

Mobility-Centric Geocasting For Mobile Partitioned Networks

Michał Piórkowski
LCA, School of Computer and
Communication Sciences
EPFL, 1015 Switzerland
Email: michal.piorkowski@epfl.ch

Abstract—We describe our design of a geocast service for mobile partitioned networks (MPNs). We focus mainly on minimizing the delivery latency. Our approach exploits the time-stability of the collective mobility pattern.

In MPNs, in contrast to MANETs, the end-to-end path is frequently not available. Thus, communication in such networks becomes problematic. To overcome this difficulty, researchers propose a solution in which the node's mobility is exploited. This paradigm is often called mobility-assisted forwarding. In order to design routing protocols for MPNs, researchers study key mobility metrics, for example the inter-contact times between nodes. Based on the analysis of a real-life mobility trace, we show that the inter-contact time distribution is spatially dependent. This is a result of spatially heterogeneous mobility pattern that appears to be stable in time. We demonstrate that any georouting protocol designed to work in MPNs can benefit from knowing such an underlying mobility pattern. We propose an abstraction called mobility map that represents the collective mobility pattern. We also present how mobility maps can be used for georouting in MPNs and we also show a simple mechanism for the collaborative discovery of mobility maps. Finally, we propose a geocast protocol for MPNs - GeoMobCast - that explicitly uses mobility maps and is designed to minimize the expected message delay while maximizing the message delivery. We empirically evaluate the protocol by using simulations and we observe the improved performance, compared to other approaches.

I. INTRODUCTION

Recently, many research groups focused on MPNs, for which an end-to-end path between the source and the destination is often impossible to set up, disconnection and reconnection is common and link performance is highly variable. These networks are also known as Delay Tolerant Networks (DTNs) or Intermittently Connected Mobile Networks (ICMNs). In these networks, information between nodes is exchanged opportunistically, when they are within communication range. Very often the messages are transported physically by nodes. For these reasons different forwarding schemes proposed for this kind of networking are called mobility-assisted [3], [7], [8], [14], [19], [22], [23].

In order to design communication protocols for MPNs, researchers study key mobility metrics, such as contact and inter-contact times between mobile nodes. They appear also to be critical in determining the delay and the capacity of the network [7]. Very often, in the design of communication protocols for mobile partitioned networks, the spatial characteristics of mobility are abstracted away. We, however, believe that it is

very important to consider such information when designing protocols for MPNs. This might help to further improve the efficiency and effectiveness of these protocols, especially when the spatial mobility pattern appears to be time-stable.

The networking service responsible for sending messages to all nodes in a given geographic region is called geocast [24]. It belongs to a broader class of georouting protocols, where nodes' addresses correspond to their geographical locations. Georouting appears to be attractive when applied to large scale mobile ad hoc networks or sensor networks with mobile nodes. Today many wireless devices are, by default, equipped with a positioning device that has even access to a digital map, hence positioning in mobile networks is no longer a nightmare and an implementation of georouting protocols becomes less problematic. This class of protocols has been studied mostly in the context of MANETs [10]. The georouting protocols, specifically geocast, can be used to support location-based applications, such as road traffic management, traveler information support or location-based publish-subscribe systems, also for MPNs.

Efficient geocasting in MPNs is not easy. First, because one cannot rely on multihop communication in such networks due to frequent disconnectivity. Second, to design the optimal path for a message, one needs to know in advance the trajectories of all the nodes. Third, flooding the network with messages is costly.

In this paper we focus on the design of a new class of mobility-centric georouting protocols that operate in MPNs with a time-stable mobility pattern. Specifically, we address the problem of minimizing the message delay for mobility-centric georouting protocols. Our main contribution is a geocast protocol, called GeoMobCast, that exploits collective mobility pattern. This protocol implements the best effort geocast service and is designed to minimize the expected message delay while maximizing the message delivery.

To summarize, we make the following contributions:

- We present the guidelines for the design of georouting protocols that operate in mobile partitioned networks with a time-stable mobility pattern.
- We analyze the real mobility trace of a large taxi fleet to show the spatial heterogeneity of inter-contact time distribution.
- We introduce the concept of a mobility map that captures

the collective mobility pattern. We propose a distributed method for collaborative mobility map discovery.

- We present a geocast service for mobile partitioned networks called GeoMobCast, which explicitly uses mobility maps.

The paper is organized as follows. In Section II we define the system model and we present the design guidelines for georouting protocols operating in MPNs. In Section III we present evidence for the spatial heterogeneity of inter-contact time based on the analysis of a real mobility trace. Next, in Section IV we introduce the concept of a mobility map and we propose a decentralized method for collective discovery of the underlying mobility pattern. In Section V we present a mobility-centric geocast protocol for partitioned networks called GeoMobCast. We study its performance in Section VI. In Section VII we present the related work. Finally in Section VIII we conclude our work.

II. GEOROUTING IN MOBILE PARTITIONED NETWORKS

We now present an informal argument about the georouting protocols operating in mobile partitioned networks, which helps us to establish design guidelines for such protocols.

A. System Model and Assumptions

There are N mobile nodes moving independently on some connected and bounded region A . Each node is equipped with a positioning device (e.g. GPS) and - at any point in time - it knows its position on A . We assume that A can be divided into M subregions A_m ($m \in [1, M]$), where each A_m is a connected topological set (cf. Figure 1). All the nodes that belong to subregion A_m stay there forever and they follow the same mobility process, specific to A_m . As authors in [7] we consider a class of popular random mobility models, i.e., random waypoint, random direction and random walk. Each node has a short-range wireless communication device. All wireless devices can transmit at the same maximum power. We say that two nodes i and j are in contact if the distance between them is smaller than a radio communication range r_c . We assume contacts to be instantaneous. We call the inter-contact time $\tau_{i,j}$ the time between two successive contacts. As authors in [7] we assume that $\tau_{i,j}$ can be approximated by a random variable that is exponentially distributed with parameter λ and mutually independent. The expected inter-contact time is given then as $E[\tau_{i,j}] = \frac{1}{\lambda}$. Because all nodes that move within A_m follow the same mobility pattern then each such subregion can be characterized by a specific expected inter-contact time $E[\tau_m]$. Note also that such a setting can lead to heterogeneous node distribution, where there exist certain locations at which it is more likely to encounter nodes than elsewhere. This is a real-world phenomena and it has been reported already in the literature [1], [9], [20].

B. Minimum Expected Delay Georouting in MPNs

We consider the following setting. Assume that each subregion A_m has the same area $|A_m|$ and it is characterized by two parameters: ρ_m - the node density and $E[\tau_m] = \frac{1}{\lambda_m}$ -

the expected inter-contact time. Then, a best effort georouting protocol designed for MPNs, achieves the minimum expected message delay $E[\mathcal{T}_{m^*}]$ when packets are forwarded within the region A_{m^*} for which the ratio $\frac{E[\tau_{m^*}]}{\rho_{m^*}}$ is the smallest. To justify this argument, we study the following example. Consider an arbitrary node s that wants to send a message to a distant destination D , e.g. a geocast region. Assume that s sends M distinct copies of one message to the destination D in such a way that each distinct copy is forwarded in one subregion A_m . From [7] we know that for nodes following a mobility process as defined in II-A, the expected message delay under a multicopy protocol can be approximated by $E[\mathcal{T}] = \frac{1}{\lambda N} \sum_{i=1}^N i \simeq \frac{\log(N)}{\lambda N}$. In our setting for each A_m we approximate a corresponding message delay by $E[\mathcal{T}_m] \simeq E[\tau_m] \frac{\log(\rho_m |A_m|)}{\rho_m |A_m|}$. Given that $|A_m|$ is the same for all A_m , we can modify the above formula: $E[\mathcal{T}_m] \simeq E[\tau_m] \frac{\log(\rho_m)}{\rho_m}$. Note that in such a setting the copy of the message that is delivered first to the destination should traverse the subregion A_{m^*} for which the ratio $\frac{E[\tau_{m^*}]}{\rho_{m^*}}$ is the smallest ($m^* = \arg \min_{m \in [1, M]} \frac{E[\tau_m]}{\rho_m}$). Thus, the minimum expected message delay $E[\mathcal{T}_{m^*}]$ can be achieved if the message is forwarded within A_{m^*} .

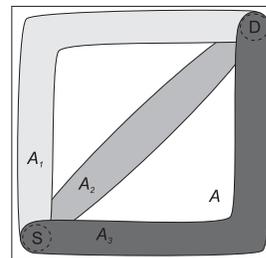


Fig. 1: Region A is divided into three subregions A_1 , A_2 and A_3 each having different mobility characteristics. The darker the subregion the lower the ratio $\frac{E[\tau_m]}{\rho_m}$. According to the argument from II-B messages have to be forwarded within A_3 .

The above example verifies the intuition that the more nodes within a certain region and the more frequently they are in contact, the higher the chance of minimizing the message delay. Thus any georouting protocol that aims to minimize the message delay has to know (i) how the forwarding subregion is connected topologically, (ii) what the node density within this subregion is and (iii) what the mobility characteristics there are.

III. EVIDENCE FROM A REAL MOBILITY TRACE

A. Dataset

For the purpose of our study, we use the GPS-based mobility traces of taxi cabs in San Francisco, USA. This data set (available upon request) contains the GPS coordinates of 665 taxis collected over 30 days in the Bay Area. Each taxi is equipped with a GPS receiver and sends a location-update (timestamp, identifier, geo-coordinates) to a central server. It appears that the location-updates are very frequent - the time

interval between the two consecutive location updates is less than 10 seconds, which allows us to accurately interpolate the cab location between the two close real location-updates. Note that this dataset does not contain information about node's connectivity, thus contacts need to be inferred. As in [11] we assume that two vehicles are in contact if the distance between them is less than or equal to a parameter r_c that corresponds to the radio communication range at maximum power. For our experiments we choose $r_c = 300$ meters (a typical maximum possible communication range for IEEE 802.11).

B. Spatial Heterogeneity of Inter-Contact Time Distribution

In this subsection we test whether the distribution of the inter-contact times is spatially dependent. We divide the whole region of San Francisco into a set of squared cells of equal size, measuring $4km^2$, each corresponding to the area of the block group (a geographical unit composed of multiple city blocks). Next, for each cell v we find the spatial sample of the inter-contact times by taking into consideration only the nodes that have visited a location v and were in contact with other nodes while visiting v . Consider nodes i and j visiting location v at time t_1 . The distance between i and j is small enough, such that they are in contact at time t_1 . Next, they may disconnect and potentially leave the location v . If later they visit v at time t_2 and they are again in contact, then the inter-contact time for pair (i, j) at location v is $\tau_{i,j}^v = t_2 - t_1$. Note that $\tau_{i,j} = \min_{\forall v} \tau_{i,j}^v$.

We select eight specific cells where the density of the nodes is higher than average, namely Aquatic Park, SF General Hospital, Shopping Center, Airport, Yellowcabs SF, the Bay Bridge, the Golden Gate Bridge and the University of SF. For each of those locations we find the inter-contact time distribution. In Figure 2 we present the empirical complementary CDF of the inter-contact times, specific for the eight preselected locations, recorded during one working day.

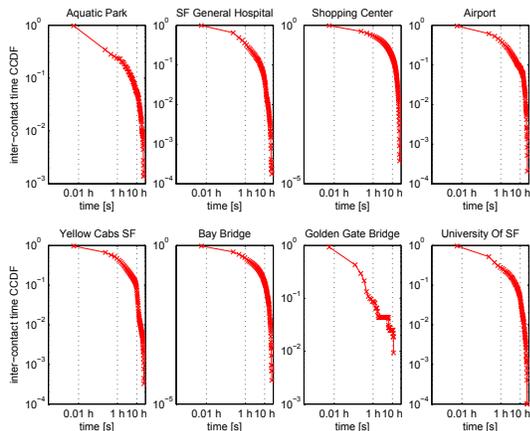


Fig. 2: The empirical distribution of the inter-contact times at different locations.

A visual inspection allows us to draw a conclusion that each location has its own specific inter-contact time pattern. We also perform a statistical test to see whether indeed all locations have different inter-contact time characteristics. We

use the Mann-Whitney U statistical test [5] to check whether the empirical samples of the inter-contact times measured at different locations come from the same distribution. The null hypothesis is that the two independent samples are drawn from a single population, and therefore that their probability distributions are equal. We perform this test for each pair of the eight preselected locations. Only in two cases out of 28, we could not reject the null hypothesis (SF General Hospital versus Airport and Airport versus Bay Bridge). Hence, we conclude that the inter-contact time distribution is spatially inhomogeneous.

C. Discussion

The above analysis proves our intuition that the inter-contact time is spatially dependent. It is correlated with the fact that certain locations appear to be more popular than others. We believe that by exploiting such an observation we can design more effective communication protocols for MPNs. For example, we could imagine a protocol that steers the message forwarding through highly populated locations that are characterized by the shortest expected inter-contact time. Note that it is not enough to rely on the information about the spatial distribution of nodes and the inter-contact time at specific locations only to perform successful message forwarding. For example, consider the situation when in between two distant highly populated locations the density is very sparse, such that the only way of transmitting packets from one dense location to the other is through mobility. We could say that between these two locations of high node density there exists a mobility link. When such links are time-stable then communication between high density locations could rely on such links. Note that a mobility link is more robust if more nodes are commuting between locations, as each commuting mobile node gives an extra opportunity to forward the packet. Hence, in order to find a proper connected area, within which the message should be propagated (cf. Section II-B), it is also important to know how, in terms of mobility, the highly populated locations are connected. Once this information is available to nodes, then they can find a sequence of high density locations, connected with mobility links, that need to be visited by the packet on its way to the destination. In the next section we introduce a concept called a mobility map, which provides nodes with information about stable mobility links between high density locations.

IV. MOBILITY MAP

A. Definition

A mobility map is a directed graph \mathcal{M} , whose vertex set $V(\mathcal{M})$ corresponds to the set of locations, e.g. city blocks, and whose edge set $E(\mathcal{M})$ to the set of mobility links between locations. We assume that there exists a mobility link between two distant locations u and w only if a node visits location u at time t and location v at time $t+T$: $\{X_i(t) \subset u \wedge X_i(t') \subset w\} \Rightarrow e = (u, w) \in E(\mathcal{M})$, where $T = t' - t$ is the travel time between locations u and w , and the $X_i(t)$ is the position of node i at time t . With each mobility link we associate

a transition probability $\pi_{(u,w)}^T$ that represents the probability that a random node moves from u to w in time T . We assume that it is given by: $\pi_{(u,w)}^T = \frac{1}{L} \sum_t \frac{|N_u(t) \cap N_w(t+T)|}{|N_u(t)|}$, where $N_u(t)$ is the set of nodes at time t present in location u , $N_w(t+T)$ is the set of nodes at time $t+T$ present in location w and L is a normalizing constant. We also assume that the transition probability associated with each mobility link is time-stationary. This assumption is crucial for the mobility-assisted forwarding in MPNs that relies on the time-stable mobility pattern. For example, a node knowing that at certain location it is very likely to encounter nodes that move towards a specific destination would try to either carry the message or to forward it to that location.

For every fixed lag T , the mobility map $\mathcal{M}(T)$ can be represented as a stochastic matrix that describes the time-homogeneous Markov chain over the finite space of states corresponding to the set of possible locations. Thus each node's mobility trace corresponds to a sample path of certain length of a Markov chain. If a time-stable mobility map is used to steer message forwarding process for a georouting protocol as suggested above, it has to have a specific property which we define here.

Explorability - when for any two locations u and v from a mobility map \mathcal{M} , both with non-zero node density, there exists almost surely a sample path, then we say that \mathcal{M} is explorable. The sufficient and necessary conditions for a mobility map to be explorable can be formulated using the Markovian taxonomy as follows. The sufficient condition for \mathcal{M} to be explorable is that $V(\mathcal{M})$ has to be an irreducible set of \mathcal{M} . The necessary condition for \mathcal{M} to be explorable is the following - if $\exists R \subset V(\mathcal{M})$ that is an irreducible set of \mathcal{M} , then each $v_k \in V(\mathcal{M}) \setminus R$ has to be accessible from at least one location $v_l \in R$ and each $v_k \in V(\mathcal{M}) \setminus R$ has to be able to access at least one location $v_l \in R$. Note that if a mobility map is explorable then the corresponding Markov chain is irreducible.

B. Evidence

1) Time-stability: To verify if the real-life mobility maps are time-stable we perform the following test on the collected mobility traces from San Francisco. We again superimpose a grid of squared cells, each measuring 500x500 [m] (a size of an average city block). Next, we generate mobility maps for different time lags $T \in \{10, 20, 50, 100, 200, 500, 1000, 2000, 5000\}$ [s]. Next, for each T we check if the corresponding mobility map $\mathcal{M}(T)$ is time-stable. To do so we define a reachable location w from location u to be a location for which pair $(u, w) \in E(\mathcal{M})$. Let the reachable location be a random variable Y_u with a specific probability density function. For every day d we define a set of reachable locations from location u , i.e., $L_u^d = \{w : (u, w) \in E(\mathcal{M}^d)\}$. The set of reachable locations L_u^d can be interpreted as a sample drawn from the corresponding distribution at day d . Thus, in order to verify if the mobility map is stable in time we need to check if on different days the set of reachable locations L_u^d comes from the same distribution. We again rely on the Mann-Whitney

U statistical test. The null hypothesis is that the two samples come from a single population, and thus their CDFs are equal. In our case for each location u we check if for every two days $d' \neq d''$ the samples $L_u^{d'}$ and $L_u^{d''}$ are drawn from the same distribution. For each location u that was visited at least by one mobile node during the period of 30 days we extract from the corresponding mobility map $\mathcal{M}^d(T)$ the set of reachable locations $L_u^d(T)$. For each time interval T we seek locations for which we cannot totally reject the null hypothesis, i.e., if there is at least one case when the null hypothesis has to be rejected for a particular $v \in V(\mathcal{M})$ then we do not consider v to be a time-stable location (a location that witnesses the same mobility pattern on different days). We represent this as η - the ratio of time-stable locations to all locations that were visited by mobile nodes within a period of 30 days. For example for $T = 100$ [s] approximately 10% of all visited locations can be considered as time-stable locations. In the table below we show the result of our test.

T [s]	10	20	50
η [%]	50.07	45.02	22.94
T [s]	100	200	500
η [%]	9.67	4.62	3.48
T [s]	1000	2000	5000
η [%]	0.87	0.15	0.44

We find that for short T , there are more time-stable locations. It is clear that the shorter the time span the easier to predict future node's location, given the current location. However, we observe that there exist a sufficient number of locations with a stable mobility pattern for which the reachable locations cover almost the whole San Francisco region (cf. Figure 3 (left vs. right)).

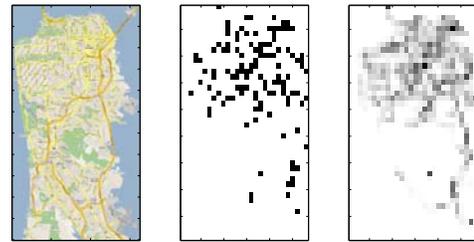


Fig. 3: (left) Real map of San Francisco; (center) All time-stable locations for $T = 100$ [s]; (right) All reachable locations from time-stable locations - the darker the pixel the higher the transition probability between a time-stable location and a reachable location.

2) Explorability: To verify if a mobility map is explorable one can use for example the edge contraction or the depth first search algorithm. We find that all $\mathcal{M}^d(T)$ are partially explorable, i.e., there exist locations that cannot communicate with the largest irreducible set R_{max} of a given $\mathcal{M}^d(T)$ - on average only 50% of all locations belong to R_{max} . However, we find that each R_{max} contains locations of the highest density, i.e., most of the cabs commute between locations from R_{max} . As the taxi trace is only a sample of the urban mobility pattern, we believe that in a mobility trace of different in type

mobile nodes (buses, taxis, pedestrians, etc.) we could find an explorable mobility map.

C. Application

Here, we present an example on the usability of the mobility map in mobility-assisted forwarding. In fact, this method is one of the components of the GeoMobCast protocol presented in Section V.

Let us consider a path $p_{(u,w)}$ on the mobility map between two distinct locations u and w . We define a collection of paths between u and w as $\mathcal{P}_{(u,w)} = \{p_{(u,w)}\}$. Let us define a cost function that associates a positive value with each path $p_{(u,w)}$:

$$C_{p_{(u,w)}} = \begin{cases} T \sum_{(v_k, v_l) \in p_{(u,w)}} \frac{(1 - \pi_{(v_k, v_l)}^T)}{\bar{N}_{v_l}} & \text{if } p_{(u,w)} \neq \emptyset \\ \infty & \text{if } p_{(u,w)} = \emptyset \end{cases} \quad (1)$$

Where \bar{N}_{v_l} is the time-average number of nodes at location v_l . Note that messages traversing paths of low cost have a much higher chance of making progress towards the destination in comparison to messages traversing paths with high costs. This is because the low cost paths contain mobility links with high transition probabilities between locations of high node density.

We define the minimum cost forwarding path: $p_{(u,w)}^*$ to be the path between two locations: u and w for which the cost function (cf. Eq. 1) takes the minimum, i.e.,

$$p_{(u,w)}^* = \operatorname{argmin}_{p_{(u,w)} \in \mathcal{P}_{(u,w)}} C_{p_{(u,w)}} \quad (2)$$

The minimum cost forwarding path $p_{(u,w)}^*$ can be used to forward messages using the mobility-assisted approach. This is because within the locations from the path $p_{(u,w)}^*$ it is very likely to encounter highly mobile nodes that move towards the location w . Hence, if an arbitrary node wants to send a geocast message: m from u to w , it can first consult its mobility map to see which path among all possible paths is the best for forwarding m .

Note that forwarding messages along $p_{(u,w)}^*$ has to be successful because (i) this path defines the connected space that contains both the source and destination (cf. Section II-B), (ii) the mobility pattern at the locations from $p_{(u,w)}^*$ is the most appropriate for moving messages from u to w and (iii) the density of nodes at this locations is high. If there are more than one $p_{(u,w)}^*$ a node could choose one of them randomly.

D. Collaborative Discovery of Mobility Maps

Here, we present a simple distributed method for a collective discovery of mobility maps. In order to build the mobility map in a decentralized manner, we assume that nodes can rely on an opportunistic gossiping method [6]. This method is simple and very robust for collective learning of common time-stable mobile network characteristics, e.g. a collective mobility pattern. This mechanism can be embedded in the neighborhood discovery service - one of the main building blocks for any DTN protocol, e.g. [27].

Recall that in order to estimate accurately the transition probability $\pi_{(u,w)}^T$ it is required that for every time t a node knows all the nodes present in location w at time $(t - T)$, as well as nodes from location u at time t . In real settings it might be hard to gain such knowledge as nodes are not synchronized and the neighbor discovery process may fail due to harsh radio communication conditions. Because of this, in the decentralized scheme, instead on $N_u(t)$, we rely on the $B_u^i(t)$ that represents the two-hop neighborhood of an arbitrary node i in location u at time t . Hence, we write $\hat{\pi}_{(u,w)}^T = \frac{|B_u^i(t-T) \cap B_u^i(t)|}{|B_u^i(t-T)|}$. Our assumption is that at any single location u the node distribution is homogeneous such that the $|N_u(t)|$ can be approximated by $|N_u(t)| \approx |B_u^i(t)| \frac{|u|}{4\pi r_c^2}$, where $|u|$ is the area of u and $4\pi r_c^2$ is the area covered by the two hop neighborhood centered at position of node i .

At first, the transition probabilities of all the mobility links from the mobility map at each node are initialized to zero. The mobility map is updated by each node independently. The neighborhood discovery service relies on a periodic broadcast of *hello* messages. Apart from the regular information (e.g. node ID) we assume that each *hello* message contains the following information:

- $\{i, \dots, j\}$ - list of node's one-hop neighbors
- $\{X_i(-T), \dots, X_j(-T)\}$ - their locations T time units ago
- $\mathcal{M}_i(T) = \{\hat{\pi}_e^{*T}\}_i$ - node's estimate of mobility map

Every node updates its map in rounds. The duration of the time interval between two consecutive rounds is Δt . At the beginning of each round, each node performs the following operations.

Using all received *hello* messages in the last round a node i constructs its two-hop neighborhood $B_u^i(t)$, where $X_i(t) \subset u$. Next, for every pair of locations $e = (v_k, u)$ visited by nodes from $B_u^i(t)$, node i finds the new estimate of the instantaneous transition probability of the mobility link, i.e., $\tilde{\pi}_e^T = \frac{|B_{v_k}^i(t-T) \cap B_u^i(t)|}{|B_{v_k}^i(t-T)|}$.

Then the estimate of the transition probability for a corresponding mobility link is updated using the exponentially weighted moving average:

$$\hat{\pi}_e^T(t + \Delta t) \leftarrow \alpha \tilde{\pi}_e^T + (1 - \alpha) \hat{\pi}_e^T(t) \quad (3)$$

Finally, for every received $\hat{\pi}_e^{*T}$ from each one-hop neighbor, the estimate of the transition probability for the corresponding mobility link is updated as follows:

$$\hat{\pi}_e^T(t + \Delta t) \leftarrow \beta \hat{\pi}_e^{*T}(t) + (1 - \beta) \hat{\pi}_e^T(t) \quad (4)$$

Thus each estimate of the mobility link is updated, at most, twice at every round - once, based on the individual observation of the node, and second based on the observations made by one-hop neighbors. The α and β constants (both $\in [0, 1]$) are protocol specific parameters and can be tuned accordingly.

Eventually, after enough periodic broadcasts and encounters, each node will have a common view on the mobility characteristics of the region in which they move. Due to the

lack of space, we omit the detailed evaluation of this method in this paper.

V. GEOMOBCAST

Here, we propose a geocast protocol for mobile partitioned networks, called GeoMobCast, designed to minimize the expected message delay by exploiting the nodes' collective mobility characteristics while maximizing the delivery success.

A. Protocol Overview

We assume that the mobility map, discussed earlier (cf. Section IV), is known to every node, i.e., that nodes have exchanged *hello* messages for long enough such that they are able to construct a consistent mobility map. When an arbitrary node i , located at u ($X_i(t) \subset u$), wants to send a geocast message m to a distant location w , it first finds the minimum cost forwarding path: $p_{(u,w)}^*$, as defined in Section IV-A. This forwarding path specifies the sequence of locations to be visited by message copies in their journey to the destination. If a message is forwarded along $p_{(u,w)}^*$ then the expected message delay is minimized, as explained in Section IV-C. The minimum cost forwarding path is then placed in the header of the message together with the corresponding counters that store information about the number of copies made at each location from $p_{(u,w)}^*$. Next, the originating node broadcasts this message to all nodes in the vicinity. A node that receives such a message first adds it to its buffer and then makes a new copy of this message with a certain probability P_{TX} and schedules it for sending at the next broadcast. This probability is proportional to (i) the number of copies made at the locations from $p_{(u,w)}^*$ (ii) the current distance to these locations and (iii) the distance between the previous and the current message copy holder. In general, the closer the current message holder is to a location from $p_{(u,w)}^*$ at which no messages were disseminated (so far) and the farther it is from the previous message holder, the higher the chance of broadcasting the message is. Upon reception of a geocast message, each node updates the counters for message copies made at a certain location from $p_{(u,w)}^*$ (both in its buffer and in the message header). Each node periodically checks the status of all messages in its buffer. All those messages that diffuse away from $p_{(u,w)}^*$ (i.e., if the distance between the node's reception location and $p_{(u,w)}^*$ is smaller than the distance between the current node's location and $p_{(u,w)}^*$) are removed from the message buffer. A message is said to be delivered when at least one copy enters the destination geocast region.

To summarize, if a message has to be delivered to the geocast region, that is out of reach for a message holder, it is pushed towards an intermediate location, where a stable mobility pattern of required characteristics is observed. The sequence of intermediate locations specifies the unique $p_{(u,w)}^*$ for message m along which the transmission should occur. Note that any intermediate location does not necessarily have to be geographically closer to the geocast region than the message holder is. This forwarding scheme benefits from the

stable collective mobility pattern, which allows the geocast message to "hitchhike" the proper mobile nodes. This is the main idea behind our protocol. Note that the proposed scheme is similar to the IP source routing [21] as the sender specifies the exact geographical route the packet must take.

B. Protocol Details

Message Format

Before we describe how the forwarding is performed we first define the format of the GeoMobCast message m . Its header contains the following fields:

- $m.ID$ - geocast message unique identifier, used to remove obsolete copies of the message from the network
- $m.D$ - the geocast region
- $m.P$ - the minimum cost forwarding path, i.e., $p_{(u,w)}^*$
- $m.K$ - the list of counters associated with $m.P$

Message Buffer

Each node that receives copy of m stores it in the dedicated message buffer. Whenever a duplicate of m is received, each node updates the appropriate fields in $m.K$ and/or in its buffer. A message is removed from the buffer only when (i) it is delivered directly by the node to the destination region or when (ii) the node receives a copy of the message that has visited all locations as specified in the message header or (iii) when a node moves away from $p_{(u,w)}^*$. Otherwise it is kept in the buffer.

Forwarding Rules

The forwarding decisions are probabilistic and are taken independently by each node. They depend on the status of the current message copy holder i , previous message copy holder j and the copy of the message m . Specifically, the probability of making a new copy is given by:

$$P_{TX} = \begin{cases} P_{p^*} \frac{|X_i - X_j|}{r_c} & \text{if } |X_i - X_j| \leq r_c \\ P_{p^*} & \text{otherwise} \end{cases} \quad (5)$$

Where $P_{p^*} = \max_{v \in m.P} \exp(-(\frac{|X_i - X_v|}{r_c} + \frac{\gamma_v}{\gamma^*}))$ and $|X_i - X_v|$ is the distance between the node i and the center of location $v \subset m.P$, $\gamma_v \subset m.K$ is the total number of message copies made at location v , the $\gamma^* \in [1, \gamma_{max}]$ is the maximum number of copies at a specific location made so far, and $|X_i - X_j|$ is the distance between the previous (j) and the current message copy holder (i).

When each new copy in the network is made with probability P_{TX} , then the following applies:

A message is propagated along the minimum cost forwarding path: It is more likely for a node to make a new copy of a message m next to a location from $m.P$ than elsewhere, because the probability of making new copies is proportional to the distance to such locations.

A message is more likely propagated towards destination: Whenever a copy of the message m is received by a new node at one of the locations from $m.P$, then the corresponding counter in the message header is incremented. Given that, at every location from $m.P$, the number of copies cannot exceed

the predefined maximum number: γ_{max} - new copies are made mostly at or close to locations not yet visited by the message m . Note that γ_{max} can be adjusted according to the local node density (the higher the density, the lower the γ_{max}). Finding the optimal value for γ_{max} is out of the scope of this paper.

A message is replicated more likely by a distant node: Nodes located farther from the previous message copy holder have a higher chance of making a new copy, compared to the nodes that are closer to the previous message copy holder. This reduces the overhead and accelerates message forwarding in dense regions.

Removing Obsolete Copies

In GeoMobcast there is no explicit acknowledgment sent back to the message originator because of no end-to-end connectivity. Instead, in order to suppress the number of message copies, the protocol relies on the following mechanisms.

Suppressing by counting: Whenever a message m visits a location specified in $m.P$ then the message holder marks this location as visited by updating the corresponding counter in $m.K$. If a duplicate of a message is received, then the $m.K$ is copied to the message buffer - only if this particular copy of m has made progress of which the node is not aware. In the opposite case, the node copies the values of the counters to $m.K$, such that more up-to-date information about message m will be propagated. If the number of copies at all locations from $p_{(u,w)}^*$ reaches γ_{max} at each location then the message is removed from the buffer.

Messages that do not make progress are dropped: Each node periodically checks if in its buffer there are messages that have not made any progress towards the destination. Thus, if a geocast message diffuses away from the $p_{(u,w)}^*$ it is removed from the buffer. This is done by checking every T_{buff} time units if the distance between the message copy holder to the $p_{(u,w)}^*$ has decreased since the reception of the message copy. This is important because obsolete messages take up unnecessary space in the message buffer.

Operational Example

Consider the scenario depicted in Figure 4. The messages are sent from region S to region D . The minimum cost forwarding path $p_{(S,D)}^*$ defines the connected region A_{m^*} in which the message should be propagated. The darker the part of the subregion A_{m^*} is, the more copies of the message there are. Here, we emphasize four nodes g, h, i, j at two time instants $t - T$ and t , each having copy of the message m at time $t - T$. As node g enters $p_{(S,D)}^*$, where the number of copies has exceeded the allowed limit, it will not make new copies of the message there. The node h at time t drops the message from its buffer as it moves away from A_{m^*} . The node j will have a much higher chance of disseminating a new copy of the message, compared to node i , as it moves to the region where the number of copies made is small and it is closer to D .

VI. EVALUATION

In this section we study the performance of the proposed geocast protocol. We first present the evaluation methodology

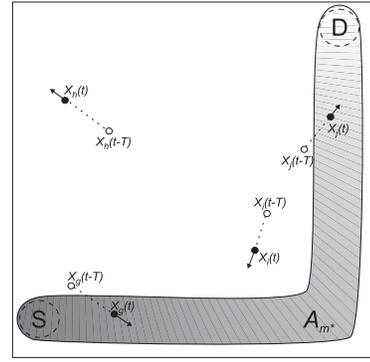


Fig. 4: Visualization of the GeoMobCast protocol operation. The shaded subregion represents the minimum cost forwarding path. The darker the subregion A_{m^*} is, the more copies of the message there are. A geocast message m is sent from S to D .

and define the performance metrics. Then, we describe the simulation setup and we present the results.

Methodology

We compare our method against the Mobility-based Adaptive Greedy Forwarding (MAGF) georouting protocol designed to operate in DTNs [15]. This is a single-copy protocol that takes advantage of individual mobility to enhance greedy forwarding in georouting. In a nutshell the idea behind the protocol is the following. Before making the forwarding decision the current message holder needs to select the most appropriate relay from its one-hop neighborhood. Nodes from one-hop neighborhood are divided into two groups - those that belong to the progressive region and those that belong to the potential region (cf. Figure 5). Nodes from the progressive region are geographically closer to the destination than the sender.

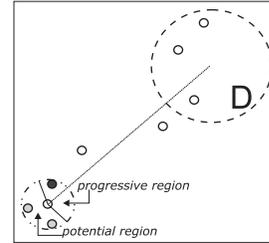


Fig. 5: The main idea behind the Mobility-based Adaptive Greedy Forwarding protocol that is used as a base for the comparison with GeoMobCast protocol.

In MAGF, by default, the message is forwarded to a node from the progressive region. In case there is no node in the progressive region, the message is passed to one of the nodes from the potential region. The node for which the motion potential is the highest is chosen as the relay. The higher the speed of a node from the potential region and the proper the node's heading (towards the destination) is, the higher the motion potential of that node is. For further details we refer the reader to the original work [15]. To the best of our knowledge this protocol is one of the first protocols that uses the mobility-centric design approach for georouting in MPNs. We choose the MAGF for the following reasons. First, it considers the

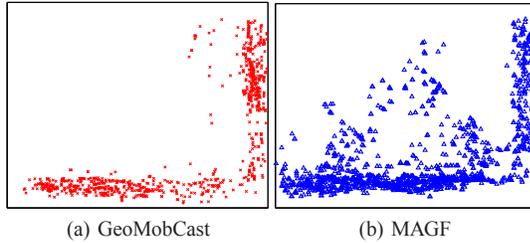


Fig. 6: Sanity check for both protocols. Each marker represents a location at which a geocast message, addressed to the region \mathcal{D} (cf. Figure 4), was received by a mobile node.

same setting, i.e., georouting in mobile partitioned networks. Second, comparing our multi-copy geocast protocol against a single-copy protocol allows us to quantitatively verify the tradeoff between the overhead of GeoMobCast versus its robustness.

We implement both schemes in the realistic wireless ad hoc network simulator JiST/SWANS [25]. We modify the MAGF protocol such that it supports geocast, i.e., the destination is an arbitrary subregion. We perform exhaustive simulations with realistic PHY (pathloss, fading) and MAC (IEEE 802.11) models. The maximum radio communication range is 300 meters. For GeoMobCast we use the MAC layer operating in the promiscuous mode. For both protocols we implement a specific networking layer that supports geographical addressing, based on the extensions to the original JiST/SWANS simulator [26]. As we consider high mobility scenarios, where speed can vary from 1 to 31 [$\frac{m}{s}$], we use frequent periodic broadcast of *hello* messages - a new *hello* message is sent every 1.5 second. Nodes move according to the Restricted Random Waypoint Mobility (RRWP) model [2]. RRWP is capable of modeling the inhomogeneous spatial node distribution as observed in reality. Specifically, it allows nodes to pick up certain waypoints more often than others. Hence, it is convenient for the generic modeling of node mobility in regions with obstacles, e.g. students on a campus or cars on a road network. Note that a particular mobility map can be viewed as a specific realization of this model. Thus, for the purpose of the evaluation we use a simple mobility map that defines a specific realization of the RRWP model. In Figure 4 we visualize the mobility map used for evaluation. The whole region of size 5000 x 5000 [m] is divided into two subregions. Mobile nodes select their waypoints preferably from the shaded subregion A_{m^*} . In the time-stationary regime 80% of nodes are moving within the subregion A_{m^*} . To make sure that a geocast message can be carried physically by a node moving directly from the source to the destination region at an average speed (15 [$\frac{m}{s}$]) we setup $T_{buff} = 500$ [s] for GeoMobCast and $T_{cache} = 500$ [s] for MAGF.

For the sanity check in Figure 6 we show an exemplary trace of successful message receptions for both protocols. Each marker corresponds to a location at which a geocast message sent from \mathcal{S} to \mathcal{D} was successfully forwarded. For GeoMobCast the message is propagated only along the minimum cost

forwarding path. This is not the case for the MAGF protocol as it is not aware of the underlying mobility pattern.

Performance Metrics

Here we specify the metrics used to assess the quality of the proposed protocol.

- message delay - the time needed to transmit a geocast message from the source to the geocast region.
- message delivery - the total number of delivered messages divided by the total number of sent messages.
- extinction time - the interval between the time when the message was delivered and the time when the last copy of the message was made.
- message overhead - the average number of message copies made.

Results and Discussion

In Figure 7a we compare the latency of MAGF and GeoMobCast protocols. We find the median, the lower and upper quartile value of the message delay for both protocols. In both cases the average message delay decreases with the number of nodes. The average message delay is approximately the same for both protocols, which suggests that they are equally good. However, when we consider the message delivery (cf. Figure 7b) it becomes clear that GeoMobCast outperforms MAGF significantly. This is for two reasons. First, GeoMobCast is aware of the underlying mobility pattern and pushes the message towards the regions where it has both a higher chance of encountering the right ferry and/or where the node's density is high (cf. Figure 6). Second, GeoMobCast relies on multi-copy forwarding and this increases the chance of successful delivery.

Unfortunately the GeoMobCast protocol pays a certain price for its effectiveness. To validate its efficiency we consider two metrics: extinction time and message overhead. The extinction time metric shows for how long, after the successful delivery, the message is still propagated in the network excluding the geocast region. The message overhead metric shows the average number of message copies made per one geocast message.

In Figure 7c we show the average message extinction time. We find the median, the lower and upper quartile values of the extinction time. Ideally the extinction time has to be ≤ 0 . In all cases the average extinction time is close to 0 and it does not depend on node density. This is because each copy of the message contains up-to-date information about the number of copies made at locations from $p_{(u,w)}^*$ and the protocol fully controls the copy-making process. In the case of the message overhead (cf. Figure 7d) we observe that the average number of message copies grows linearly with the number of nodes. However, the number of copies made is relatively small compared to the number of nodes present in the network. For example, in the network of 400 nodes, where 80% of them move along $p_{(u,w)}^*$, only 40 copies (on average) are made during the forwarding process. The message delay is approximately the same for both protocols, as both try to push the message towards the destination at the same speed by favoring nodes that make progress towards the

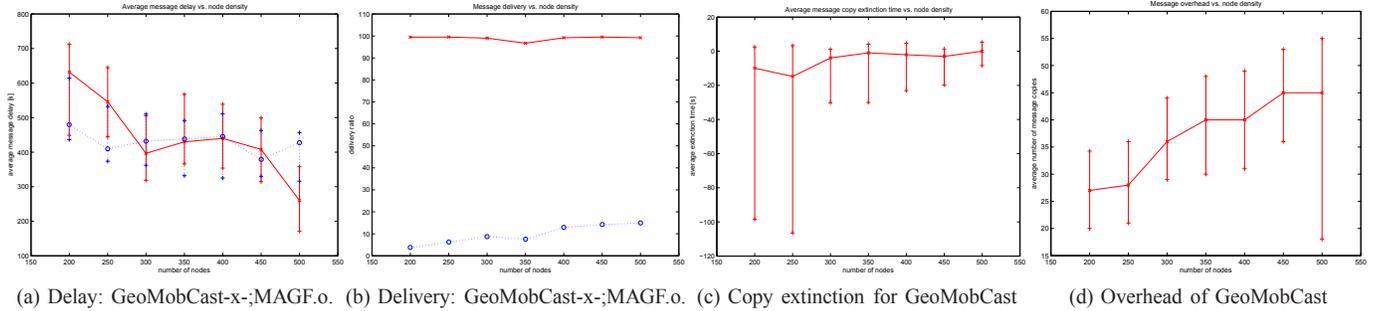


Fig. 7: Performance Evaluation of GeoMobCast versus MAGF

destination. In terms of the message delivery the GeoMobCast protocol outperforms MAGF, mainly because it is aware of the collective mobility pattern. The message overhead might be improved, for example, by adaptively adjusting the γ_{max} parameter.

VII. RELATED WORK

An exhaustive survey on georouting, geocasting and position-based routing protocols can be found in [17], [18]. Here we focus on geocasting protocols that could be used in MPNs that either relay on a certain type of limited flooding or make use of the mobility-centric design paradigm.

Flooding reduction in geocasting algorithms is achieved mainly by finding so-called restricted forwarding zones. They contain both the source and destination subregions. For example in Location Based Multicast (LBM), proposed by Ko and Vaidya [12], the minimum rectangular region containing the source and destination subregions is chosen as a forwarding zone. Then the restricted flooding is performed by nodes within this zone. Another approach, as an extension to the LBM protocol, is proposed by the same authors in [13]. This approach differs from LBM in the way the forwarding zone is chosen - the physical location information is used to reduce the overhead of geocast delivery. In contrast to these two methods the GeoMobCast protocol chooses the forwarding zone more carefully, by taking into account the time-stable collective mobility characteristics.

The idea of forwarding messages towards the destination along specific locations was already proposed by Blazevic et al. in [2], as well as by Lim et al in [16]. In [2], Blazevic et al. propose a protocol called Terminode Remote Routing that uses anchored paths, a list of locations used as loose source routing [21]. Anchored paths are discovered and managed by sources using dedicated protocols, e.g. the Geographical Map-based Path Discovery, which is similar to our method used in GeoMobCast for discovering the minimum cost forwarding path. However, in our case the cost value associated with each path depends specifically on the mobility characteristics and not on the geographical distances. Moreover, we also propose a method for collaborative mobility map discovery and based on the real-life mobility trace analysis we show that such maps are time-stable. In [16], Lim et al. propose the Landmark Guided Forwarding protocol that gives a hybrid solution of topological

and geographical routing algorithms. Although they address a different problem, the use of so-called geographical landmarks is similar to our proposal, i.e., if the packet destination resides outside the local scope it is routed towards a geographically determined optimal landmark node.

All the above geocasting and georouting schemes are designed to operate mostly in dense networks, even not necessarily mobile. In this work we focus on mobile partitioned networks. In the past few years many research groups proposed different DTN routing protocols, e.g. [3], [8], [19], [22]. In order to design efficient protocols, researchers study the mobility characteristics quantified by the inter-contact time distributions [3], [4], [8], [11]. In [4] Conan et al. made an interesting observation that more attention has to be paid to the inhomogeneity of the inter-contact time as it allows for the design of even more efficient mobility-assisted forwarding protocols. Following this observation we focus on the spatial dependency of the inter-contact time in mobile partitioned networks. To the best of our knowledge this is the first approach to analyze the inter-contact time distribution in the spatial context.

The concept of mobility-centric protocol design has been present in the networking community for some time. In [28], Wu et al. propose a Mobility-Centric Data Dissemination Protocol. The proposed protocol is designed for vehicular networks to operate efficiently and reliably despite the highly mobile nature of such networks. In [19], Merugu et al. focus on designing a routing protocol for DTNs with predictable mobility. They propose to construct space-time routing tables where the next hop node is selected from the current, as well as the future, neighbors. In [14], Leguay et al. propose a DTN routing algorithm based on the use of a high-dimensional Euclidean space, called MobySpace, constructed upon nodes' individual mobility patterns. In contrast to all the above proposals in this work we focus on exploiting the collective mobility pattern.

VIII. CONCLUSION AND FUTURE WORK

In this paper we present the GeoMobCast, a geocast protocol designed to operate in mobile partitioned networks with time-stable mobility characteristics. We also propose a method for collaborative mobility map discovery, which allows nodes to estimate mobility characteristics specific to a certain region.

Based on the real-life mobility trace of a large taxi fleet, we demonstrate that real mobility maps are time-stable and moreover can be used to find so-called forwarding paths used to perform source routing-like message forwarding. We also give evidence that the distribution of the inter-contact time (a metric used to characterize mobility in mobile partitioned networks) is spatially dependent. We suggest that this observation needs to be taken into account when designing georouting protocols for mobile partitioned networks. We also show that georouting protocols designed for mobile partitioned networks, which aim to minimize the expected message delay, should be aware of mobility characteristics in the regions where messages are forwarded.

For our future work we plan to improve the GeoMobCast protocol by reducing its overhead. A natural step would be to combine GeoMobCast with the MAGF protocol. For example, the forwarding decisions could be taken in two steps. First, the minimum cost forwarding path, which consists of intermediate geographical landmarks of certain mobility characteristics, can be chosen as it is done for the current version of GeoMobCast. Second, the protocol could relay on the MAGF protocol to forward the packets towards a specific intermediate landmark using a fanout of mobile nodes that either are geographically closer to the landmark location or have high mobility potential.

It has been shown recently by Yan et al. [29] that worm propagation in mobile opportunistic networks significantly depends on the mobility pattern. Hence, we believe that mobility maps can be used also to design mobility-centric worm mitigation techniques. For example mobility maps can be used to steer patches towards regions of the highest density and/or mobility, to protect the network against worm spreading. We plan to address this problem in our future work.

We also plan to study in detail the performance of the collaborative mobility map discovery. For example it will be interesting to know what the convergence time of this method is - "How long does it take for nodes to build a consistent mobility map, which is approximately the same at each node?". Also, we plan to verify the accuracy of the mobility map that is built using the proposed technique. It is also interesting to study how inaccurate or incomplete mobility maps influence the message forwarding in MPNs.

ACKNOWLEDGMENT

The authors would like to thank Matthias Grossglauser, George Theodorakopoulos, Natasa Sarafijanovic-Djukic and Holly Cogliati for help in improving the manuscript. We are indebted to the Exploratorium, the Museum of Science, Art and Human Perception for making the GPS traces publicly available through the cabspotting¹ project.

The work presented in this paper has been supported (in part) by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5006-67322.

¹<http://cabspotting.org>

REFERENCES

- [1] M. Balazinska and P. Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *ACM MobiSys*, 2003.
- [2] L. Blazevic, J.-Y. L. Boudec, and S. Giordano. A location-based routing method for mobile ad hoc networks. *IEEE TMC*, 4(2):97–110, 2005.
- [3] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of Human Mobility on Opportunistic Forwarding Algorithms. *IEEE TMC*, 6(6):606–620, 2007.
- [4] V. Conan, J. Leguay, and T. Friedman. The Heterogeneity of Inter-contact Time Distributions: Its Importance for Routing in Delay Tolerant Networks, 2007. arXiv:cs/0609068v2.
- [5] W. J. Conover. *Practical Nonparametric Statistics*. Wiley, 1999.
- [6] R. Friedman, D. Gavidia, L. Rodrigues, A. C. Viana, and S. Voulgaris. Gossiping on MANETs: the Beauty and the Beast. *SIGOPS Oper. Syst. Rev.*, 41(5):67–74, 2007.
- [7] R. Groenevelt, P. Nain, and G. Koole. The Message Delay in Mobile Ad Hoc Networks. *Perform. Eval.*, 62(1-4):210–228, 2005.
- [8] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket Switched Networks and Human Mobility in Conference Environments. In *ACM SIGCOMM WDTN'05*.
- [9] W. j. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy. Modeling Time-Variant User Mobility in Wireless Mobile Networks. In *IEEE INFOCOM'07*.
- [10] X. Jiang and T. Camp. A Review of Geocasting Protocols for a Mobile Ad Hoc Network. In *GHC'02*.
- [11] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnovic. Power law and Exponential Decay of Inter Contact Times between Mobile Devices. In *ACM MobiCom'07*.
- [12] Y.-B. Ko and N. Vaidya. Geocasting in Mobile Ad Hoc Networks: Location-Based Multicast Algorithms. In *IEEE WMCSA'99*.
- [13] Y.-B. Ko and N. H. Vaidya. Flooding-Based Geocasting Protocols for Mobile Ad Hoc Networks. *Mobile Networks and Applications*, 2002.
- [14] J. Leguay, T. Friedman, and V. Conan. Evaluating Mobility Pattern Space Routing. In *IEEE INFOCOM'06*.
- [15] J. Li and S. M. Shatz. Toward Using Node Mobility to Enhance Greedy-Forwarding in Geographic Routing for Mobile Ad Hoc Networks. In *MODUS'08*.
- [16] M. Lim, A. Greenhalgh, J. Chesterfield, and J. Crowcroft. Landmark Guided Forwarding. *IEEE ICNP'05*.
- [17] C. Maihofer. A Survey of Geocast Routing Protocols. *IEEE Comm. Surveys and Tutorials*, 6(2):32–42, 2004.
- [18] M. Mauve, J. Widmer, and H. Hartenstein. A Survey on Position-Based Routing in Mobile Ad Hoc Networks. *IEEE Network*, 15(6):30–39, 2001.
- [19] S. Merugu, M. Ammar, and E. Zegura. Routing in Space and Time in Networks with Predictable Mobility. Technical Report GIT-CC-04-7, Georgia Institute of Technology, 2004.
- [20] M. Piórkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. On Clustering Phenomenon in Mobile Partitioned Networks. In *ACM MobilityModels*, 2008.
- [21] (RFC791) Internet Protocol Specification. <http://tools.ietf.org/html/rfc791>.
- [22] N. Sarafijanovic-Djukic, M. Piórkowski, and M. Grossglauser. Island Hopping: Efficient Mobility-Assisted Forwarding in Partitioned Networks. In *IEEE SECON'06*.
- [23] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Spray and Wait: an efficient routing scheme for intermittently connected mobile networks. In *ACM SIGCOMM WDTN'05*.
- [24] I. Stojmenovic. Geocasting with Guaranteed Delivery in Sensor Networks. *IEEE Wireless Communications*, 11(6):29–37, 2004.
- [25] JiST/SWANS. <http://jist.ece.cornell.edu>.
- [26] Extended version of JiST/SWANS. <http://vanet.info>.
- [27] W. Wang, V. Srinivasan, and M. Motani. Adaptive Contact Probing Mechanisms for Delay Tolerant Applications. In *ACM MobiCom'07*.
- [28] H. Wu, R. Fujimoto, R. Guensler, and M. Hunter. MDDV: A Mobility-Centric Data Dissemination Algorithm for Vehicular Networks. In *ACM VANET'04*.
- [29] G. Yan, H. D. Flores, L. Cuellar, N. Hengartner, S. Eidenbenz, and V. Vu. Bluetooth Worm Propagation: Mobility Pattern Matters! In *ACM ASIACCS*, pages 32–44, 2007.