MOBILE AD HOC NETWORKING: THE CUTTING EDGE DIRECTIONS

MOBILE AD HOC NETWORKING: THE CUTTING EDGE DIRECTIONS

Stefano Basagni, Marco Conti, Silvia Giordano and Ivan Stojmenovic

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright ©2012 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created ore extended by sales representatives or written sales materials. The advice and strategies contained herin may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department with the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication Data:

Mobile Ad Hoc Networking: The Cutting Edge Directions, Second Edition

by Stefano Basagni, Marco Conti, Silvia Giordano and Ivan Stojmenovic
Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

CONTRIBUTORS

THRASYVOULOS SPYROPOULOS, Mobile Communications Department, EURECOM, France

ANDREEA PICU, Communication Systems Group, ETH Zürich, Switzerland

CONTENTS

CHAPTER 11

OPPORTUNISTIC ROUTING

THRASYVOULOS SPYROPOULOS[1] AND ANDREEA PICU[2]

[1] Mobile Communications Department, EURECOM, France
[2] Communication Systems Group, ETH Zürich, Switzerland

Abstract— ***Opportunistic or Delay Tolerant Networks (DTNs) may be used to enable communication in case of failure or lack of infrastructure (disaster, censorship, remote areas) and to complement existing wireless technologies (cellular, WiFi). Wireless peers communicate when in contact, forming an impromptu network, whose connectivity graph is highly dynamic and only partly connected. To cope with frequent, long-lived disconnections,*** **opportunistic routing** ***techniques have been proposed in which, at every hop, a node decides whether it should forward and/or store-and-carry a message. Despite a growing number of such proposals, there still exists little consensus on the most suitable routing algorithm(s) in this context. One of the reasons is the large diversity of emerging wireless applications and networks exhibiting such "episodic" connectivity. These networks often have very different characteristics and requirements, making it very difficult, if not impossible, to design a routing solution that fits all.***

In this chapter, we start by describing some key characteristics of DTN environments. We first discuss generic **network characteristics** ***that are relevant to the routing process (e.g. network density, node heterogeneity, mobility patterns), and identify three major obstacles that most opportunistic routing protocols have to overcome: (a) the uncertainty of future connectivity, (b) the ever-present structure in human and vehicular mobility, and (c) the***

Mobile Ad Hoc Networking: The Cutting Edge Directions, Second Edition. By Stefano Basagni, Marco Conti, Silvia Giordano and Ivan Stojmenovic

heterogeneity of node resources and mobility. We then describe replication techniques aiming to cope with the first obstacle, and utility-based forwarding techniques which aim at exploiting the second. Finally, we discuss hybrid, "state-of-the-art" routing schemes, that combine both replication and contact prediction to achieve good performance in a variety of settings and environments.

11.1 INTRODUCTION

Traditionally, communication networks (wired or wireless) have always been assumed to be connected all or most of the time. Networks are connected in the sense that there (almost) always exists at least one end-to-end path between every pair of nodes in the network. When partitions occur, they are considered transitory failures and core network functions such as routing react to these failures by attempting to find alternate paths. Even in wireless multihop ad hoc networks (or MANETs), where links are more volatile due to wireless channel impairments and mobility, partitions are still seen as exceptions and assumed infrequent and shortlived.

However, advances in wireless communications as well as ubiquity of portable computing devices has spurred new, emerging applications – such as emergency response, smart environments, habitat monitoring, and vehicular networks (VANETs) – for which the assumption of "universal connectivity" among all participating nodes no longer holds. In fact, for some of these applications and usage scenarios, the network may be disconnected most of the time; in more extreme cases, there may never be an end-to-end path available between a source and a destination. Other than the application scenarios themselves, additional factors contribute to frequent, arbitrarily long-lived connectivity interruptions, such as node heterogeneity (i.e., nodes with different radio ranges, resources, battery life, etc.), volatile links (e.g., due to wireless propagation phenomena, node mobility, etc.), energy efficient node operation (e.g., duty cycling).

Networking under such intermittent connectivity is particularly challenging, as many of the assumptions made by traditional protocols (TCP, DNS, etc.) "break" in this context [35]. Nevertheless, routing is (arguably) one of the biggest hurdles to overcome. Traditional routing protocols, both table-driven or proactive (e.g., link-state routing protocols like OSPF and OLSR) and reactive ones (e.g., DSR, AODV), assume the existence of a complete end-to-end path and try to discover it, before any useful data is sent. As a result, their performance deteriorates drastically as connectivity becomes increasingly sporadic and short-lived.

To this end, the Opportunistic or Delay Tolerant Networking (DTN) model has been proposed. Opportunistic networks aim at enabling communication in case of failure of the communication infrastructure (natural disaster, censorship) or utter lack of it (rural areas, extreme environments). They are also envisioned to enhance existing wireless networks (e.g., offload cellular data traffic), enabling novel applications. DTN nodes harness unused bandwidth by exchanging data when they are in physical proximity (in *contact*), over high-speed interfaces (e.g., Bluetooth, WiFi Direct), circumventing the infrastructure (e.g., to avoid costs or improve transmissions rates) or offloading the infrastructure, in collaboration with the provider [27, 75].

Clearly, the Opportunistic or DTN viewpoint is applicable in a broad range of settings and scenarios with rather diverse characteristics regarding network density, network size, node resources, node mobility, performance requirements, as well as knowledge or predictability of future connectivity opportunities. As a result, a very large number of routing and opportunistic forwarding solutions have been proposed in the last decade, targeting one or more environments each. They can be classified into three categories: (i) *deterministic* or *scheduled*, (ii) *enforced*, and (iii) *opportunistic* routing. *Deterministic* routing solutions are used when contact information is known a priori. Jain et al. [35] showed how partial or full information about contacts, queues, and traffic can be utilized to route messages from a source to a destination in the case of disruptions. They presented a modified Dijkstra algorithm, based upon information on scheduled contacts and compared the proposed approach to an optimal Linear Programming formulation. *Enforced* routing solutions introduce special-purpose mobile devices – message ferries [81] or data mules [62]) – which move over predefined paths in order to provide connectivity and deliver messages to otherwise disconnected parts of network (islands). *Opportunistic* routing, in its simplest form, performs epidemic packet dissemination [71] as follows. When nodes A and B encounter, node A passes to B replicas of all messages A is carrying, which B does not have, and viceversa. In other words, epidemic routing is to episodically connected environments what flooding is to “traditional”, (well-)connected networks. While epidemic routing guarantees minimum delivery delay, it may be prohibitively expensive, since it consumes considerable network resources, due to the excessive amount of message duplicates generated.

Our focus here will be mainly on the latter *opportunistic* approaches to DTN routing, i.e., where no contact information is known a priori and no network infrastructure (e.g., special-purpose nodes with controlled trajectories) exists to provide connectivity. The majority of the plethora of opportunistic routing schemes employ one or more of the following basic mechanisms:

(i) Storing and carrying a message for a long period of time, until the respective node has a communication opportunity (“mobility-assisted”).

(ii) Making local and independent forwarding decisions. These decisions may be taken either in utter ignorance of the destination’s position and of future contacts or based on knowledge collected through online measurements and observation. This knowledge is subsequently analyzed statistically and used to make decisive predictions. The goal of each forwarding decision is to bring the message probabilistically closer to each destination (i.e., to increase the probability that the destination will be contacted/met).

(iii) Propagating multiple copies of the same message in parallel (“replication”), to increase the probability of at least one being delivered.

(iv) Source coding and network coding techniques, which often reduce performance variability and improve resource usage.

The rest of the chapter is organized as follows. In the next Section, we identify and discuss the three main characteristics of Opportunistic Networks, around which all forwarding schemes are designed and optimized, namely (i) the volatility and uncertainty of connectivity, (ii) the inherent structure and patterns in node mobility, (iii) and the available network and node resources. In Section 11.3, we focus on the first aspect, and see how replication and coding techniques are used to improve delivery performance, without explicit knowledge about past or future contacts. Then, in Section 11.4, we discuss how contact patterns can be inferred by observing and maintaining some statistics about past contacts, and how simple or more sophisticated techniques can be used to predict future ones, based on collected knowledge. In both these Sections, network and node resources are an ever-present, driving concern, not entirely orthogonal to the other two main DTN features, volatile connectivity and structured mobility. As a natural extension, many opportunistic routing protocols combine both redundancy and prediction techniques, in order to provide better and robust performance. Section 11.5 discusses some examples of opportunistic routing scheme, which we consider to be state-of-the-art, that exploit both replication and prediction techniques. We conclude this chapter in Section 11.6 with a discussion of open issues in opportunistic forwarding research.

11.2 THE CORNERSTONES OF OPPORTUNISTIC NETWORKS

We start with a discussion of DTN environments and their characteristics, that play an important role in designing efficient forwarding algorithms. This taxonomy of scenarios has three main dimensions, namely *connectivity*, *mobility*, and *network and node resources*. Based on the analysis of the different properties of a scenario, we observe that all opportunistic forwarding schemes must, at the foundation, deal with three key elements in DTNs: (i) the *uncertainty of future connectivity* and the stochastic nature of many processes involved, (ii) the *patterned nature and heterogeneity of node mobility* and (iii) the heterogeneous (non-)availability of *network and node resources*.

11.2.1 Connectivity

As discussed earlier, connectivity, or rather the lack or instability of connectivity, is the starting point for considering alternative routing paradigms. Two well-known definitions of network connectivity are: (i) the probability that a path exist between two randomly chosen nodes [79], or (ii) the percentage of nodes connected to the largest connected component [79]. Although these two definitions are slightly different, they have similar implications from a macroscopic point of view. It has been recently recognized that, due to node mobility, wireless channel impairments, limited node capabilities, etc., connectivity in many envisioned wireless networks, like the ones mentioned in Section 11.1, will be consistently below 100%. As a result, the whole spectrum of possible connectivity values, all the way from 0 (very sparse

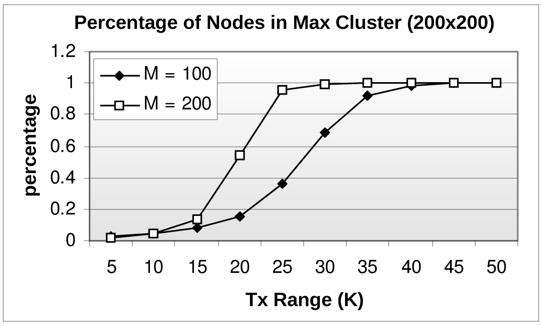


Figure 11.1 Expected percentage of total nodes in largest connected component, as a function of the number of nodes (M) and transmission range (K) (200×200 grid).

networks) to 1 (connected networks), must be considered when designing routing algorithms.

It is well-known from percolation theory that, in networks consisting of randomly placed (or randomly moving) nodes, connectivity exhibits a *phase transition* behavior [44] as depicted in Fig. 11.1. Specifically, if connectivity is scaled by changing the nodes' transmission range, then the following can be observed [45]: (i) for (many) low transmission range values, connectivity values are relatively low: there is no large cluster, but rather very small clusters (some with 1 node), whose sizes are exponentially distributed; (ii) when transmission range crosses some threshold value, connectivity increases rapidly and quickly enters a region where a giant component is formed, containing a large percentage of nodes, while the rest of the nodes form smaller clusters (again exponentially distributed in size). A more pessimistic picture is painted if one considers connectivity on a line, or more realistically on a long narrow band. This model is relevant to vehicular connectivity, for example, on freeways. In this case, percolation theory predicts that, with uniform placement of nodes, the network is asymptotically always partitioned [16].

This phase transition behavior has some important implications: *random networks*, i.e., those formed by randomly placed nodes (e.g., sensors scattered uniformly in the field) or randomly moving nodes (e.g., random direction), will be either *sparse* or *almost connected*, in most cases. But, if the transmission range or the number of nodes are low, the situation can arise, where nodes tend to form clusters (or connectivity islands), due to their mobility patterns. Thus, in the following, we discuss three different kinds of networks according to their connectivity, namely: *almost connected networks*, *sparse networks*, and *connectivity islands*.

Almost connected networks: The connectivity graph, while relatively dense, often exhibits partitions. Furthermore, due to node mobility or link quality fluctuations, a good percentage of pairs are connected end-to-end at any time, yet the paths might not be long-lasting. Traditional proactive (e.g., link-state) or reactive routing protocols (e.g., DSR, AODV) could still deliver a part of the traffic successfully (although with a higher overhead for route maintenance and more frequent route discoveries). Yet, they are unable to deliver any traffic between nodes that lie in different partitions. Opportunistic forwarding can complement such path-based protocols to "push" a message through the right sequence of partially and temporarily overlapping clusters [28].

Sparse networks: In these networks no large clusters exist. Nodes are isolated most of the time or have at most a few neighbors. Every now and then, two such nodes come into contact, at which time they can exchange data or other useful information, *for a limited contact duration*. Traditional end-to-end protocols clearly would fail, and any attempt to maintain multihop neighborhood information and paths has little value and high cost. As a result, in these networks a message must be routed predominantly by opportunistic forwarding. When a new candidate relay is encountered, the forwarding scheme must decide whether it should handover the message, forward a (coded or uncoded) copy, or do nothing. We will see that this decision can be fully random, greedy – based on some utility, or probabilistic. Another important implication of sparse networks is that whenever two nodes encounter each other, there is only a small probability that other nodes are also within range. As a result, there is little contention, on average, at the MAC layer for each transmission, and there is also little (in-channel) interference. This suggests that available bandwidth per contact and/or node buffer space are the limiting factors as far as performance is concerned. What is more, it also suggests that forwarding or scheduling techniques that aim to choose the right neighbor (e.g., transmit to the best neighbor according to some utility function) [68] or combine packets for different neighbors (e.g., opportunistic network coding [40]) will not offer much here.

Clusters or Connectivity Islands: It has been observed that in the real world node mobility is non-uniform. While the phase transition phenomenon described earlier implies that random networks are either sparse or almost connected, in real world different connectivity structures may be observed. Vehicular nodes tend to gather around different concentration points (e.g., traffic lights, junctions, toll, etc.) [61], humans may gather in popular or communal locations (e.g., on campus, people move within their own departments [29]) and both are bound by geography: streets, paths, buildings. Even with virtually no geographic constraint, many animal species move together in herds [37]. Further examples of real world DTNs, with non-uniform mobility include *First Mile Solutions* [72] and *VLINK* [41]. A snapshot of the connectivity graph for these networks exhibits significant clustering and community structure [58], with well defined islands of (good) connectivity, and few or no contemporary paths between clusters.

11.2.2 Mobility

Node density, transmission range, and node placement (uniform or clustered) dictate how *a single snapshot* of the connectivity graph looks like. Mobility, on the other hand, is to a large extent responsible for *how the connectivity graph evolves over time*. In other words, it defines how a sequence of connectivity graph snapshots may look like, how the properties of subsequent snapshots depend on the current one, etc. The mobility process is not only responsible for the amount of variability between different realizations of the network over time, but also for the ability to predict how the experienced realization will evolve and result in future contacts. It therefore deserves a central place in the study of opportunistic forwarding algorithms which, in lack of end-to-end connectivity and path information, can only try to understand the stochastic properties of the mobility process and use them to their advantage.

Briefly, the following fundamental properties of the mobility process underlying the network are relevant to opportunistic forwarding.

Mobility Intensity: An interesting property of a node's mobility is the magnitude of the surface area it traverses in a given amount of time. Intuitively, the larger this area, the more the contact (and thus forwarding) opportunities this node will have. On the downside, the shorter the duration of these occurring contacts as well. Mobility intensity is generally related to the *absolute* speed of the node and the frequency and duration of pauses.

Mobility Locality: This property relates the mobility of a node to the total network area. A given node may move within only a small subset of locations and never or infrequently visit the rest of the network. Recent studies of collected mobility traces reveal that most nodes do show such skewed location preferences, visiting a small percent of locations for a large percent of the time [32]. As a result, many recent mobility models attempt to reproduce this behavior by assigning "home locations" to each node [33, 6, 48]. A different flavor of mobility locality is related to the mixings properties of the model. Informally, a node may, in the long run, visit every location in the network, but it may still move locally with a slow drift towards other locations. A relevant metric is then the magnitude of *new area* it covers in a given amount of time. More formally, if one considers the mobility process as a Markov chain over all network locations[1], then mobility locality is related to the *mixing time* of this Markov chain [1]. As a simple example, a random walk model has much higher mixing time than random waypoint mobility in the same area. Consequently, the random walk mobility is much more local than random waypoint mobility.

Mobility Regularity: In addition to the number of locations (or percentage of total area) a node tends to visit, different models (or nodes) may exhibit different degrees of regularity in the *sequence* and/or *timing* of such visits. For example, a node may visit the same locations $A_1, A_2, \ldots, A_n$ every day, but in a random sequence each time. By contrast, another node may go to A_2 after A_1, to A_3 after A_2, etc., with high probability and/or at the same time of the day. Preliminary results, as well as an understanding of human routine, suggest that strong periodicity and pattern is to be expected [23].

Mobility Heterogeneity: In a number of real mobility scenarios, nodes may not only exhibit considerable structure in their individual mobility behavior, but also significant differences with other nodes' patterns. While all nodes may exhibit similar qualitative properties in a scenario, e.g. preference towards only a small subset of locations, they often differ in the actual choice of locations (e.g., two "nodes" working

---

[1] Note that we can still introduce location preference in this model.

or living at different parts of a city), as well as in mobility intensity, locality, and regularity.

Mobility Correlations: Finally, while not all nodes in a given mobility scenario are expected to exhibit uniform mobility characteristics, subsets of nodes will be subject to higher correlations in their preferred locations and visiting patterns. Consider, for example, classmates or colleagues, who tend to be frequently collocated due to their common working or studying environment, or friends, who deliberately choose to collocate frequently. Such correlations have been clearly observed in collected mobility traces [31]. The extent to which these correlations are a result of social relationships driving mobility actions or side-effects of location associations and geography is still a subject of research.

In general, the larger the amount of *average* node mobility, the better the performance of routing protocols that rely on such mobility. Furthermore, in a number of situations (e.g., uncorrelated mobility uniform across nodes), it holds that the higher the average node mobility, the less sophisticated the design of a protocol must be. This seems to be in contrast with the traditional viewpoint that node mobility has a negative effect on routing protocol performance.

11.2.3 Node Resources

Although network and node resources are becoming less and less of an issue in wired networks, it is not typically the case for their wireless counterparts. Depending on the application, network and node capabilities such as bandwidth, storage, and battery lifetime may vary largely. Resource availability or lack thereof should play an important role in the design and performance of a routing protocol.

Bandwidth: As cellular operators facing an exponential increase in data traffic would quickly confirm [21], in the wireless arena, the available bandwidth is always a valuable and often scarce resource. If bandwidth is limited, then routing protocols should be efficient, especially in terms of signaling and control information exchange. In addition, the more limited the available bandwidth, the more prudent the choice of forwarding opportunities must be.

Storage: Sensor networks are the typical case where available memory at nodes may be limited, relative to the amount of information that must be stored locally. Besides affecting the choice of the routing algorithm to be used, storage limitation also influences relevant routing protocol parameters (e.g., Time-To-Live (TTL) of a packet) as well as mechanisms such as buffer replacement policies and garbage collection [80, 63].

Battery Lifetime: Power awareness is usually an important feature in routing protocols for wireless networks[2]. In the case of DTNs, it becomes even more critical, especially in the case of deployments in remote, hard to access regions where nodes may be left unattended for extended periods of time. In order to minimize the energy waste in DTNs, optimal searching or probing intervals are calculated using statistical

---

[2] There are of course some notable exceptions, e.g., VANETs.

information of contact opportunities [38, 73, 2] and energy efficient sleep scheduling mechanisms are constructed [77, 9].

Heterogeneous Node Capabilities: Finally, similarly to the case of mobility behavior, nodes may also have largely varying resource capabilities. Vehicular wireless nodes (with little or no energy and storage limitations) may coexist in a DTN with small smartphones carried by pedestrians. In such a scenario, a routing protocol should be able to identify the more capable nodes, as they are possibly better candidates for relaying traffic than smaller, battery-operated nodes.

11.2.4 Efficient Opportunistic Forwarding: A Challenge and an Opportunity

From the above discussion, it is not difficult to see that, regardless of the specific implementation details and choices of different opportunistic forwarding algorithms (which we will see in more detail in the next section), all DTN routing schemes must deal efficiently with three key characteristics.

On the one hand, there is the uncertainty and stochastic nature of *future connectivity*. This stochastic nature implies that different realizations of the processes underlying the network (e.g., mobility) could lead to very different performances; the same sequence of forwarding decisions may result in highly variable performance for consecutive messages, with some of them not ever being delivered. To cope with randomness and guarantee good performance over a maximum of possible realizations, some type of redundancy (*coding*) is usually employed. In its simplest form, this means routing multiple replicas of the same message, in parallel. However, more advanced coding techniques can be used, as we will see.

On the other hand, *node mobility*, a key ingredient of the store-*carry*-and-forward approach, is not completely random. Communicating devices (laptops, smart phones, and even sensors) are usually carried by humans or vehicles. Humans exhibit considerable structure and predictability in their mobility decisions, as these decisions are guided by habit, social links, and locality [23, 31]. This also implies that node mobility is not uniform. Different nodes move to different locations, with different intensities, and meet different other nodes. Similar arguments (albeit different patterns) can be made for vehicular mobility as well. These characteristics imply a largely heterogeneous environment where specific nodes can be better “next hops” for a given destination or for many destinations.

Finally, whether coping with volatile connectivity or exploiting structure in mobility, *network and node resources* must be used efficiently. For example, excessive replication can easily saturate the medium as well as nodes’ storage units, while looking for the best relay may always end up choosing the same (highly mobile) nodes, resulting in fast battery depletion and node outage. Like mobility, node resources (e.g., battery, processing power, communication capabilities, etc.) may also be different among nodes in a DTN network (e.g. a simple mobile phone versus a wireless router installed in a vehicle). For efficiency, they should therefore be differently exploited.

Summarizing, the dynamicity and unpredictability of future connectivity, as well as the relative scarcity of network and node resources present a *challenge* that an opportunistic forwarding algorithm needs to overcome, while mobility patterns and heterogeneities present an *opportunity* that can be exploited to make more intelligent forwarding decisions.

11.3 DEALING WITH UNCERTAINTY: REDUNDANCY-BASED ROUTING

As explained in the previous sections, due to DTNs' sparse and often highly dynamic connectivity, no end-to-end path between a source and a destination can be known in advance. In addition, with the store-carry-and-forward model, a path is no longer a sequence of links, all existing concurrently. Instead, it may be described by two sequences: links $e_1, e_2, \ldots, e_n$ and some potential link traversal times $t_1, t_2, \ldots, t_n$, respectively. *If* these times are non-decreasing $t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n$, and *if*, for all i, link e_i is up at time t_i, then the two sequences form a *space-time path*. This implies that, when a message is fowarded over link e_i at time t_i, link e_{i+1} may not be up. In fact, unless future connectivity is both deterministic and known in advance (e.g., a scheduled satellite link), at time t_i, there is no guarantee that link e_{i+1} will ever be up.

In other words, a salient feature of the stochastic connectivity (e.g., due to mobility) environment prevalent in DTNs, is that exact space-time paths between node pairs can only be recognized *a posteriori* (after their realization). Concretely, when a node A carrying a message for destination D, encounters a node B, node A cannot know which of A and B will be the first to establish a direct (or multihop) link to D. Suppose A decides to forward the message to B, and A later encounters another node C. It may turn out that C actually meets D an order of magnitude sooner than B, or than any other nodes that B later encounters. In the worst case, where no knowledge about at least some statistics of future contacts can be assumed, uncertainty accumulates and different sequences of forwarding decisions may lead to vastly different end-to-end delivery delays.

In such a situation, where choosing one next hop over another does not seem to offer any guarantees or even simple improvement in the (expected) delivery delay, it is no surprise that the first proposed DTN routing solutions simply decide not to choose, but rather to *replicate* the message to any node encountered, in an attempt to *exploit in parallel all possible space-time paths leading to a destination.*

11.3.1 Flooding-based Schemes

Epidemic routing: The first DTN routing scheme using replication to exploit multiple space-time paths in parallel was *epidemic* routing [71]. Epidemic routing is essentially flooding or broadcasting, adapted to the sparse DTN environment with infrequent contacts. Periodic broadcast (e.g., as in MANET flooding) would be wasteful (too many transmissions with no nodes in range) and inefficient (contacts may occur between two broadcasts). As a consequence, epidemic routing works as follows. Each node maintains a *message vector* indicating which messages the current node is storing in its buffer. When two nodes encounter[3], they first exchange and compare their message vectors. Then they exchange *all* messages not in common, so that, at the end of the contact, both nodes are storing exactly the same messages. In this manner, whenever a node is carrying ("is infected with") a given message, that node transfers the message to any other node it encounters, hence the name "epidemic".

Epidemic routing takes the concept of exploiting multiple space-time paths in parallel to the extreme. On the positive side, it is easy to see that epidemic routing is *guaranteed* to find the *shortest* space-time path between any source and any destination (in terms of end-to-end delay), as it will follow *all* paths. On the other hand, this implies an immense overhead per message. Consider a network of 1000 nodes. If the shortest space-time path between two nodes is, e.g., 10 hops, epidemic routing routing will perform 990 wasted transmissions, in addition to the ones absolutely necessary for multihop routing. This essentially corresponds to an efficiency of 1%, a value that is unacceptable in most engineering applications. In the more general case, it is known that the shortest path in a network of N nodes scales as $\log N$ or slower (e.g., in "small-world" graphs). This implies that the overhead of epidemic routing grows to infinity as $\frac{N}{\log N}$ unless some special measures are taken.

What is worse, if node buffer space and contact bandwidth are limited, epidemic routing is no longer optimal. Message copies find full buffers and are dropped, and transmissions are delayed when not all intended messages can be transferred during a single contact. In fact, depending on the buffer management and scheduling policies, new messages may kick out old messages before the latter are delivered, leading to congestion collapse phenomena [67].

Fig. 11.2 shows a simulation-based comparison of epidemic routing and an optimal scheme, which has full knowledge of future contacts, and sends a single message copy over the shortest space-time path. On the left side plot, the large overhead of epidemic routing is apparent. The increase in transmissions with higher traffic is due to increased contention and retransmissions. On the right side plot, the effect of resource constraints on the delay of epidemic is shown. For low traffic, the delay is optimal; but, as traffic increases, the delay of epidemic routing diverges, due to contention and queueing phenomena.

As a result, all research in the field of DTN routing has since focused on achieving the optimal delivery delay and probability guaranteed by epidemic routing, while using network and node resources much more efficiently (i.e., reducing the overhead).

Reducing the overhead of epidemic routing: A number of schemes attempt to reduce the overhead of epidemic while still remaining flooding-based in nature. Haas and Small [25] import the immunization and vaccination concepts from epidemiology and apply them to epidemic routing. In the proposed *IMMUNE_TX* and *VACCINE* recovery schemes, after a destination receives a message, it propagates

---

[3] An encounter is established through a MAC layer beaconing process, which constantly looks for nodes in range.

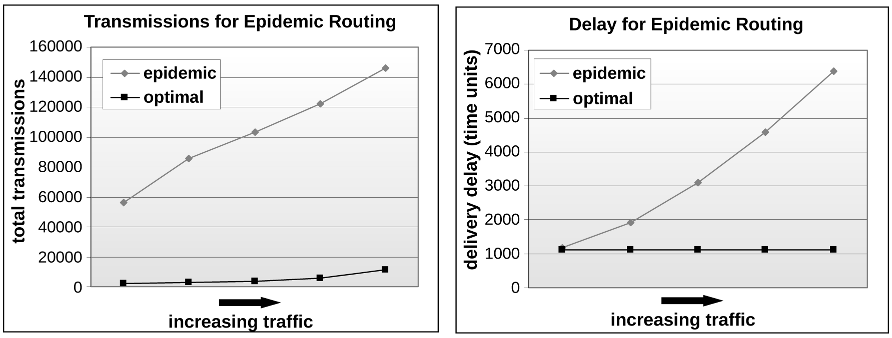


Figure 11.2 Simulation-based performance comparison of epidemic routing and an optimal, oracle-based scheme.

"anti-messages" to infected nodes and, respectively to all nodes. Since anti-messages have much smaller sizes (only message ID), the overhead is effectively reduced.

Gossiping or *Randomized Flooding* [80] will copy a message during a contact with a probability p less than 1, as opposed to epidemic routing where $p = 1$. This offers a knob to tune the "aggressiveness" of the message spreading protocol. Another way to achieve this could also be, for example, to allow each relay to copy the message to at most k other nodes [67]. While with very careful tuning of the replication probability p (e.g., as the function of TTL, number of nodes, etc.), similar delivery ratios with epidemic routing (albeit somewhat longer delays) could be achieved, this is very scenario-specific, and requires a priori knowledge of global network parameters. In practice, all these schemes result in almost every node ($\mathcal{O}(N)$) receiving a copy of each message, given a large enough message TTL.

An alternative way to limit replication, *Limited-time Flooding* [69], is to use a threshold on the epidemic routing time. Before the threshold timer expires, messages are spread using epidemic routing. When the timer expires, any node with the message may only transmit it directly to the destination. Similarly to gossip, careful selection of the timer is needed in order to achieve performance targets and overhead limits, making this strategy as well rather impractical. Limited-time flooding is however, very useful for analytical purposes.

Another proposal is to allow only the source of a message to create and forward replicas. A relay can then only forward the message to the destination. This is often referred to as the *2-hop scheme*. In a uniform mobility environment (independent identically distributed (IID) inter-contact times), the 2-hop scheme results in an average number of $\frac{N-1}{2}$ transmissions per message. However, only paths of at most 2 hops can be used, which can be quite restrictive (if not insufficient) in clustered, very local mobility scenarios.

Finally, while the above routing schemes indirectly control (to some extent) the congestion potentially caused by epidemic routing (by reducing the rate of replication), *SLEF* (Self-Limiting Epidemic Forwarding [18]) tries to directly deal with congestion. The basic mechanism of SLEF consists in reducing the number of hops each message is allowed to traverse, as a function of the perceived congestion. As a result, when very few messages compete for nodes' resources, messages can spread to the entire network, as in epidemic routing. On the other hand, if the number of messages spread in parallel is high, each message is only locally spread, within a few hops around the source. Obviously, an online mechanism to infer local congestion is necessary for SLEF to operate. Furthermore, while this scheme can be a sensible solution for applications like the broadcasting of location-related content (e.g., advertisements), it is not suitable for generic, end-to-end unicast or multicast.

11.3.2 Controlled Replication Schemes

A common underlying characteristic of the schemes discussed thus far is that the number of transmissions (i.e., copies generated) per message is not fixed and directly controllable. Instead, it is a function of the network size (i.e., the total number of nodes). The larger the network, the more transmissions per message, with this relation being in most cases linear. This raises important scalability concerns. To this end, controlled replication schemes were proposed that enforce a fixed (usually small) number of transmissions per message, independently of network size. Controlled replication is often referred to as *Spray and Wait* [65, 63].

The goal of Spray and Wait is to ensure that each message is delivered with at most L transmissions, where $L \ll N$. Different flavors of spraying achieve this goal with different policies during the first ("spraying") phase:

Source Spraying: The source of a message alone may create additional copies and forward them to encountered relays. This is similar to the 2-hop scheme. However here, the source stops after having distributed $L-1$ copies or sooner, if the destination was among the first $L-1$ nodes encountered. The source and the relays may then only forward the message to the destination. This latter phase is the "wait" phase.

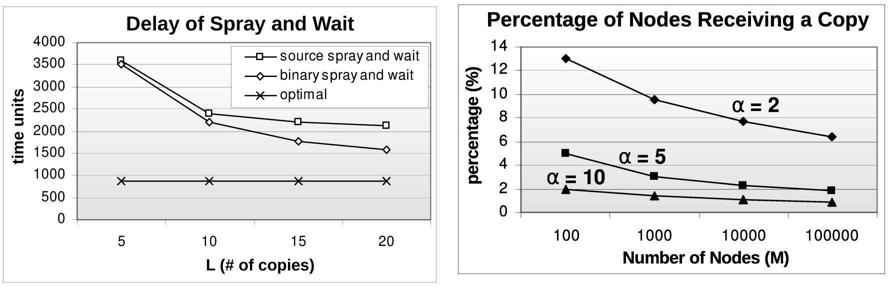
Binary Spraying: To speed up the spraying phase, relays can be also allowed to spread copies further. However, to ensure that the total number of copies remains $\leq L$, a quota system is implemented as follows: (i) the source starts with one message copy and a quota of L allowed replications; (ii) when a node (source or relay) with a message copy and a quota $i > 1$ encounters a relay without a copy, it forwards a message copy and half of the quota; in other words, after replication, both nodes have a copy and one has a quota of $\lceil i/2 \rceil$ and the other $\lfloor i/2 \rfloor$; (iii) a node with a copy and a quota of 1 may only forward the message to the destination.

It can be proven [67] that, under independent and identically distributed (IID) mobility (e.g., all nodes move according to the random waypoint model), binary spraying has the shortest delay among all quota-based spraying methods. However, source spraying may be preferred in some scenarios where only the source can be *relied on* to spread copies (e.g., in scenarios where relays may decide to drop their copy, and waste their assigned quota [57]).

In general, the performance of Spray and Wait largely depends on the mobility environment. Referring back to Section 11.2, if the mobility model is characterized by high mobility intensity and low locality, Spray and Wait can achieve similar delays to epidemic routing (e.g., a slowdown of only $2\times$) with almost an order of magnitude fewer transmissions. Fig. 11.3(a) shows the delay of both Binary and Source Spray and Wait as a function of the number of replicas used. As can be seen there, adding a few extra replicas quickly brings the delay close to the optimum, even in this small network. Furthermore, the performance increase, when using more and more replicas, obeys a *law of diminishing returns*.

Another important property of Spray and Wait is that, for IID uniform mobility, one can analytically solve for the number of copies to achieve a desired performance compared to the optimal (epidemic under no resource constraints). This turns out to be *independent of the actual mobility model and independent of the size of the geographic area in which network nodes are moving.*. In other words, one can easily choose the number of copies to achieve a delay α times the optimal ($\alpha \geqslant 1$) only as a function of the number of nodes N [67]. Different online distributed estimation algorithms could be used to obtain N if it is not known a priori.

Finally, the relative performance improvement of Spray and Wait against epidemic routing, actually increases with network size, making a strong case for scalability. Fig. 11.3(b) depicts the required number of replicas (as a percentage of the total number of nodes) in order to achieve a delay at most α times that of epidemic routing. While for a small network of 100 nodes, about 13% of the nodes must act as relays for the delay to be bounded by $2\times$ the optimal, for a larger network of 10000 nodes, this percentage drops to less than 8%.



(a) Delay of Spray and Wait on a network with 100 nodes performing random walks on a 100 × 100 lattice.

(b) Number of replicas L, as a percentage of the total number of nodes M, needed to achieve an end-to-end delay of α times the optimal delay.

Figure 11.3 Spray and Wait: delay–number of replicas relationship

While controlled replication algorithms excel in ideally uniform, non-local mobility models, their performance can drop rapidly when nodes show strong location preference, correlated mobility, etc. Consider for example a scenario where clusters of n nodes each are assigned to different, non-overlapping home locations. From each cluster, only a single node moves around the whole network, while the rest of the nodes move only locally inside their home location. Assume now that a message generated in cluster X is destined to a node in a different cluster Y. Assume further that Spray and Wait with L copies is used. It is easy to see that, even with an infinite TTL, only $\frac{L}{n}$ will be delivered. By increasing n, the delivery ratio of Spray and Wait in this scenario can become arbitrary low. In this case, an efficient forwarding algorithm would try to discover and use the node in the source's cluster who also visits the destination cluster, and avoid wasting copies to "local" nodes, which have very similar mobility pattern with the source. We discuss such algorithms in detail in Section 11.4.

11.3.3 Coding-based Schemes

Replication algorithms like the ones described thus far attempt to route a message over multiple space-time paths in parallel, to ensure that one of the will reach the destination soon. In other words, they attempt to increase the *expected performance* (e.g., delay or delivery probability). Even if one copy is wasted on a path never leading to the destination, there is a low probability that all of them will[4]. However, replication schemes are only one (simple) type of *coding*. It is well-known that they are not necessarily the most efficient coding schemes. More sophisticated codes can be devised, to efficiently generate the necessary amount of redundancy, in order to cope with the few or many (depending on the scenario) bad space-time paths. This is somewhat analogous to the space-time codes used in wireless communication [70]. As a result, coding approaches have also been proposed for DTN routing.

Source Coding: Even if the expected delay for a given scheme (e.g., Spray and Wait) is adequate for some applications, performance might still exhibit high *variance* with some messages being delivered with very short delays and other with rather large ones. Alternatively, if the performance metric is not delay, but rather the number of messages being delivered within some Time-To-Live (TTL) after which the message is considered stale, then high variance in the delay, from a message to the next one, may be undesirable. Source coding aims at increasing delivery reliability and reducing worst-case delay. A notable example is *erasure coding* [74], in which the coding is performed by the source, a coded part of a message is further treated as any other message in the network, and there is no specific implications on routing and forwarding.

A variation of source coding known as *Distributed Source Coding* (DSC) tries to minimize the propagation of redundant information in the network, and thus reduce overhead. Sensor networks, which are aimed at a variety of monitoring applications (e.g., environmental and habitat monitoring), are the typical target scenario for DSC [78]. The basic idea behind DSC is to take advantage of the data's inherent spatial and temporal locality to suppress propagation of unnecessary information. For example, in a sensor network tasked to measure the temperature field of a given region, nodes that are in close proximity to one another are expected to report similar temperature values. Through DSC strategies, nodes can identify such redundancies and perform *in-network aggregation* to reduce the volume of data transmitted in the network [64]. Another example of DSC are growth codes [39], which use coding redundancy at neighbors to avoid the impact of loss.

---

[4] Obviously, we are assuming an IID mobility process. If this is not the case, then all copies *could* in fact take bad, correlated paths, as in the toy mobility example described above.

Network Coding: Network coding has been proposed as a way to increase the capacity of wireless networks [76, 40]. The main idea behind network coding is to allow mixing of messages at intermediate nodes in the network. In this way, a receiver reconstructs the original message, once it receives enough encoded messages. Network coding is shown to achieve maximum information flow in a network, which is not attainable with traditional routing schemes.

Linear network coding has been shown to achieve the capacity of information networks [49]. In this coding scheme, nodes can apply a linear transformation to a vector (a block of messages over a certain base field) before passing it further in the network. It can be used to reduce the time to deliver a given flow, maximize the throughput, reduce the number of transmissions (and thus energy expended), etc.

Random network coding, where coding coefficients are chosen by each node randomly from a large enough field (often $\mathbb{Z}^8$) and in a distributed manner, is an efficient method to implement network coding in practice (coding coefficients are sent as part of the packet, with only a small overhead) [15]. To take advantage of the benefits of network coding in a wireless, often "challenged" environment, the following modification of greedy replication have been proposed [76]: instead of transmitting single packets, linear combinations of packets are generated and transmitted; assume a node A has a set of linear combinations of N packets $S^{(A)}_{\text{msg}} = \{\hat{m}_1, \hat{m}_2, \dots, \hat{m}_m\}$ and encounters another node B. Then, it creates a linear combination of all its messages in the queue:

$$\hat{m}_{new} = \sum_{i=1}^{m} c_i \hat{m}_i. \tag{11.1}$$

Here, the addition is *modulo* the given base field chosen for network coding. When enough independent combinations ($\geqslant N$) of the N messages, belonging to a given coding generation, have been received, a node can *decode* them to get the original N messages.

One key problem with the network coding approach described above is that coding *every* single message together may result in never collecting enough independent combinations of messages to successfully decode, especially when the network is sparse or when the nodes' degree is low. Some control is needed on how many and which messages will be coded together. This is known as generation control. Coding messages from many different sessions and from large time or sequence number windows (large generations) might result in high delivery delays. On the other hand, using small generations limits the amount of gains achievable by network coding. Finally, even controlling the generations in a distributed manner, may pose significant challenges. For these reasons, it has been suggested to implement network coding hop-by-hop, in an *opportunistic* fashion [40]. Opportunistic network coding simply takes advantage of favorable traffic patterns to locally save some transmissions, without requiring any generation control or imposing additional delays, but its performance still suffers in very sparse networks.

11.3.4 Discussion of Replication-based Forwarding

All the schemes we have discussed in this section are *randomized* replication schemes. They do not differentiate between nodes in the network. All nodes are equal, and some of them are chosen *randomly* (e.g., the first ones encountered) to act as relays. The only difference among the various schemes is the sheer number of relays (ranging from all nodes to a small subset of them), and the deterministic or probabilistic decision to replicate on an encounter. Coding simply allows the propagated messages to be coded packets rather than raw messages.

Randomized replication can be a good policy when node mobility is IID, that is, statistically all nodes move in a similar manner, but independently and only differ in the particular sample paths realized. Randomized replication may also be appropriate when the mobility pattern is unknown and unpredictable. One such example is when the mobility process is non-stationary, with key statistics and behaviors changing faster than an online estimation and learning algorithm could infer and exploit them.

Nevertheless, as we have stressed earlier, human mobility (and mobility in general) exhibits significant structure and predictability. Different nodes may be significantly better next hops than others for some destinations, or even all destinations. What is more, with more or less local observations/measurements, one can often infer these nodes. Because scenarios like these, with skewed location preference, communities, and heterogeneous mobility behaviors, are the rule rather than the exception in real mobility scenarios, sophisticated schemes have been proposed that attempt to infer and exploit whatever pattern and heterogeneity might exist. These are the subject of the next section.

11.4 CAPITALIZING ON STRUCTURE: UTILITY-BASED FORWARDING

Replication-based schemes, presented in Section 11.3, route multiple message replicas in parallel to combat uncertainty in future contacts. The replica bearers (relays) for each replica are randomly chosen in all these schemes, regardless of node-specific mobility patterns or resource constraints.

Nevertheless, as explained earlier, mobility is often highly structured, exhibiting some amount of predictability. Mobility is also characterized by heterogeneity among most nodes, but also correlations within small subsets of nodes. Knowledge about generic properties of the model, and more specifically about the mobility characteristics of individual nodes, can be an important guideline in making smarter decisions than purely random forwarding.

As a result, a number of sophisticated opportunistic forwarding schemes are based on the following basic functions:

(a) collect and analyze some statistics about past contacts among (all or a subset of) nodes;

(b) assign a utility for each node based on these statistics; this utility may be *destination-dependent* or *destination-independent* and aims at quantifying the ability of a candidate next hop to deliver the message probabilistically closer to a destination (itself, or through subsequent intermediate nodes).

(c) perform a deterministic or probabilistic decision as a function of the current relay's and the candidate relay's utilities (and perhaps additional parameters).

Such algorithms are often referred to as *utility-based* schemes. While multi-copy algorithms (i.e., like the ones in Section 11.3, using more than one replica per message), could also utilize a utility-based mechanism, in this section we will focus on *single-copy* schemes. Utility-based single-copy algorithms do not spawn additional copies of a message. When the node currently holding the message (the source or a relay) encounters another node, the utility of the two nodes (often with respect to a specific destination) is evaluated. If the new node has a higher utility than the current one, the message is handed over and no local copy is retained[5].

We note that some of the utility-based forwarding schemes we will discuss in this section were, in fact, initially proposed as multi-copy schemes. Nevertheless, a single-copy version of all such schemes can be defined and we will use this to isolate our discussion from redundancy-related issues. In Section 11.5, we will turn our attention to more sophisticated schemes, employing both replication and utilities.

Various parameters differentiate two nodes and can be used in calculating their respective utilities. These parameters can be broadly categorized to *contact-related* (i.e., related to the mobility properties of nodes involved) and *non-contact-related* (e.g., related to node resources, social relationships, etc.).

11.4.1 Contact-based Utility

Mobility plays a key role in DTNs, both as an *enabler* (in the store-carry-and-forward paradigm) and *differentiator* between nodes' future contact probabilities. Consequently, a very large number of opportunistic forwarding solutions attempt to collect and process information about past contacts (e.g., between a given relay and all other possible destinations) and derive some useful statistic/predictor about them (e.g., expected inter-contact time).

11.4.1.1 Pair-based Contact Utilities A number of utility-based schemes proposed are optimized around the contact properties of *individual node pairs*. In most cases, the node pairs of interest are formed of an intended destination and a node being evaluated as a candidate relay for that destination. Different properties of the contact process between a relay and a destination can and have been considered. We discuss some of them here.

Age of Last Encounter: One of the first utilities to be suggested was the time elapsed since two nodes last encountered each other (i.e., were in range and aware of it) [17]. In fact, the original proposal was targeting MANET environments and was arguing that this time contained indirect location information. Because nodes

---

[5] One notable exception is the source node, which may retain a local copy even in single-copy schemes, in order to implement an end-to-end (transport-layer) retransmission mechanism.

tend to move in a continuous manner (i.e., they do not perform jumps in space), a smaller timer value often implies a smaller distance to the destination, assuming that the average speed of nodes does not vary too much. In addition to the last encounter time, a node can record its encounters with another node by noting the position at the time of encounter as well [24]. This can be useful in predicting the destination node's current location, even though a past encounter is no guarantee for future encounters with the destination. In DTNs, when nodes are heterogeneous in terms of their characteristics and capabilities, additional parameters should be used in combination with the age of last encounter, in order to choose a suitable relay node. Furthermore, note that if node mobility is relatively uniform, the age of last encounter only offers (some) benefits in denser connectivity environments.

The relation of the age of last encounter to the *residual time* until the next encounter is known to depend on the *inter-contact* process. If all pairwise inter-contact times are drawn from the same probability distribution, then different processes have different implications for routing. For example, if inter-contact times are exponentially distributed, knowledge of the age contains no information about residual times and is thus irrelevant for making a forwarding decision (i.e., the protocol would degenerate to random forwarding). If, however, the inter-contact process is heavy-tailed, then a longer age implies a longer residual time, making age information relevant [26]. While inter-contact times and their distribution in real mobility scenarios have been very extensively studied [8, 32, 11] and are still the subject of interesting new insight [59], we are not aware of any opportunistic forwarding scheme explicitly using the age-residual time relation as a function of inter-contact time distribution.

History of Past Encounters: The age of last encounter utility only takes into account a single past contact. An opportunistic forwarding scheme could choose to keep track of a longer history of past contacts and their statistics. Out of this measured data for a node pair's contacts, it can then derive different utilities. One such option is to maintain an estimate of the ***frequency*** of encounters (or inversely, the mean inter-contact time) between two nodes. If a node meets the destination frequently, it can be a reasonable relay. A utility function that takes into account both the age of last encounter and frequency is proposed in PRoPHET [50]. While the original protocol is a modification of epidemic routing (and thus flooding-based), the proposed utility can still be considered for single-copy forwarding. Another interesting property of past contacts is the average ***contact duration***. While frequent contacts may be important to ensure a short delay (for small messages), frequent but very short contacts may be quite useless when large amounts of data must be transferred or the node association process is long and wastes a large chunk of the contact duration (as is the case, for example, in 802.11).

Figure 11.4 shows a scatterplot of contact duration and contact frequency for all node pairs in a collected mobility trace. While there seems to be a correlation between the two metrics, there are pairs whose contact duration is much stronger than the contact rate and vice versa. In order to provide a single *scalar* utility for each pair, that combines information about both contact rate and duration, the *principal component* over all $\{$rate,duration$\}$ tuples can be used [31]. The direction of the

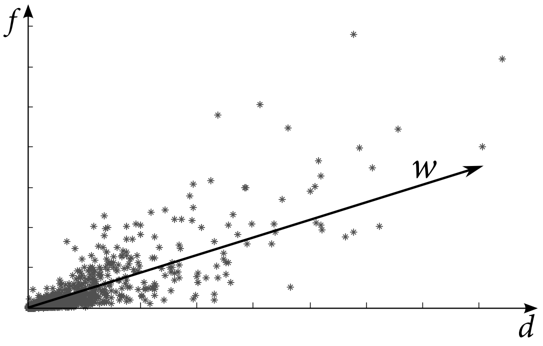


Figure 11.4 Scatterplot of contact duration (x-axis) and contact frequency (y-axis) for different node pairs in a collected mobility trace. Vector w corresponds to the principal component direction and w_{ij} is the PCA utility for pair $i \leftrightarrow j$

principal component is shown as **w** in Figure 11.4 and the utility for each node pair is the projection of the $\{$rate,duration$\}$ vector on **w**.

More Complex Pair Contact Predictors: More sophisticated utility functions or predictors can be built based on past contact history. A Kalman filter could be used for more accurate predictions in case of highly structured contact processes. For example in [52], the authors use a Kalman filter to predict the future utility (i.e., delivery probability) of each relay, based on each past reported values. If delivery probability is calculated as a function of contact properties (frequency, age, duration), then this method offers a more sophisticated way for finer grain prediction. Contrary to this, most of the aforementioned methods are simple, first order auto-regressive predictors. Finally, higher moments of contact statistics (e.g., inter-contact times) could be used. For some applications, a relay with slightly less frequent, on average, but highly regular contacts with the destination could be preferable to a relay with smaller average inter-contact times but higher variance.

Pattern of Visited Locations: In the real world, mobile users move with certain purposes in mind (e.g., going to work, going to a class, going from work to lunch, etc). Additionally, they may follow specific paths in between these locations due to geographical constraints. As a result, people tend to follow a *movement pattern* in their daily activities. These patterns are a function of a variety of parameters including professional activity, work and home location, etc. What is more, most people also tend to spend the majority of their time in a small subset of *preferred* locations, as opposed to indiscriminately roaming everywhere (unless, this is part of their job, e.g., taxi driver, salesman, etc). *Location preference* as well as the periodic nature of human mobility (diurnal and weekly patterns) have been consistently demonstrated in a variety of real mobility traces [29]. Mobility patterns (known a priori or learnt online by collecting appropriate statistics) could help identify a *profile* for a given node; nodes with a mobility profile matching or similar to the destination can be considered good candidate relays for messages to that destination [47, 46, 22]. While this method does not directly measure (and match) contacts, the profile of locations visited is used as an indirect measure of past contacts (and thus future contact probabilities) between two nodes. We therefore include this method here, in the pairwise contact-based utilities.

Maintenance and Overhead of Contact Statistics: Keeping track of more detailed information about past contacts could help identify more accurately good candidate next hops. On the other hand, keeping more information about encounters increases the overhead in terms of context data that needs to be stored. Another consideration is how long to keep this history about a certain destination at a node as it may not be useful, or even misleading after a certain threshold of time depending upon the dynamics and mobility pattern of participating nodes. As a result, sliding windows or exponentially weighted time averages (EWMA) are more often used.

11.4.1.2 Contact Graph Utilities All forwarding schemes discussed so far in this section only consider pairwise contact metrics to identify the utility of a relay (e.g., the contact frequency and/or duration of a candidate relay and a destination). While sophisticated protocols have been proposed based on pairwise properties, mobility patterns exhibit significant complexity and correlations between *subsets of nodes*. These correlations as well as any macroscopic mobility patterns cannot be (easily) captured using pairwise contacts. As a very simple example, a given node X may be a good next hop for a destination D, even if X rarely meets D. This may be the case, for example, if X meets another node Y often, and Y meets D often. As a second example, X may meet many nodes in general (even if not D itself), increasing thus the chances that it will soon meet nodes that do meet D often.

These patterns and correlations are not visible in the instantaneous connectivity graph, which is sparse and changes fast over time. In the case of pairwise contact metrics, the statistics over many past connectivity snapshots (between two nodes) were collected (e.g., all past contacts during a time window) and *aggregated* into a single scalar value (e.g., average inter-contact time, total contact duration, etc.). To visualize, understand, and exploit more complex mobility patterns, it has been proposed to aggregate complete connectivity snapshots (the instantaneous connectivity graph over different time instants) into a single static graph. This graph is often referred to as the *contact graph* or *social graph*. The reason for the latter name is, on the one hand, that this graph captures long-term behaviors (habits), often stemming from social behaviors, and on the other hand, because these graphs exhibit complex structure typical to the field of *complex networks* and *social networks* [54].

There are two key steps when designing an opportunistic forwarding algorithm utilizing properties of the contact graph.

1) Create the contact graph out of a sequence of past (instantaneous) connectivity graphs.

2) Use contact graph properties to compose a utility function that efficiently identifies “good” next hops.

A large number of different proposals exist for Step 2. Utilities based on contact graph paths (essentially space-time path probabilities), centrality metrics, community

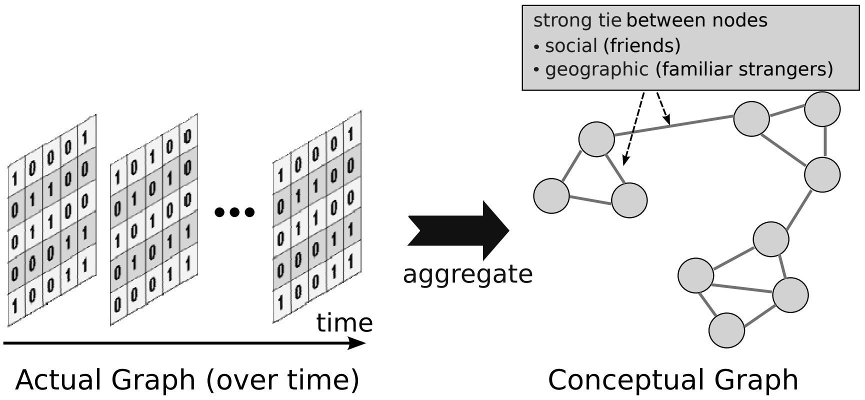


Figure 11.5 Aggregation of a sequence of (instantaneous) connectivity matrices into a single “social” or “contact graph”.

membership, etc have been proposed. Contrary to this, much less attention has been given to Step 1, the creation of the contact graph and the implications and information loss of the chosen methodology.

Contact Aggregation: Figure 11.5 depicts the problem of contact aggregation (i.e., contact graph creation). A sequence of binary matrices **A(t)** corresponds to the connectivity at each time instant t (assume t discrete for simplicity). Based on the matrices' ij entries, **A(t)**$_{ij}$ (in consecutive snapshots), we need to decide whether to include a link, and possibly a weight, within the contact graph between nodes i and j. The contact graph is undirected, and it can either be a weighted or an unweighted graph.

Weighted graph. A scalar weight w_{ij} is derived as a function of the **A(t)**$_{ij}$, for some past time window (e.g., $[t_1, t_2]$). That is, $w_{ij} = f(\mathbf{A}(\mathbf{t_1})_{ij}, \mathbf{A}(\mathbf{t_1+1})_{ij}, \ldots, \mathbf{A}(\mathbf{t_2})_{ij})$. This function normally aims to capture the *strength* of the contact process between i and j, that is, the future contact probability between i and j.

Unweighted graph. If the contact graph is unweighted, then a link may either exist (implying a high future contact probability) or not exist (implying this node pair's link is not that useful in the routing process). This could be achieved for example by introducing a cutoff threshold for weights: if the link weight w_{ij} is below this value, then it is removed from the contact graph; if it is higher, then a link (with no weight) is included.

While a weighted graph contains more (and more specific) information, it is also much more cumbersome to process (e.g., to derive utility metrics), especially in an online fashion. If for example, we are considering a network of $10\,000$ nodes, this implies a $10\,000 \times 10\,000$ contact graph matrix (complete mesh), and 10^8 link weights. Inversion, spectral analysis, etc become considerably slower. It is often the case that only a small subset of contact node pairs have significant weight and are useful for routing, and the rest can be ignored (at least for non-flooding protocols). This implies that a very sparse $10\,000 \times 10\,000$ binary matrix can be used instead, for the unweighted graph. However, an important problem in the unweighted case is the choice of aggregation threshold. While this is usually done with preselected or empirical values and window sizes, [30] shows that the choice of threshold should be done carefully, as there are often more wrong choices than right ones, potentially producing misleading results. The same work proposes an efficient algorithm to choose this threshold in an automatic "blind" way, under mild assumptions.

Contact graphs for numerous synthetic mobility models, and collected mobility traces have been studied [30, 31, 34]. There are some key properties that seem to underlie many, if not most, mobility scenarios:

Community structure. Contact graphs seem to exhibit considerable community structure with subsets of nodes well-connected to each other, with fewer or weaker links between subsets (or communities).

Small world. Contact graphs exhibit small world properties, namely very short paths between any two nodes usually exist. This implies the existence of short space-time paths. However, it does not imply that these paths can be easily found.

Skewed degree distribution. Contact graph weight distributions and node degree distributions exhibit considerable heterogeneity.

Having discussed how to create the contact graph, we now turn our attention to the type of contact graph properties used as utilities for opportunistic forwarding.

Centrality-based Utility: Node centrality is a metric that has been considered for DTN routing. The *betweeness centrality* of a node i is defined as the number of shortest paths between any network nodes going through node i. It has been argued that nodes with high betweeness centrality can serve as "bridges" between communities relaying the messages from the community the source lies in to the community the destination lies in. Nevertheless, betweeness centrality cannot easily be calculated as it requires global network information. SimBet [13] approximates it using *ego-centrality*.

Another centrality metric that can be *locally* calculated is *degree centrality*. Degree centrality is essentially the degree of the node in the contact graph (or the sum of link weights, in the weighted case). Degree centrality is essentially related to the amount of mobility of a given node (see Section 11.2) and the total rate of meetings of that node with all other network nodes. In other words, high degree centrality implies a node that moves a lot and meets lots of other nodes within a window of time. Degree centrality is used in smart spraying schemes like [69, 53], as well as a utility in delegation-based forwarding [20]. It is also used in the second explicitly "social" protocol, BubbleRap [34].

Similarity-based Utility: Another contact graph property of interest in DTNs is the similarity between two nodes (also known as *structural equivalence* in the fields of complex and social networks [54]). Unlike utility metrics considering contacts between only a node i and a node j, two nodes are *similar*, if the have a lot of common neighbors in the contact graph. This, in turn, implies that two nodes with high similarity might also belong in the same community and serve as good relays of each other directly or through one of their neighbors. An additional reason why similarity is important is because contact graphs are based on (slowly) collected statistics and may be incomplete or obsolete. Consequently, a weak link between i and j may be merely a sampling artifact, if the two nodes exhibit high similarity nonetheless. Similarity is also used in the SimBet protocol [13]. In a slightly different manner, BubbleRap [34] identifies communities on the contact graph explicitly and assumes two nodes are similar (and thus good relays for each other) if they belong in the same contact graph community.

Complete Social Network Analysis-based Schemes: Two important opportunistic forwarding schemes have been recently proposed, that stipulate the contact graph approach and use it to design utilities that seem to outperform existing DTN schemes, at least in the scenarios considered. SimBet uses a per node utility that takes into account both the similarity of a given relay i with the intended destination d (denoted $Sim_i(d)$ here), as well as the ego-centrality of the same relay, Bet_i. A utility $U_i(d)$ is then defined as: $U_i(d) = \alpha Sim_i(d) + \beta Bet_i$. The original proposal is to weigh the two values linearly, with arbitrary weights. The underlying idea of the protocol is to utilize bridging (high betweeness) nodes to push the message outside the source's community. Then, similarity is used to "home in" to the destination's community.

BubbleRap uses an approach to routing similar to SimBet. Again, betweenness centrality is used to find bridging nodes until the content reaches the destination community. Communities are explicitly identified by a community detection algorithm, instead of implicitly by using similarity. Once in the right community, content is only forwarded to other nodes of that community: a local centrality metric is used to find increasingly better relays within the community.

Probabilistic Path-based Utilities: As mentioned earlier, a given node i may be a good relay for a destination d, not because it meets d frequently but because it meets another node j that meets d. Taking this further, node i may be a good next hop because it is in the beginning of a space-time path that has a high chance of realization. Unfortunately, these cannot be captured by a utility accounting only for individual pair contacts and their statistics. Intuitively, this means that contact utilities may have some transitivity properties that should be considered. A number of protocols have been proposed, that considered the probability of future realization of complete (or partial paths) as opposed to individual contacts. While contact graphs were introduced later than these and the papers themselves do not explicitly refer to contact graphs, they best belong in this section, as they define path metrics based on the link metrics on the contact graph, implicitly or explicitly.

PRoPHET [50], discussed earlier, does not only consider the contact frequency and age between a relay and a destination. It also introduces a transitivity component, so that the utility of a relay is increased when it meets often with another relay that has a high utility for the destination. Path metrics are also considered in [12]. Finally, one of the simpler policies introduced in an early DTN paper [35], MED (minimum expected delay), essentially assigns the link weight (in the contact graph) to be the expected inter-contact time between two nodes. A path metric is then composed of link delays to obtain a path delay and a normal routing algorithm on the contact graph can be used to obtain such paths and the best next hop. MEED (minimum estimated expected delay) [36] is a proposal for a practical implementation of MED, since the properties of contacts for "far away" nodes are not known a priori and reliably.

11.4.2 Non-contact-based Utility

In addition to mobility related properties, a number of other node characteristics can be considered when making an opportunistic forwarding decision. These may include node resources, social relations, security and trust related parameters, user interests and geography.

Node Resources: When forwarding a message to a node, the resources and capabilities of that node should be considered. Even if a certain node has some ties to the destination (e.g., close friendship), giving a message copy to that node might be a waste of resources, if it is almost out of battery. Chances are it will either turn itself off or run out of battery before it gets a chance of delivering the message. Similarly, if a candidate relay has its buffer almost full, it might be more prudent to prefer another node instead. This may not only result in smaller queuing delays, but may also reduce the probability of the message getting dropped later. Consequently, nodes should maintain the current status of their resources, which can be used to identify nodes that are "good" (or "bad") relays independent of the destination.

In this direction, one research thread considers DTN routing from a resource allocation point of view. The idea is to forward or replicate a message to a relay, based upon the available resources in order to maximize the likelihood of message delivery, when two nodes meet. RAPID [3, 4] is the first protocol which treats DTN routing as a resource allocation problem. Follow up work improving the utility and proposing an efficient distributed implementation method can be found in [42, 43]. In these protocols, utilities are defined based on the total buffer occupancy per message. Messages then are ordered with respect to their utilities, keeping in view the goal of optimizing specific quantities (e.g., delay), which allows computation of desired performance metrics such as worst-case delivery delay and packet delivery ratio. The protocol translates a routing metric to per-packet utilities and at every transfer opportunity, it is verified if the marginal utility of replication justifies the resources used.

Erramilli and co-authors [19] have studied the idea of prioritizing messages to better manage network resources in a resource-constrained environment. They have used delegation forwarding [20] as their forwarding algorithm. Another protocol using the resource allocation concept is ORWAR (Opportunistic Routing with Window-Aware Replication) [60]. ORWAR differentiates among messages in function of their utilities and allocates more resources to high utility messages. Utilities are expressed as "utility per bit", such that they can be used to optimize for buffer space and bandwidth. ORWAR replicates messages in the order of high utilities first, and drops messages in the reverse order, if needed. This is also a replication routing scheme, but the replication decision depends on pre-estimated available bandwidth values and the number of allowed replicas per message depends on the message utility.

Social Relations: Humans are involved in complex social relationships (networks). As a consequence, people who are socially-related to each other (e.g., friends, students in the same class, and colleagues in the same department) are expected to interact more often with each other. These social features have important implications for networks formed by communication devices operated or carried by humans (e.g., vehicles, PDAs, laptops). Knowledge about existing social links may allow one to choose a data relay that has a much better chance of encountering the destination soon. Some recent schemes propose to use explicitly social properties for routing [7, 51].

Other Information: Finally, additional information can be relevant for routing. Geographical information such as the home city or postcode could be used, as well as other user profile information [5, 55]. Also, the willingness or trustworthiness of a node might be an important factor to consider, to avoid relays that might later drop the assigned replica(s).

11.5 HYBRID SOLUTIONS: COMBINING REDUNDANCY AND UTILITY

Utility-based forwarding schemes can be very efficient in discovering the right relays, and considerably improving the quality of forwarding decisions (compared to random). In mobility scenarios with enough structure and heterogeneity, such schemes can collect contact statistics locally, exchange these regionally or globally, and apply sophisticated machine learning, time-series, and complex network analysis-based algorithms to infer patterns and predict future contacts.

Nevertheless, the DTN environment remains stochastic. While forwarding decisions may be better than random (i.e., providing a guaranteed increase in future meeting probability), this is still a *probability*. Even good forwarding decisions in a large network and far away from the destination may choose relays who still have relatively low (even if slightly better than the previous hop) delivery probability. As a result, even the most sophisticated utility-based algorithms discussed in the previous section are not guaranteed to provide good performance for all messages routed. Uncertainty of future contacts remains a fact, and “betting all your money” on a single (albeit promising) space-time path can result in poor performance. To this end, current state-of-the-art schemes combine the power of replication and utility-based forwarding to achieve good and robust performance in numerous mobility environments.

There are three main flavors of this combination: (i) flooding-based schemes, that spread the message only to nodes with higher utility; the number of replicas is not explicitly limited; (ii) spray schemes, that start by spraying a fixed number of copies (e.g., binary spraying) and then route each copy further, using a utility-based forwarding policy; (iii) smart replication schemes, where an explicitly limited number of replicas is used, but each of them is, from the beginning, only handed over to “appropriate” relays, instead of handed out randomly (e.g., to whomever is encountered first).

11.5.1 Utility-based Flooding

Epidemic routing is efficient in exploiting all possible paths, including the best one. Yet it comes with an immense overhead in medium and large networks. To achieve similar performance, yet use much fewer space-time paths per message (and thus fewer resources), a number of proposals exist for *utility-based flooding*. A utility is defined and maintained for each pair of nodes in the network. Each node i maintains a value for the utility function $U_i(j)$ for every other node j in the network. If a node i carrying a message copy for a destination d encounters a node j with no copy of the message, then a new copy may be created and forwarded to j, depending on its utility towards d. Two types of utility-based forwarding rules can be used.

Rule 1) **Absolute utility criterion:** $U_j(d) > U_{\text{thresh}}$, for some U_{thresh} threshold value.

Rule 2) **Relative utility criterion:** $U_j(d) > U_i(d) + U_{\text{thresh}}$.

Some of the existing utility-based flooding proposals are the following. PRoPHET [50] is a utility-based flooding scheme, whose utility has been described in Section 11.4. In principle, PRoPHET's $U_i(d)$ has the following properties:

- it increases when i meets d,
- it decreases with time, when i is not in contact with d,
- it increases when i meets another node j, with a non-zero utility for d.

The increase and decrease rates as well as their weights are empirically chosen. There is no further analytical study or understanding of their exact effect.

BubbleRap [34] is also flooding-based in design. Using the contact graph, communities and the nodes contained in each community are identified online, through a community detection algorithm. Then, when a message replica is outside the destination's community, a potential relay is evaluated based on its betweenness centrality as the utility. This relay utility is used to decide on creating and forwarding one more replica or not. Once a message replica has reached the destination's community, it is only forwarded to other members of that community: a local centrality metric (degree centrality) is used to find increasingly better relays within the community.

11.5.2 Spray and Utility-based Spraying

As mentioned in Section 11.3.2, controlled replication or spraying algorithms excel in uniform and high mobility environments. However, in scenarios with local mobility and heterogeneity, all copies may get stuck with the wrong relays (e.g., nodes in the same community or location as the source). To cope with such scenarios, a source could spray the limited budget of copies quickly after message creation, and then allow each copy to be further forwarded (handed over, not copied) using an appropriate utility-based scheme.

Spray and Focus [66, 67] performs binary spraying of L copies, as in the Spray and Wait case. However, after the replication quota for a relay node reaches 1, it can still hand over its copy to another, better relay, if it encounters one. The utility used in Spray and Focus is a simple pairwise contact utility, similar to the one in

PRoPHET [50]. Different versions with and without utility transitivity have been tested.

While SimBet [13] was originally proposed as a single-copy scheme, it was later improved with a controlled replication component [14]. There, a small number of copies is generated and distributed to encountered relays. Then, each of these copies is routed independently according to the basic SimBet utility function described in Section 11.4.

11.5.3 Smart Replication

While quite efficient, the above hybrid schemes do not directly control, nor can they predict the total number of transmissions per message. This is an often undesirable feature, as depending on the mobility properties, such multi-copy schemes can become unstable and thus unscalable as the number of network nodes increases. In order to maintain the advantages of controlled replication (fixed number of copies, and thus resource usage, per message) *and* exploit the patterns and heterogeneity of real mobility environments, *smart replication* schemes were proposed [68, 53].

[68] uses explicit “labels” or a degree centrality estimate (by measuring the number of unique nodes met during a time window) as the utility. Then, binary or source spraying is employed, with copies forwarded only to relays that either have a higher utility (*Rule 2* above) or have a high enough utility (higher than a threshold – *Rule 1* above).

Encounter-Based Routing (EBR) [53] is another example of controlled, utility-based replication, in which the future rate of node encounters is predicted using a moving average of the number of past encounters. An encounter metric is computed locally at each node. An existing relay grants a new relay node a number of replicas proportional to the ratio between the advertised encounter metrics of the two nodes.

11.5.4 Hybrid DTN–MANET Environments

We conclude this chapter on hybrid solutions with a short discussion of algorithms for hybrid DTN-MANET environments. It is often the case, especially in urban scenarios, that the experienced connectivity, while not fully end-to-end and stable, is also not as sparse as assumed in the DTN setting. This is the case, for example, with *almost connected networks* and *islands of connectivity*, discussed in Section 11.2. In these cases, DTN routing may often be too pessimistic (and slow), while MANET routing solutions often perform satisfactorily. In such scenarios, it may be more sensible to first attempt to maintain complete path information and search a destination using traditional distance vector or link-state routing schemes. Only when the destination cannot be found in this manner, should DTN modules be integrated in the algorithm, to cope with the occasional disconnection or sparse regions of the network.

A simple approach is to maintain path information (e.g., using a link-state protocol such as OLSR [10]) inside connected components. If the destination is found in this routing table, the message follows the usual MANET way towards the destination.

If, on the other hand, the destination cannot be found within the current connected component, a DTN scheme takes over (e.g., Spray and Wait) to route the message to other connected components and ultimately to the remote destination [56].

Another approach is to use old routing table information [28]. Even if the link layer reports a disconnection on the path that used to reach a given destination D, the routing layer ignores this and still routes the message over that path, towards D. The motivation for this is that, since the path did exist, partially forwarding a message along that direction (until it reaches the broken link, where it is stored), is still a good forwarding decision. A high chance exists for the two disconnected components to merge again near the border of the cluster, where the path broke. If fresh information about a new, connected path arrives, then the aged routing information is discarded.

Finally, when many islands of connectivity exist, one approach is to have an intra-island routing mechanism based on MANET principles, and an inter-island routing mechanism which exploits the mobility of nodes among islands [61]. DTN techniques such as the use of redundancy can be employed for inter-island routing to improve performance.

11.6 CONCLUSION

In this chapter, we have discussed some key properties of opportunistic networking scenarios, and have presented a large number of forwarding schemes, which aim to cope with and/or exploit these properties. From the above discussion it is apparent that, due to the large variety of scenarios and characteristics, no single routing scheme is optimal for every imaginable DTN scenario. Mobility properties, node density, node resources, and performance requirements are only some of the salient features of the targeted scenario(s), the designer of opportunistic routing schemes should consider, in order to produce the best “recipe” for the challenges at hand. We hope that the exposition in this chapter offers sufficient guidelines to choose the right routing protocol for a given scenario or to modify and improve an existing one.

While a lot of research has been devoted to the topic of opportunistic forwarding, a number of interesting problems remain. Even though contact graphs have been shown to provide a useful handle in future contact prediction, the following questions remain unanswered: How much information is lost in the aggregation phase and how does this affect the performance of Complex Network Analysis-based schemes? Under what conditions is the contact graph (and thus the underlying dynamic connectivity graph) *navigable*? Does real (not inferred by estimation) information about social relationships help in opportunistic forwarding and when? Another open problem is performance analysis under realistic mobility assumptions and the protocol optimization that can ensue from the results of such analyses. Finally, with a number of foreseen opportunistic network applications being content-centric, group communication schemes (e.g., multicast, anycast, publish-subscribe systems etc) will become more important than unicast routing. To what extent can existing solutions for unicast and/or the acquired insight from their study be successfully applied to group communication scenarios remains unknown.

REFERENCES

1. David Aldous and James Fill. *Reversible Markov Chains and Random Walks on Graphs*. 2001.
2. Eitan Altman, Amar Prakash Azad, Tamer Basar, and Francesco de Pellegrini. Optimal Activation and Transmission Control in Delay Tolerant Networks. In *Proceedings of IEEE INFOCOM*, 2010.
3. Aruna Balasubramanian, Brian Neil Levine, and Arun Venkataramani. DTN Routing as a Resource Allocation Problem. In *Proceedings of ACM SIGCOMM*, 2007.
4. Aruna Balasubramanian, Brian Neil Levine, and Arun Venkataramani. Replication Routing in DTNs: A Resource Allocation Approach. *IEEE/ACM Transactions on Networking*, 2010.
5. Chiara Boldrini, Marco Conti, Jacopo Jacopini, and Andrea Passarella. HiBOp: a History Based Routing Protocol for Opportunistic Networks. In *Proceedings of IEEE WOWMOM*, 2007.
6. Chiara Boldrini, Marco Conti, and Andrea Passarella. User-centric Mobility Models for Opportunistic Networking. In *Proceedings of Springer BioWire*, 2008.
7. Augustin Chaintreau, Pierre Fraigniaud, and Emmanuelle Lebhar. Opportunistic Spatial Gossip over Mobile Social Networks. In *Proceedings of ACM WOSN*, 2008.
8. Augustin Chaintreau, Pan Hui, Jon Crowcroft, Christophe Diot, Richard Gass, and James Scott. Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms. In *Proceedings of IEEE INFOCOM*, 2006.
9. Bong Jun Choi and Xuemen Shen. Adaptive Asynchronous Clock based Power Saving Protocols for Delay Tolerant Networks. In *Proceedings of IEEE GLOBECOM*, 2009.
10. Thomas Clausen and Philippe Jacquet. *Optimized Link State Routing Protocol, draft-ietfmanet-olsr-11.txt*, 2003.
11. Vania Conan, Jeremie Leguay, and Timur Friedman. Characterizing pairwise inter-contact patterns in Delay Tolerant Networks. In *Proceedings of ICST Autonomics*, 2007.
12. Vania Conan, Jeremie Leguay, and Timur Friedman. Fixed point opportunistic routing in delay tolerant networks. *IEEE Journal on Selected Areas in Communications*, 26(5):773–782, 2008.
13. Elizabeth M. Daly and Mads Haahr. Social network analysis for routing in disconnected delay-tolerant MANETs. In *Proceedings of ACM MobiHoc*, 2007.
14. Elizabeth M. Daly and Mads Haahr. Social Network Analysis for Information Flow in Disconnected Delay-Tolerant MANETs. *IEEE Transactions on Mobile Computing*, 8(5):606–621, 2009.
15. Supratim Deb, Clifford Choute, Muriel Médard, and Ralf Koetter. Data Harvesting: A Random Coding Approach to Rapid Dissemination and Efficient Storage of Data. `http://tesla.csl.uiuc.edu/~koetter/publications/supratim.pdf`, 2005.
16. Olivier Dousse, Patrick Thiran, and Martin Hasler. Connectivity in ad-hoc and hybrid networks. In *Proceedings of IEEE INFOCOM*, 2002.
17. Henri Dubois-Ferriere, Matthias Grossglauser, and Martin Vetterli. Age matters: efficient route discovery in mobile ad hoc networks using encounter ages. In *Proceedings of ACM MobiHoc*, 2003.

18. Alaeddine El Fawal, Jean-Yves Le Boudec, and Kavé Salamatian. Self-Limiting Epidemic Forwarding. Technical report, EPFL, Lausanne, 2006.
19. Vijay Erramilli and Mark Crovella. Forwarding in Opportunistic Networks with Resource Constraints. In *Proceedings of ACM CHANTS*, 2008.
20. Vijay Erramilli, Mark Crovella, Augustin Chaintreau, and Christophe Diot. Delegation Forwarding. In *Proceedings of ACM MobiHoc*, 2008.
21. Kevin Fitchard. AT&T: Improving 3G network the major focus of 2010. `http://connectedplanetonline.com/3g4g/news/ATT-improving-3G-focus`.
22. Joy Ghosh, Sumesh J. Philip, and Chunming Qiao. Sociological orbit aware location approximation and routing in MANET. In *Proceedings of IEEE BroadNets*, 2005.
23. Marta C. Gonzalez, Cesar A. Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.
24. Matthias Grossglauser and Martin Vetterli. Locating Mobile Nodes with EASE: Learning Efficient Routes from Encounter Histories Alone. *IEEE/ACM Transactions on Networking*, 14(3), 2006.
25. Zygmunt J Haas and Tara Small. A new networking model for biological applications of ad hoc sensor networks. *IEEE/ACM Transactions on Networking*, 14(1):27–40, 2006.
26. Mor Halchor-Balter. *Performance Modeling and Design of Computer Systems*. preprint, 2011.
27. Bo Han, Pan Hui, V.S. Anil Kumar, Madhav V. Marathe, Guanhong Pei, and Aravind Srinivasan. Cellular traffic offloading through opportunistic communications. In *Proceedings of ACM CHANTS*, 2010.
28. Simon Heimlicher, Merkourios Karaliopoulos, Hanoch Levy, and Thrasyvoulos Spyropoulos. On Leveraging Partial Paths in Partially-Connected Networks. In *Proceedings of IEEE INFOCOM*, pages 55–63, 2009.
29. Tristan Henderson, David Kotz, and Ilya Abyzov. The changing usage of a mature campus-wide wireless network. In *Proceedings of ACM MobiCom*, 2004.
30. Theus Hossmann, Thrasyvoulos Spyropoulos, and Franck Legendre. Know thy neighbor: towards optimal mapping of contacts to social graphs for DTN routing. In *Proceedings of IEEE INFOCOM*, pages 866–874, 2010.
31. Theus Hossmann, Thrasyvoulos Spyropoulos, and Franck Legendre. Putting Contacts into Context: Mobility Modeling beyond Inter-Contact Times. In *Proceedings of ACM MobiHoc*, 2011.
32. Wei-jen Hsu and A. Helmy. On Modeling User Associations in Wireless LAN Traces on University Campuses. In *Proceedings of ACM MobiHoc*, 2006.
33. Wei-jen Hsu, Thrasyvoulos Spyropoulos, Konstantinos Psounis, and Ahmed Helmy. Modeling Time-Variant User Mobility in Wireless Mobile Networks. In *Proceedings of IEEE INFOCOM*, 2007.
34. Pan Hui, Jon Crowcroft, and Eiko Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. In *Proceedings of ACM MobiHoc*, page 241, New York, New York, USA, 2008.
35. Sushant Jain, Kevin Fall, and Rabin Patra. Routing in a Delay Tolerant Network. In *Proceedings of ACM SIGCOMM*, 2004.

36. Evan P. C. Jones, Lily Li, and Paul A. S. Ward. Practical Routing in Delay-Tolerant Networks. In *Proceedings of ACM WDTN*, 2005.
37. Philo Juang, Hidekazu Oki, Yong Wang, Margaret Martonosi, Li Shiuan Peh, and Daniel Rubenstein. Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with ZebraNet. In *Proceedings of ACM ASPLOS*, 2002.
38. Hyewon Jun, Mostafa Ammar, and Ellen Zegura. Power Management in Delay Tolerant Networks:A Framework and Knowledge-Based Mechanisms. In *Proceedings of IEEE SECON*, 2005.
39. Abhinav Kamra, Vishal Misra, Jon Feldman, and Daniel Rubenstein. Growth Codes: Maximizing Sensor Network Data Persistence. In *Proceedings of ACM SIGCOMM*, 2006.
40. Sachin Katti, Hariharan Rahul, Wenjun Hu, Dina Katabi, Muriel Medard, and Jon Crowcroft. XORs In The Air: Practical Wireless Network Coding. In *Proceedings of ACM SIGCOMM*, 2006.
41. S. Keshav. KioskNet (VLINK). `http://blizzard.cs.uwaterloo.ca/tetherless/index.php/VLink`.
42. Amir Krifa, Chadi Barakat, and Thrasyvoulos Spyropoulos. Optimal Buffer Management Policies for Delay Tolerant Networks. In *Proceedings of IEEE SECON*, pages 260–268, 2008.
43. Amir Krifa, Chadi Barakat, and Thrasyvoulos Spyropoulos. Message Drop and Scheduling in DTNs: Theory and Practice. *IEEE Transactions on Mobile Computing*, 2011.
44. Bhaskar Krishnamachari, Stephen Wicker, and R Bejar. Phase Transition Phenomena in Wireless Ad Hoc Networks. In *Proceedings of IEEE Globecom*, pages 2921–2925, 2001.
45. Bhaskar Krishnamachari, Stephen Wicker, Rámon Béjar, and Marc Pearlman. Critical Density Threholds in Distributed Wireless Networks. In *Springer Communications, Information and Network Security*. 2002.
46. Jérémie Leguay, Vania Conan, and Timur Friedman. Evaluating MobySpace-based Routing Strategies in Delay-Tolerant Networks. *Wiley Wireless Communications and Mobile Computing*, 2007.
47. Jérémie Leguay, Timur Friedman, and Vania Conan. {DTN} Routing in a Mobility Pattern Space. In *Proceedings of ACM WDTN*, 2005.
48. Dan Lelescu, Ulaş C. Kozat, Ravi Jain, and Mahadevan Balakrishnan. Model t++: an empirical joint space-time registration model. In *Proceedings of ACM MobiHoc*, pages 61–72, 2006.
49. Shuo-Yen Robert Li, Raymond W. Yeung, and Ning Cai. Linear Network Coding. *IEEE Transactions on Information Theory*, 2003.
50. Anders Lindgren, Avri Doria, and Olov Schelen. Probabilistic routing in intermittently connected networks. *ACM Mobile Computing and Communication Review*, 7(3), 2003.
51. Abderrahmen Mtibaa, Martin May, Christophe Diot, and Mostafa H. Ammar. PeopleRank: Social Opportunistic Forwarding. In *Proceedings of IEEE INFOCOM*, 2010.
52. Mirco Musolesi, Stephen Hailes, and Cecilia Mascolo. Adaptive Routing for Intermittently Connected Mobile Ad Hoc Networks. In *Proceedings of ACM WoWMoM*, 2005.
53. Samuel C Nelson, Mehedi Bakht, and Robin Kravets. Encounter-Based Routing in DTNs. In *Proceedings of IEEE INFOCOM*, 2009.

54. Mark E. J. Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–256, 2003.
55. Hoang Anh Nguyen, Silvia Giordano, and Alessandro Puiatti. Probabilistic Routing Protocol for Intermittently Connected Mobile Ad hoc Network (PROPICMAN). In *Proceedings of IEEE WoWMoM*, 2007.
56. Jörg Ott, Dirk Kutscher, and Christoph Dwertmann. Integrating DTN and MANET routing. In *Proceedings of ACM CHANTS*, pages 221–228, 2006.
57. Antonis Panagakis, Athanasios Vaios, and Ioannis Stavrakakis. Study of two-hop message spreading in DTNs. In *Proceedings of IEEE WiOpt*, 2007.
58. Juyong Park and Mark E. J. Newman. Statistical mechanics of networks. *Physical Review E*, 70(6):15, 2004.
59. Andrea Passarella and Marco Conti. Characterising Aggregate Inter-contact Times in Heterogeneous Opportunistic Networks. In *Proceedings of IFIP Networking*, pages 301–313, 2011.
60. Gabriel Sandulescu and Simin Nadjm-Tehrani. Opportunistic DTN routing with Window-Aware Adaptive Replication. In *Proceedings of AINTEC*, 2008.
61. Natasa Sarafijanovic-Djukic, Michal Piorkowski, and Matthias Grossglauser. Island Hopping: Efficient Mobility-Assisted Forwarding in Partitioned Networks. In *Proceedings of IEEE SECON*, 2006.
62. Rahul C. Shah, Sumit Roy, Sushant Jain, and Waylon Brunette. Data MULEs: Modeling and analysis of a three-tier architecture for sparse sensor networks. In *Proceedings of IEEE SNPA Workshop*, 2003.
63. Tara Small and Zygmunt J. Haas. Resource and Performance Tradeoffs in Delay-Tolerant Wireless Networks. In *Proceedings of ACM WDTN*, 2005.
64. Ignacio Solis and Katia Obraczka. Effcient Continuous Mapping in Sensor Networks Using Isolines. In *Proceedings of MobiQuitous*, pages 325–332, 2005.
65. Thrasyvoulos Spyropoulos, Konstantinos Psounis, and Cauligi S. Raghavendra. Spray and Wait: Efficient Routing in Intermittently Connected Mobile Networks. In *Proceedings of ACM WDTN*, 2005.
66. Thrasyvoulos Spyropoulos, Konstantinos Psounis, and Cauligi S. Raghavendra. Spray and Focus: Efficient Mobility-Assisted Routing for Heterogeneous and Correlated Mobility. In *Proceedings of IEEE ICMAN*, 2007.
67. Thrasyvoulos Spyropoulos, Konstantinos Psounis, and Cauligi S. Raghavendra. Efficient routing in intermittently connected mobile networks: the multiple-copy case. *IEEE/ACM Transactions on Networking*, 16:77–90, 2008.
68. Thrasyvoulos Spyropoulos, Thierry Turletti, and Katia Obraczka. Utility-based Message Replication for Intermittently Connected Heterogeneous Networks. In *Proceedings of IEEE AOC*, 2007.
69. Thrasyvoulos Spyropoulos, Thierry Turletti, and Katia Obraczka. Routing in Delay Tolerant Networks Comprising Heterogeneous Node Populations. *IEEE Transactions on Mobile Computing*, 8(8):1132–1147, 2009.
70. David N C Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

71. Amin Vahdat and David Becker. Epidemic routing for partially connected ad hoc networks. Technical Report CS-200006, Duke University, 2000.
72. United Villages. First Mile Solutions. `http://www.firstmilesolutions.com`.
73. Wei Wang, Vikram Srinivasan, and Mehul Motani. Adaptive Contact Probing Mechanisms for Delay Tolerant Applications. In *Proceedings of ACM MobiCom*, 2007.
74. Yong Wang, Sushant Jain, Margaret Martonosi, and Kevin Fall. Erasure Coding Based Routing for Opportunistic Networks. In *Proceedings of ACM WDTN*, 2005.
75. John Whitbeck, Marcelo Amorim, Yoann Lopez, Jeremie Leguay, and Vania Conan. Relieving the wireless infrastructure: When opportunistic networks meet guaranteed delays. In *Proceedings of IEEE WoWMoM*, 2011.
76. Jörg Widmer and Jean-Yves Le Boudec. Network Coding for Efficient Communication in Extreme Networks. In *Proceedings of ACM WDTN*, 2005.
77. Yong Xi, Mooi-Choo Chuah, and Kirk Chang. Performance Evaluation of a Power Management Scheme for Disruption Tolerant Network. *Springer Mobile Networks and Applications*, 12(5-6):370–380, 2007.
78. Zixiang Xiong, Angelos D. Liveris, and Samuel Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine*, 21(5), 2004.
79. Dan Yu and Hui Li. On the Definition of Ad Hoc Network Connectivity. In *Proceedings of IEEE ICCT*, pages 990–994, 2003.
80. Xiaolan Zhang, Giovanni Neglia, James Kurose, and Don Towsley. Performance Modeling of Epidemic Routing. In *Proceedings of IFIP Networking*, 2006.
81. Wenrui Zhao, Mostafa Ammar, and Ellen Zegura. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *Proceedings of ACM MobiHoc*, 2004.