therefore, it has higher success rate than Random and PROPHET. However, Cenwits has lower success rate than DSearching. This is because Cenwits only simply utilizes the recent appearances of the target node but neglects the mobility pattern of the target node. On the contrary, DSearching combines the two information to enable more efficient and accurate node searching.

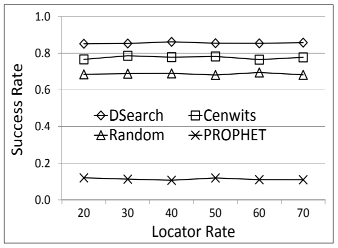
We also find that except DSearching, other three methods have obvious higher success rate in the test with the DNET trace than in the test with the DART trace. This is because the DART trace represents a network with a lot of sub-areas (i.e., 159) while the DNET trace is a small scenario with 18 sub-areas. Then, in the test with the DNET trace, though locators in Random, PROPHET, and Cenwits have no or limited information about the mobility of the target nodes, they can meet the target nodes easily. DSearching has similar success rate in the tests with both traces because the locator always moves towards the most possible sub-area that the target node would be. Then, even when the number of sub-areas increases, it can stably find most target nodes. Such a result validates the effectiveness of the mobility information distribution in DSearching in networks with different sizes.

2) Average Delay: Figure 6(b) and Figure 7(b) present the average delays of the four methods in the tests with the DART trace and the DNET trace, respectively. We see that the average delays follow: DSearching<Cenwits<Random<PROPHET.

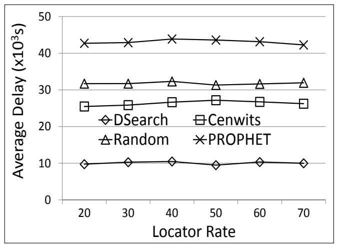
PROPHET has the highest average delay because the node followed by the locator may stay in certain sub-areas for a long time, resulting in an extremely long delay. Random also has a large average delay since nodes search randomly. Cenwits actively searches where the target node has shown recently. Since nodes usually would stay in a sub-area for a while, Cenwits has a small average delay. For DSearching, it further reduces the delay by utilizing both recent visiting records and long term MPT to guide node searching, which can help predict where the target node would be more accurately, leading to the least average delay.

3) Average Path Length: Figure 6(c) and Figure 7(c) show the average path lengths of the four methods in the tests with the DART trace and the DNET trace, respectively. We see that the average path lengths of the four methods follow: PROPHET<DSearching<Cenwits<Random.

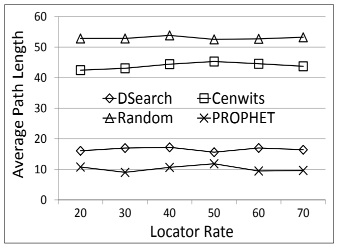
PROPHET has the lowest average path length. This is because locators in it often stay in a sub-area for a long time.



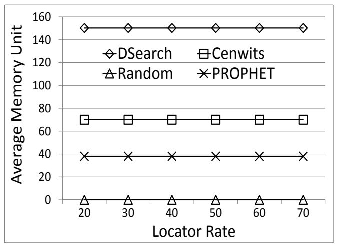
(a) Success rate.



(b) Average delay.

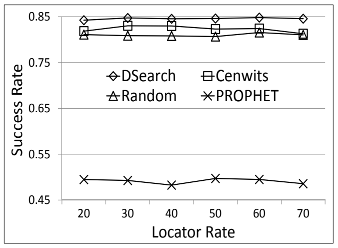


(c) Average path length.

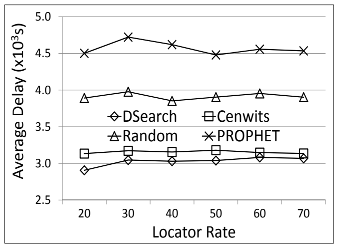


(d) Average node memory usage.

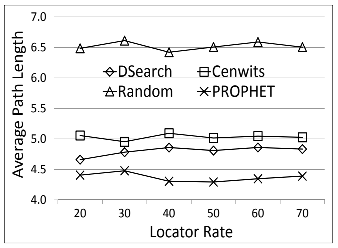
Fig. 6: Performance with different locator rates using the DART trace.



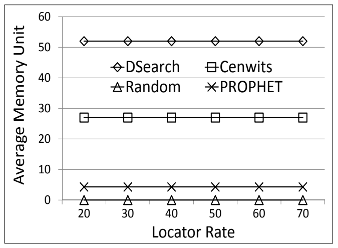
(a) Success rate.



(b) Average delay.



(c) Average path length.



(d) Average node memory usage.

Fig. 7: Performance with different locator rates using the DNET trace.

Therefore, PROPHET searches very slowly, which means successful locators only search a few sub-areas, leading to the shortest search path. DSearching has the second least average path length because the locator movement in it has the highest possibility of encountering the target node by utilizing both transient visiting records and long term mobility pattern of the target node. For Cenwits, it only utilizes the transient appearance records of the target node to guide the node searching, resulting in less efficient locator movement and longer average searching path length than DSearching. Random has the highest path lengths because the locator searches randomly in the network.

4) Average Node Memory Usage: Figure 6(d) and Figure 7(d) show the average node memory usages of the four methods in the tests with the DART trace and the DNET trace, respectively. We see that the average memory units of the four methods follow: Random<PROPHET<Cenwits<DSearching.

Random has 0 average memory unit since nodes in it do not need to store any information for node searching. In PROPHET, each node needs to store its meeting probabilities with all other nodes, leading to a small amount of memory units. Cenwits has higher average memory units than PROPHET since it needs to store a large amount of node appearance records on each node. For DSearching, each node stores both visiting records and MPT of other nodes, resulting in the highest memory usage.

Though DSearching has the most memory usage among the four methods, the absolute amount of memory usage is acceptable. We find that the average memory unit on each node is about 150 and 50 in the tests with the two traces. Recall that each unit is about 50 bytes (i.e. one row of the MPT and four visiting records). This means that the average memory usage on each node is only about 7.5 KB and 2.5 KB in the two tests, which can easily be satisfied in modern devices. Therefore, we conclude that the designed mobility information distribution algorithm is memory efficient in DTNs.

We also see that the average memory units of the four methods are constant with different locator rates. This is because the memory usage on each node is only decided by node mobility and is irrelevant with the number of locators.

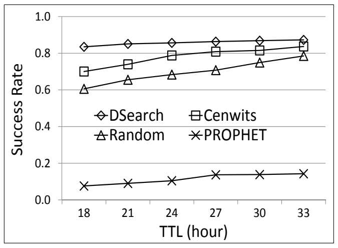
D. Experiments with Different Locator TTLs

We varied the TTL of each locator to see how different methods scale to locator TTL. Considering the DART trace is much longer than the DNET trace (119 days vs 20 days) and has relative slow node movement (student vs bus), we varied the TTL from 18 hours to 24 hours and from 2 hours to 7 hours for the DART trace and the DNET trace, respectively.

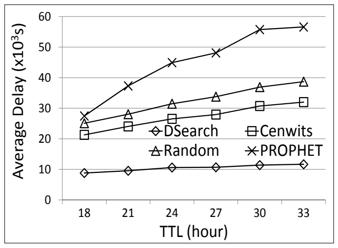
1) Success Rate: Figure 8(a) and Figure 9(a) present the success rates of the four methods in the tests with the DART trace and the DNET trace, respectively. We see that the success rates of the four methods follow: DSearching> Cenwits>Random>PROPHET. This is the same as that in Figure 6(a) and Figure 7(a) for the same reasons.

We further find that when the TTL increases, Cenwits and PROPHET have closer and closer success rate with DSearching. This is because that the major difference between DSearching and the two methods is the node searching efficiency. When each locator is allowed to search longer, Cenwits and PROPHET can eventually find most target nodes, leading to a high success rate. However, this comes at the cost of increased average delay, as shown in the next section.

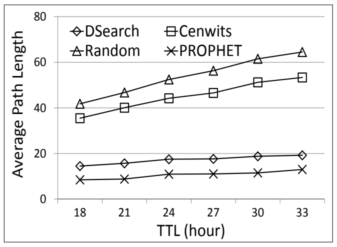
2) Average Delay: Figure 8(b) and Figure 9(b) present the average delays of the four methods in the tests with the DART trace and the DNET trace, respectively. We find that the average delays of the four methods follow the same relationship as in Figure 6(b) and Figure 7(b): DSearching <Cenwits<Random<PROPHET. The reasons are the same with those in the tests with different locator rates.



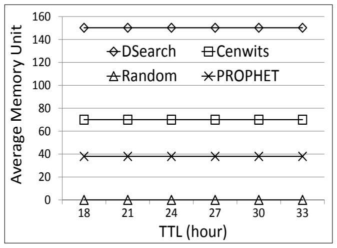
(a) Success rate.



(b) Average delay.

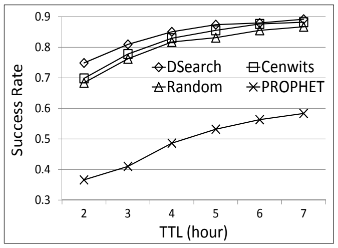


(c) Average path length.

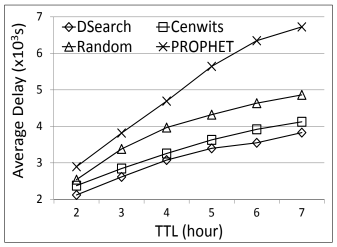


(d) Average node memory usage.

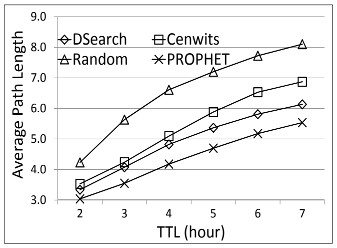
Fig. 8: Performance with different locator TTLs using the DART trace.



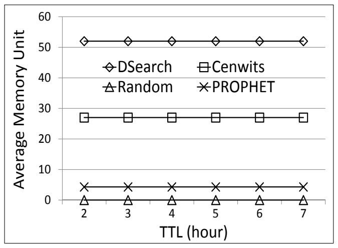
(a) Success rate.



(b) Average delay.



(c) Average path length.



(d) Average node memory usage.

Fig. 9: Performance with different locator TTLs using the DNET trace.

We also see that when the TTL increases, the average delay increases. This is because when the TTL increases, locators that may fail to find their target nodes when the TTL is small can find their target nodes. Then, the average delay increases due to more successful locators with a large delay.

3) Average Path Length: Figure 8(c) and Figure 9(c) present the average path lengths of the four methods in the tests with the DART trace and the DNET trace, respectively. We find that the average path lengths follow: Random< DSearching<Cenwits<PROPHET, which is the same as in Figure 6(b) and Figure 7(b). The reasons are also the same.

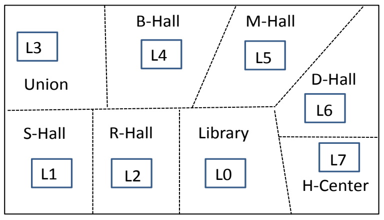
We also see that when TTL increases, the average path length increases. This is caused by the same reason as in Section IV-D1 and IV-D2: when TTL increases, more locators can find their targets after searching many sub-areas, leading to increased average search path length.

4) Average Node Memory Usage: Figure 8(d) and Figure 9(d) present the average memory units a node uses in the tests with the DART trace and the DNET trace, respectively. We see that the results are the same with Figure 6(d) and Figure 7(d). This is caused by the fact that the memory usage is independent with the locator TTL.

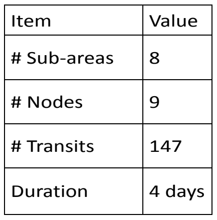
Combining all above results, we conclude that DSearching presents superior performance compared to other methods with different locator rates and TTLs. Such a result justifies our design goal: efficient node searching with acceptable cost.

E. Test With Real Environment Data

We further applied the DSearching to the mobility data of 9 students on our campus. The 9 students are from 4 departments in our university. We selected 8 sub-areas, each of which is represented by a frequently visited building. Such mobility data is obtained based on the GPS records on mobile phones, which is more accurate than the AP association records in



(a) Maps for sub-area division.



Item	Value
# Sub-areas	8
# Nodes	9
# Transits	147
Duration	4 days

(b) Statistical data.

Fig. 10: Configuration in the real environment.

the two real traces. The test environment and the mobility information are summarized in Figure 10(a) and 10(b).

Since DSearching shows stable performance with different locator rates in previous tests, we only varied the locator TTL from 20 minutes to 70 minutes and set the R_p to 40. Since the distance between the test buildings are not far, we assume that a locator averagely takes 5 minutes to move from one sub-area to a neighboring sub-area.

TABLE VII: Results with real environment data

TTL (min)	20	30	40	50	60	70
Success Rate	0.62	0.75	0.83	0.88	0.89	0.93
Ave. Delay (min)	7.6	10.2	10.8	12.2	13.8	17.6
Ave. Path Length	1.4	1.9	2.0	2.3	2.6	3.1
Ave. Node Memory Usage	5	5	5	5	5	5

The test results are shown in Table VII. We find that when the locator TTL increases, success rate, average delay and average path length increase. This is the same as our observation in previous experiments with the two real traces. The reasons are the same that when the TTL increases, more locators can find the target nodes after a large delay and a long searching path.

We also see that when the TTL was set to 70 minutes, a successful locator takes only about 17 minutes (or 3 transits between sub-areas) to find the target node on average. Further, each node only needs 5 units of memory on average to support the node searching, which is very low and can easily be satisfied. In conclusion, DSearching is effective and efficient in searching mobile nodes in realistic DTNs.

F. Evaluation of the Advanced Extensions

We further evaluate the effect of the extensions proposed in Section III-E. Considering that DSearching presents stable performance in the tests with different locator rates, we only conducted tests with different TTLs. Specifically, as in Section IV-D, we again varied the TTL from 18 hours to 24 hours and from 2 hours to 7 hours for the DART trace and the DNET trace, respectively. Other settings are the same as mentioned in Section IV-B. In this test, we use DSearch to denote the original DSearching without any extensions, DSearch_Pr to denote DSearching with the improved MPT, DSearch_Rt to represent DSearching with the routine table, and DSearch_Both to represent DSearch with both the improved MPT and the routine table.

1) Success Rate: Figure 11(a) and Figure 12(a) present the success rates of DSearch and the three extensions in the tests with the DART trace and the DNET trace, respectively. We see from the two figures that in both tests, the success rates of the four methods follow: DSearch_Both>DSearch_Rt≈DSearch_Pr>DSearch. In DSearch_Pr, the advanced MPT can predict the next transit more accurately by considering the previous transit. In DSearch_Rt, the routine table can guide the locator to move to the places where the target visits regularly. Therefore, both DSearch_Pr and DSearch_Rt have higher success rate than the original DSearch. Since the two extensions improve the original DSearch from two different aspects, it is hard to distinguish which one is better. We can also see that they present similar success rate in our test. Furthermore, DSearch_Both results in the highest success rate since it integrates both the improved MPT and the routine table.

2) Average Delay: Figure 11(b) and Figure 12(b) show the average delays of DSearch and the three extensions in the tests with the DART trace and the DNET trace, respectively. We find from the two figures that the average delays of the four methods follow DSearch_Both<DSearch_Rt≈DSearch_Pr<DSearch. Such a result matches with the relationship on success rate of each method, as shown in the previous section. This is because by utilizing the improved MPT and the routine table, DSearch_Rt and DSearch_Pr can enable locaters to find the target more quickly, thereby reducing the average search delay. Similarly, DSearch_Both has the lowest average delay because it enables both the improved MPT and the routine table, both of which can improve node searching efficiency.

3) Average Path Length: Figure 11(c) and Figure 12(c) show the average path lengths of DSearch and the three extensions in the tests with the DART trace and the DNET trace, respectively. We observe from the two figures that the average path lengths of the four methods follow DSearch_Both<DSearch_Rt≈DSearch_Pr<DSearch. In DSearch_Pr, the improved MPT can guide the locator to search along the path of the target more accurately. In DSearch_Rt, the routine table can guide the locator to move to the regularly visited places of the target node. Both extensions effectively avoid searching unnecessary sub-areas. Thereby, they lead to shorter average path length than DSearch. Again, DSearch_Both leads to the shortest average search path because both extensions are included.

Combining above results, we conclude that the two extensions can improve the node searching efficiency.

4) Average Node Memory Usage: Figure 11(d) and Figure 12(d) illustrate the average node memory usages of DSearch and other three extensions in the tests with the DART trace and the DNET trace, respectively. We see from the two figures that the average node memory usage of the four methods follow DSearch_Both>DSearch_Rt>DSearch_Pr> DSearch. In DSearch_Pr, the MPT table is extended to include more fine-grained transit probabilities, resulting in more memory usage than DSearch. For DSearch_Rt, the routine table is created and distributed in the same way as the MPT table. Therefore, it generates even more memory usage than DSearch_Pr. Then, DSearch_Both has the highest node memory usage by including both extensions.

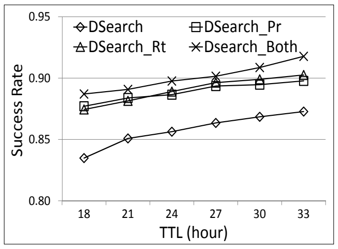
Though these extensions lead to more memory usage, they bring about improvement on the node searching efficiency. On the other hand, the increased memory usage is on the same level with the original memory usage. This means that the two extensions do not lead to a significant burden on memory usage. The system designer can decide which extensions can be enabled based on the actual system need to achieve a balance on efficiency and memory usage.

V. Conclusion

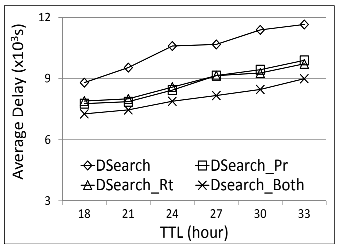
In this paper, we propose DSearching, a distributed mobile node searching scheme in DTNs where nodes present certain mobility patterns. In DSearching, the whole network is split into sub-areas. A node's mobility information includes both transient sub-area visiting records and long-term transition patterns between sub-areas. Each node distributes its visiting record for a sub-area to nodes that are likely to stay in the previous sub-area. The long term mobility information of a node in a subarea is distributed to a limited number of long-staying nodes in the sub-area. Such information enables the locator to search along the path the target node traverses, resulting in efficient node searching. Advanced extensions that utilize sub-area visiting correlation and routine information are also proposed in the paper to further enhance the node searching efficiency. Extensive experiments with both real traces and on-campus DTN trace validate the high effectiveness and high efficiency of DSearching. In the future, we plan to investigate how to fully utilize searching agents to further improve node searching efficiency.

Acknowledgements

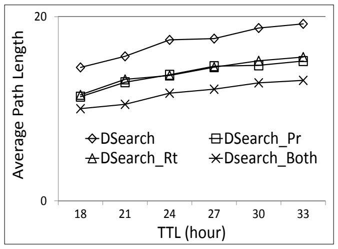
This research was supported in part by U.S. NSF grants IIS-1354123, CNS-1254006, CNS-1249603, CNS-1049947, CNS-0917056 and CNS-1025652, Microsoft Research Faculty Fellowship 8300751.



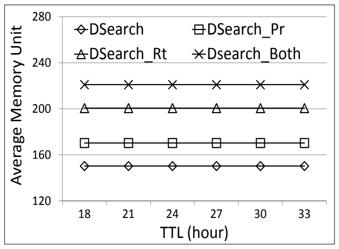
(a) Success rate.



(b) Average delay.

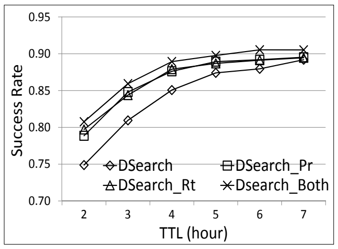


(c) Average path length.

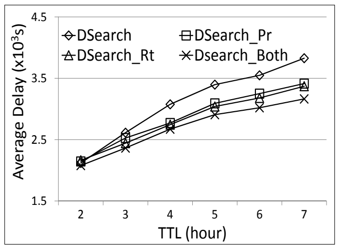


(d) Average node memory usage.

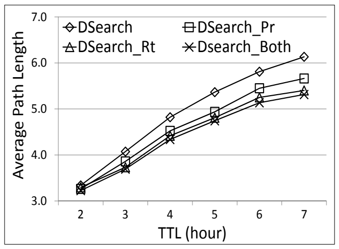
Fig. 11: Performance of each extension with different locator TTLs using the DART trace.



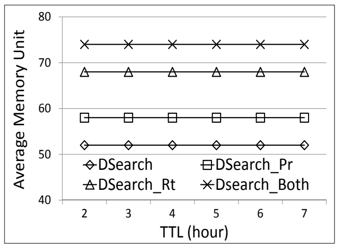
(a) Success rate.



(b) Average delay.



(c) Average path length.



(d) Average node memory usage.

Fig. 12: Performance of each extension with different locator TTLs using the DNET trace.

References

[1] S. Jain, K. R. Fall, and R. K. Patra, "Routing in a delay tolerant network," in *Proc. of SIGCOMM*, 2004.

[2] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet," in *Proc. of ASPLOS-X*, 2002.

[3] J. Zhao, Y. Zhu, and L. M. Ni, "Correlating mobility with social encounters: Distributed localization in sparse mobile networks." in *Proc. of MASS*, 2012.

[4] S. Lu, Y. Liu, Y. Liu, and M. Kumar, "Loop: A location based routing scheme for opportunistic networks." in *Proc. of MASS*, 2012.

[5] K. Chen and H. Shen, "Leveraging social networks for p2p content-based file sharing in disconnected manets." *IEEE TMC*, 2013.

[6] B. Thorstensen, T. Syversen, T. Walseth, and T.-A. Bjørnvold, "Electronic shepherd - a low-cost, low-bandwidth, wireless network system." in *Proc. of MobiSys*, 2004.

[7] J. Huang, S. Amjad, and S. Mishra, "Cenwits: a sensor-based loosely coupled search and rescue system using witnesses." in *Proc. of SenSys*, 2005.

[8] J.-H. Huang, L. Jiang, A. Kamthe, J. Ledbetter, S. Mishra, A. Cerpa, and R. Han, "Sensearch: Gps and witness assisted tracking for delay tolerant sensor networks," in *Proc. of Ad Hoc-Now*, 2009.

[9] A. Balasubramanian, B. N. Levine, and A. Venkataramani, "DTN routing as a resource allocation problem." in *Proc. of SIGCOMM*, 2007.

[10] K. Lee, Y. Yi, J. Jeong, H. Won, I. Rhee, and S. Chong, "Max-Contribution: On optimal resource allocation in delay tolerant networks." in *Proc. of INFOCOM*, 2010.

[11] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: social-based forwarding in delay tolerant networks," in *Proc. of MobiHoc*, 2008.

[12] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets," in *Proc. of MobiHoc*, 2007.

[13] J. Link, D. Schmitz, and K. Wehrle, "GeoDTN: Geographic routing in disruption tolerant networks," in *Proc. of GLOBECOM*, 2011.

[14] I. Leontiadis and C. Mascolo, "GeOpps: Geographical opportunistic routing for vehicular networks," in *Proc. of WOWMOM*, 2007.

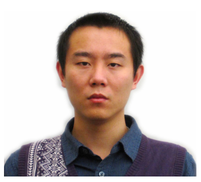
[15] L. Song, U. Deshpande, U. C. Kozat, D. Kotz, and R. Jain, "Predictability of wlan mobility and its effects on bandwidth provisioning." in *Proc. of INFOCOM*, 2006.

[16] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Proc. of MOBICOM*, 2004.

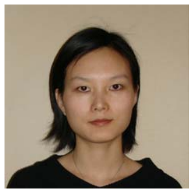
[17] K. Chen and H. Shen, "Dtn-flow: Inter-landmark data flow for high-throughput routing in dtns." in *Proc. of IPDPS*, 2013.

[18] A. Balasubramanian, B. N. Levine, and A. Venkataramani, "Enhancing interactive web applications in hybrid networks," in *Proc. of MOBICOM*, 2008.

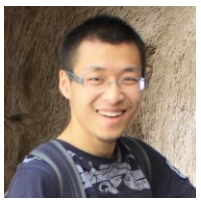
[19] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks." *Mobile Computing and Communications Review*, vol. 7, no. 3, 2003.



Kang Chen Kang Chen received the BS degree in Electronics and Information Engineering from Huazhong University of Science and Technology, China in 2005, the MS in Communication and Information Systems from the Graduate University of Chinese Academy of Sciences, China in 2008, and the Ph.D. in Computer Engineering from Clemson University. He is currently a Postdoctoral Research Fellow in the Department of Electrical and Computer Engineering at Clemson University. His research interests include mobile ad hoc networks, delay tolerant networks and vehicular networks.



Haiying Shen Haiying Shen received the BS degree in Computer Science and Engineering from Tongji University, China in 2000, and the MS and Ph.D. degrees in Computer Engineering from Wayne State University in 2004 and 2006, respectively. She is currently an Associate Professor in the Department of Electrical and Computer Engineering at Clemson University. Her research interests include distributed computer systems and computer networks, with an emphasis on P2P and content delivery networks, mobile computing, wireless sensor networks, and grid and cloud computing. She is a Microsoft Faculty Fellow of 2010, a senior member of the IEEE and a member of the ACM.



Li Yan Li Yan received the BS degree in Information Engineering from Xi'an Jiaotong University, China in 2010, and the M.S. degree in Electrical Engineering from University of Florida in 2013. He currently is a Ph.D. student in the Department of Electrical and Computer Engineering at Clemson University, SC, United States. His research interests include wireless networks, with an emphasis on delay tolerant networks and sensor networks.