University College Dublin

School of Computer Science and Informatics

Transfer Report

Exploiting human networks for delay tolerant environmental sensing

Supervisors:

Paddy Nixon and Simon Dobson

Candidate:

Matthew Stabeler

April 2009

Contents

1 Exploiting human networks for delay tolerant environmental sensing

By exploiting the *features* and *idiosyncrasies* of *human* **social, ad-hoc, mobile networks**, and understanding *human* **temporal** and **spatial movement patterns**; *opportunistic* routing of *sensing* and *messaging* data can be *efficient* and *delay tolerant***,** *without* reliance upon installed *infrastructure*.

Since before the advent of the "power in our pockets" provided by personal mobile devices, Mobile Ad-Hoc Networks (MANETs) have been the subject of much research. Many applications do not require the end-to-end connectivity provided by a pure MANET, as they are resilient to delays and have low Quality of Service (QoS) requirements. Examples being email, SMS, peer-to-peer file sharing and publish-subscribe mechanisms [4, 3]. These kinds of applications may be able to exploit the characteristics of social/mobile networks, such as those of human beings, by forwarding data hop-by-hop, without the need for a complete end-to-end path. Miklas et al [16] have shown that having some knowledge of a social network improves the performance of mobile systems, and investigate its effect on routing messages within social networks.

Our interest also lies in utilizing a sub set of MANETs, known as Opportunistic Networks (OppNets), specifically Delay Tolerant Networks (DTN), and Pocket Switched Networks (PSN) for mobile devices, as carried by human beings, for environmental sensing.

We are interested in developing an autonomous environmental sensing system that reacts to phenomena in a local area, and to global instructions and queries, which utilizes the power (mobile devices) found in the pockets of a large number of the population for routing sensor data, for example, to monitoring stations. However, we believe an environmental monitoring system does not currently exist that can efficiently exploit the features and idiosyncrasies of delay tolerant human networks to route the data, that does not rely on installed infrastructure.

2 Background

In this section we briefly describe MANETs in general, then more specifically a number of solutions for delay tolerant networking in Intermittently Connected Networks and a briefly overview the main aspects of delay tolerant networks.

2.1 MANETS

Conti *et al* [4] say that *mobile ad-hoc networks* (MANETs) have been the subject of much research, and there is a large amount of literature regarding many aspects of them. There have been a number of highly specialised applications which use MANETs such as disaster situation and battlefield communications, but widespread adoption in industry of general-purpose MANETs has not occurred. Conti *et al* [4] also suggest that, without a clear, understandable problem that is solved by having pure MANETs, users are not convinced of its benefits, making the research a purely academic venture, which in turn means that research is not conducted for convincing real-world scenarios. They suggest that the most similar technology to pure MANETs that has a clear application, is MESH computing for internet connectivity which has been developed in various forms, such as *RoofNet* (an MIT project), the *metro-scale broadband city network* (in Cerritos, California,USA), or the *Mad City Broadband* (in Madison, Wisconsin, USA). Another similar technology is Vehicular Ad-hoc NETworks (VANETs) which have received industry interest [4] in their use for Intelligent Transport Systems (ITS).

"Pure indicates that the MANET paradigm is strictly followed: no infrastructure is assumed to implement the network functions, and no authority is in charge of managing and controlling the network. General-purpose denotes that these networks are not designed with any specific application in mind, but rather, they support any legacy TCP/IP application" - [4]

2.2 Route discovery

Route discovery in MANETs and their sub-classes can be categorised into two main types. The MANET Working Group (in [4]) defines the following:

Proactive routing protocols are derived from legacy Internet distance-vector and link-state protocols. They attempt to maintain consistent and updated routing information for every pair of network nodes by propagating, proactively, route updates at fixed time intervals.

Reactive routing protocols establish the route to a destination only when requested. The source node in the route discovery process usually initiates the route request. After a route is established, it is maintained until either the destination becomes inaccessible or until the route is no longer used.

Also identified are a number of classes of routing protocol which improve network performance at least in some specific scenarios. *Hybrid Protocols* use both *proactive* and *reactive* approaches to combine their advantages. *Location-aware routing protocols* use the location of a node to route packets in the direction of the destination, via its neighbours on the most direct route. *Energy-aware routing protocols* take consideration the energy available in the network to select paths. The goal may be to minimize energy consumed, or to maximise the overall network lifetime by preserving network connectivity. In *hierarchical routing protocols* only selected nodes may do multi-hop communication, these nodes represent a group of non-selected nodes, who communicate only with the selected node to transmit data to a node outside of the group.

2.3 Intermittently Connected Networks

Zhang [28] describes how many of the routing protocols designed to cope with the dynamic topology of MANETs, such as OLSO, AODV, DSR, LAR, EASE, ODMRP and others, implicitly assume that there is contemporaneous end-to-end path between any source and destination pair. Zhang [28] says that "In MANETs, when nodes are in motion, links can be obstructed by intervening objects. When nodes must conserve power, links are shut down periodically. These events result in intermittent connectivity. At any given time, when no path exists between source and destination, network partition is said to occur. Thus, it is perfectly possible that two nodes may never be part of the same connected portion of the network." These types of networks are defined as Intermittently Connected Networks. Examples of these types of networks are interplanetary/satellite communications, sensor networks; where sensor do not have the power to transmit all of the time or over long distances, and military networks where nodes move randomly and are subject to being destroyed.

These types of networks must have tolerance for delays, as there can be undefined lengths of time between co-location/transmission range, when routing protocols take this into account, these types of networks are called Delay Tolerant Networks (DTNs), also known as Opportunistic Networks (OppNets).

Zhang [28] categorises a large number of DTN protocols into two sets; the ***Deterministic Routing*** set contains those which know the state of the network or has some global knowledge in advance of sending a message. The ***Stochastic or Dynamic*** *Networks* set includes protocols that react to network behaviour that is random or not known, these protocols depend on decisions regarding where and when to forward messages.

Pelusi et al [20] also define two categories of OppNet routing types; In ***Infrastructure-Less Routing***, *Dissemination-Based Routing* performs delivery of a message to a destination by simply diffusing it all over the network. The heuristic behind this policy is that, since there is no knowledge of a possible path towards the destination nor of an appropriate next-hop node; the message should be sent everywhere. *Context-based Routing* exploits information about the context in which nodes are operating so as to identify suitable next hops towards the eventual destinations (e.g., the home address of a user is a valuable piece of context information to decide the next hop).

In ***Infrastructure-Based Routing***, *fixed infrastructure* means that a source node wishing to deliver a message generally keeps it until it comes within reach of a base station belonging to the infrastructure, then forwards the message to it. Base stations are generally gateways towards less challenged networks (e.g., they can provide Internet access or be connected to a LAN). Hence, the goal of an opportunistic routing algorithm is to deliver messages to the gateways, which are supposed to be able to find the eventual destination more easily. In Carrier-Based Routing, nodes of the infrastructure are mobile data collectors. They move around in the network area, following either predetermined or arbitrary routes, and gather messages from the nodes they pass by. These special nodes are referred to as carriers, supports, forwarders, MULEs, or even ferries. They can be the only entities responsible for messages delivery, when only node-to-carrier communications are allowed, or they can simply help increasing connectivity in sparse networks and guaranteeing that also isolated nodes can be reached.

2.4 OppNets and DTNs

There are a number of OppNet and DTN protocols of note, the basic mechanism being flooding, where the sending party sends a copy of its the message to every neighbour it meets, for some period of time, then in turn, each neighbour transmits the message until some time to live (TTL) or hop-count is reached. This method is effective and can have a very high delivery rate, but at the cost of huge overhead. This method is very inefficient in terms of power consumption and network usage, and not always suitable for nodes that are for example, battery powered.

2.4.1 Epidemic

One of the most interesting [17] works in the area of Delay Tolerant Networks is Vahdat et al's [26] *Epidemic Routing*. The overall goal of Epidemic Routing is to maximize message delivery rate and minimize message delivery latency, while also minimizing the aggregate system resources consumed in message delivery. The authors base their system on a variant of epidemic algorithms which states that given random exchange of data among replicas, all updates will be seen by all replicas in a bounded amount of time; The message for a particular destination copied to other nodes that come into contact based on some probability, furthermore, messages are given a limited hop-count (TTL) and nodes are limited in the amount of buffer space in which they will carry another hosts message, so that old messages are dropped. This is an example of an *Infrastructure-less*, *Dynamic network* with *Reactive*, *Dissemination-Based Routing.* There are many algorithms based on the epidemic routing mechanism, which are said to be *epidemic-like*. Eugster et al [9] suggest the concept of an epidemic like overlay network, suitable for Peer-to-Peer (P2P) networks to operate in.

2.4.2 Li and Rus

The seminal paper (Cited in [17]) describing asynchronous communication in intermittently connected mobile ad-hoc networks, by Li and Rus [15] describes a system which guarantees message delivery within a minimum time, but where nodes must modify their trajectory to transmit messages, such as squads of soldiers on a battlefield. Nodes must have some knowledge of the global network making it a *Deterministic* and *Infrastructure-less* system with *Proactive*, *Context-Based routing*.

2.4.3 Spray and wait

The Spray and Wait routing scheme [23] is an epidemic-like, or semi-epidemic algorithm, which attempts to limit overhead, by using two phases of message passing. The spray phase is defined: for every message originating at a source node, L message copies are initially spread – forwarded by the source and possibly other nodes receiving a copy – to L distinct "relays". Each forwarding node shares L/2 of its copies of the message to the neighbouring node, until it has only one copy left. The wait phase is defined: if the destination is not found in the spraying phase, each of the L nodes carrying a message copy performs direct transmission (i.e. will forward the message only to its destination).

2.4.4 Gossip

Hass et al's [11] Gossip-Based routing works in a similar way to Epidemic routing, in that it improves upon flooding messages within a system by limiting the number of messages disseminated or gossiped by specifying a probability of onward propagation of the message by neighbouring nodes. It is based on the theory of Percolation, and suggests that gossiping exhibits a kind of bimodal behaviour, in that in almost all executions, either hardly any nodes receive the message, or most of them do. GOSSIP is an example of an *Infrastructure-less*, *Dynamic network* with *Reactive*, *Dissemination-Based Routing*, which incorporates some *Deterministic* elements.

The paper proposes four mechanisms for onward propagation of messages:

GOSSIP1(p, k) - The source and intermediate nodes gossip with the probability p to neighbouring nodes, k determines that for the first k hops, the message is gossiped with probability 1, so that the message does not die out prematurely. GOSSIP1(1,1) is the same as flooding.

GOSSIP2(p, k, p2, n) is as GOSSIP1, and additionally tries to react to networks that are poorly connected, when a node has less than n neighbour nodes, it gossips with the higher probability of p2.

GOSSIP3(p, k, m) - if a node broadcasts a message (as with GOSSIP1) but does not receive a duplicate message back from (m) neighbours within a reasonable time period, then it broadcasts to all neighbours. The authors found that the timeout period can be small and does not have a significant effect upon latency.

GOSSIP4(p, k, k') – Uses the concept of zones, where a node has a number of neighbour nodes, it pro-actively tries to keep a routing table for all nodes that are within k' hops away from it, which is its zone. A node can transmit directly to any node in its zone, zones therefore become 'collectors' for messages destined for any node within the zone.

2.4.5 CAR and SCAR

In Musolesi and Mascolo's [17] *Context Aware Routing (*CAR) each node uses a *multi-attribute utility* to determine the probability of forwarding a message onto a neighbour. Delivery probabilities are synthesized locally from context information. The authors define context as the set of attributes that describe the aspects of the system that can be used to drive the process of message delivery. The example used is the *change rate of connectivity*, i.e., the number of connections and disconnections that a host experienced over the last T seconds. This parameter measures relative mobility and, consequently, the probability that a host will encounter other hosts. CAR pro-actively keeps a record of nodes in the locally connected network and uses *Dynamic Destination-Sequenced Distance-Vector* (DSDV) for synchronous routing to nodes. CAR is optimised by using predicted future values of the context attributes for making routing decisions. Predicted values are computed using techniques based on Kalman filters. These techniques do not require the storage of the entire past history of the system and are computationally lightweight, making them suitable for a resource-scarce mobile setting [17]. The SCAR [19] system is based on CAR, but is optimised for sensor networks, by adding battery power as one of the attributes to drive the process of delivery to 'sinks' within the network. SCAR demonstrates the configurable nature of CAR, where any context can be used to tune the algorithm to drive the process of message delivery.

In the situation where a messages is intended for a node not connected to the local network; the node in the local network with the highest probability of delivering the message, is sent the message, as the topology of the network changes over time, messages are passed to the nodes with the highest probability of delivery until the destination node is connected to a network with the node carrying the message. In the CAR protocol it is assumed that only one copy of the message exists at any one time.

2.4.6 HAGGLE

The HAGGLE project [21], funded by the European Union is a large project touted as a clean-slate design for mobile ad-hoc networks, it presents a set of architectural principles based on the needs of mobile users as characterised by the Pocket Switched Networking environment (PSN, discussed below). Scott et al [21] propose a set of interrelated principles which they believe to be fundamental implications of the situation faced by mobile users.

- *Forward using application layer information* instead of using IP addresses, as a user may communicate via many interfaces, devices and protocols.
- *Asynchronous operation* - applications should not have to deal with low level communications - they just need to send a message, and some communications service should take on the role of delivering the message by whichever means suitable (WiFi, Bluetooth exchange, SMS, etc.).
- *Empower immediate nodes* - intermediate nodes (to a destination) may want to keep a copy of the data, in case the data is needed by the user wishes to find the same data in the future for example a popular news website. The authors suggest that this is effectively ad-hoc multi-casting.
- *Messages switching* - instead of concerning the application layer with packets, data should be transmitted as a single message unit, low-level inter-node communications can still use packets, but messages should not be fragmented across different routing nodes.
- *All user data should be network visible*, marked with metadata about its user-level properties, such as access authorisation, creation/modification/expiry times, this is so that 1) data can be shared by multiple devices owned by one person, or to a group of others, 2) Prevent duplication of transmission across multiple interfaces.
- *Build request-response into the network* - many user level tasks (applications) implement request-response semantics - e.g. web browsing / file sharing. By building it into the network there is no reliance upon a particular application to perform the task.
- *Exploit all data transfer methods* – the authors define three opportunities for forwarding; *neighbourhood connectivity* (e.g. Bluetooth, 802.11 ad hoc, infrared), *infrastructure connectivity* (WiFi, 3G, GPRS, SMS, etc) and *user mobility* (Shared mobility pattern, recent sightings of destination node) which can all be utilised to transmit messages.
- *Take advantage of brief connection opportunities* - urgent data should be prioritised, to make most efficient use of bandwidth.
- *Empowered and informed resource management* - controlling the use of resources; storage, networking, battery power.

- *Use and integrate with existing infrastructure* where possible so that users who have adopted Haggle can still communicate with those that haven't.

Haggle uses Application Data Units (ADUs), which are both a unit of data for communications, and a unit of data for applications, they are composed of many attributes, consisting of type-value pairs.

The key aspect of haggle is in the mapping of names to addresses (email, Bluetooth, etc), an ADU may contain a number of 'addresses' which may be understood by different applications. Addresses can be flagged as being nearby (e.g. if connected to an access point, all domain names may be nearby, or a list of the Bluetooth device addresses in the vicinity). The resource manager in haggle evaluates the value and cost of completing tasks (e.g. delivering a message to a neighbour) based on resources consumed and their cost. The goal is to interface with existing applications at a much higher level than a socket connection, and provide proxies to existing services where needed (e.g. email).

2.4.7 Pocket Switched Networks

Pocket Switched Networks [2] (PSNs) are a sub-set of DTNs which deal only with networks involving human networks, specifically, networks created by mobile devices carried by people. Described by Scott et al as follows:

"Pocket Switched Networking (PSN) is the term [used] to describe the situation faced by today's mobile information user. Such users have one or more devices, some/all of which may be with them at any time, and they move between locations as part of a normal schedule. In so moving, the users can spend some (or much) of their time in "islands of connectivity", i.e. places where they have access to infrastructure such as 802.11 access points (APs) which they can use to communicate with other nodes via the Internet. They also occasionally move within wireless range of other devices (either stationary or carried by other users) and are able to exchange data directly with those devices" [21]

2.4.8 BUBBLE Rap

Hui et al's [13] BUBBLE Rap is based on the heterogeneity of human interaction, which causes hubs (popular individuals) and communities to emerge from human social networks. The authors aim is to use these structures to design forwarding algorithms for *Pocket Switched Networks*. In BUBBLE Rap, two assumptions are made; each node belongs to at least one community (single node communities are allowed), and, each node has a global ranking (global centrality) across the whole system, and a local ranking within its local community (it may belong to multiple communities). Messages are 'bubbled' up the hierarchical tree formed by the global ranking of centrality by comparing their ranks, if the candidate node is in the same community as the destination node, then the message is passed to it. At this point, the local ranking is used to determine which nodes should be passed the message; in this way the messages reach the most popular (highest centrality) nodes within a community, who are more likely to come into contact with the destination node. The distributed version of BUBBLE; DiBuBB is designed to calculate the local centrality automatically based on a variable time window, (6 hours was optimum), in which the number of encounters with other nodes was used to calculate a node's centrality (They called this S-Window), another approach was to calculate the centrality based on the average contacts in previous time windows, (dubbed C-Window). The authors used K-Clique and Weighted Network Analysis to pre-calculate communities within their trace datasets. They also found that both S-Window and C-Window were able to approximate the pre-calculated centrality quite well.

2.4.9 Publish Subscribe

Costa et al [5] describe SocialCast, a routing protocol for publish-subscribe that exploits predictions based on metrics of *social* interaction (e.g., patterns of movements among communities) to identify the best information carriers. SocialCast uses the same forecasting techniques as CAR [17] (Kalman Filters), but uses different method for group communication, which is not supported in CAR. Any node within the network may act as publisher or subscriber, when a message is published it is tagged with a related interest. The goal is to deliver the message to nodes with at least one interest matching those in the message. The authors assume that mobility of users is driven by their social behaviour, which is determined by their common interests and that users with common interests are more likely to meet with each other that other people.

Routing in SocialCast is determined by a utility function which is based on the probability of co-location with users with similar interests and the changed degree of connectivity of the user. The authors describe three phases of the protocol [5];

Interest Dissemination, where each node broadcasts a control message with the list of its interests to 1-hop neighbours, along with a list of locally computed utility values for mobility and co-location. This information is key to the protocol, and is stored in the routing table of its neighbours. Identifiers for the most recently received messages are also piggybacked with this control message.

In *Carrier Selection*, the local node compares its own utility for a given message, with that of it's 1-hop neighbours, if one of the neighbours has the highest utility, beyond some threshold (to prevent bouncing back and forth between nodes with similar utility), then the neighbour node is selected as the next carrier.

In *Message Dissemination* each nearby neighbour who has advertised an interest that matches that of a message in the local nodes buffer, are sent a copy of the message. If a neighbouring node was selected as the best carrier, it is also sent the message, and the local copy deleted.

The authors plan to implement SocialCast within the HAGGLE platform.

2.4.10 Directed Diffusion

Intanagonwiwat et al [14] discuss directed diffusion for sensor networks, they describe how all data in their system is named, and all nodes in the system are application aware, enabling diffusion to achieve energy savings by choosing empirically good paths, which are adaptively re-enforced, and by caching and processing data within the network. They best describe their system in the context of human operators posing questions to a sensor network, such as "How many pedestrians do you observe in region X", as follows:

The human operator's query would be transformed into an interest that is diffused towards the nodes in the region X or Y. When a node in that region receives in interest, it activates its sensors which begin collecting information about pedestrians, this information returns along the reverse path of interest propagation. Intermediate nodes might aggregate the data e.g. more accurately pinpoint the pedestrians location by combining reports from several sensors. An important feature of directed diffusion is that interest and data propagation and aggregation are determined by localised interactions (message exchanges between neighbours or nodes within some vicinity) [14]

2.5 Areas of further research in DTN

Zhang's survey paper [28] identifies 10 areas in Delay Tolerant Networks for further research as follows:

1. What is the proper objective function in designing a protocol in DTNs: short delay or high throughput, or others? Related questions are how to define the system capacity in such an intermittently connected network.
2. Methods to determine how many nodes to forward [to] should be developed. There is a trade-off. The larger the number of nodes forwarded to, the better the chance for packets to reach their destination, but the more network resources (bandwidth and buffer space) are needed. Analytical models should be developed, if possible, and simulation results should be obtained to quantify the trade-off.
3. When multiple copies of the packets are in the network, duplication of packets occurs and such duplication requires a method of eliminating unnecessary copies to reduce the buffer occupancy. Where should the duplication reduction be done, at the destination or intermediate nodes and how? When original packets are received successfully at the receiver, how should intermediate nodes be informed to discard these packets? Informing intermediate nodes requires extra resources. Again, there is a trade-off between efficiency and additional overhead.

4. Scheduling becomes much more complex in DTNs than in IP-centric networks, because connections in DTNs are intermittent while in normal IP networks they are not. Appropriate buffer management schemes (which packets to discard when full) and scheduling should be developed. One possible approach is to have separate queues for different outgoing links. Those packets whose destination will be disconnected soon (if known) should be scheduled to transmit first.
5. Whenever possible, information about node location and future movement should be utilized in designing the protocols. The forwarding protocols should leverage simple and accurate link availability estimation methods to make intelligent decisions, if feasible. [...](See citations in [28] for papers dealing with link availability). How to define user profiles and how to use them to estimate the delivery probability is also an open issue.
6. New security mechanisms must be developed, as techniques that rely on access to a centralized service cannot be used, or the assumption that all intermediate nodes are trusted is not valid.
7. Self-learning and automation algorithms should be developed so the underlying network is cognitive, and thus intelligent decisions on scheduling and forwarding can be made automatically.
8. Open spectrum allows secondary users to opportunistically explore unused licensed band on a non-interfering basis. New algorithms to utilize those unused channels (resulting in intermittent connectivity) dynamically and efficiently should be developed.
9. Transmissions in networks with directional antennas are often pre-scheduled and may result in intermittent connectivity. Power management in energy-aware network (range and/or wake/sleep periods control) may also result in intermittent connectivity. Therefore, scheduling transmissions with directional antennas and power management should take into account the DTN requirements or characteristics.
10. As the mobility of nodes in a mobile ad hoc network might lead to network partitions, and directional antennas can transmit over longer distances, [there are proposals](Cited in [28]) to use directional antennas to bridge such partitions when needed. The basic idea behind this method is to use the capability of a directional antenna to transmit over longer distances, but to adaptively use this capability only when necessary for selected packets. Methods to close broken links in ad hoc networks should be developed to cope with partial connectivity.

Chaintreau et al [2] suggest two areas for further research as follows:

1. Current mobility models (e.g. random waypoint, uniformly distributed locations) do not have the characteristics observed in our human mobility experiments. New mobility models are therefore required in order to facilitate evaluation of potential opportunistic data transmission schemes.
2. Little work has been done in the area of informed design of opportunistic forwarding algorithms. This remains an area ripe for study. Suitable directions for work might involve the sharing of recent contact information between nodes, leading to a more careful selection of potential relay nodes which are likely to have a short path to the destination, while also being independently moving as compared to other chosen relay nodes.

Scott et al [21] identify that future work will be needed to develop and evaluate solutions to various challenges in PSN, including forwarding algorithms, security policies, usability aids, and resource management policies.

2.5.1 DTN Literature Overview

From the literature, it is evident that there are a number of factors involved when considering new delay tolerant routing protocols for mobile ad-hoc networks; How many copies of a messages should be sent; how far should the message be sent (TTL); should the route be discovered proactively, reactively or using a hybrid of both; how many what meta-data about known routes should be kept in routing tables; should pair-wise communication take place based on: probability of co-location with, or the geographic location of, or the probability of onward delivery by, a neighbouring node; how big should a message be, and what is the ideal atomic data-unit (message or packet switched); is data dissemination based on one-to-one (SMS like), one-to-many (Publish-Subscribe), many-to-many (reactive Sensor Networks) or many-to-one (reporting Sensor Networks) messaging, how should messages be acknowledged; is there infrastructure (Data Mules, Sinks, Access Points); how should addressing of other nodes be handled; what types of applications will be supported?

It is also clear that there are aspects of networks that can be exploited, such as the relationship between nodes in the network [17][21][13] the patterns of node movement [15] [10] and temporal activity [10].

2.6 Social Networks

Before we can start to prototype some protocol for communication based on the features of human networks, it is important to understand the features of the social networks created by humans, and try to determine aspects of it that can be exploited.

The use of mobile devices by everyday users has increased substantially within the last few years, Wallace [27] discusses statistics from the Informa Telecoms and Media report showing that by 2006, 40 countries were expected to reach 100% mobile phone penetration. We are particularly interested in investigation and exploiting the networks created by the movement and social connections of human being by using these devices as nodes.

Some existing protocols use some aspects of the social network as part of their algorithm; CAR and SCAR use the probability of co-location with another node as its context. Pocket Switched Networks also use this type of probability to route messages onwards.

Li and Rus's algorithm assumed that nodes have a knowledge of the approximate location of other nodes in the space, but assumes nodes will be actively attempt to change direction to transmit the message.

2.7 The Small World Problem

"What is the probability that any two people, selected arbitrarily from a large population, such as that of the United States, will know each other?" [25]. Travers and Milgram [25] examine the Small World Problem and try to explore social connectedness, suggesting that an interesting aspect, takes account of the fact that, while persons *a* and *z* may not know each other directly, they may share one or more mutual acquaintances. They asked 296 people from different areas of the country (US) to send a letter to one person (in Boston), with only a small amount of information about the person (name, industry, town), constraining them to only send it to people they know on first name terms, who they think will be able to get it another step towards the recipient. They found that pattern of chains emerged, where 48% of all chains pass through certain socio-metric 'stars', these 'stars' can be considered to be highly connected individuals, or individuals with high centrality. The authors also found that messages tended to be sent through similar classes of society, and business types (financial) to that of the recipient, they also discovered that men were more likely to pass to another man, and women tended to pass on to either sex. The average chain length was 5.2.

2.8 Human Mobility Patterns

Understanding mobility patterns has two benefits for Delay Tolerant Networks, firstly, understanding how nodes move, may allow us to develop techniques that exploit features of these patterns, secondly, for testing our developed systems, we need a reliable and accurate method to evaluate their effectiveness in real-world situations, and a model of mobility for whatever nodes we are studying (human, animal, robots etc.) means that we can simulate systems efficiently, before expending the effort of real-world testing.

Gonzalez et al [10] studied the movement of 100,000 individuals over a period of six months. They claim that our understanding of the basics laws governing human motion remains limited owing to the lack of tools to monitor the time-resolved location of individuals. They compared the data against the models of Levy Flight, and Random Walk; human movement patterns often used in the literature to simulate movement. They found that the data shows a high degree of temporal and spatial regularity i.e. Individuals have a significant probability to visit the same locations. They also show that each individual has a time independent characteristic travel distance e.g. some may travel only locally most of the time, but some will also travel longer distances regularly. After correcting for differences in travel distances and the inherent anisotropy [(naively put: direction of movement)] of each trajectory, the individual travel patterns collapse into a single spatial probability distribution, indicating that, despite the diversity of their travel history, humans follow simple reproducible patterns. [10]

Museli and Mascolo [17] question the value of the models of movement used in the literature, for example the Random Waypoint mobility model, suggesting that they generate purely random movements that are very different from the ones observed in the real world and produce meaningless co-location patterns. They describe how they used a mobility model based on social network theory, the Community based mobility model (Cited in [5, 17]), which relies on the simple observation that *mobile networks are social networks after all*, since mobile devices are carried by individuals. The model is able to generate movements that are based on the strength of the relationships between the people carrying the devices.

They identify that the key problem is generation of a synthetic social network with realistic characteristics in terms of clustering and average path lengths between members of the communities. They base their approach on the Caveman model (cited in [17]) to generate a social network characterised by a realistic clustering degree. The social network is built starting from a certain number of fully connected graphs representing communities living in isolation, like primitive men in caves. According to this model, every edge of the initial network in input is re-wired to point to a node of another cave with a certain probability p. The re-wiring process is used to represent random interconnections between the communities. A weight modelling the importance of the relationship between two individuals is associated to each link (i.e., edge of the graph) of the network; edges between individuals of the same community are higher than the others. [17] The work in Costa et al [5] from the same group also adopts this community based mobility model for evaluating their approach to socially aware routing for publish-subscribe delay tolerant networks.

3 Proposed solution

We propose to address the areas of further research identified by Zhang; points 1, 5 and 7, Chaintreau et al's first point and Scotts et al's suggestion that DTN routing protocols need investigation.

Our goal is to develop a distributed environmental sensing application which uses a framework for opportunistic, delay tolerant networking to support autonomous behaviour of sensing nodes, which react to environmental phenomena. We will provide the underlying functionality of this system by concentrating on communications framework and autonomous behaviour aspects of such a system, for deployment to mobile devices carried by human beings. Such a system should make best use of, but should not rely upon installed infrastructure.

Our system will place the sensing where the people are, and will be able to track the interesting areas dynamically, whilst at the same time providing some beneficial service to users; i.e. the incentive for them to use our system.

We believe that the best solution will take a number of the approaches taken to address the problems in DTNs and combine them to create a framework that uses the context derived from social networks, probability of co-location and an understanding of human spatial and temporal movement patterns. Such a system should support multiple types of messaging, for example, the many-to-one characteristic of sensor networks (e.g. to support sensor data), the one-to-one characteristic of SMS text messaging (e.g. to support control messages) and the one-to-many style of publish-subscribe (e.g. to alert peers to local phenomenon). The closest project to our vision of a DTN framework, is the HAGGLE project [21].

As identified by Conti et al [4] it is important for any system to have clear and tangible benefits in order for it to be adopted by members of the public in any practical way. In addition to creating a general framework for DTN communication, we would provide applications that use it, for example a mobile phone based P2P file-sharing network, and SMS-like communications system.

One approach to the underlying problem of routing in delay tolerant networks is to find a link between local properties of a graph, and global properties of the whole network - for example, can the degree of N local neighbours be representative of the graph as a whole, or the cluster of which it is a member, and therefore, can we make assumptions about who best to transfer the message to?

We would like also like to determine if a link exists between the movement patterns defined by Gonzalez et al [10], and the degree of connectivity/centrality of a node (person) within a social network.

We will investigate whether the type of relation between individuals (e.g. friend, relative, colleague, stranger) can be used as a metric for onward forwarding of messages and use this meta-information to build a probability based routing table, perhaps by incorporating the work done by Musolesi and Mascolo [17]. Practically, it may be possible to bootstrap information about users from existing social network information available through social networking websites.

For any new system for delay tolerant networking to be considered credible, it must be evaluated against realistic data, ideally by real-world testing. From the literature it is clear that some thought has been given to generating realistic movement models for simulators, therefore, we intend to evaluate and test our system against the works discussed in the literature, and ideally with a real-world deployment.

4 Summary of progress to date

In this authors initial background research, context awareness in pervasive systems was the subject of investigation, of particular note were the works by Abowd et al [1] that defines the four primary types of context for pervasive systems as Identity, Activity, Location and Time (who, what, where and when), and makes the following statements about context; Context is any information that can be used to characterize the situation of an entity. A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the users task. Henriksen and Idulska [12] also define four flavours of context; sensed, static, profiled and derived, they also discuss the issues surrounding the accuracy of data; unknown, ambiguious, imprecise and erroneous data.

This prompted a project on sensor fusion, which showed that by fusing or 'mashing' the readings from multiple installed sensor systems, it was possible to get a better understating of the context inherent in the system. The paper was published [18] as below:

> Steve Neely, Matthew Stabeler, and Paddy Nixon SensorMash*: **Exploring System Fidelity through Sensor Mashup**. Adjunct Proceedings of the Sixth International Conference on Pervasive Computing, 2008.*

We explored ways to investigate ways to elicit information from users in the environment, and discovered that people were often happy to give up their profile information if it appeared to have some benefit to them. We completed a project which allowed users to register their basic details, and give pointers to more profile information stored in many different places on the web. We were able to harvest these profiles and provide access to the data via API. We demonstrated show that developers of pervasive and ubiquitous system would find this aggregated information highly useful, and that it would bootstrap the development of pervasive applications if users profile information did not need to be collected for every application. The details of this project were published [24] as follows:

Matthew Stabeler and Graeme Stevenson and Simon Dobson and Paddy Nixon ***Basadaeir: harvesting user profiles to bootstrap pervasive applications*** . *Adjunct Proceedings of the Seventh International Conference on Pervasive Computing*, 2009.

We also showed how this type of system has a very real application; pervasive information services based in user profiling. In Shannon et al [22] we proposed using this type of system to link user's real-world presence to their online profiles, to provide personalised services and advertisements. Published as follows:

Ross Shannon, Matthew Stabeler, Aaron Quigley and Paddy Nixon ***Profiling and targeting opportunities in pervasive advertising***. 1st Workshop on Pervasive Advertising, Pervasive 2009.

In Dobson et al [7] we briefly discuss how our Basadaeir project aligns with our research group's Construct project, in which this author is a member of the development team. Construct is an open-source framework that brings together *best of breed* technologies, including gossiping for data distribution, semantic web techniques for data representation and manipulation, and Zero configuration for discovery. Published as follows:

Simon Dobson and Graeme Stevenson and Graham Williamson and Stephen Knox and Matthew Stabeler and Lorcan Coyle and Steve Neely and Paddy Nixon ***An Open-Source Infrastructure for Pervasive Computing***. *PerAda Magazine, 2008.*

We have also conducted work towards gathering datasets, in the case of Coyle et al [6], for activity recognition, but in our work on DTNs, we will be able to use these high-quality datasets to analyse social networks, mobility patterns and temporal activities. Published as follows:

Lorcan Coyle, Juan Ye, Susan McKeever, Stephen Knox, Matthew Stabeler, Simon Dobson and Paddy Nixon. ***Gathering datasets for activity identification.*** In Proceedings of the Workshop on Developing Shared Home Behaviour Datasets to Advance HCI and Uniqiutous Computing Research at CHI 2009. Boston, MA., 2009.

5 Future work

We intend to concentrate on the fundamental aspects of our approach to solving the problem; namely, developing robust DTN protocols that enable our vision of a DTN application framework. Specifically, we will concentrate on exploiting features the networks of interest that we have identified, to develop a suite of routing protocols, or one unified protocol that will cater for the needs of a number of data types.

In collaboration with other members of the group, we will analyse existing datasets to find links between the local features of social networks and connectivity patterns to global properties that can be exploited for informed routing about these networks. Specifically, we will investigate the links between human social network, their real-world connectivity patterns and movement patterns.

We will experiment with different routing types for different data situations – e.g. sensor data vs one-to-one messages. From this we will develop a robust algorithm for node to node communication that exploits communications opportunities, which can be used to control nodes for environmental sensing when necessary.

We have started planning for a project to realise an accurate and robust simulator for plug-in mobility models, and this author has proposed an ODCSS summer project to take place at UCD to develop the software that will enable this.

We intend to simulate our work to evaluate it against the relevant works presented in this document, before building an empirical study involving a real-world deployment of this system.

We will also strive to derive a formalisation of dynamic networks from our understanding gained from our analysis.

5.1 Datasets

There are a number of datasets available which can be analysed to derive movement models, social networks and temporal patterns, for example the Reality Mining project [8] collected data from 100 mobile phone over 9 months; Bluetooth devices in proximity, cell tower IDs, application usage, and phone status. The authors were able to measure information access and use in different contexts, recognize social patterns in daily user activity, infer relationships, identify socially significant locations, and model organizational rhythms.

An as yet unpublished data-set from within UCD and TCD is currently being collected, it is similar in nature to the Reality Mining dataset, but is being conducted over a longer time period collecting data about WiFi, Bluetooth devices in proximity and cell tower ID.

5.2 Timescale

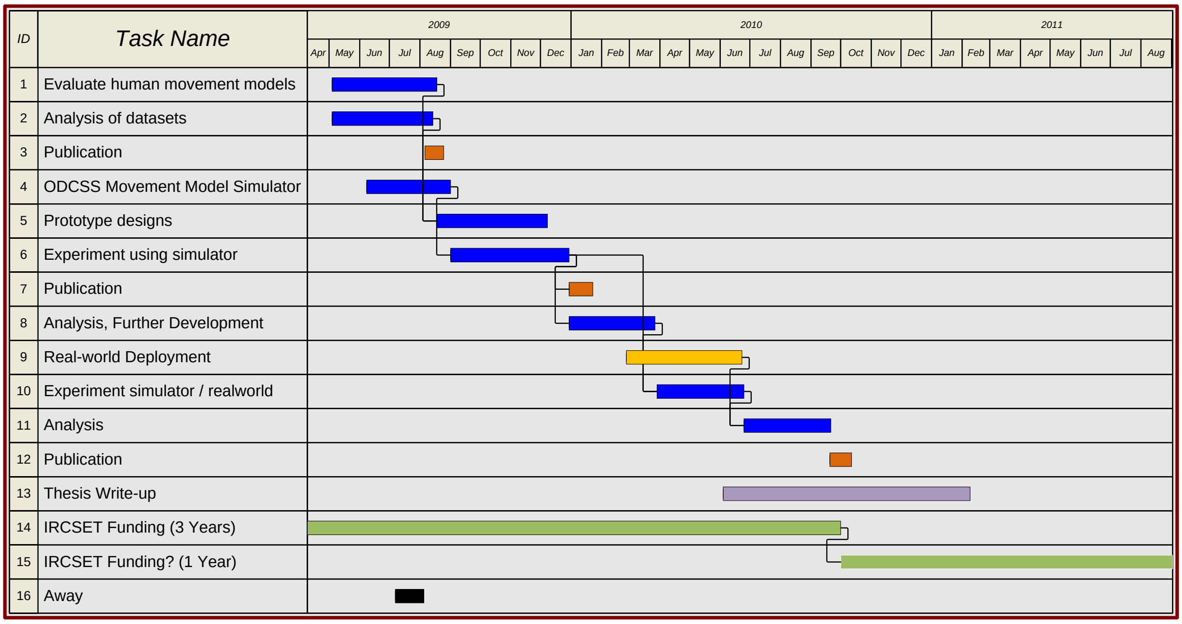

Table 1 – Gantt chart showing estimated working time for each project part.

Our future work is split into the following packages, also shown as a Gantt chart in Table 1 above.

- Dataset analysis
- Movement Model Simulator
- Prototype design and experimentation
- Analysis of initial results
- Real-world deployment (if time)
- Further experimentation
- Final Analysis

We have also identified the following publication targets, in which we intend to present our work.

- ACM Transactions on Sensor Networks
- Ubicomp
- Personal and Ubiquitous Computing
- IEEE/ACM Transactions on Networking
    - IEEE Conference on Computer Communications (INFOCOM)

6 References

[1] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, and Pete Steggles Mark Smith. Towards a better understanding of context and context-awareness. In *"Handheld and Ubiquitous Computing: First International Symposium, HUC'99, Karlsruhe, Germany, September 1999. Proceedings"*, volume "Volume 1707/1999", page 304. "Springer Berlin / Heidelberg", 1999.

[2] Augustin Chaintreau, Pan Hui, Jon Crowcroft, Christophe Diot, Richard Gass, and James Scott. Pocket switched networks:real-world mobility and its consequences for opportunistic forwarding. Technical Report 617, University of Cambridge, 15 JJ Thomson Avenue Cambridge CB3 0FD United Kingdom phone +44 1223 763500, Feb 2005.

[3] M. Conti and S. Giordano. Multihop ad hoc networking: The reality. *Communications Magazine, IEEE*, 45(4):88–95, April 2007.

[4] M. Conti and S. Giordano. Multihop ad hoc networking: The theory. *Communications Magazine, IEEE*, 45(4):78–86, April 2007.

[5] P. Costa, C. Mascolo, M. Musolesi, and G.P. Picco. Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks. *Selected Areas in Communications, IEEE Journal on*, 26(5):748–760, June 2008.

[6] Lorcan Coyle, Juan Ye, Susan McKeever, Stephen Knox, Matthew Stabeler, Simon Dobson, , and Paddy Nixon. Gathering datasets for activity identification. volume In Proceedings of the Workshop on Developing Shared Home Behaviour Datasets to Advance HCI and Uniquitous Computing Research at CHI 2009, 04/04/2009 2008.

[7] Simon Dobson, Graeme Stevenson, Graham Williamson, Stephen Knox, Matthew Stabeler, Lorcan Coyle, Steve Neely, and Paddy Nixon. An open-source infrastructure for pervasive computing. *PerAda Magazine*, 10/2008 2008.

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

[9] R.; Kermarrec A.-M.; Massoulie L. Eugster, P.T.; Guerraoui. Epidemic information dissemination in distributed systems. *Computer*, 37(5):60–67, May 2004.

[10] Marta C. Gonzalez, Cesar A. Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(453):779–782, June 2008.

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

[12] K. Henricksen and J. Indulska. Modelling and using imperfect context information. *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, pages 33–37, March 2004.

[13] Pan Hui, Jon Crowcroft, and Eiko Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. In *MobiHoc '08: Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, pages 241–250, New York, NY, USA, 2008. ACM.

[14] Chalermek Intanagonwiwat, Ramesh Govindan, and Deborah Estrin. Directed diffusion: a scalable and robust communication paradigm for sensor networks. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 56–67, New York, NY, USA, 2000. ACM.

[15] Qun Li and Daniela Rus. Sending messages to mobile users in disconnected ad-hoc wireless networks. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 44–55, New York, NY, USA, 2000. ACM.

[16] Andrew G. Miklas, Kiran K. Gollu, Kelvin K. W. Chan, Stefan Saroiu, Krishna P. Gummadi, and Eyal de Lara. *Exploiting Social Interactions in Mobile Systems*, volume 4717/2007 of *Lecture Notes in Computer Science*, chapter UbiComp 2007: Ubiquitous Computing, pages 409–428. Springer Berlin / Heidelberg, 2007.

[17] Mirco Musolesi and Cecilia Mascolo. Car: Context-aware adaptive routing for delay tolerant mobile networks. *IEEE Transactions on Mobile Computing*, 99(1), 2008.

[18] Nixon P Neely S, Stabeler M. Sensormash: Exploring system fidelity through sensor mashup. In J. Kay-G. Kortuem S R. Mayrhofer, A Quigley, editor, *Adjunct Proceedings of the Sixth International Conference on Pervasive Computing*, pages 83–86. Ó, May 2008.

[19] B. Pasztor, M. Musolesi, and C. Mascolo. Opportunistic mobile sensor data collection with scar. *Mobile Adhoc and Sensor Systems, 2007. MASS 2007. IEEE Internatonal Conference on*, pages 1–12, Oct. 2007.

[20] Luciana Pelusi, Andrea Passarella, and Marco Conti. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. *Communications Magazine, IEEE*, 44(11):134–141, November 2006.

[21] James Scott, Jon Crowcroft, Pan Hui, and Christophe Diot. Haggle: a networking architecture designed around mobile users. pages 78–86, 01 2006.

[22] Ross Shannon, Matthew Stabeler, Aaron Quigley, and Paddy Nixon. Profiling and targeting opportunities in pervasive advertising. In *1st Workshop on Pervasive Advertising*, volume 1, Nara, Japan, May 2009. Pervasive 2009.

[23] Thrasyvoulos Spyropoulos, Konstantinos Psounis, and Cauligi S. Raghavendra. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In *WDTN '05: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 252–259, New York, NY, USA, 2005. ACM.

[24] Matthew Stabeler, Graeme Stevenson, Simon Dobson, and Paddy nixon. Basadaeir: harvesting user profiles to bootstrap pervasive applications. In *The Seventh International Conference on Pervasive Computing - Late Breaking Results*, Nara, Japan, 5 2009.

[25] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.

[26] Amin Vahdat and David Becker. Epidemic routing for partially-connected ad hoc networks. Technical report, Duke University, 2000.

[27] Bob Wallace. 30 countries passed 100% mobile phone penetration in q1. Online, June 2006. Telecommunications Online, http://www.telecommagazine.com/newsglobe/article.asp?HH_ID=AR_2148, last accessed Mar 2009.

[28] Zhensheng Zhang. Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and challenges. *Communications Surveys & Tutorials, IEEE*, 8(1):24–37, Quarter 2006.