Sociological Orbit aware Location Approximation and Routing (SOLAR) in DTN

Joy Ghosh, Chunming Qiao, Sumesh J. Philip, Hung Ngo, Seokhoon Yoon
Department of Computer Science and Engineering
University at Buffalo, The State University of New York
201 Bell Hall, Buffalo, NY 14260-2000
Email:{joyghosh, qiao, sumeshjp, hungngo, syoon4}@cse.buffalo.edu

***Abstract*— Routing in delay tolerant networks poses a challenging problem compared to a conventional data network due to the uncertainty and time varying nature of network connectivity. Initial research in this area has considered algorithms based on deterministic mobility of the nodes in DTN. While the assumption of deterministic mobility can lay the groundwork for a theoretical understanding of DTN, such knowledge may not be applicable to mobile ad hoc networks. In this work, we introduce a novel concept of a partially repetitive "orbital" pattern of mobile users (nodes) involving a set of "hubs", that may be better suited for a semi-deterministic mobility modeling of DTN users. This partially deterministic movement pattern is both practical as well as useful in the sense that the hub list information can be useful for locating nodes and routing packets to them in a DTN.**

***Index Terms*— Mobility framework, Routing protocol, Delay tolerant networks, Performance analysis**

I. Introduction

A Delay Tolerant Network (DTN) architecture [6] focuses on asynchronous communication within an intermittently connected network. A few of the constraints imposed by such a network are requirements for persistent storage within nodes, frequent interruptions, lack of any end-to-end path between a pair of source and destination nodes, presence of heterogeneity, in addition to the constraints imposed by mobile wireless networks such as varying channel conditions, lack of infrastructure, etc. A DTN can be modeled as a multi-graph, where the edge capacities and durations between nodes are time varying due to the mobility of mobile users.

Routing in a DTN has received significant attraction from the research community recently. Several routing-related issues in DTN were addressed in [15], which focused mainly on networks with known connectivity patterns, such as satellites with fixed paths, or busses with fixed routes. The authors developed several algorithms to analyze the knowledge to performance relationship in different protocols and demonstrated that their algorithms performed better with more network knowledge. However, for the availability of such global knowledge they assumed the presence of certain "knowledge oracles" that may not be applicable to most mobile ad hoc networks.

Routing issues were also addressed in general intermittently connected networks in [32], where the authors proposed an Epidemic Routing protocol that relies on data buffering and node mobility to spread messages in the network. The so-called "summary vectors" were used for nodes to selectively exchange data packets, in order to limit the number of data transmissions. Similar work ([5], [12]) on data dissemination was also done for sensor and ad hoc networks.

Epidemic Routing was extended in [23], where the authors proposed a probabilistic routing scheme whereby each node maintains the so-called "delivery predictability" to each known destination, and uses this metric to make routing decisions. Similarly in [25], the authors proposed a context-aware adaptive routing algorithm that takes into account the suitability of a node for carrying a message based on context information of the node at multiple dimensions. More recently, the authors in [21] suggested an algorithm that relies on vehicles to act as mobile routers, which connect disconnected sensor networks to a known destination. Around the same time, the authors in [14] and [37] respectively studied the effects of controlled message flooding and controlled mobility in large scale DTN.

Although node mobility is known to affect routing protocol performance [2], none of the work done on DTN with unpredicted and opportunistic connectivity takes any mobility pattern information into account for routing. The work in [13] exploited node mobility in MANET as a type of multi-user diversity, and showed that node mobility may actually help increase the theoretical capacity of a MANET. In addition, research has also been done ([30], [31], [33]) on MANET to take advantage of mobility information obtained via continuous location tracking and "micro-level" mobility prediction. However, such methods and solutions are not applicable to DTN due to its inherent intermittent nature.

In this paper, we introduce a novel protocol framework called Sociological Orbit aware Location Approximation and Routing (SOLAR), which takes advantage of the "macro-mobility" information obtained from the sociological movement pattern of mobile DTN users. This mobility information, also referred to as the "mobility profile", is extracted from our observation that the movement of a mobile user exhibits a partially repetitive "orbital" pattern involving a set of "hubs". We study how to leverage the sociological orbit based mobility of users in routing packets within DTN.

In a conventional MANET (as opposed to intermittently connected networks), it has been shown that SOLAR is not only general enough to be realistic, but is also specific enough to be useful [10], [11]. In particular, the proposed SOLAR protocol can be practically implemented without a need for constant location updates (or tracking) and flooding, that makes it equally attractive to DTN settings. More specifically, the novelty of this work lies in the fact that sociological orbits are partially deterministic (unlike the satellites and busses studied in [15]), and are suitable for unpredicted DTN with opportunistic connectivity.

To the best of our knowledge, this is the first work to exploit the sociological "orbital" concepts in DTN networks. The main contributions of this work is to first extend the SOLAR algorithm (earlier proposed for MANET) to SOLAR-DIST, which uses store-and-forward routing to suit the DTN. In addition, we propose two new SOLAR based algorithms, one called SOLAR-PROB and the other SOLAR-KSP, which assume varying levels of hub based probability information. A more detailed description of each protocol is presented in Section III-B. We also compare the three algorithms with Epidemic Routing and determine the protocol performance in terms of throughput and overhead.

The rest of the paper is outlined as follows. In Section II, we motivate our work by discussing the sociological movement pattern of mobile DTN users, and describe an example Random Orbit model. In Section III, we provide the details of the proposed Sociological Orbit aware Location Approximation and Routing (SOLAR) protocol and elaborate on the different versions of SOLAR in Section III-B. In Section IV, we evaluate the performance of SOLAR through simulations, and showcase its simplicity and superiority in terms of higher throughput and lower control overhead. We conclude this work in Section V.

II. Sociological Movement Pattern

In the real world, users routinely spend a considerable amount of time at a few specific place(s) that we refer to as hub(s). For example, a graduate student in school may visit and spend some significant amount of time in his/her laboratory, a seminar room, or the cafeteria. Although it is hard (or may be even against privacy policies) to keep track of an individual at all times, one can still take advantage of the fact that most users' movements are within and in between a list of hubs. In these situations, it is often possible to estimate/measure hub-visit probabilities and inter-hub movement patterns of an individual. This information then constitutes a part of the users' *mobility profiles*. For example, even if we do not know the exact location of the graduate student at any given time, given his/her mobility profile we can most probably find him/her in either the laboratory, or the seminar room, or the cafeteria, without having to look all over the building/campus. The more "periodic" the movement pattern is, the more we can take advantage of the mobility profile.

This orbital movement pattern is also observed in a time and space based hierarchy. For example, on a typical weekday, the graduate student could leave home for school in the morning, visit the gymnasium in the evening, and return home at night. Similarly, the student may stay in his home town for a few weeks and visit friends and family in other cities over some weekends, forming yet another higher level nation-wide orbit. This hierarchical concept is illustrated in Figure 1.

In practice, hubs can be identified in a variety of ways. GPS service is the obvious first choice. Signal strengths of wireless Ethernet packets can be used for location sensing and real-time tracking [20]. In the broader contexts of *pervasive/ubiquitous computing* [1], and *Ambient Intelligence* (AmI) [29], localization in a cosmopolitan area will be even more readily available.

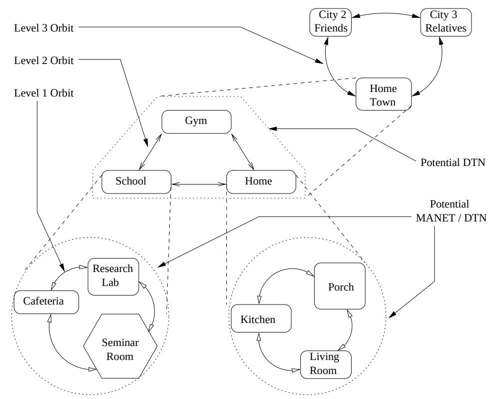


Fig. 1. A hierarchical view of sociological orbits

Interestingly, an orbit is one of the most natural forms of motion observed in the microscopic world of molecules, as well as in the planetary universe. However, such natural orbits are mostly deterministic, and their continuous motion does not have the notion of special places like hubs.

Note that our orbital movement pattern differs from existing mobility patterns studied in the literature in that it neither models the motion of the users at a micro-level (i.e., on small time scales or within small distances), nor simply predicts user locations via historical/statistical tracking information ([30], [31], [33]). It also differs from the deterministic mobility patterns assumed within DTN, where either exact locations of a node can be predicted with an appropriate "oracle", or no location information is available. To the best of our knowledge, no prior work has explored the implication of such a macro-level partially deterministic sociological mobility pattern and its application to location approximation and routing in DTN, despite its practicality.

A. An Example Random Orbit Model

To illustrate the concept of the sociological orbital movement, we first construct a simple yet practical orbital model called the Random Orbit. The Random Orbit model allows for the creation of a certain number of hubs within the simulation terrain for all the nodes, as specified by the parameter *Total Hubs*. These hubs are located at random places within the terrain, and as a result they may or may not overlap with each other. Each node can visit a subset of randomly chosen hubs creating a *Random Orbit*. The list of hubs a node visits is bounded by *Hub List Size*, and the time it spends in each hub is specified by *Hub Stay Time*. Together, these two parameters define an Inter-Hub Orbit (IHO). We also allow for an occasional change in the specific list of hubs assigned to a node in its IHO by defining an *IHO Timeout*, upon which a node is assigned a fresh list of hubs to visit.

The mobility pattern of individual nodes shall comprise of two parts: movement inside a hub, and movement in between hubs. For convenience, the movement inside each hub, which shall also be referred to as the Intra-Hub Movement (IHM), was chosen to follow a modified Random Waypoint mobility model, whose speed range is denoted by *Intra-Hub Speed* (with a non-zero minimum as suggested by [35]), and whose pause time is denoted by *Intra-Hub Pause*. For movement in between hubs, we define a Point-to-Point Linear (P2P Linear) model. In this model, when a node wants to leave one hub for another, it randomly selects a point within the destination hub and moves towards it linearly from its current position with a velocity defined by the range *Inter-Hub Speed*. ***Note that for each of the two parts, any known practical mobility models may be chosen.***

Figure 2 illustrates the Random Orbit model. Such a model is suitable for modeling wireless devices carried by users working in an office building, attending a convention, or around a campus. As users move around, devices either automatically, or with the user's permission/assistance may record the hubs visited most often, and share the hub-based orbital mobility profile with trusted "acquaintances". Such mobility profile can then help improve routing as described next.

Note that, this example Random Orbit model does not simply integrate two common mobility models (Random Waypoint, and P2P Linear), but most importantly also introduces the practical orbital movement amongst hubs.

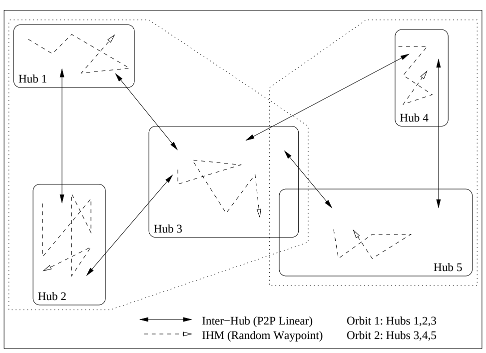


Fig. 2. The Random ORBIT Model

III. Sociological Orbit aware Location Approximation and Routing (SOLAR)

We now briefly describe our Sociological Orbit aware Location Approximation and Routing (SOLAR) protocol. SOLAR uses a concept of "acquaintance" similar to that in our prior work in Acquaintance Based Soft Location Management (ABSoLoM) protocol [9], as well as to some degree the concept of *peer collaboration* (among 'acquaintances') in [3]. However, to the best of our knowledge, SOLAR is among the first to make use of macro-level sociological mobility profiles of DTN users in obtaining approximate location information of mobile users, as well as in improving routing.

A. Protocol Overview

In SOLAR, when two mobile nodes meet, each node will send the list of hubs it goes to, as a part of its "hello" packet, to the other node, and therefore becoming "acquainted" with each other by caching each other's hub list. When a source node needs to send data to a destination node, it first checks if the destination node's hub list is known (cached) or not. If so, data packets are sent to the hubs on the list. If not, the source node will send a "query" packet to its acquaintances asking if any of them knows the hub list of the destination. An acquaintance will send a positive "response" packet (containing the hub list of the destination) back if it knows, and otherwise, may ask its own acquaintances for further help.

A transmission from a node to its acquaintance will be referred to as a *logical hop* here after, which often comprises of multiple physical hops. A query packet will be dropped after a limited number of logical hops. If the source does not receive a positive response packet in time, it can always resort to flooding the first data packet. When a destination receives the data packet, it can send a positive response containing the current hub it is in back to the source.

For simplicity, we assume that each node knows the (approximate) coordinates of the hubs, and its own location. In addition, a simple greedy geographic forwarding algorithm [22] is used to route a packet to a hub by picking an approximate center point of the hub as the destination. To adapt the basic greedy geographic forwarding to the intermittent connectivity of DTN, we incorporate the notion of store-and-forward at the intermediate nodes.

The algorithm above was chosen mainly for its simplicity which serves our proof-of-concept purpose. In the future, more sophisticated geographic algorithms (such as GOAFR [19]), or, in cases where geographic locations are not available, *virtual* geographic routing protocols [26], [28] will be thoroughly investigated, experimented and incorporated into our framework.

The basic scheme described above can have several variations and improvements. For example, a node may send (receive) a "hello" packet to (from) all nodes within two physical hops of its radio range to increase the number of acquaintances. The source could cache the hub list of the destination node, and vice versa. Additionally, a node may also cache the hub list of other remote nodes by e.g., snooping into the query/response packets. In a highly mobile environment, a node can also quickly build up the number of its acquaintances by exchanging the hub lists of its acquaintances, along with its own hub list with nodes that are within its radio range.

As the number of the acquaintances of a node becomes large, one may select only a subset of the acquaintances to send a query packet to. In fact, as long as the subset is selected in such a way that all the hubs that would have been visited by all the acquaintances are already visited by the subset of acquaintances, the inquirer will unlikely miss a positive response due to the reduction in the number of query packets to be sent. This selection of a suitable subset of acquaintances is reduced to a minimum SET COVER problem, which is known to be $\mathbf{NP}$-hard [8]. To find an exact solution, we have adopted the Quine-McCluskey algorithm [24], [27] in our implementation of SOLAR.

The Quine-McCluskey algorithm finds an exact minimum set cover in worst-case exponential time. For large instances of the problem, certainly the Quine-McCluskey algorithm should be replaced by efficient approximation algorithms (e.g, the greedy algorithm [16], or the primal-dual schema algorithm [4]). Note also that querying acquaintances physically far away will induce larger overhead and delays. Hence, acquaintances farther away can be "weighted" higher, and our problem becomes the WEIGHTED SET COVER problem, which can also be approximated with the algorithms cited above. This direction, along with methods to estimate physical distances of acquaintances, will be explored in more details in a future work.

When sending a packet to a node with a known hub list, the packet can be sent to the nearest hub first (or the current hub if known), which may forward the packet to the other hubs in some pre-determined sequence later (using a series of unicast). Or, one copy of the packet is sent to each of the hubs (as in simulcast). Other forwarding methods such as multicast and anycast may also be used, along with common "caching" techniques whereby the packet is stored at one or more intermediate nodes for some time.

In the presence of so-called "routing holes", packets may be either dropped, or cached (i.e. store-and-forward routing), or routed using more sophisticated algorithms [17]. If the hub size is larger than the radio range, a local flooding (within 2-3 physical hops) may be used with minimum overhead to reach the destination within the hub. Note that, unlike in other geographical routing schemes, SOLAR can potentially have lower control overhead in terms of *location updates*, because the hub list of a node stays valid for quite some time due to its sociological orbital movement pattern. This is in contrast with the conventional schemes that do not take advantage of the orbital patterns, wherein the exact x-y coordinates of a node changes often and thus may require many *location updates* (which are most likely needed after a node moves from one hub to another, even with some hierarchical location management). In SOLAR, a node may choose to notify each other by "location updates" only when there is a change in its hub list (as a result of an occasional IHO Timeout).

B. Knowledge to Performance Relation in SOLAR

Noting that the level of mobility knowledge has a direct impact on protocol performance, we propose three different variations of SOLAR: SOLAR-DIST, SOLAR-PROB and SOLAR-KSP. In the following subsections, we outline each scheme in detail.

1) SOLAR-DIST - Only Hub List Sharing: In this variation of SOLAR, nodes are assumed to start off with "zero knowledge" about each other. It is assumed that the nodes know their approximate locations in addition to their hub lists and approximate tiling of the terrain in terms of hub coordinates. This is the most basic version of SOLAR and is similar to the general description of the protocol in Section III-A. When any packet is to be forwarded to a destination whose hub list is known, a single copy of the packet is unicast towards the hubs in a sequential manner, while only minimizing the total distance traveled by the packet. The goal of this simple protocol is to direct the packet towards the next unvisited hub in the list, which is the closest to the current location of the packet, till either the destination is found, or all hubs are visited.

2) SOLAR-PROB - Hub List and Associated Probability Sharing: This version of SOLAR assumes that a node knows of the probabilities associated with each of its hubs in its hub list, in addition to the assumptions made in SOLAR-DIST. The probability associated with each hub in the node's hub list specifies the likelihood of the node visiting (or staying in) that hub. With this additional information, when the source has (or learns of) the destination's hub list, it orders the hub list in a descending order of the associated probabilities and geographically forwards the message to the hub with the highest probability.

3) SOLAR-KSP - Hub List, Probability and K-Shortest Paths: In this version of SOLAR, each node shares its hub list as well as the associated probability information with the complete subset of nodes that share at least one hub with it. This can be done by either exchanging summary vectors of the information collected when any pair of nodes come in contact with each other as in Epidemic Routing [32] or by efficient broadcasting. Once this information is obtained, each node can locally compute the probability of contacting every other node, either directly or indirectly via other nodes, in a distributed manner.

Hub list information of all nodes can be formally represented as a weighted graph $G = (V, E)$, where V is the set of all the nodes, and E is the set of weighted edges between every pair of nodes that have at least one hub in common. Let $P(u, v)$ be the probability that nodes u and v come in contact. Then the weight of edge (u, v) is given by:

$$w(u, v) = \log\left(1/P(u, v)\right).$$

In this weighted graph, each node applies a variation of the Dijkstra's Shortest Path algorithm [7] to find K paths (KSP) to every other destination, such that:

1) a path with the minimum total weight is chosen first
2) each path has a different node as the next hop.

Note that due to condition 2, a node may have less than K shortest paths. In any case, a node orders the K paths for each destination in descending order of data delivery (or contact) probability.

The weight-function choice can be explained as follows. Given two nodes u and v that have at least one hub in common, there are two ways to interpret the probability $P(u, v)$: (a) it is the probability that u and v *ever* come to contact; and (b) it is the probability that u and v come to contact at a randomly chosen time. Interpretation (b) gives more information, but also requires more work to maintain temporal hub data. If nodes are moving between hubs in a more or less periodic manner, then (b) is the interpretation of choice. However, interpretation (a) fits our *Random Orbit* model described earlier.

For each node u, let $\mathcal{H}(u)$ be the set of u's hubs. For each $H \in \mathcal{H}(u)$, let $P(u, H)$ be the probability that u visits H, which u itself can estimate. Then, $P(u, v)$ can be estimated by

$$P(u, v) = \sum_{H \in \mathcal{H}(u) \cap \mathcal{H}(v)} P(u, H) P(v, H).$$

Our objective is, for each source s and destination t, to devise an algorithm that maximizes the probability that s can reach t via at most k down-stream neighbors. Let $\overline{P}(s,t)$ denote this probability. Noting that maximizing a product of probabilities is the same as minimizing the sum of the $\log$ of their inverses, the K-shortest path version of Dijkstra's algorithm makes sense. Let $u_1, \ldots, u_k$ ($k \leq K$) be the neighbors of s "closer" to t in this metric, i.e. $\overline{P}(u_i,t)$ has been evaluated and they are larger than the current estimate of $\overline{P}(s,t)$. The probability that s can reach t via at least one of these neighbors can be re-estimated as

$$\begin{aligned} \overline{P}(s,t) &= 1 - \prod_{i=1}^{k}\big(1 - P(s,u_i)\overline{P}(u_i,t)\big) \\ &\geq 1 - \left(\frac{k - \sum_{i=1}^{k} P(s,u_i)\overline{P}(u_i,t)}{k}\right)^k \end{aligned}$$

The inequality follows from the arithmetic-geometric means inequality. Consequently, maximizing the sum $\sum_{i=1}^{k} P(s,u_i)\overline{P}(u_i,t)$ will push $\overline{P}(s,t)$ larger. The k-shortest path Dijkstra's algorithm aims to accomplish this by choosing at most K best terms $P(s,u_i)\overline{P}(u_i,t)$.

Once the K shortest paths are constructed, a node only needs to maintain the next hops for each of the paths (maximum of K entries per destination nodes). Packet routing in SOLAR-KSP is then carried out as follows.

When the source has a packet to send to the destination, it first checks if the destination is within radio range, in which case the packet is directly delivered. Else, it caches a copy of the packet for the destination and looks for the next hop on the path with the highest delivery probability. If that next hop node is within radio range, the packet is forwarded to it. If not, the packet is cached at the source for that next hop node. After a specified timeout period, if the source has not yet come in contact with either the next hop node or the destination, it selects the next path in the order of delivery probability and considers the next hop node in that path as the best candidate for packet forwarding.. Each node in the path repeats this same process as that packet gets forwarded towards the destination.

Note that for each of the three SOLAR versions, although several different variations (with separate tradeoffs) are possible, these basic schemes are sufficient to explore the knowledge to performance relationship in the context of our orbital mobility model.

IV. Performance Analysis

In this section, we describe our extensive simulation study that we carried out to compare the performance of the SOLAR protocols using the GloMoSim [36] simulator. We included Epidemic Routing protocol [32] (referred to as EPIDEMIC from here on), in our comparisons because of its simple yet efficient performance in face of general intermittent networks. In our implementation of SOLAR-DIST and SOLAR-PROB, we chose the maximum value for logical hops (for query packets) to be 2. In addition, to account for "routing holes" we adopted the simple notion of store-and-forward routing integrated into our greedy geographic forwarding, wherein in the event of a local maxima, the packet gets cached at the last node (and forwarded in time) instead of getting dropped immediately. In SOLAR-KSP, we consider a value of $K = 3$ and refer to the protocol as SOLAR-KSP3 from here on. For the simulation scenario, we considered a DTN built within a corporate campus consisting of several buildings (hubs). Corporate employees spend most of their time within the hubs and intermittently move in between hubs. To model realistic speeds of mobile users within such a network, we considered the work in [18], [34] and fixed the ORBIT Inter-Hub and Intra-Hub speed parameters, along with the other simulation parameters as shown in Table I. We chose two metrics to evaluate the performance of each protocol as described below:

<u>Data Throughput</u>: This metric is defined as the ratio of the total number of data packets received correctly by all destinations, to the total number of data packets generated by all sources, for the entire duration of the simulation.

<u>Relative Control Overhead</u>: This metric is defined as the amount of control information (measured in bytes) that each node sends for each successfully received data packet in the network. In SOLAR-DIST and SOLAR-PROB, the control packets are *Hello, Hub List Query, Hub List Response,* and *Location Update* packets. In SOLAR-KSP, the control overhead is due to the initial exchange of hub list information (through *Hello* messages) to build the k-shortest paths. In Epidemic Routing, the overhead is due to the exchange of *Hello* and *Summary Vector* messages.

In what follows, we will examine how different parameters such as total number of hubs (given a fixed terrain), hub size, radio transmission range, and the total number of nodes affect the protocol performance. To this end, we vary one of these four factors while fixing all other parameters. To obtain the results for a steady state of the DTN, we run the simulation for 1000 seconds without traffic at the start, followed by 1000 seconds of traffic generation, and conclude by another 1000 seconds of no new traffic. Each plot point in the results is averaged over 5 different simulation runs with varying random seeds.

A. Variation in Total number of Hubs

The number of hubs in the terrain affects protocol performance due to its direct impact on the expected node density within hubs, and the hub list sizes of each node.

Data Throughput: Figure 3(a) shows the data throughput of all the protocols with varying number of hubs. SOLAR-KSP3 performs the best, followed by SOLAR-PROB, SOLAR-DIST, and then EPIDEMIC. Due to the availability of complete hub list and its associated probability information, each node in SOLAR-KSP3 is able to learn about all other nodes in due time, and is able to locally compute the K shortest paths to all destinations. Thus, each node has up to K next hop nodes that it may contact to deliver the packet before giving up. Alternatively in SOLAR-PROB, although every node shares its hub list and associated probability information, there is no guarantee that each node will have a complete knowledge of the hub lists of all the nodes that have at least one hub in common. Thus, first of all each source node may need

TABLE I

SIMULATION PARAMETERS

GENERAL PARAMETERS			
Simulation Duration (each run)	3000s	Terrain Size	1000m x 1000m
Number of Nodes (*Users*)	*Vary*, (Default= 100)	Radio Range	*Vary*, (Default= 100m)
Cache Size	15 Data Packets	Cache Timeout	None
MAC Protocol	IEEE 802.11	Mobility Model	Random Orbit (RW + P2P)
ORBIT PARAMETERS			
Total Hubs (*Rooms*)	*Vary*, (Default= 15)	Hub Size	*Vary*, (Default= 50m x 50m)
Hub Stay Time	50s-100s	Orbit Timeout	250s-500s
Hub List Size	2 to Total Hubs	Inter-Hub Speed	10m/s-30m/s
Intra-Hub Pause	1s	Intra-Hub Speed	1m/s-10m/s
TRAFFIC PARAMETERS			
CBR connections	200 (5 packets each) Random	Data Payload	512 bytes per packet

to go through a hub list query and response phase, and then send the packet to the different hubs in the destination's hub list in the order of decreasing probability. Nodes in SOLAR-DIST are exempt from all hub probability information. Thus source nodes send a packet to the hubs in the destination's hub list in an order that minimizes the total distance traversed by the packet. However, a packet may take longer to reach the destination, or may not reach at all within the given simulation period, thereby incurring marginally lower throughput than SOLAR-PROB. Nodes in EPIDEMIC are unaware of any mobility information and only rely on mobility to effectively disseminate data. However, under the orbital mobility pattern, the probabilistic meeting of nodes do not favor EPIDEMIC, where each node may take substantial time to visit new hubs and meet other nodes. Moreover, the additional constraint of cache size seriously limits the capability of nodes in EPIDEMIC to store an appreciable amount of data to greedily "infect" each other, thereby leading to a poor performance.

With an increase in the number of hubs, the average number of nodes per hub decreases, and reduces the average number of neighbors for a node in a hub. In SOLAR-KSP3, this results in a node frequently caching a packet meant for a next hop node in the chosen shortest path. Since the next hop is based on the hub list probability of each node, given sufficient time, this node will find the next suitable hop and the packet will get delivered eventually. However, given the fixed simulation time, the throughput of SOLAR-KSP3 decreases due to the above reason. On the other hand, SOLAR-PROB and SOLAR-DIST use greedy geographic forwarding to send packets towards a hub. As the number of hubs increases, the nodes get spread out over the terrain, which is favorable for geographic forwarding. This is reflected in the marginal increase in throughput of both these protocols. In EPIDEMIC, the decrease in the average number of neighbors in a hub (with an increase in the number of hubs) also reduces the rate at which data is exchanged between new pairs of nodes, and thus further degrades the protocol performance.

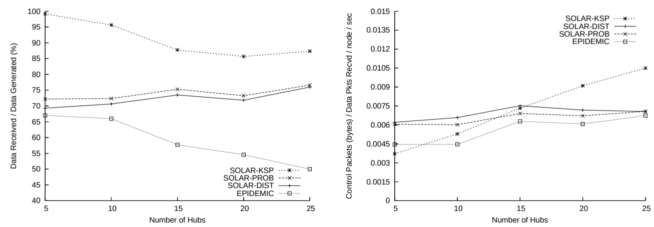
Relative Control Overhead: From Figure 3(b), we note that the relative control overhead incurred by the protocols shows an inverse relationship with that of their throughput performance. In SOLAR-PROB and SOLAR-DIST, the marginal changes in throughput with varying number of hubs does not reflect significantly in their relative control overhead performances. In EPIDEMIC, although the decrease in throughput with an increase in the number of hubs was quite significant, the control overhead does not increase much due to its efficient use of summary vectors, which help in performing data exchange selectively. On the other hand, in SOLAR-KSP3 as the number of hubs increases, the hub list size of nodes increases, which in turn increases the information every node shares with each other (i.e., probability information of a larger number of hubs in each node's list). This leads to an increase in the overall relative control overhead.

B. Variation in Hub Size

We study the effects of the hub size on the protocol performance in this section. In the following simulations, the hubs were considered to be square regions with varying sizes.

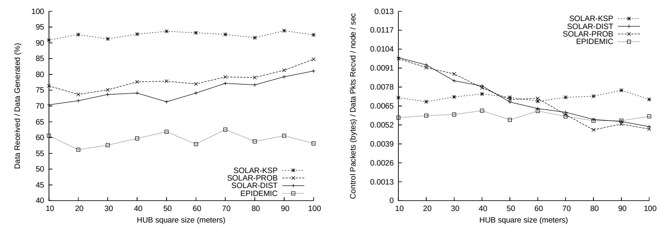
Data Throughput: Figure 4(a) shows results similar to Figure 3(a) and for similar reasons. The relative performance of the three SOLAR protocols are still consistent with the "knowledge to performance" relation as described in Section III-B. As hub size increases, SOLAR-KSP3 is not much affected due to the fact that the default radio range is still sufficient for nodes to keep track of the other nodes in the hub as neighbors. However, in SOLAR-PROB and SOLAR-DIST, as hub sizes increases, nodes get an opportunity to spread over larger regions in the terrain, that favors geographic forwarding by increasing the chances of one node finding another node to forward a packet to. Thus, the throughput in these latter two protocols are seen to increase with an increase in hub size. The spreading of nodes also fosters the process of data dissemination in EPIDEMIC. However, with limited buffer at each node, it does not have much effect on the data throughput.

Relative Control Overhead: As seen in Figure 4(b), the relative control overhead seems to vary inversely with the throughput of the protocols. For example, in SOLAR-PROB and SOLAR-DIST we see a steady decrease in the relative control overhead with an increase in hub size, that reflects the



(a) Data Throughput (%) (b) Relative Control Overhead

Fig. 3. Protocol Performance vs. Total Hubs



(a) Data Throughput (%) (b) Relative Control Overhead

Fig. 4. Protocol Performance vs. Hub Size (length of each side)

associated increase in throughput. On the other hand, SOLAR-KSP3 and EPIDEMIC are not much affected for reasons similar to its throughput performance.

C. Variation in Number of Nodes

In this section, we study the effect of varying the number of nodes in the network on the performance of the protocols. Given a fixed terrain size and radio transmission range, the total number of nodes directly impact the network connectivity.

Data Throughput: Once again, the relative performance of the three SOLAR protocols shown in Figure 5(a) are as expected. With an increase in the number of nodes, the network connectivity increases, which in turn favors all the protocols in different ways. In SOLAR-KSP3, a larger number of nodes is indicative of a larger number of average neighbors for each node. This increases the probability of a node finding one of the next hops on its K-Shortest Paths to a particular destination, in its neighborhood. In SOLAR-PROB and SOLAR-DIST, a larger number of nodes aids in geographic forwarding, where every intermediate node has a higher probability of finding a neighbor closer to the destination than itself. In EPIDEMIC however, a larger number of nodes only indicates higher amounts of data exchange, that quickly fills up the limited buffer space in each node, rendering the process of data infection ineffective. This leads to a rapid decrease in the data throughput with an increase in the number of nodes.

Relative Control Overhead: As shown in Figure 5(b), for all the protocols, the relative control overhead reduces with an increased number of nodes due to reasons discussed for their associated throughput comparison. In addition, for SOLAR-PROB and SOLAR-DIST, with a greater number of nodes, there is a higher probability of intermediate nodes responding to the queries for hub lists, which in turn decreases the amount of new queries which may have been generated by the source's acquaintances (in the event of zero knowledge of the queried destination). Thus, the relative control overhead for SOLAR-PROB and SOLAR-DIST reduces significantly with an increase in the number of nodes. In EPIDEMIC however, as the number of nodes increases leading to full buffers within nodes, the use of summary vectors leads to a very low number of packet exchanges, that results in the relative control overhead to decrease significantly as seen in Figure 5(b). In SOLAR-KSP3, a larger number of nodes marginally increases the subset of nodes each node shares a hub with. However, that increase in the absolute overhead is far less than the increase in data throughput, wherein the relative control overhead is seen to reduce rapidly with an increase in the number of nodes.

D. Variation in Radio Range

The effect of varying hub sizes given a fixed radio range has been discussed in Section IV-B. In this section, we study the effect of varying radio ranges given a fixed hub size.

Data Throughput: The results as seen in Figure 6(a) are intuitive. With a larger radio range, nodes in each protocol discover a larger number of average neighbors, which in turn helps all the protocols for reasons explained for Figure 5(a).

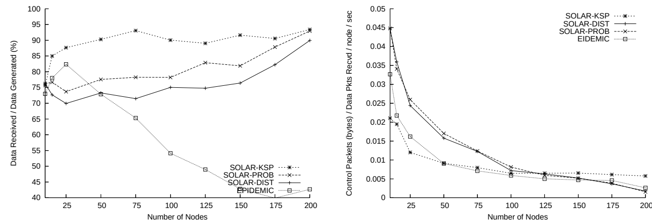
Relative Control Overhead: The relative control overhead is seen to decrease steadily in Figure 6(b) with an increase in the radio transmission range. This change is more significant in SOLAR-PROB and SOLAR-DIST than in SOLAR-KSP3, or EPIDEMIC. In SOLAR-KSP3, the absolute overhead in gaining complete knowledge is a constant, where each node needs to share with every other node their hub list and associated probability information. A larger radio range will only aid in accelerating this process, as nodes will have a larger number of neighbors on average. In EPIDEMIC, larger radio ranges initially fosters the exchange of data packets, but eventually the performance gets constrained by the limited buffer space within the nodes. Alternatively, in SOLAR-PROB and SOLAR-DIST, larger radio range helps nodes dynamically exchange their hub list (and associated probabilities for SOLAR-PROB) with a larger number of nodes, which in turn aids in the query process. In addition, greater number of neighbors also help in geographic forwarding. Thus, we see a more significant reduction in the relative control overhead for SOLAR-PROB and SOLAR-DIST with an increase in the radio range.

V. Conclusion

To summarize, based on the above study, we show that the use of sociological orbit aware hub list information can prove beneficial to routing in DTN, and that added information (in different degrees) regarding the associated hub probabilities can significantly improve the performance of the routing protocols.

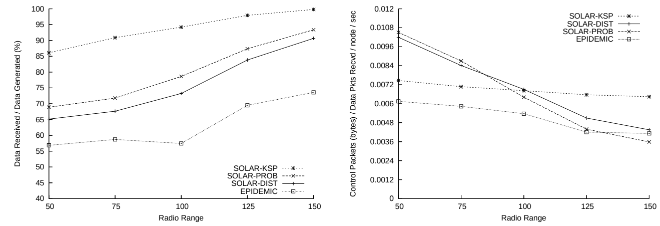
References

[1] Abowd, G., and Mynatt, E. Charting past, present, and future research in ubiquitous computing. *ACM Transaction on Computer Human Interaction 7*, 1 (March 2000), 29–58.



(a) Data Throughput (%) (b) Relative Control Overhead

Fig. 5. Protocol Performance vs. Number of Nodes



(a) Data Throughput (%) (b) Relative Control Overhead

Fig. 6. Protocol Performance vs. Radio Range

[2] Bai, F., Sadagopan, N., and Helmy, A. Important: a framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks. *Proceedings of IEEE INFOCOM '03 2* (March 2003), 825–835.

[3] Banerjee, S., and Chrysanthis, P. Peer support in a mobile world. *Proceedings of NSF Workshop on Context-Aware Mobile Database Management* (January 2002).

[4] Bar-Yehuda, R., and Even, S. A linear-time approximation algorithm for the weighted vertex cover problem. *J. Algorithms 2*, 2 (1981), 198–203.

[5] Beaufour, A., Leopold, M., and Bonnet, P. Smart-tag based data dissemination. *First ACM International Workshop on Wireless Sensor Networks and Applications (WSNA02)* (June 2002).

[6] Cerf, V., Burleigh, S., Hooke, A., Torgerson, L., Durst, R., Scott, K., Fall, K., and Weiss, H. Delay tolerant network architecture. *Internet Draft draftirtf-dtnrg-arch-02.txt (work in progress)* (July 2004).

[7] Dijkstra, E. W. A note on two problems in connexion with graphs. *Numerische Mathematik 1* (1959), 269–271.

[8] Garey, M., and Johnson, D. Computers and intractability: A guide to the theory of np-completeness. *W.H. Freeman and Company* (June 1979).

[9] Ghosh, J., Philip, S., and Qiao, C. Acquaintance Based Soft Location Management (ABSoLoM) in MANET. *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC) '04* (March 2004).

[10] Ghosh, J., Philip, S. J., and Qiao, C. Orbit mobility framework and orbit based routing (obr) protocol for manet. *Tech Report, CSE Dept. of University at Buffalo, Buffalo, NY, TR-2004-08* (July 2004).

[11] Ghosh, J., Philip, S. J., and Qiao, C. Sociological orbit aware routing in manet. *Tech Report, CSE Dept. of University at Buffalo, Buffalo, NY, TR-2005-09* (March 2005).

[12] Glance, N., Snowdon, D., and Meunier, J.-L. Pollen: using people as a communication medium. *First ACM International Workshop on Wireless Sensor Networks and Applications (WSNA02) 35*, 4 (March 2001), 429–442.

[13] Grossglauser, M., and Tse, D. N. C. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Transactions on Networking 10*, 4 (August 2002), 477–486.

[14] Harras, K. A., Almeroth, K. C., and Belding-Royer, E. M. Delay tolerant mobile networks (dtmns): Controlled flooding in sparse mobile networks. *To appear in Proceedings of Networking '05, Waterloo, Ontario, Canada* (May 2005).

[15] Jain, S., Fall, K., and Patra, R. Routing in a delay tolerant network. *Proceedings of ACM SIGCOMM'04* (September 2004).

[16] Johnson, D. S. Approximation algorithms for combinatorial problems. *J. Comput. System Sci. 9* (1974), 256–278. Fifth Annual ACM Symposium on the Theory of Computing (Austin, TX., 1973).

[17] Karp, B., and Kung, H. T. GPSR : Greedy perimeter stateless routing for wireless networks. *Proceedings of ACM/IEEE MobiCom '00* (August 2000), 243–254.

[18] Knoblauch, R. L., Pietrucha, M. T., and Nitzburg, M. Field Studies of Pedestrian Walking Speed and Start-Up Time. *Transportation Research Board Records*, 1538 (1996).

[19] Kuhn, F., Wattenhofer, R., and Zollinger, A. Worst-case optimal and average-case efficient geometric ad-hoc routing. In *MobiHoc '03: Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing* (New York, NY, USA, 2003), ACM Press, pp. 267–278.

[20] Ladd, A., Bekris, K., Rudys, A., Wallach, D., and Kavraki, L. On the feasibility of using wireless ethernet for indoor localization. *IEEE Transactions on Robotics and Automation 20*, 3 (June 2004), 555–559.

[21] LeBrun, J., Chuah, C. N., Ghosal, D., and Zhang, H. M. Knowledge-based opportunistic forwarding in vehicular wireless ad hoc networks. *To appear in IEEE Vehicular Technology Conference* (Spring 2005).

[22] Li, J., Janotti, J., Couto, D. S. J. D., Karger, D. R., and Morris, R. A Scalable Location Service for Geographic Ad Hoc Routing. *Proceedings of ACM/IEEE MobiCom '00* (August 2000), 120–130.

[23] Lindgren, A., Doria, A., and Schelen, O. Poster: Probabilistic routing in intermittently connected networks. *Proceedings of The Fourth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2003)* (June 2003).

[24] McCluskey, E. J. Minimization of boolean functions. *Bell Systems Technical Journal 35*, 5 (1956), 1417–1444.

[25] Musolesi, M., Hailes, S., and Mascolo, C. Adaptive routing for intermittently connected mobile ad hoc networks. *Proceedings of IEEE 6th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM'05), Taormina, Italy* (June 2005).

[26] Nicol, D. M., Goldsby, M. E., and Johnson, M. M. Simulation analysis of virtual geographic routing. *To appear in Proceedings of the Winter Simulation Conference* (2004).

[27] Quine, W. V. A way to simplify truth functions. *American Mathematical Monthly 62*, 9 (1955), 627–631.

[28] Rao, A., Papadimitriou, C., Shenker, S., and Stoica, I. Geographic routing without location information. In *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking* (New York, NY, USA, 2003), ACM Press, pp. 96–108.

[29] Remagnino, P., and Foresti, G. Ambient intelligence: A new multidisciplinary paradigm. *IEEE Transactions on Systems, Man and Cybernetics 35*, 1 (Jan 2005), 1–6.

[30] Samal, S. Mobility pattern aware routing in mobile ad hoc networks. *MS Thesis, Virginia Polytechnic Institute and State University* (May 2003).

[31] Su, W., Lee, S.-J., and Gerla, M. Mobility prediction and routing in ad hoc wireless networks. *International Journal of Network Management 11*, 1 (February 2001), 3–30.

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

[33] Wang, W., and Akyildiz, I. F. On the estimation of user mobility pattern for location tracking in wireless networks. *Proceedings of IEEE Globecom '02* (November 2002), 619–623.

[34] Williams, A. Cycling Speed. *Article in Peak Performance Online (http://www.pponline.co.uk/encyc/0065.htm)*, 65.

[35] Yoon, J., Liu, M., and Noble, B. Random Waypoint Considered Harmful. *Proceedings of IEEE INFOCOM '03 2* (March 2003), 1312–1321.

[36] Zeng, X., Bagrodia, R., and Gerla, M. Glomosim: a library for parallel simulation of large-scale wireless networks. *Proceedings of the 12th Workshop on Parallel and Distributed Simulations (PADS) '98* (May 1998), 154–161.

[37] Zhao, W., Ammar, M., and Zegura, E. Controlling the mobility of multiple data transport ferries in a delay-tolerant network. *In Proceedings of IEEE INFOCOM '05* (March 2005).