3-2010

Moving Targets: Geographically Routed Human Movement Networks

Adam J. Aviv
University of Pennsylvania, aviv@cis.upenn.edu

Micah Sherr
University of Pennsylvania, msherr@cis.upenn.edu

Matt Blaze
University of Pennsylvania, blaze@cis.upenn.edu

Jonathan M. Smith
University of Pennsylvania, jms@cis.upenn.edu

Recommended Citation

Adam J. Aviv, Micah Sherr, Matt Blaze, and Jonathan M. Smith, "Moving Targets: Geographically Routed Human Movement Networks", . March 2010.

Moving Targets: Geographically Routed Human Movement Networks

Abstract
We introduce a new communication paradigm, Human-to-human Mobile Ad hoc Networking (HuManet), that exploits smartphone capabilities and human behavior to create decentralized networks for smartphone-to-smartphone message delivery. HuManets support stealth command-and-control messaging for mobile BotNets, covert channels in the presence of an observer who monitors all cellular communication, and distributed protocols for querying the state or content of targeted mobile devices.

In this paper, we introduce techniques for constructing HumaNets and describe protocols for efficiently routing and addressing messages. In contrast to flooding or broadcast schemes that saturate the network and aggressively consume phone resources (e.g., batteries), our protocols exploit human mobility patterns to significantly increase communication efficiency while limiting the exposure of HuManets to mobile service providers. Our techniques leverage properties of smartphones – in particular, their highly synchronized clocks and ability to discern location information – to construct location profiles for each device. HuManets' fully-distributed and heuristic-based routing protocols route messages towards phones with location profiles that are similar to those of the intended receiver, enabling efficient message delivery with limited effects to end-to-end latency.

Keywords
Human-to-Human networks, HumaNet, Geographic Routing, Botnets, Mobile Botnets

Disciplines
Other Computer Sciences | Software Engineering

Moving Targets: Geographically Routed Human Movement Networks*

Technical Report: MS-CIS-10-12
March 2010, University of Pennsylvania

Adam J. Aviv   Micah Sherr   Matt Blaze   Jonathan M. Smith

{aviv,msherr,blaze,jms}@cis.upenn.edu

Abstract

We introduce a new communication paradigm, *Human-to-human Mobile Ad hoc Networking* (HUMANET), that exploits smartphone capabilities and human behavior to create decentralized networks for smartphone-to-smartphone message delivery. HUMANETs support stealth command-and-control messaging for mobile BotNets, covert channels in the presence of an observer who monitors all cellular communication, and distributed protocols for querying the state or content of targeted mobile devices.

In this paper, we introduce techniques for constructing HUMANETs and describe protocols for efficiently routing and addressing messages. In contrast to flooding or broadcast schemes that saturate the network and aggressively consume phone resources (e.g., batteries), our protocols exploit human mobility patterns to significantly increase communication efficiency while limiting the exposure of HUMANETs to mobile service providers. Our techniques leverage properties of smartphones – in particular, their highly synchronized clocks and ability to discern location information – to construct *location profiles* for each device. HUMANETs' fully-distributed and heuristic-based routing protocols route messages towards phones with location profiles that are similar to those of the intended receiver, enabling efficient message delivery with limited effects to end-to-end latency.

1 Introduction

The convergence of computing and communications technologies has resulted in the "smartphone" – a highly-portable communications device which is *also* a computing device with substantial processing capability and storage capacity. This transformation has been so effectively achieved that users routinely download and install new software on their phones as if it were a desktop or laptop computer. In this paper, we show that a combination of human behavior and smartphone features enables novel networking capabilities, with implications for privacy and security. In particular, we utilize several features of humans and smartphones:

First, the presence of multiple communication channels, such as 802.11 and Bluetooth, permits messages to be exchanged with physically proximate peers without resorting to the use of the traditional cellular infrastructures and carriers. For example, the Texas Instruments OMAP 4 [16] System-on-a-Chip (SoC) to support smartphones (Figure 1) illustrates many key architectural features of communications/computing convergence, including multiprocessing and a variety of accelerators, in addition to plentiful communications. Although smartphones are designed for use in highly centralized and tightly controlled cellular

*Errata published February 2011: www.cis.upenn.edu/~aviv/papers/targets-errata.pdf

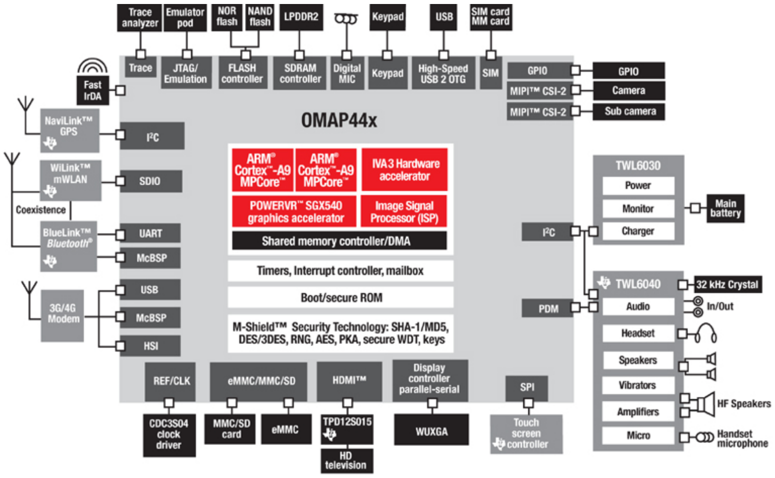


Figure 1: Block diagram of TI OMAP 4430, Courtesy of Texas Instruments

networks, we introduce methods for constructing decentralized and unmonitored communication channels through phone-to-phone message passing.

The second feature we exploit is that current-generation phones are programmable devices that include geo-location features (GPS and E911) and highly accurate real-time clocks. These features make it possible for software running on a handset to periodically discern its physical location with a resolution that is typically less than a city block. Using basic machine learning techniques, smartphones can construct location *profiles* that characterize possible future locations a given handset may frequent.

Third, cellular telephones are tightly coupled to individual people, perhaps more than any other computing or communication devices in common use. The handsets are generally left on (except during specific times where their use is prohibited, such as air travel or during lectures), and they often travel with their owners throughout their daily activities. That is, telephones go where their owners go. The existence of human travel patterns provides a valuable heuristic by which messages can be efficiently routed towards their destinations.

Finally, human travel patterns have two useful, if seemingly contradictory, properties that are critical for our purposes: they are often both *regular* and *chaotic*. They are regular in the sense that humans often go to the same places over and over. They are chaotic in the sense that human contact patterns are often perturbed by randomness, and so a given person may contact many different people in a given geographic area over time.

This paper presents the case for, and design of, unmonitored and fully decentralized networks that are outside the purview of the cellular network. This new communication paradigm, which we call *Human-to-human Mobile Ad hoc Networking* (HUMANET), enables novel and unintended uses of smartphones. As one example, such phone-to-phone networks support stealthy command-and-control messaging for mobile BotNets. HUMANETS' messaging delivery mechanisms allow a BotNet operator to direct commands toward particular phones or phones within geographic areas to attack critical cellular infrastructure [30]. As a second example, HUMANETS may be used to covertly deliver messages in repressive regimes in which cellular voice and data communication is being monitored. Finally, HUMANETS could be deployed to direct queries toward targeted smartphones to retrieve personal information such as contact addresses and emails, and even surreptitiously recorded audio.

We introduce techniques for constructing such phone-to-phone networks, and present messaging protocols that *efficiently* route messages towards their intended receivers. Flooding or broadcast schemes do not scale beyond a small number of simultaneous senders, incur significant bandwidth and storage overhead, and may be detectable by the smartphone owner simply by observing increases in battery depletion. In contrast, our protocols exploit human mobility patterns to significantly increase communication efficiency, and in many instances, incur a fixed bandwidth cost per message.

Our techniques leverage properties of smartphones – in particular, their highly synchronized clocks and ability to discern location information – to construct location profiles for each mobile device. HUMANETs' fully-distributed routing protocols route messages towards phones with location profiles that are similar to those of the intended receiver, enabling efficient message delivery with limited effects on end-to-end latency.

In trace driven simulation, we show that HUMANET-based smartphone routing in city-wide areas successfully routes 85% of the messages to their intended destinations. The effects of our techniques on latency are nominal: 75% of the messages reach their destination within a day. In comparison to epidemic flooding in which 60% of all phones in the network store a copy of a single sent message, our techniques incur fixed storage costs, requiring only a small fixed-sized subset of the network to carry message copies. Our routing algorithm is highly scalable, permitting many more concurrent messages in the system than allowed by flooding and gossiping techniques.

2 Overview

HUMANETs are *decentralized, peer-to-peer, metropolitan-area* networks compromised of mobile smartphone devices that function independently of the highly centralized and tightly controlled cellular networks for which such smartphones are designed.

Two distinguishing properties of the HUMANET environment are *mobility* and *decentralization*. The former means that the nodes – "smart" mobile phones – move in predictable but yet also chaotic ways within the network's geography. The latter property reflects the lack of centralized control in such networks. There is no infrastructure in which node devices are registered, assigned network addresses or names, or given authority to become part of the network. While the protocols exploit various services of the cellular networks to which nodes are attached, the network operates without the explicit knowledge or cooperation of the mobile telephone system. In fact, a specific design goal is that cellular network operators be unable to detect or prevent the deployment of a HUMANET.

These properties – regular and chaotic mobility and lack of infrastructure – prohibit the use of standard routing and addressing techniques that are common in most ad-hoc networks. Without a central routing infrastructure, to which nodes should a message addressed to a receiver R be forwarded? And how is R unambiguously named without a central naming authority?

The HUMANET routing protocol, described in more detail in Section 6, enables a sender to route a message towards a receiver based on knowledge of the receiver's location profile (i.e., places that the receiver is likely to be).

Our basic algorithm operates in two phases:

1. In the first phase, messages are passed from phone-to-phone towards locations specified in the receiver's profile. We utilize a greedy algorithm that forwards a message to a nearby phone if the profile belonging to that phone is more similar to the receiver's profile than the node that currently carries the message. To enable our system to simultaneously carry multiple messages with limited storage and communication overhead, we do not utilize epidemic routing techniques to route messages towards particular locations.

2. Under the assumption that people return to places that they have previously frequented, the greedy

property described above will deliver a message with high probability to regions specified in the intended receiver's profile. Since not all humans come into contact with each other, messages are then flooded within the target local area. To stop viral infection, messages are quarantined within the boundaries of receiver's profile.

Using these techniques, HUMANETs provide three messaging primitives: *unicast* (directed messaging to a particular phone), *anycast* (communication targeted at any individual within a location), and *multicast* (the transmission of messages to all parties within a location). The protocol implements heuristics that exploit the observation that humans tend to return to the places that they have previously frequented, enabling efficient heuristic-driven routing in such decentralized and highly mobile networks.

We emphasize that the HUMANET protocols are optimized for good performance within a medium-size geographic region such as a single metropolitan area. The number of simultaneous copies in the network of a given message is small, bounded by the number of copies originated by the sender. Flooding occurs only once a message has arrived within the receiver's local area. Additionally, messages are transferred only to nodes likely to travel directly to the receivers' local areas, with no "backbone" nodes or multi-hop source routing. This simplifies the design of the routing scheme, and, as we will see, performs well in the metropolitan areas in which HUMANETs are intended to be deployed.

In the sections that follow, we describe HUMANETs in more detail, and analyze our routing heuristic using trace data of human movements. Further, we demonstrate that our routing protocol reliably delivers messages with little overhead, and compare our techniques against general and probabilistic flooding algorithms as well as random walks.

3 Related Work

Vahdat *et al.* proposed *epidemic routing* as a means to connect isolated network points using mobile users called *carriers* [31]. When two carriers come into contact, they synchronize their messages. Message delivery occurs when a carrier meets its message's intended destination. Vahdat *et al.* showed that such network design can be reliable, but incurs a significant bandwidth cost [31]. To reduce bandwidth overhead, Haas *et al.* proposed *gossip routing* protocols in which the exchange of message between two nearby peers occurs probabilistically [13]. Their technique imposes only modest latency overhead, and utilizes significantly less bandwidth than epidemic-based routing protocols. The use of such epidemic and gossip techniques are most appropriate for disseminating information to a large group of network nodes. In contrast, HUMANETs attempt to minimize the number of phones that must carry a message in order for it to be delivered to its intended destination. Each send event incurs a fixed cost in a HUMANET network; by comparison, the number of nodes that carry a message using epidemic and gossip based routing increase exponentially over time.

HUMANETs draw on previous work in the area of geographic [17, 21, 18, 20] and position based routing [32, 28]. Similarly to these techniques, HUMANETs direct messages aggressively towards targeted geographic areas. In contrast to geographic routing, we do not route based on the location of a neighbor nor do we require that neighbors' positions remain static. HUMANET are targeted for highly mobile networks in which its constituents (i.e., people) exhibit frequent movement. Unlike most existing position-based routing techniques that rely on fixed neighborsets, HUMANET are highly dynamic networks with no fixed infrastructure. Routing towards a particular geographic position is therefore difficult, as a phone that is "closer" to the desired target location at one point in time is at the whim of its owner, and may move further away from the intended destination depending upon the actions taken by its human operator. Instead, HUMANET route messages towards contacts who are more likely to be in geographic areas frequented by the message's intended receiver.

Human movement based store-and-forward networks have been investigated in the context of wearable computing. Davis *et al.* evaluated protocols for mobile wearable computer networks [8]. Their work focused on routing technique where the carrying device has a fixed size buffer for message storage. They found that dropping messages addressed to nodes that are encountered least often is an efficient strategy for message delivery. In HUMANETs, instead of dropping message destined for locations that are infrequently encountered, phones pass messages off to others who are more likely to travel to the destination than themselves.

Another form of wearable computing, *pocket switched networks* [14, 7, 6], has been used in real world human mobility experiments (at the scale of a conference) with varying results. Although similar to HUMANET, these pocket switched networks are designed to overlay traditional IP routing in delay tolerant networks and do not incorporate geographic addressing. HUMANETs are designed for a much wider scale (i.e., the size of a city) and support greater mobility.

An important difference between HUMANETs and wearable computing is that the former is designed for smartphones – fairly ubiquitous computing devices – while the latter requires hardware that has not yet witnessed widespread adoption. HUMANETs also utilize smartphones' capacity to discern locality information, using such information to efficiently route messages.

Existing literature has explored the efficacy of phone-to-phone message propagation. Fleizach *et al.* investigated different malware propagation rates based on a variety of possible infection vectors [11]. However, their propagation model often requires the use of the cellular network. In contrast, a principal design goal of HUMANETs is to avoid the use of centralized cellular infrastructure, and instead pass messages between phones using point-to-point messaging paradigms. Zyba *et al.* investigated defenses against phone-to-phone infections by using phone-to-phone virus signature dissemination [34]. Their simulation methodology is based on analytic models and synthetic traces constructed using Levy Walks [26]. Unfortunately, such simulation techniques cannot be directly applied at the scale HUMANETs are designed to operate. Moreover, HUMANETs rely on the tendency for humans to return to the same places they have been before. Simulations based on random Levy Walk – although accurate for human walking patterns – do not reflect the routine travel patterns exhibited by human beings.

Traynor *et al.* investigated the use of mobile phones to attack the underlying cellular network [10]. In related work [30], it was demonstrated that coordinated mobile phones can launch devastating denial of service attacks against critical cellular infrastructure. HUMANETs' routing techniques can be a means for bot masters to coordinate mobile phones to launch such attacks at specific locations. Singh *et al.* studied the use of Bluetooth to coordinate command and control functionality in mobile botnets [27]. They demonstrated that in high density locales, successful command propagation can occur with high probability. However, their techniques partially employ the cellular network and incorporate fixed infrastructure. In contrast, HUMANETs are entirely decentralized and do not leverage cellular services. The inherent difficulty of tracking messages in HUMANET make it an ideal choice for BotNet command and control.

As is true for all computing devices, smartphone software and hardware are susceptible to design and implementation errors that lead to security vulnerabilities [24, 19]. Smartphones, by design, house personal information and are attractive targets for potential attacks. The general phone-to-phone messaging primitives presented in this paper provide a possible mechanism that attackers may utilize to covertly communicate with compromised phones.

4 Human Mobility Datasets

Although mobility patterns can be generated heuristically using Levy Walks [26] and other analytic techniques, our routing protocols rely on sociological patterns: specifically, the tendency of humans to frequent particular locations (e.g., their homes and offices). Such human behavior cannot be easily represented using synthetic modeling. Although synthetic movement patterns may express human-like characteristics (for example, the contact rates between individuals), we are not familiar with any algorithm that produces the diurnal movement patterns and routines exhibited by human beings.

To evaluate the feasibility and effectiveness of HUMANET routing protocols, we therefore depend on human movement datasets – traces of actual human movement collected over periods of time. Each dataset is of varying length and presents movement patterns at different granularities.

Unfortunately, published datasets of human movement are not perfect. Many of the available datasets cover too short a timespan (e.g., less than a day) and consequently prevent us from evaluating our geographic routing protocols (since movement patterns cannot be derived from such a short timescale). We therefore confine our evaluation to the following larger datasets:

- The **Cabspotting Dataset** [25] contains GPS coordinates and timestamps of 536 taxicabs in the San Francisco area. The dataset spans 20 days: from May 20, 2008 until June 7, 2008. The Cabspotting Dataset provides the finest movement granularities of all of the analyzed datasets and is used in our trace driven simulation (see Section 7). It should be noted that although the movements of taxis are not representative of the general population (taxis are arguably more mobile than the average person), simulations using this dataset can be interpreted as a representing a HUMANET network composed of the taxi drivers' smartphones.

- The **VAST Dataset** [1] contains anonymized pseudoidentities of mobile phone users, the times of their phone calls, and the cellular tower from which the call originated. The data is composed of 400 participants and consists of measurements from a one week period. The cellular tower identifiers have been obfuscated, preventing accurate reconstruction of physical locations.

- The **Reality Mining Dataset** [9] contains timing information about mobile phone users, the calls they make, and the tower from which calls originate. The dataset collection period is roughly one year. The study was composed of 100 participants and was originally used to model social interaction networks. Again, tower identifiers are obfuscated and cannot be correlated to real physical, hindering our ability to perform accurate distance measurements.

In the following section, we describe how location information is used to develop compact profiles that accurately represent the locations most frequented by smartphone users.

5 Location Profiles

HUMANETs are highly decentralized networks with no fixed infrastructure and are composed of participants who frequently change location. Our routing protocols achieve efficient routing in such highly dynamic and unstructured networks by exploiting human mobility patterns. Since humans tend to return to places that they have frequented in the past, messages may be routed towards such strategic locations (assuming they are known). However, even if the sender of a message knows the precise location of the intended receiver, messages must be passed from phone to phone, requiring transfers to just those phones that are more likely to travel towards the receiver's location. To permit such intelligent routing decisions, each phone maintains a *location profile* that compactly defines the geographic areas in which its owner tends to locate. By examining the location profiles of its nearby peers, a smartphone can transfer a message to another phone that is more likely to intercept the message's target. In this section, we describe how smartphones may efficiently (and, in the case of BotNets, covertly) construct accurate location profiles of their owners.

Generating a location profile requires classifying movements into three categories: regular, irregular, and exceptional. Regular movements are those that occur daily or nearly daily (e.g. commuting habits). Irregular movements, on the other hand, occur on the order of days but do not happen daily (e.g., weekend outings). Finally, exceptional movements are those that occur infrequently (e.g., vacations). Location profiles are constructed to recognize all such movement patterns.

Participants in a HUMANET utilize GPS or E911 capabilities to periodically discover and record their locations. Since routing decisions in HUMANETS are based on profiles of movement patterns, these location data must be accurately grouped into geographic regions.

To construct profiles, a smartphone applies the k-means clustering algorithm to its recorded locations. Described in more detail in Section 5.1, the k-means algorithm assigns one of k possible "colorings" to each location, grouping closely located measurements into the same cluster. After applying k-means, phones convert these clusters into geographic regions. Each region constitutes a phone's *home* and is represented using a polygon (i.e., an ordered set of geographic coordinates). A phone's profile is defined as a collection or subset of its known homes.

Due to software and hardware limitations on smartphones, measured locations may sometimes be inexact or even inaccurate, causing outliers to exist in phones' location transcripts. The clusters produced by the k-means algorithm cannot therefore be directly mapped into homes, as such outliers will skew their shape. Thus, it may not be feasible to represent all points within a cluster as a contiguous geographic region. Techniques for converting clusters to homes are described in Section 5.2. Once homes (polygons) are computed, phones must decide what subset of homes comprises their profiles. Mechanisms for selecting homes are introduced in Section 5.3.

There are obvious computational demands placed on the smartphone during the profile construction process. However, unlike polling and message transfer (see Section 6) that occur while the phone is in motion and can place significant demand on the battery (see Section 8), profiles can be computed daily or semi-daily. More importantly, profiles can be constructed whenever the phone is charging so that the computation does not drain the battery.

5.1 k-means

The k-means algorithm [12, 22] divides a set of n observations into k clusters such that each observation belongs to the cluster whose mean center (centroid) is closest. The general algorithm proceeds as follows. First, k random points are selected as the initial mean centers (centroids) for the clusters. Next, each observation is placed in the cluster to which it is closest, with distances measured to the cluster's centroid. The mean centers are then recomputed based on the new assignments. If the newly computed centroids move more than a threshold amount when compared to their previous location, then the means have not yet stabilized. In such cases, the closest mean for each data point is recomputed, each observation is placed in the cluster whose centroid they are now closest to, and the means are again recomputed. The procedure repeats until either the clusters have moved less than the threshold amount, or a specified number of iterations of the algorithm have occurred.

The initial assignment of the k centroids may significantly impact the output of the algorithm. However, since we are clustering human movement, we can bootstrap the selection of the centroids by using the centroids computed from a previous run of the algorithm. Here, we utilize the observation that people tend to revisit places they have been before. With the exception of the initial run (in which no previous location data can be used to kickstart the algorithm), using this heuristic significantly reduces the number of iterations required to locate stable centroids.

Unlike existing techniques in which grids or connected graphs are used to develop location profiles for message dissemination [20, 17, 21, 18], our k-means approach is highly adaptive and is not bound by

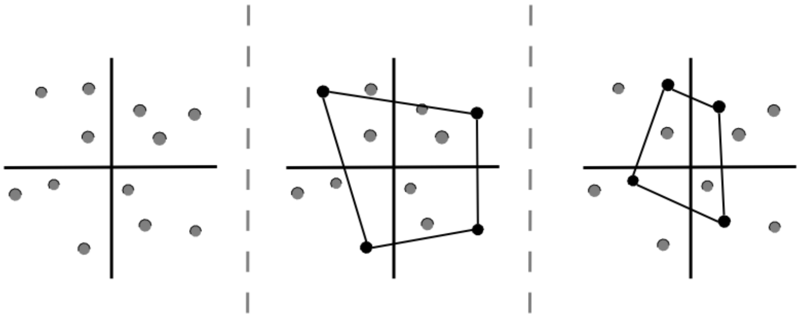


Figure 2: Converting k-means clusters into geographic homes. *Left:* Points belonging to the same cluster. The origin represents the centroid of the cluster. *Center:* Defining a polygon using the farthest point in each quadrant. *Right:* Defining a polygon using the median point in each quadrant.

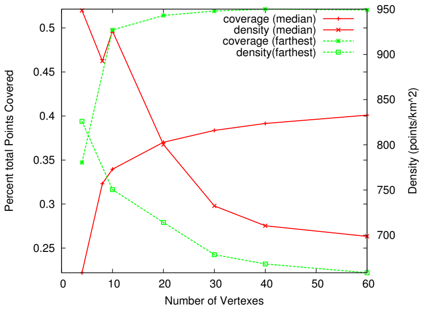


Figure 3: The coverage and density curves for median and farthest vertex selection.

any indexing scheme or choice of graph vertex locations. Grids reduce the ability to dynamically define regions and are too rigid for the scale of human movements HUMANET is designed for. GPS coordinates and polygonal regions have been previously proposed as an addressing scheme [15], but only in the context of routing in static networks.

5.2 Constructing Homes

The k-means algorithm partitions geographic points into clusters. We now show how to convert a given cluster into a polygon.

For each cluster, the phone divides the cluster radially about its centroid into equal-sized angular divisions. A single point from each angular division will be selected as a vertex in the polygon. That is, the number of angular divisions equals the number of vertices in the polygon. Figure 2*(left)* depicts an example cluster in which the two-dimensional space is divided into four angular units.

We investigate two strategies for selecting the vertex in each quadrant. To produce wide area polygons, the phone may select the farthest point from the centroid (Figure 2*(center)*). Alternatively, the median distance points can be selected (Figure 2*(right)*) providing a good compromise, weighing the distance between the farthest and the closest points in each angular division.

In Figure 3, we compare the above two home construction techniques based on their achieved *coverage* (the number of points within the polygon) and *density* (the coverage divided by the square area) for differing number of vertices. Our evaluation is based on the Cabspotting Dataset described in Section 4. As can be discerned from the figure, using the farthest points provides the best coverage, but results in the least density. However, farthest based selection curves intersect at approximately 8 vertices with a higher density and coverage than the median techniques which intersects at 21 vertices.

The number of vertices in a home determines the number of bits required to encode it. Homes with a large number of vertices more precisely define a geographic area than polygons of lesser degree, but do so at the expense of storage and communication cost. Profiles (collections of homes) should fit within a single packet MTU (usually 1500 bytes for WiFi). Using two 4-byte doubles to represent each vertex, using 8-sided polygons permits profiles that contain up to 23 home regions.

To demonstrate the feasibility of constructing homes based on actual measurements from a smartphone device, we implemented a simple location tracking program on the G1-Android phone. Figure 4 (*left* and *center*) provides examples of the output of our home construction algorithms using real-world data. Given the small size of our collected data, we do not argue that the homes depicted in the leftmost and center maps are representative of the home sizes, shapes, or locations that would be expected for the general population. Rather, our results demonstrate that home construction using phone-provided GPS traces is feasible. The leftmost map depicts the travel patterns of an author who divides his time between his residence and the University. The center map accurately reflects the locations frequented by another author who commutes between his home in northern New Jersey and the University of Pennsylvania in Philadelphia. For comparison, the rightmost map shows the home regions of a taxicab using data from the Cabspotting dataset. The home construction algorithm accurately captures the travel routines of the taxi – a significant fraction of its time is spent at the airports in the Bay Area.

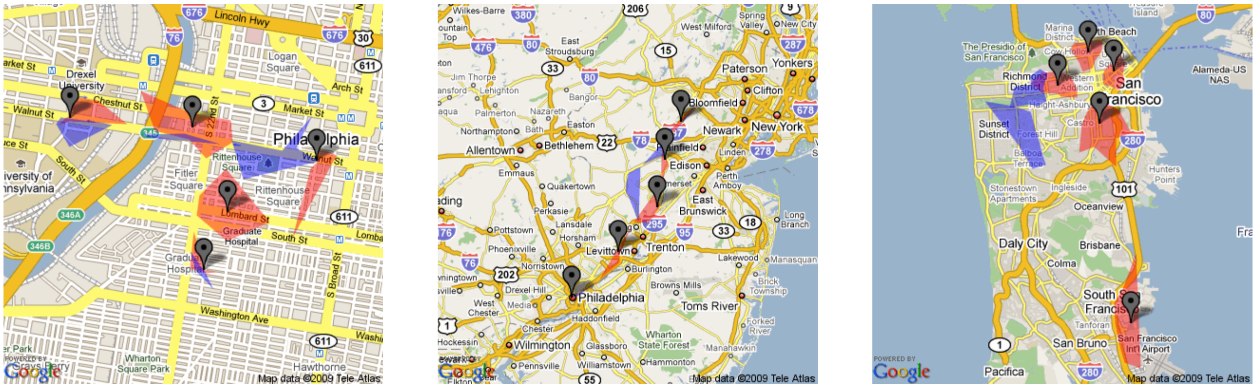
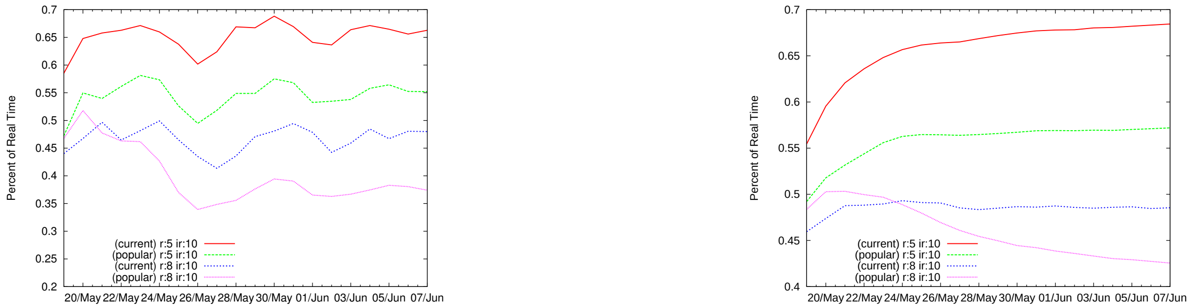


Figure 4: Home computation for real location movements collected using the G1 Android smartphone (*left* and *center*) and the Cabspotting dataset (*right*). The markers denote centroids of polygons. Red polygons indicate regular homes. Blue polygons represent irregular homes.

5.3 Home Selection

We explore two strategies for selecting which homes should be added to a profile. Our goal is to ensure that the homes that constitute the profile accurately reflect the locations commonly frequented by the phone's human carrier.

The first technique, *popularity selection*, makes profile membership decisions based on how frequently a geographic region is mapped into a home. If the centroid of a newly computed polygonal P falls within a previously computed home H, then P is considered to be *reinforced* by H, and H is replaced by P in the phone's profile. The phone records a *popularity timeline* for each home in its profile. The popularity



(a) Fraction of time (per day) spent within a profile's homes

(b) Fraction of measured GPS coordinates that reside within a profile's homes

Figure 5: The percent of points and real time per day spent in the previous days computed homes. In both graphs, two different home selection routines are used, most popular selection and most current selection. r is the number regular and ir is number of irregular homes and. May 26th was memorial day.

timeline records the times at which a home has been reinforced over the previous week, and the number of records in the timeline reflects the number of times over the previous week that a home has been reinforced. If a home's timeline becomes empty (the home has not been reinforced over the past week), then the home expires and is ejected from the phone's profile. Homes in the profile are sorted based on the number of entries in their timeline. The r homes with the highest number of entries are called the phone's regular homes, while the remaining ir homes are the irregular homes.

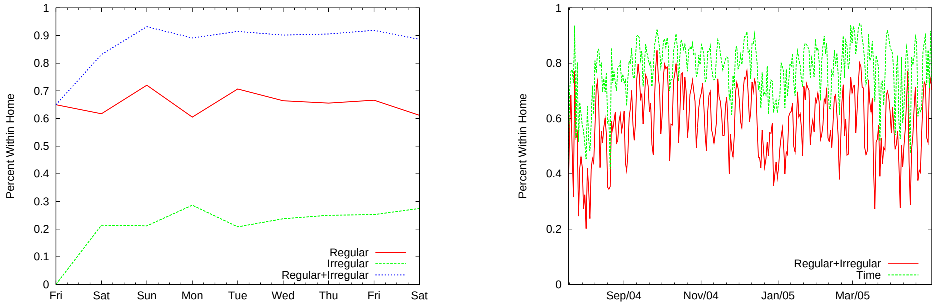
Alternatively, in the *current selection* strategy, all homes computed from the previous day's tracking data are considered regular regardless if they are reinforced. Yesterday's homes that are not reinforced move to the irregular list of homes, and are ejected after 7 days. The set of irregular homes that are included in the profile is then chosen based on *sparsest principles*: the subset of irregular homes whose sum of pairwise distances, measured from centroid-to-centroid, is greatest. We employed a greedy algorithm for this selection process that runs in linear time and produces good approximations of the optimal result.

Figure 5*(a)* shows the fraction of the day spent within a profile's home regions for the two home selection policies. Although both home selection routines produce similar curve shapes, the current selection procedure outperforms the popularity selection in all cases. Popular selection weighs events too far in the past, and thus has lower totals for points found within homes. This is clearly demonstrated in the effect that Memorial Day (May 26th) had in Figure 5*(a)* on 8 regular and 10 irregular homes for popular selection. All curves experienced a significant drop in their predictions entering the weekend due to exceptional movement caused by the holiday. However, popular based selection with 8 regular and 10 irregular homes "remembered" the exception movement for longer, and as a result, continued to decrease (Figure 5(b)) while the others adjusted, recovered and showed increases in their predictions (Figure 5(a)).

5.4 The *Return-to-Home* Principle

HUMANET routing protocols exploit behavior patterns of humans. An underlying assumption of our routing techniques is that homes are good indicators of future location. We call our hypothesis the *Return-to-Home Principle*.

To test whether our hypothesis holds true in practice, we utilize the Cabspotting dataset to measure the percentage of a day's collection of GPS coordinates that reside within the previous day's home regions. The



(a) VAST dataset

(b) Reality Mining

Figure 6: Percent of calls occurring from previous days profile, measure from the VAST dataset

results are presented in Figure 5. The figure shows that 65% of GPS points fell within the previous day's profile and the phone resided in its profile's homes for 65% of the day. In our worst performing case, 39% of recorded locations were inside home regions, and 45% of the day was spent within these homes.

Note that the Return-to-Home principle is upheld more often when profiles contain smaller numbers of daily computed homes. This is to be expected: with fewer computed homes per day, each home will cover a wider geographic area, and thus a phone is more likely to return to it. However, the same intuition would suggest that fewer daily computed homes would perform worse in terms of routing performance again due to the wider geographic area, but as we demonstrate in Section 7, using profiles with a fewer number of regular homes resulted in higher message delivery rates when compared to using profiles with more homes computed daily.

To further support our Return-to-Home principle, we validated our results with the VAST [1] and Reality Mining [9] datasets. Since neither dataset provided geographic coordinates, we made the simplifying assumption that celltower IDs represented home regions. We utilize the popular home selection technique since the sparsest principle cannot be evaluated with distance information. Our results (Figure 6) provide additional evidence for the Return-to-Home principle. 77% and 87% of time was spent within the previous day's profile regions using the Reality Mining and VAST datasets, respectively.

6 Routing Protocols

HUMANETs are decentralized phone-to-phone networks with no fixed routing infrastructure. Previous work has examined the use of epidemic [31] or gossip [13] routing techniques to deliver messages in similarly (un)structured mobile ad hoc networks. However, such approaches incur significant bandwidth overhead, as a single message must be duplicated and carried by a sizable fraction of the network for it to be delivered. HUMANETs aim to support multiple simultaneous senders, and provide messaging functionality without overburdening the storage capacities and batteries of its constituent smartphones. In this section, we describe protocols for efficiently routing messages towards their intended targets.

6.1 Messaging Primitives

HUMANETs support unicast, multicast, and anycast message delivery. The `sendToLocation` primitive routes messages towards a targeted location, and either delivers the message to a particular party within the location (anycast) or all parties within the specific area (multicast). Such functionalities are useful when the

Time (secs)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Stage	Profile announcements					Processing					Transfer				

Figure 7: The synchronous timing scheme used to broadcast profiles, process announcements, and deliver messages to nearby peers. Listed times reflect offsets (in seconds) from the start of an interval.

location of the receiver(s) matter more than their individual identity(-ies).

HUMANETs also enable senders to address messages to individual receivers. The `sendToReceiver` primitive provides unicast functionality. To route messages, the sender must have knowledge of the locations that are frequented by the receiver. The more information the sender possesses about the receiver's mobility habits, the more likely messages are to be successfully delivered.

For both messaging schemes, HUMANETs provide best-effort unreliable message delivery. Reliability protocols may be deployed on top of HUMANET messaging primitives, although the development of such protocols is outside the scope of this paper.

6.2 Synchronous Phone-to-Phone Message Delivery

Since smartphones are battery-constrained devices, asynchronous message delivery over WiFi is infeasible. To preserve battery charge, HUMANET routing protocols operate in synchronized rounds in which phones announce their presence, exchange profile information, and relay messages between devices.

Figure 7 shows the stages of a single round of our protocol. Rounds begin at regular intervals (every five minutes) and last for a set duration (fifteen seconds). To communicate, phones participate in an ad hoc network, assigning themselves IP addresses chosen uniformly at random from the 10.0.0.0/8 address space[1]. For added stealth, the frequency and SSID of the ad hoc network can vary over time and be derived cryptographically using a shared secret and the current time (for example, by basing both parameters on an HMAC over the number of seconds past the epoch).

During the *profile announcement* stage, each phone broadcasts an *(id,profile)* tuple, where *id* is a 20 byte random nonce that may vary between rounds and *profile* is a concise representation of a phone's location profile. The *(nonce,profile)* message should fit into an MTU-sized IP packet and is transmitted via UDP broadcast.

Following the profile announcement stage, message carrying phones consider the profiles advertised by their peers and determine whether their messages should be relayed to their neighbors. This *processing* stage represents the "intelligence" of our routing design, and is described in detail in Sections 6.3 and 6.4.

Finally, messages are exchanged during the *transfer* stage. To deliver a message, a sender broadcasts a message of the form *(*id_X*,msg)* where id_X is the nonce broadcast by the intended receiver during the profile announcement stage.

The broadcasting of location information during the profile announcement could be viewed as a privacy violation. If the HUMANET is deployed on a voluntary basis (that is, individuals opt-in to participate), a reasonable preference setting is to not reveal certain travel destinations during the profile announcement stage. An application running on the mobile device could allow the operator to disable location polling or censor certain aspects of his/her profile. However, we note that many individuals are not averse to voluntarily sharing their location information. Several location sharing services such as Google Latitude [2] enable its users to share their locations with their friends. If, on the other hand, the HUMANET was deployed as a

[1] Such an address space holds approximately 2^{24} possible addresses (one address is reserved for broadcast); the probability that no collisions occur if x phones participate in the local ad hoc network is $\frac{2^{24}!}{(2^{24}-x)!2^{24^{x}}}$.

msg_type	D	msg_id	mode	until	message_data

msg_type	$\mathcal{B}_{ID}$	R_{B_i}	msg_id	until	message_data

Figure 8: HumaNet message formats for `sendToLocation` *(top)* and `sendToReceiver` *(bottom)* communication primitives.

command and control structure for mobile BotNets, the owners of the smartphone would be unaware that their location profiles were being broadcast.

Time Synchronization To successfully receive the messages of their peers, phones must be highly synchronized (i.e., having clock skew of no more than a few seconds). Although a sufficient granularity of synchronization can be achieved by polling NTP servers, such communication may be monitored by the upstream Internet provider (e.g., the mobile service provider in the case that the request is made over the data service) and may be used to reveal the identities of HumaNet participants. Fortunately, GSM and CDMA networks provide mechanisms for tight time synchronization across mobile devices. Modern cellphones use the station's *Network Identity and Time Zone* (NITZ) service to synchronize their time with their base station when powered on [29]. (Smartphones may also request synchronization on demand, for example, on a daily basis if the device is left running for extended periods of time.) Due to 3G requirements, base stations across a cellular carrier are highly synchronized. For example, the UMTS data service utilized by GSM networks requires that the clock drift between base stations to be less than 0.05 parts per million [3] or 4.3ms over 24 hours [23]. CDMA has more stringent requirements, mandating that drift not exceed 7.5μs per day [23]. To minimize skew, base stations typically use high-precision GPS receivers to determine time-of-day. For example, CDMA specifies a maximum error of 3μs [5], far within the tolerances required by our fifteen second window.

6.3 Routing to a Location

To direct messages towards a particular location, our routing protocol exploits human behavior patterns. Specifically, we rely on the Return-to-Home Principle: a person (and the phone he/she carries) is likely to visit in the future the places that have been visited in the past. Hence, the observed profiles of peers provides heuristic direction information that may be used to forward packets. A neighbor whose profile includes regions that overlap that of the targeted destination is (under our underlying hypothesis) more likely to visit that destination in the future than a randomly selected neighbor.

The `sendToLocation` communication primitive operates as follows:

Let $P_X =< H_{X_1}, H_{X_2}, ..., H_{X_k} >$ represent $\mathcal{X}$'s profile, where each H_{X_i} denotes a home region (i.e., an area bounded by a polygon) frequented by $\mathcal{X}$. Each smartphone $\mathcal{X}$ also maintains a 20-byte Bloom Filter [4] BF_X initialized to 0 (the use of the Bloom Filter is explained below).

Let $\mathcal{A}$ be a mobile device that currently carries a message m addressed to a destination D (a polygon represented by an ordered set of geographic coordinates). We refer to $\mathcal{A}$ as the *carrier* of message m. The format of m is shown in Figure 8 *(top)*, and consists of the `sendToLocation` `msg_type`, a destination D, a 20-byte message identifier `msg_id`, a specifier `mode` that indicates whether the message should be delivered to a single phone (anycast) or all phones within the specified destination (multicast), a timestamp `until` that indicates the future date and time (in UTC) at which the message should be dropped if it is not delivered, and a payload `message_data`. For simplicity, we assume that the entire message may be transmitted in a single UDP packet (longer messages may be supported by adding sequence identifiers to the message format).

During the profile announcement stage, $\mathcal{A}$ observes the broadcast profiles of her peers. Let $\alpha(P_k, W)$ be a function that computes the number of homes (polygons) in profile P_k that intersect with polygon W, $\mathcal{C}$ be a peer whose profile P_C was received by $\mathcal{A}$ during the profile announcement phase, and $\wedge$ denote the logical *and* operator. $\mathcal{A}$ relays the message m to $\mathcal{C}$ during the transfer phase if and only if the following two properties are met:

Greedy Property: $\alpha(P_C, D) > \alpha(P_A, D)$

Acyclic Property: $BF_C \wedge \texttt{msg\_id} \neq \texttt{msg\_id}$

Additionally, to prevent the message from spreading epidemically (and consuming storage and bandwidth), $\mathcal{A}$ *relays the message at most once*.

Intuitively, the Greedy Property requires that $\mathcal{C}$ has more *similarity* to the targeted area D than $\mathcal{A}$, where similarity is defined as the number of overlaps between a phone's profile and the desired destination. In the case that $\alpha(P_{C_i}, D) \leq \alpha(P_A, D)$ for all nearby phones $\mathcal{C}_i$, then $\mathcal{A}$ does not transmit the message and retains her copy.

When a phone $\mathcal{C}$ receives a message m during the transfer phase, it adds the corresponding `msg_id` to its Bloom Filter. The Acyclic Property prevents the formation of cycles in the routing path.

The above greedy routing algorithm moves the message closer to phones that possess home regions that intersect the targeted destination. However, the algorithm does not necessarily guarantee progress. To illustrate, consider the case in which a carrier of a message has a profile that multiply intersects the destination. Due to the Greedy Property, the phone will not transfer the message to a peer that has fewer intersections. If the carrier does not move to the destination, the message is effectively pigeonholed in a suboptimal location.[2]

To prevent such effects, phones record the time t_{m_i} at which a message m_i is received. A *local timeout* occurs when phone $\mathcal{C}$ stores a message m_i for longer than some threshold value. If a local timeout occurs, $\mathcal{C}$ transfers m_i to the next phone that it encounters, provided that the Acyclic Property (but not necessarily the Greedy Property) holds. Finally, to prevent messages that do not reach their destination from continuously traversing the network, a *global timeout* occurs if the current time exceeds `until`. In such cases, the carrier discards the message, and the message is permanently lost. [3]

When the carrier of a message m enters the location defined by D and the `mode` flag specifies anycast, then (by definition) the message has been successfully delivered. If $\mathcal{C}$ is located within D and the message is to be multicast, then $\mathcal{C}$ changes the `mode` specifier to *Flood*. In such a case, both the Greedy and Acyclic Properties are ignored, and the carrier relays the messages to all phones it contacts within area D. In turn, receiving phones duplicate the message to its nearby peers, but delete the message as soon as they leave D. That is, the message is flooded to all phones, but the spread of the message is confined to D. The message is continuously passed within D until the global timeout expires, at which time all phones discard the message.

6.4 Routing to an Individual

The `sendToReceiver` primitive is a special case of multicast mode `sendToLocation`. Here, we consider a phone $\mathcal{A}$ that wants to deliver a message m to a phone $\mathcal{B}$. By assumption, $\mathcal{A}$ is able to identify

[2] Although the carrying phone's profile will eventually reflect new homes that do not intersect the targeted destination, profile updates occur too infrequently (once every 24 hours) to permit efficient message transfer in such cases.

[3] Loss may also occur if a phone that carries the message is lost (becomes immobile) or destroyed. At the cost of a multiplicative increase in messaging cost, the protocol may be extended to support a fixed number of copies (per region) to offset the probability of loss. Amending the protocol to handle loss is straightforward, and is omitted for brevity.

possible locations of $\mathcal{B}$ (for example, the locations he often frequents). We label such guesses as to $\mathcal{B}$'s location as *regions* $R_B =< R_{B_1}, ..., R_{B_k} >$.

For each region $R_{B_i} \in R_B$, $\mathcal{A}$ constructs a message m_i of the form depicted in Figure 8 *(bottom)*, where $\mathcal{B}_{ID}$ is a unique ID associated with $\mathcal{B}$ (for example, his telephone number or MAC address). The `msg_id` is consistent among all k copies of the message (where $k = |R_B|$).

Messages are passed between phones using the Greedy and Acyclic Properties defined above. Each message copy is addressed to a particular region, and is passed greedily towards phones whose home locations intersect with that region. Unless a carrier enters a region R_{B_i}, it relays at most one copy of the message. Until messages reach their specified regions, there will be at most k copies of the message in the network at any given time.

As with multicast delivery, the flooding procedure described above commences when a carrier enters the region specified in the message header. Such flooding is quarantined to the specified region (that is, phones that contain a flooded message will discard it once it leaves the message's designated destination area).

7 Trace Driven Simulation

To evaluate the efficacy of our routing techniques, we constructed a HUMANET-routing simulator. Since it provides the finest location granularity, we performed simulation using the Cabspotting dataset. Due to the limited size of the dataset (536 cabs), we did not implement the flooding stage of our routing protocol. We therefore consider a message sent via the `sendToReceiver` primitive to be successfully delivered if it is directly received by its intended target.

7.1 Alternative Routing Algorithms

We compare the performance of HUMANET against three alternative techniques: *epidemic flooding*, *probabilistic epidemic flooding*, and *probabilistic random walk*.

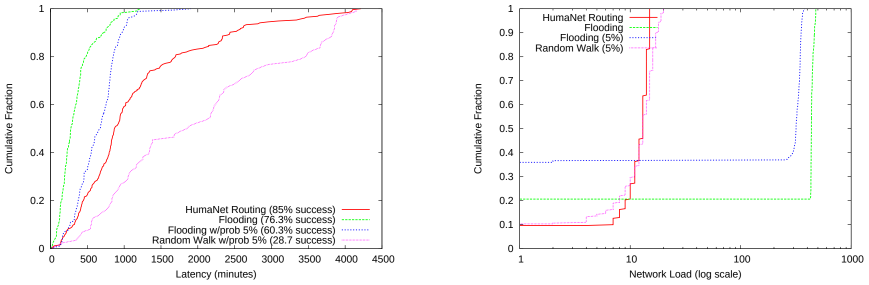
In contrast to our HUMANETs' profile based approach, the epidemic flooding technique transfers messages to all phones that come in contact with a carrier. Local and global timeouts are still observed; when a local timeout is triggered, a phone will not accept the message again. Probabilistic epidemic flooding follows the same timeout rules, but message transfers occur between phones with a fixed probability $p_E \in (0, 1)$.

Probabilistic random walk is useful to isolate the effect of routing based on profiles. Random walks function like the HUMANET routing protocol, but they do not utilize location information. A carrier of a message will transfer that message at most once. When it encounters another phone, it transfers the message with some probability $p_W \in (0, 1)$. To provide a fair point of comparison, a sender who utilizes the random walk technique sends the same number of message copies as would have occurred if the HUMANET-routing algorithm were used. That is, since our approach transmits k copies of a message – each addressed to a home in the receiver's profile – random walks are also initialized with k copies.

In all instances, all phones discard their carried messages when the global timeout occurs.

7.2 Metrics

We measure two primary statistics during simulation: *Latency* is defined as the time required for a message to successfully reach its destination, relative to the time at which the message was first sent. *Network load* is a measure of the total number of message copies that reside in the network during the message's lifespan (i.e., the period between being sent and the global timeout). Epidemic flooding and probabilistic epidemic flooding both lead to exponential growth in network load (since messages are duplicated rather



(a) Cumulative Distribution for Latency (b) Cumulative Distribution for Network Load

Figure 9: Cumulative distribution of latency and network load (measured as total number of messages in the network) for HUMANET compared with flooding, probalistic flooding (5%), and random walk (5%). Success rates in the lefthand figure represent the percentage of messages that were successfully received by the receiver before the global timeout.

than transferred). In the case of random walk and HUMANET-routing, both techniques have fixed costs per message (the number of homes in the receiver's profile).[4]

7.3 Simulation Results

We ran simulations for a number of different profile configurations, including those found to be most efficient in Section 5. Experiments were conducted using five and eight regular homes and ten irregular homes for both current and popular profile selection criteria. The probabilistic epidemic flooding and probablistic random walk approaches both transferred messages to nearby phones with probability 0.05. Each simulation consisted of 300 independent runs in which the sender and receiver were chosen uniformly at random. The local and global timeouts are 10 hours and three days, respectively.

Figure 9(a) presents the cumulative distribution of latencies among all 300 simulation runs. The y-axis represents the fraction of paths that had latencies at least that of the corresponding x-axis value. The graph plots the latency of *successful* message deliveries (i.e., cases in which the message is delivered); the latency of undelivered messages is infinity.

As expected, epidemic flooding and probabilistic epidemic flooding deliver messages with less latency than the random walk and HUMANET techniques. In 95% of successful runs, epidemic flooding delivered the message within 14 hours, and probabilistic epidemic routing delivered within 17 hours. By comparison, 95% of successful HUMANET simulations delivered the messages within 54 hours, but 75% of successful runs deliver the message within 24 hours even without using the protocol's flooding phase. Probabilistic random walk performs significantly worse; here, less than 50% of the messages are delivered within one day.

Interestingly, HUMANET outperforms all other techniques by a large factor in terms of successful delivery. 85% of the simulations resulted in successful delivery when using HUMANET compared to 76.3% for epidemic flooding. Probabilistic random walk only delivered 28% of the messages successfully, highlighting the benefits of routing based on location profiles.

[4] The flooding technique used by HUMANETS' unicast and multicast delivery primitives incurs exponential network load, but the spread of the message is quarantined to a geographic area. The maximum number of phones that may carry the message is determined by the area's population.

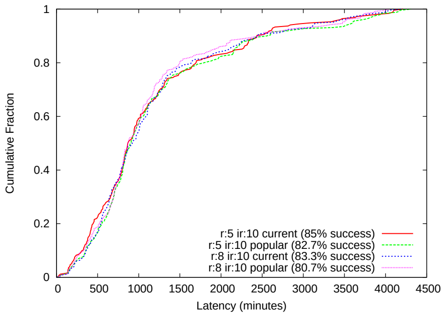


Figure 10: Cumulative distribution of latency for different profile selection routines: 5 regular and 10 irregular homes for current and popular based selection, and 8 regular and 8 irregular homes for current and popular based selection.

The impact of HUMANET is even stronger when comparing network load. Figure 9(b) (log scale) presents the cumulative distribution of the maximum network loads that occur for a single sent message. Since the maximum number of homes in the receiver's profile is 15 (five regular and ten irregular homes), neither HUMANET nor probabilistic random walk ever imposed a network load of more than 15 messages. By comparison, epidemic routing and probabilistic epidemic routing both incur significant network load, and in all simulations a large majority (more than 60% for probabilistic and 80% for epidemic routing) of the phones carried the message at some point. Such loads would render the network unusable for even a limited number of multiple sender and receiver pairs.

We were also interested in verifying the results from Section 5 in which the Return-to-Home Principle is used to predict that utilizing five regular and ten irregular homes should outperform other profile configurations. Although the tested techniques produce similar distributions of achieved latency (for successful deliveries), the success rates of the five-regular-ten-irregular approach are greater than that of using eight regular homes. Additionally, as predicted in Section 5, using the current selection strategy outperforms popular-based selection. It should be emphasized however, that in all cases, HUMANETs deliver messages more reliably than epidemic flooding techniques, and do so with little latency overhead.

8 Detection, Observability and Disruption

The highly decentralized and dynamic nature of HUMANETs provides unique detection and observability protection. Clearly, such phone-to-phone networks operate outside the view of the cellular network. HUMANETs have the additional advantage that they offer limited protection from active and passive adversaries who either infiltrate the network or stalk (literally!) network participants. Since achieving a global view of the network requires monitoring each mobile participant, such wide-scale monitoring is likely infeasible.

A passive adversary – stationary or mobile – may monitor nearby HUMANET profile broadcasts, but there is no guarantee that a transfer would follow, making it difficult to discern whether or not a HUMANET participant is currently carrying a message. The passive adversary could stalk a smartphone that is sending beacons until a message is transferred and then view the message headers and the (possibly encrypted) payload. However, such an attack requires significant time and effort, and may be overtly conspicuous to the targeted smartphone owner.

An active adversary, after viewing a message header's destination address, could disrupt that particular

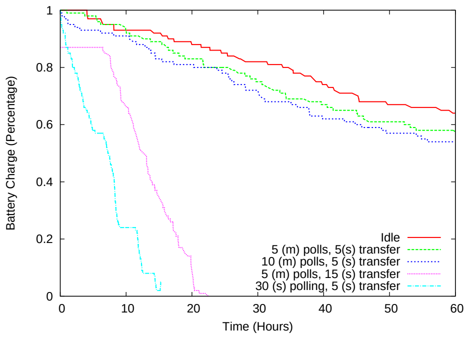


Figure 11: Battery charge vs. Time (hours) for different polling and transfer rates.

message by advertising a profile that matches the destination, causing the message to be transferred to the adversary. Still, multiple message copies may exist in the network since the sender directs a message copy to each of the receiver's homes. To completely eliminate that message from the network, an adversary must employ multiple mobile confederates to track and intercept every copy of the message. It should be noted that in centrally controlled HumaNet (i.e., those established by BotNet operators), encryption may be used to obfuscate message headers. Without knowledge of the decryption key, an attacker could not forge convincing profile advertisements and would be unable to cause the carrier to transfer its message.

Detection protection by the smartphone owner is also an important property to maintain, especially if a HumaNet is deployed for use in a BotNet. The code that runs a HumaNet can be protected in the traditional ways, via obfuscation and polymorphism. However, the unique relationship an owner has with his/her smartphone also implies an intimate knowledge of its usage statistics, namely how long the smartphone runs on a single charge under normal usage. If the beacon and transfer procedure is too aggressive, it will adversely effect the battery consumption, and the smartphone owner may detect and remove the HumaNet software.

HumaNet provide two "knobs" that can be tuned to balance battery usage with effective messaging: the interval between synchronous communication sessions and the duration of each session. Figure 11 graphs battery charge (as a percentage) over time with different polling and transfer periods on an HTC Android G1 running Cupcake 1.5 and the modified CyanogenMod v4.0.2 kernel [33]. We measure based on transfer periods (rather than beacon and processing periods) because data transmission requires the most radio resources, and thus provides a conservative bound of battery consumption. Additionally, GPS location information was collected during the same transfer time block to simulate the cost of geographic tracking. The experiment assumes a worst-case scenario in which a transfer must occur after every profile announcement stage. Additionally, our experiment assumes that data is transferred for the duration of the time window.

A smartphone loaded with third party software that is being actively used for voice and data communication as well as to run applications may see different and more dramatic trends, thus masking HumaNet's presence further. Nonetheless, it is clear that with reasonable polling and transfer periods, either ten or five minute messaging intervals with five second transfer windows, the battery consumption of a HumaNet participant is on par with smartphones that do not participate.

9 Conclusion

This paper demonstrates new networking opportunities enabled by the convergence of computing and communications technologies as embodied by the smartphone. The unique relationship people have with their cellular devices is unprecedented. No other computationally expressive, easily programmable, and highly connected device is so tightly bound to its owners movements. Moreover, smartphones remain powered on for indefinite periods of time, can accurately determine their locations, and maintain highly synchronized clocks. Yet, no other computer device operates in a more centralized and controlled network environment than cellular devices. HUMANETs provide a mechanism for constructing decentralized phone-to-phone networks that avoid the tightly controlled (and monitored) cellular network. In developing HUMANETs, we contribute novel routing techniques for opportunistic networks specifically designed for human-to-human contact networks.

We have presented methods for constructing highly accurate *location profiles* based on k-means clustering and polygonal groupings that serve as the basis of our routing protocol. Our techniques leverage repetitive properties in human movement patterns to predict future travel. In comparison to flooding and gossip-based routing (in which messaging costs grow exponentially over time), our routing protocols that human movement patterns incur fixed delivery costs to geographic regions. Using trace-driven simulations, we show that our techniques are bandwidth efficient and deliver messages with low latency: 75% of successful messages deliveries occur within 24 hours, with a surprisingly high success rate of 85%.

Unfortunately, we are aware of no large-scale (e.g., country-wide) human mobility dataset that accurately captures human movement patterns (i.e., routines) over the long term. However, our simulation results suggest that HUMANET will scale. Verifying such a claim can only be done via the deployment of a real HUMANET network of smartphones, and is an intended area of future work.

The relationship people have with their computing devices, and in particular their mobile phones, is a particularly interesting area of future research. The proliferation of smartphone devices permits novel human-to-human contact networks in which network communication is disseminated on top of societal structures. It is this technological kinship that we share with our devices that enables such networks, and we wish to explore the implications of this partnership to computing.

References

[1] IEEE VAST 2008 Challenge. http://www.cs.umd.edu/hcil/VASTchallenge08/.

[2] Google latitude. http://www.google.com/latitude/intro.html.

[3] 3rd Generation Partnership Project. Universal mobile telecommunications system (UMTS); synchronization in (UTRAN) stage 2. Technical Specification Group Services and System Aspects 3GPP TS25.402 v8.1.0, 3rd Generation Partnership Project, July 2009.

[4] B. H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Commun. ACM*, 13(7):422–426, 1970.

[5] S. Bregni. *Synchronization of Digital Telecommunications Networks*. Wiley, 2002.

[6] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Pocket switched networks: Real-world mobility and its consequence for opportunistic forwarding. Technical Report 617, University of Cambridge, 15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom, Febuary 2005.

[7] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606–620, 2007.

[8] J. A. Davis, A. H. Fagg, and B. N. Levine. Wearable computers as packet transport mechanisms in highly-partitioned ad-hoc networks. *Wearable Computers, IEEE International Symposium*, 0:141, 2001.

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

[10] W. Enck, P. Traynor, P. McDaniel, and T. L. Porta. Exploiting Open Functionality in SMS-Capable Cellular Networks. In *CCS '05: Proceedings of the 12th ACM Conference on Computer and Communications Security*, pages 393–404, New York, NY, USA, 2005. ACM Press.

[11] C. Fleizach, M. Liljenstam, P. Johansson, G. M. Voelker, and A. Mehes. Can you infect me now?: malware

propagation in mobile phone networks. In *WORM '07: Proceedings of the 2007 ACM workshop on Recurring malcode*, pages 61–68, New York, NY, USA, 2007. ACM.

[12] E. Forgy. Cluster anlaysis of multivariate data: Efficiency vs. interpetability of classifications. *Biometrics*, 21(3):768, 1965.

[13] Z. J. Haas, J. Y. Halpern, and L. Li. Gossip-based ad hoc routing. *IEEE/ACM Trans. Netw.*, 14(3):479–491, 2006.

[14] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket switched networks and human mobility in conference environments. In *WDTN '05: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 244–251, New York, NY, USA, 2005. ACM.

[15] T. Imielinski and J. C. Navas. Geographic addressing, routing, and resource discovery with the global positioning system. *Communications of the ACM Journal*, 1997.

[16] T. Instruments. Omap 4 mobile applications platform. Technical report, February 2009. http://focus.ti.com/lit/ml/swpt034/swpt034.pdf.

[17] B. Karp and H. T. Kung. Gpsr: greedy perimeter stateless routing for wireless networks. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 243–254, New York, NY, USA, 2000. ACM.

[18] D. Kim and N. Maxemchuk. Simple robotic routing in ad hoc networks. In *Network Protocols, 2005. ICNP 2005. 13th IEEE International Conference on*, pages 10 pp.–168, Nov. 2005.

[19] S. Lemon. Apple may patch serious sms vulnerability on iphone. *InfoWorld*, July 2 2009. "http://infoworld.com/d/mobilize/apple-patching-serious-sms-vulnerability-iphone-934".

[20] J. Li, J. Jannotti, D. S. J. De Couto, D. R. Karger, and R. Morris. A scalable location service for geographic ad hoc routing. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 120–130, New York, NY, USA, 2000. ACM.

[21] W.-H. Liao, J.-P. Sheu, and Y.-C. Tseng. Grid: A fully loation-aware routing protocol for mobile ad-hoc networks. *Tekecommnication Systems*, 18(1):37–60, September 2001.

[22] J. MacQueen. Some methods for classifcation and anlaysis of multivariant observations. In L. M. L. Cam and J. Neyman, editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics*, number 1. Univeristy of California Press, 1967.

[23] P. Mann. Timing synchronization for 3G wireless. *EE Times Asia*, December 2004.

[24] C. Miller and C. Mulliner. Fuzzing the phone in your phone. *blackhat usa+2009*, July 25-30 2009. http://www.blackhat.com/html/bh-usa-09/bh-usa-09-speakers.html#Miller.

[25] M. Piorkowski, N. Sarafijanovoc-Djukic, and M. Grossglauser. A Parsimonious Model of Mobile Partitioned Networks with Clustering. In *The First International Conference on COMmunication Systems and NETworkS (COMSNETS)*, January 2009.

[26] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong. On the levy-walk nature of human-mobility. In *IEEE Infocom*, 2008.

[27] K. Singh, S. Sangal, N. Jain, P. Traynor, and W. Lee. Evaluating Bluetooth as a Medium for Botnet Command and Control. Technical Report GT-CS-09-11, Georgia Institute of Technology, Atlanta, GA, 2009.

[28] I. Stojmenovic. Position-based routing in ad hoc networks. *Communications Magazine, IEEE*, 40(7):128–134, Jul 2002.

[29] T-Mobile. Setting the time and date (T-Mobile G1). http://support.t-mobile.com/doc/tm52540.xml.

[30] P. Traynor, M. Lin, M. Ongtang, V. Rao, T. Jaegar, P. McDaniel, and T. L. Porta. On cellular botnets: Measuring the impact of malicious devices on a cellular network core. *ACM Conferance on Computer Security (CCS'09)*, November 9-13 2009.

[31] A. Vahdat and D. Becker. Epidemic routing for partially-connected ad hoc networks. Technical report, Duke University, 2000.

[32] S. Čapkun and M. Hamdi. Gps-free positioning in mobile ad hoc networks. *Cluster Computers*, 5(2):157–167, November 2 2004.

[33] xda-developers & others. Cyanogenmod v. 4.0.2, August 2009. http://www.cyanogenmod.com/downloads/stable-rom.

[34] G. Zyba, G. M. Voelker, M. Lilijenstam, A. Méhes, and P. Johansson. Defending mobile phones from proximity malware. In *Proceedings of the IEEE Infocom Conferaence 2009*. IEEE, April 2009.