IJERMT

International Journal of Engineering Research & Management Technology

May- 2015 Volume 2, Issue-3
ISSN: 2348-4039

Email: editor@ijermt.org www.ijermt.org

IKCE "Sharpening Skills..... Serving Nation"

Special Issue: 2nd International Conference on Advanced Developments in Engineering and Technology Held at Lord Krishna College of Engineering Ghaziabad, India

Identifying and Addressing Routing Strategies in Delay Tolerant Network

Amit Kumar Tyagi, Dr. Prof. S.K. Tyagi
Department of Computer Science and Engineering,
Pondicherry Engineering College
Puducherry

ABSTRACT
Delay-Tolerant Networks (DTNs) have the potential to interconnect devices in regions that current networking technology cannot reach. DTNs are a network of regional networks i.e. it is an overlay on top of regional networks, including the Internet. It is an attempt to extend the reach of networks. For that, idea is that; an end-to-end connection may never be present. So to make communication possible between two networks, intermediate nodes take responsibility of the data to transferred and forward it as the opportunity arises. Both nodes and links may be inherently unreliable and disconnections may be long-lived. To realize the DTN vision, routes must be found over multiple unreliable, intermittently-connected hops. Till now, various researchers have investigated this fundamental challenge in DTNs, particularly over the past five years. So keeping these points in our mind, this paper provides detailed surveys in the area of routing in DTNs and presents system for classifying the proposed routing strategies.

Keywords--- Delay Tolerant Networks, Flooding, Routing, Hop Relay, Tree.

1. INTRODUCTION

Wired and wireless networks have enabled a wide range of devices to be interconnected over vast distances for e.g. today it is possible to connect from a cell phone to millions of powerful servers around the world. As successful as these networks have been, but still they cannot reach everywhere, and for some applications their cost is much prohibitive. The reason for these limitations; assumptions, are not true in all environments like, the first and most important assumption is that an end-to-end connection exists from the source to the destination, possibly via multiple intermediaries. This assumption can be easily violated due to mobility, power saving, or unreliable networks for e.g. if a wireless device is out of range of the network then it cannot use any application that requires network communication.

A delay-tolerant network (DTN) is a network of regional networks: it is an overlay on top of regional networks, including the Internet [24]. Delay-tolerant networking (DTN) is an attempt to extend the reach of networks. It promises to enable communication between "challenged" networks, which includes deep space networks, sensor networks, mobile ad-hoc networks, and low-cost networks. The core idea is that these networks can be connected if protocols are designed to accommodate disconnection [8].The communication characteristics are relatively homogeneous in a communication region. Figure 1, shows specific task of DTN network.

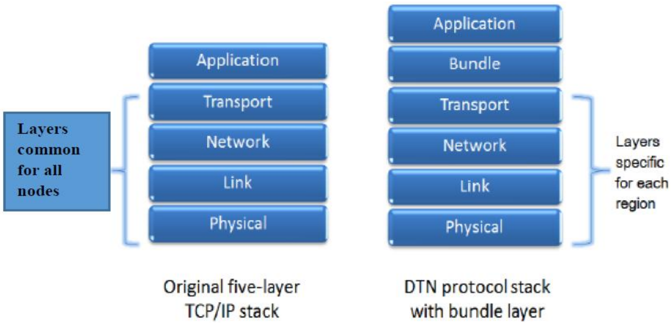

Fig1- DTN specific stack

The wireless DTN technologies may be diverse, including not only radio frequency (RF) but also free-space optical, ultra-wide band (UWB), and acoustic (sonar or ultrasonic) technologies. Each region has a unique region ID which is knowable among all regions of the DTN. DTN gateways have membership in two or more regions and are the only means of moving messages between regions. In Figure 2, laptops communicate with each other to exchange data. If the destination laptop is not present, which may occur if the student has gone home, the network stores the messages until they return.

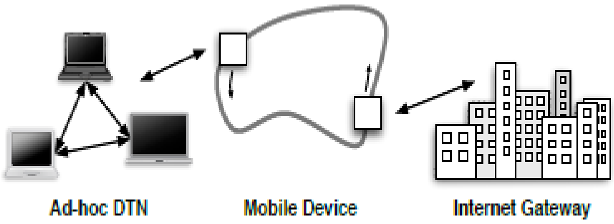

Figure 2: Laptops communicating with each other and the Internet via delay-tolerant networking

This device picks up requests from the school and delivers them to the gateway, and then provides the responses on its next trip. Now there are many other applications for delay-tolerant networks. In developing regions, applications range from education to health care to government services [3]. DTNs could also be used to gather data from everything ranging from sensors in oceans [4], to satellites in space [5].
One of the fundamental problems that arise when designing networks that handle disconnection is routing. Before network can be usable, it must be possible to get data from the source to the destination. This paper shows the routing problems arise in DTNs. The research dates back to before the term "delay tolerant "was widely used. The adjectives "intermittently connected,""sparse," and "disconnected" are also used to describe networks without constant end-to-end connections.

Now this paper presents two properties that can be used to classify delay-tolerant routing strategies: replication and knowledge. Replication describes how a routing strategy relies on multiple copies of each message, and knowledge describes how information about the network is used to make decisions. Hence organization of this paper as: section 2 discusses about characteristics of DTNs and the metrics used to evaluate each technique. Section 3, discuss about a system for classifying each strategy based on the two properties. This divide the routing strategies into two broad families, flooding and forwarding, based on their use of replication and knowledge. Section 4, analyze each family, and show how prior research fits into the classification scheme. Finally, section 6 and section 5 present a summary of the current work in routing for DTNs, and propose directions/problems for future research.

2. Network Characteristics

In order to discuss the routing problem, we need a model that describes the network. A DTN is composed of computing systems participating in the network, called nodes. One-way links connect some nodes together. These links may go up and down over time, due to mobility, failures, or other events. When the link is up, the source node has an opportunity to send data to the other end. In the DTN literature this opportunity is called a contact [8]. More than one contact may be available between a given pair of nodes for e.g. a node might have a high-performance, expensive connection and a low-performance, cheap connection that are available simultaneously for communication with the same destination. The contact schedule is the setoff times when the contact will be available. In graph theory, this model is a time-varying multi graph. The DTN architecture proposes to use this network by forwarding complete messages over each hop. These messages will be buffered at each intermediate node, potentially on non-volatile storage. This enables messages to wait until the next hop is available, which may be a long period of time.

As described by Jain et al., the amount of time for a message to be transferred from one node to another can be divided into four components: waiting time, queuing time, transmission delay, and propagation delay [19].

2.1 Challenges

Delay-tolerant networks (DTNs) present many challenges that are not present in traditional networks. Many stem from the need to deal with disconnections, which directly impacts routing and forwarding. However, because these networks enable communication between a wide ranges of devices, there are secondary problems that routing strategies may need to be aware of, such as dealing with limited resources.

2.1.1 Contact Schedules: Of the four inter-node delay components, the most significant is likely to be the waiting time, since it might range anywhere from seconds to days whereas the others are typically much shorter. Thus the most important characteristics of a DTN is the contact schedule, which depends strongly on the application area under consideration.

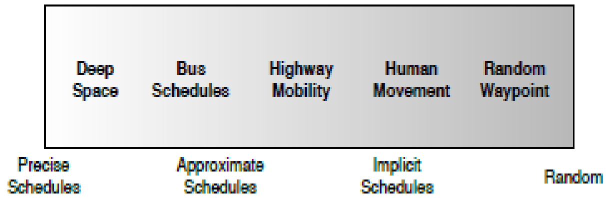

Figure 3: A spectrum of contact schedule predictability

Contact schedules can be placed on an approximate spectrum based on how predictable they are, as shown in Figure 3. At one extreme contact schedules that are very precise. Beside this, an example would be deep space networks, where disconnections are caused by movements of objects in space that can be calculated very accurately. One step less predictable would be scheduled networks with errors for e.g. city buses, work. These buses have a schedule, but it is not precise. Many human activities have implicit schedules, there is no guarantee when a person will be at work, but their schedule is fairly regular. Finally, at the other end of the spectrum are networks with completely random connectivity. These networks are widely studied in the ad-hoc networking community because the models are simple to work with.

Some work on DTN routing has investigated networks with proactive mobility, where the nodes actively move in response to communication needs [10, 23, 24, and 25]. In our model, this can be represented as contacts that can be selectively brought up or down. Networks with this type of mobility fall somewhere in the middle of the spectrum because some contacts are unpredictable, but the controlled contacts are predictable. The routing techniques discussed in this paper are equally applicable to networks with proactive mobility. However, these networks also require some type of cost/benefit optimization to make decisions about proactive movement, which is outside the scope of this literature.

2.1.2 Buffer Space: In order to cope with long disconnections, messages must be buffered for long periods of time. This means that intermediate routers require enough buffer space to store all the messages that are waiting for future communication opportunities. This means that intermediate routers require buffer space proportional to demand. An alternate point of view is that routing strategies might need to consider the available buffer space when making decisions. In the studies surveyed here, all nodes have an equal amount of buffer space and the strategies do not make decisions based on this resource.

2.1.3 Contact Capacity: A question that is closely related to the contact schedule is; How much data that can be exchanged between two nodes? This depends on both the link technology and the duration of the contact. Even if the duration is precisely known, it may not be possible to predict the capacity due to fluctuations in the data rate. At a first glance, it might appear that this is a simple issue for routing strategies to deal with. A naive approach would be to ignore the contact capacity, except in cases where the message is simply too large to be sent across the contact without fragmentation. If the volume of traffic is very small compared to the capacity of contacts in the network, then this is a reasonable approach. However, if the volume of traffic increases due to a large number of users, or due to large messages being exchanged, the contact capacity becomes very important. In this situation, the best contact could become one i.e. "inefficient" according to other criteria, but has the largest contact volume and thus is best equipped to handle large traffic demands.

2.1.4 Energy: Some nodes in DTNs may have limited energy supplies either because they are mobile, or because they are in a location that cannot easily be connected to the power grid. Routing consumes energy by sending, receiving and storing messages, and by performing computation. Hence, routing strategies that send fewer bytes and perform less computation will be more energy efficient. Additionally, routing strategies can optimize power consumption by using energy-limited nodes sparingly. While researcher's have investigated general techniques for saving power in DTNs [18], none of the routing strategies surveyed has incorporated power-aware optimizations.

2.1.5 Processing Power: One of the goals of delay-tolerant networking is to connect devices that are not served by traditional networks. These devices may be very small, and similarly have small processing capability, in terms of CPU and memory. These nodes will not be capable of running complex routing protocols. The routing strategies presented here could still be used on more powerful gateway nodes, in order to connect the sensor network to a general purpose delay-tolerant network.

2.2 Evaluation Criteria

In order to compare routing strategies, we must define some metrics for evaluating their performance. Since the exact numbers for the metrics depend on many factors, we will only discuss them in relative terms.

2.2.1 Latency: A secondary metric is the latency, the time between when a message is generated and when it is received. This metrics important since many applications can benefit from assort delivery latency, even though they will tolerate long waits. Many applications also have some time window where the data is useful for e.g. if a DTN is used to deliverer-mail to a mobile user, the messages must be delivered before the user moves out of the network.

2.2.2 Delivery Ratio: In DTN, the important network performance metric is the delivery ratio. However, in DTNs, a message is rarely actually "lost."Rather, the network was unable to deliver messages within an acceptable amount of time. Thus, this paper define the delivery ratio as the fraction of generated messages that are correctly delivered to the final destination within a given time period.

2.2.3 Transmissions: Some routing strategies transmit more messages than others, either because they use multiple copies of each message, make different decisions about the next hop, or because of protocol overhead.

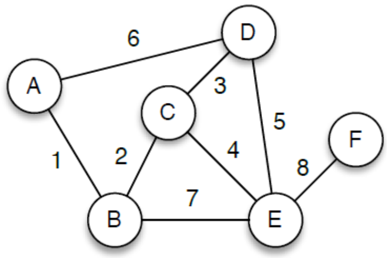

Figure 4: Example DTN scenario

The number of transmissions is a measure of the amount of contact capacity consumed by a protocol. It is also an approximate measure of the computational resources required, as there is some processing required for each message. Additionally, each transmission consumes energy, so it is also an approximate measure of power consumption. For scenario, figure 4shows throughout to illustrate the route of five among strategies. There are six nodes labelled A through F. The contacts go up and down one at a time, in the sequence shown by the label son each line. All links are bidirectional, and messages are always generated by node A.

3. Routing Strategy Properties

This survey categorizes delay-tolerant routing strategies using two properties i.e. replication and knowledge. Replication denotes show the strategy uses multiple copies of a message, and how it chooses to make those copies. Knowledge indicates how the strategy uses information about the state of the network in order to make routing decisions, and also how it obtains that information.

3.1 Knowledge

Some routing strategies require more information about the network than others. At one extreme, a node can make decisions with zero knowledge about the network, except which contacts are currently available. These strategies use static rules that are configured when the strategy is designed, and every node obeys the same rules. This leads to simple implementations that require minimal configuration and control messages, since all the rules are hard-coded ahead of time. The disadvantage is that the strategy cannot adapt to different networks or conditions, so it may not make optimal decisions. At the other end of the spectrum, a node might need to know the complete future schedule of every contact in the network. Provided that the information is accurate, this allows routing strategies to make very efficient use of network resources by forwarding a message along the best path. There is a range of values in between these two extremes for e.g. approximate information about the future contact schedules might be available. Or, a strategy might require no information in advance, but instead will learn it automatically.

3.2 Replication

Delay-tolerant networks may rely on components that are unreliable or unpredictable. To compensate for this, many routing strategies make multiple copies of each message, in order to increase the chance that at least one copy will be delivered, or to reduce delivery latency. The intuition is that having more copies of the message increases the probability that one of them will find its way to the destination, and decreases the average time for one to be delivered. This shows a clear trade-off between cost and performance. The cheapest approach is to have a single copy of the message. However, a single failure will result in the message being lost. The most reliable approach is to have each node carry a copy of the message. In this case, the message is lost only if all the nodes in carrying it are unable to deliver it. However, this consumes bandwidth and storage resources proportional to the number of nodes in the network.

A related issue is characterizing the best approach to making replicas like Jain et al. present a theoretical approach to determine which set of paths to use, provided that the path failure probabilities are known and independent [14].Erasure coding and networking coding schemes have also been investigated to attempt to keep the benefit of multiple copies while reducing the resource costs [15, 26]. These techniques appear promising, and it should be possible to integrate them with the routing strategies presented here.

4. Routing Strategy Families

DTN routing strategies divide into two families based on which property a strategy uses in order to find the destination i.e. flooding strategy and forwarding strategy. Flooding strategies, which rely primarily on replicating messages to enough nodes so the destination receives it, and forwarding strategies, which rely on knowledge about the network to select the best path to the destination?

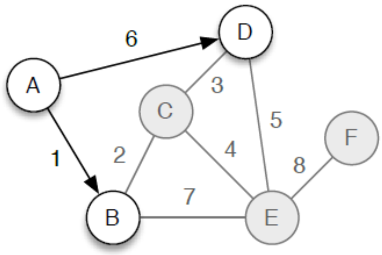

Figure 5: Direct Contact routing example

We first describe each family in general, and then describe specific examples in each family.

4.1 Flooding Strategies

Strategies in the flooding family deliver multiple copies of each message to a set of nodes, called relays. The relays store the messages until they connect with the destination, at which point the message is delivered. The earliest work-in the area of DTN routing falls into this family. Many of them date before the term “delay-tolerant” became popular. Traditionally, these strategies have been studied in the context of mobile ad-hoc networks, where random mobility has a good chance of bringing the source into contact with the destination. Message replication is then used to increase the probability that the message gets delivered. The basic protocols in this family do not need any information about the network; however more advanced schemes use some knowledge to improve performance.

4.1.1 Two Hop Relay: In this strategy, the source copies the message to the first nodes that it contacts. The source and the relays hold the message and deliver it to the destination. Since there are now n+1 copies of the message in the network, more bandwidth and storage are consumed.

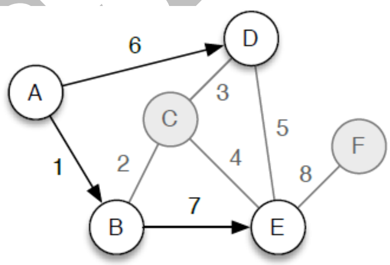

Figure 6: Two-hop routing example

However, the resource consumption is limited and can be tuned by adjusting the number of copies. This strategy has a much better chance of delivering the message than the Direct Contact strategy. If we assume that each node contacts the destination within independent probability p, then this strategy will deliver each message with probability $1-(1-p)(n+1)$, which is approximately$(n+1)p$ if p is very small. Similarly, increasing the number of copies decreases the average latency; since the message is delivered as soon as any of the $n + 1$ nodes contacts the destination. This strategy has the same fundamental limitation as Direct Contact: If the n+ 1 node never reach the destination, the message cannot be delivered. In scenarios where the mobility is random, this might be rare, but in networks with structured connectivity this could bevery common. For example, if node A has a message for node E, it would send copies to both nodes B and D, as shown in Figure6. When node B connects with node C, it would not send it the message, since node C is not the destination. Finally, the message would be delivered at time 7 when node B connects with node E. At the end, nodes A, B, D and E have all received the message. Node A can reach all other nodes via two-hop relay except node F, since it is a minimum of three hops away. Grossglauser and Tse'swork can be used to increase the capacity of mobile ad-hoc networks under ideal conditions [20]. This approach has also been studiedas a routing strategy for sensor networks [17], and for scenarios with proactive mobility [23]. It proposed as a fall back when ad-hoc routing cannot find a connected path [16] in delay tolerant network.

4.1.2 Direct Contact: This strategy waits until the source comes into contact with the destination before forwarding the data. This is the degenerate case of the flooding family, where the set of relays contains only the destination. It can also be considered a degenerate case of the forwarding family, where it always selects the direct path between the source and the destination. However, since this strategy does not require any information about the network and only uses a single hop, we will consider it to be a flooding strategy. Due to its simplicity, it does not consume many resources, and it uses exactly one message transmission. However, it only works if the source contacts the destination. The Info station architecture proposed using direct contact delivery between mobile nodes and fixed gateways as a technique for increasing wireless network throughput and decreasing cost [21]. Grossglauser and Tse show that in their mobile ad-hoc network scenario, this strategy has a capacity that approaches zero as the number of nodes increases [20].In the example scenario, node A can only deliver messages to nodes B and D, as shown in Figure 5. Additionally, it is faster for node A to deliver a message to node D via the path A-B-C-D, which it cannot do.

4.1.3 Tree Based Flooding: Tree-Based Flooding (TBF) strategies extend two-hop relay by distributing the task of making copies to other nodes. When a message is copied to a relay, there an indication of how many copies the relay should make. This is called Tree-Based Flooding because the set of relays forms a tree of nodes rooted at the source. Two-hop relay can be viewed as TBF with a depth of one.
There are many ways to decide how to make copies. A simple scheme is to allow each node to make unlimited copies, but to restrict the message to travel a maximum of n hops from the source [22].This limits the depth of the tree, but places no limit on its breadth. A refinement is to also limit the node to make at most m copies [27]. This limits both the depth and the breadth of the tree, which limits the total number of copies to a maximum of Pni=0 mn. A more complex alternative is to limit the total number of copies to N. When a node makes a copy, it distributes the responsibility for making half of its current copies to the other node, and keeps half for itself [7, 27]. This scheme has been shown to be optimal if the inter-node contact probabilities are independent and identically distributed [7].

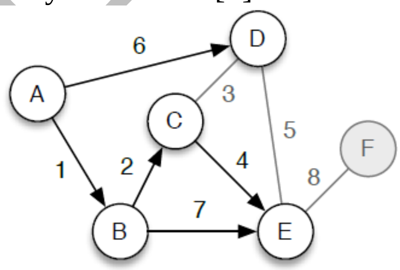

Figure 7: Tree-Based Flooding Example

TBF can deliver messages to destinations that are multiple hops away, unlike Direct Contact or Two- Hop Relay. However, tuning the parameters can be a challenge.
If they are too conservative, many extra copies will be made. Conversely, if they are too aggressive, then the message may not propagate to the destination. Consider an example scenario if node A has a message for node E, and it can make a maximum of four additional copies. At time 1, it sends a copy to node B along with directions to make one copy (b (4 − 1)/2c), shown in Figure 7. Node A keeps two copies for itself (d (4 − 1)/2e). At time 2 when node B connects with node C, B's additional copy is delivered. At time 3, node C connects to node D. However, it cannot send D a copy because it has no copies to distribute. At time 4, C

delivers the message to the E. At time 6, A sends D a copy, since it does not know that the message was already delivered. At this point, node A has one remaining copy that it will deliver if it contacts another node.

4.1.4 Epidemic Routing: Epidemic algorithms were originally proposed for synchronizing replicated databases [29]. Vahd at and Becker applied these algorithms to forwarding data in a DTN [22]. In effect, the queue of messages waiting to be delivered is the database that needs to be synchronized. Epidemic algorithms guarantee that provided a sufficient number of random exchanges of data, all nodes will eventually receive all messages. Thus, the destination node is guaranteed to have received the data. Epidemic Routing works as follows. When a message is sent, it is placed in the local buffer and tagged with a unique ID. When two nodes connect, they send each other the list of all the messages IDs they have in their buffers, called the summary vector. Using the summary vector, the nodes exchange the messages they do not have.
When this operation completes, the nodes have the same messages in their buffers. Epidemic Routing represents the extreme end of the flooding family because it tries to send each message over all paths in the network. This provides a large amount of redundancy since all nodes receive every message, making this strategy extremely robust to node and network failures.

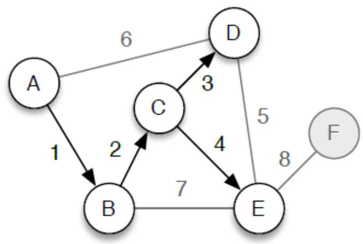

Figure 8: Epidemic Routing Example

Additionally, since it tries every path, it delivers each message in the minimum amount of time if there are sufficient resources. In the example, the message will be delivered from A to E via the fastest path (A-B-C-E), as shown in Figure 8.All nodes will receive the message except node F because node E does not replicate messages that are destined for it. Epidemic Routing is relatively simple because it requires no knowledge about the network. For that reason, it has-been proposed to use it as a fallback when no better method is available [31]. The disadvantage is that a huge amount of resources are consumed due to the large number of copies. This requires large amount of buffer space, bandwidth, and power. Many papers have studied ways to make Epidemic Routing consume fewer resources [11, 13, 27, 28, and 30]. One of the problems is that the message continues to propagate through the network, even after it has been delivered.
The original epidemic algorithms paper proposed "death certificates" to solve this problem [29]. The idea is that a new message is propagated informing nodes to delete the original message and to not request it again. Ideally, the death certificate will be much smaller than the original message, so overall the resource consumption is reduced. Researchers have explored various schemes for tuning how aggressively the death certificates are propagated. Small and Haas show that the more aggressive the death certificate propagation, the less storage is required at each node [27],while Harras and Almeroth show that the more aggressive strategies transmit more messages [35].

A critical resource in epidemic routing is the buffer. An intelligent buffer management scheme can improve the delivery ratio over the simple FIFO scheme [30]. The best buffer policy evaluated is to drop packets that are the least likely to be delivered based on previous history. If node A has met B frequently, and B has met C frequently, then Ais likely to deliver messages to C through B. Similar metrics are used in a number of epidemic protocol variants [11, 24, 28, and 30]. This approach takes advantage of physical locality and the fact that movement is not completely random. While these protocols are more efficient than the original Epidemic routing protocol, they still transmit many copies of each message.

4.2 Forwarding Strategies

The strategies in this family take a more traditional approach to routing data in a DTN. They use network topology information to select the best path, and the message is then forwarded from node to node along this path. A path can be found using location-based routing, assigning metrics to nodes or by assigning metrics to links. Some of these approaches have been explored in wired and multi-hop wireless networks. However, the protocols designed for these environments will not function in delay-tolerant networks, since they assume that links are usually connected. By definition, the strategies in this family require some knowledge about the network. They typically send a single message along the best path, so they do not use replication.

4.2.1 Gradient Routing: An alternate approach is to assign a weight to each node that represents its suitability to deliver messages to a given destination. When the custodian of a message contacts another node that has a better metric for the message's destination, it passes the message to it. This approach is called gradient routing because the message follows a gradient of improving utility function values towards the destination. The idea was first applied to ad-hoc networks in 2000 [46]. This requires more network knowledge than location-based routing for two reasons. First, each node must store a metric for all potential destinations. Second, sufficient information must be propagated through the network to allow each node to compute its metric for all destinations. The metric could be based on many parameters, such as the time of last contact between the node and the destination, remaining battery energy, or mobility. An extremely simple utility function is to forward a packet with a certain probability. This approach does not seem practical, but it has been used as a baseline for comparison with more advanced techniques [19, 45]. Gradient Routing has been shown to decrease the delay when compared to direct contact [45]. Similar schemes have been proposed for routing data towards base stations in sensor networks. For example, the history-based protocol presented for Zebra Net is a gradient routing strategy [33]. A theoretical analysis of gradient routing can be found in [45], and a discussion about predicting utility function values can be found in [38].

One of the shortcomings of gradient routing is that it can initially take a long time for a good custodian to be found, since it may take some time for the utility function values to propagate, or because the metric values in the region around the initial custodian are all equally poor. One approach that has been shown to reduce the delivery latency is to initially use random forwarding until the utility value reaches a certain threshold [45]. This hybrid approach initially allows a message to actively explore the network until it finds a good carrier, and then it uses the standard utility routing to efficiently reach the destination. Burgess et al. use a similar technique to quickly propagate a new message in their epidemic routing variant [11].

4.2.2 Location Based Routing: The forwarding approach that requires the least information about the network is to assign coordinates to each node. A distance function is used to estimate the cost of delivering messages from one node to another. The coordinates can have physical meaning, such as GPS coordinates, as has-been studied for mobile ad-hoc networks [42]. Alternatively, the coordinates can have meaning in the network topology space, instead of physical space, which has been used to estimate network latency between arbitrary nodes on the Internet [43]. In general, a message is forwarded to a potential next hop if that node is closer in the coordinate space than the current custodian.

The advantage of location-based routing is that it requires very little information about the network, eliminating the need for routing tables and reducing the control overhead. In order to determine the best path, a node only needs to know its own coordinates, the coordinates of destination, and the coordinates of the potential next hops. Given these three pieces of information, a node can easily compute the distance function and determine where the message should be sent. Location-based routing has two well-known problems. The first problem is that even if the distance between two nodes is small, there is no guarantee that they will be able to communicate.

In the case of physical coordinates, consider two wireless nodes on opposite sides of a wall that blocks all radio signals, as shown in Figure 9. Node A wants to send a message to D. It has two potential next hops: nodes B and C. Since node B is the closest to node D, it forwards the message to B. However, B cannot communicate with D due to the obstruction, whereas node C has a line of sight. The problem is that location does not necessarily correspond to network topology. This problem is somewhat alleviated by using virtual coordinates, since they are designed to closely represent the network topology. However, it is still possible that the message can fall into a local minimum and not reach the destination. In mobile ad-hoc networks, protocols have been explored which attempt to route around obstructions like this [41]. The second problem is that anode's coordinates can change. If a node moves, its physical coordinates change. If the network topology changes, a node's virtual coordinates change.

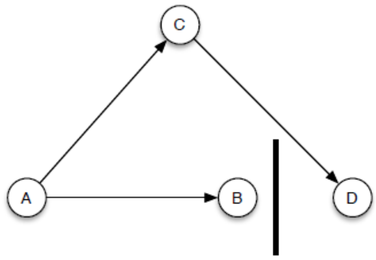

Figure 9: Location-based routing fails because of alocal minimum

This complicates routing because the source needs the coordinates of the destination node. These two problems mean that implementation location-based routing is not as simple as it appears. Lebrun et al. proposed using the motion vector of mobile nodes to predict their future location. Their scheme passes messages to nodes that are moving closer to the destination [39], which results in a better delivery ratio than two-hop routing with less overhead than Epidemic routing. Leguay et al. presented virtual coordinate routing strategycalled mobility pattern spaces [44]. In their strategy, the node coordinates are composed of a set of probabilities, each representing the chance that a node will be found in a specific location. Various distance functions are then computed on this vector. Their results show that this approach reduces the amount of resources consumed when compared with Epidemic Routing, while still delivering a substantial fraction of the messages. Neither of these works addresses the local minima or changing coordinates problems. However, they do show that these techniques are applicable to DTNs.

4.2.3 Link Metrics: Routing strategies that use link metrics resemble traditional network routing protocols. They build a topology graph, assign weights to each link and finally run a shortest path algorithm to find the best paths. This requires the most network information as each node must have sufficient knowledge to run a routing algorithm. Link weights are assigned to try and provide optimal service to the endpoints, based on some performance metric: the highest bandwidth, lowest latency, and the highest delivery ratio. In delay tolerant net works, the most important metric is the delivery ratio, since the network must be able to reliably deliver data. A secondary metric is the delivery latency. Thus, the challenge is to determine a system for assigning link metrics that maximizes the delivery ratio and minimizes the delivery latency. Some metrics may also attempt to minimize resource consumption, such as buffer space or power. The first paper that proposed using link metrics for routing in delay-tolerant networks suggests that an appropriate metric is to minimize the end-to-end delivery latency [19].

The intuition is that this minimizes the amount of time that a message consumes buffer space, and thus it should also maximize the delivery ratio since there is more space available for other messages. Their work uses a metric that is the time it will take for a message to be sent over each link. Since this value may depend on the time a message arrive sat a node, the authors present a time-varying version of Dijkstra's shortest path algorithm. Finding the path that delivers the message with the shortest delay has also been used for proactive mobile networks [25].Jain et al. present a variety of different metrics for networks with precise schedules, each of which require different amounts of information. However, all of the metrics assume that the contact schedule is precise. The first metric, called Minimum Expected Delay (MED), is the metric that requires the least amount of information. It assumes that the queuing time is zero, and that the average of the sum of transmission time, propagation delay, and waiting time is known precisely. This value is the expected delay: the average amount of time it takes for a message to go from one node to another, assuming that all arrival times are equally likely. The next metric is Earliest Delivery (ED). The queuing delay is assumed to be zero, and the propagation and transmission delays are assumed to be known precisely. Thepath is selected that will get the message to the destination at the earliest time. This requires the complete contact schedule, whereas MED only requires the average value of the waiting time. The next metric is Earliest Delivery with Local Queuing (EDLQ), which uses the buffer occupancy at each node to add an estimate of the queuing delay to the ED metric. Finally, the Earliest Delivery with All Queues (EDAQ) uses the information about the traffic demands for all nodes in order to compute the exact queuing delay. This paper shows that the protocols with more information have higher delivery ratios and lower delay. However, for some scenarios the difference between them is small.

The techniques presented by Jain et al. assume that accurate information about the complete contact schedules is known in advance. This may be feasible for scenarios where the connectivity is extremely predictable and reliable. Unfortunately, in reality, schedules may be imprecise or completely unpredictable. To address this problem, Jones et al. presented a metric called the Minimum Estimated Expected Delay (MEED), where the weights are based purely on observed connectivity [40]. They compute a metric based ona sliding window of observed connectivity. The assumptions that the future connectivity will be similar to the past.

To distribute the metrics throughout the network, they use an epidemic protocol to propagate link-state table updates. Their results show that in a scenario based on wireless LAN data [34], this technique approaches 95% of the delivery ratio of the ED metric. An interesting issue that arises when using shortest path routing with link metrics is when to make the routing decisions. The traditional choice is to make the decision about the entire path at the source (source routing), or to make the decision about the next hop when the message arrives(per-hop routing). If the link metrics do not change while the message is in transit, the paths selected by these options will be the same. This is true for networks where the end-to-end delays are very small or in networks where the contact schedules are precisely known. However, if the metrics are approximate and the messages take a long time to traverse the network, as is the case in delay-tolerant networks, the choice of when to make routing decisions may have a significant impact. Jones et al. argue that the best choice is to make decisions as late as possible, since that will allow messages to be forwarded using the most recent information [40]. To do this, they present what they callper-contact routing, where each node recomputed its routing table each time a contact becomes available, and then evaluates all the messages in its buffer to determine if they should be forwarded over the available links. This is computationally expensive, but always makes decisions with the most recent information. Their results show that these optimizations reduce the delivery latency. Handorean et al. also propose a very similar scheme, which they call a path update [31].

5. Open Research Problems

Delay and disruption tolerance is a novel emerging wireless networking paradigm that has not yet reached maturity. DTN Protocols are still in their infancy, large-scale DTN deployments do not exist at all, and small-scale deployments are still at their very early stages. These factors make real-life evaluation of present DTN modus operandi difficult. Indeed, rare are the protocols that were implemented, tested in real-life and proven to be free of lethal stealthy assumptions. But, the high costs and complexities of deploying real DTN test beds (e.g.: deep space networks) leave us currently with no other choice but to evaluate the performance of DTNs through simulation studies built around many unrealistic assumptions.

This paper surveys a carefully selected a set of recent works. Regard this, yet various DTN challenges prevail as open research topics. In what follows, this paper discusses some of the DTN-related topics that lend themselves nicely to further investigation. This paper start by listing what, are the most essential problems to which attention must be directed first. We believe that solutions to those problems will capture the essence of DTNs and allow for interesting advancements in the field, powerful and efficient DTN protocol developments.

A. Essential Challenging Problems

1) Analytical Modeling and Performance evaluation of DTNs may be one of the most important research areas. DTN characteristics vary from one environment to another. Thus the development of a generalized DTN model is quite a challenging problem.

2) None of the routing protocols proposed in the DTN open literature specifies a clear-cut procedure for setting up paths between communicating nodes. The separation between the control plane (i.e: determination of routes) and the forwarding (data) plane in the context of DTNs is clear. However, while significant efforts have been invested in handling forwarding issues, control has not yet gained a lot of attention.

3) Recall that DTN nodal contacts may be classified from highly deterministic to predictable all the way to absolutely unknown opportunistic. However, the less network information is available, the more the need for learning procedures increases. Such procedures are bandwidth consuming. The design of more intelligent, efficient and chattiness free network learning procedures is of particular interest.

4) When a bundle is received by its ultimate destination, its remaining replicas become useless. Instructing intermediate nodes to discard such copies requires additional resources. This is yet another form of the unresolved issue relating to the tradeoff between efficiency and overhead.

5) Bundle security is still at its early stages. No security standards have been defined yet. We expect a lot of future work to be done in this direction. What follows is a list of other persisting challenges that span the different DTN subsectors. None of the following problems has a priority over the others. They are listed arbitrarily with no particular order.

B. Other Persisting Problems

1) Erasure coding involves a lot of processing and hence requires more power. However, it was shown to improve the worst-case delay in [4]. This is particularly useful when applications require that bundles be delivered within a specific time interval. We expect more future investigations in this direction.

2) As a node becomes congested, incoming bundles may be dropped due to buffer overflow. This increases the dropping rate and adds network overhead as the forwarder inefficiently ends up consuming precious bandwidth to transmit of bundles that are dropped. An intelligent way to cope with this problem is to let a receiving node inform the forwarder of the probability p that inbound bundles are going to be dropped. Below a particular bandwidth occupancy threshold B, a receiving node always accepts incoming bundles. However when B is exceeded, p must increase. Analytical studies that determine an optimal value for B and that further explores the variability of p are future works of particular interest.

3) Due to short contacts and large inter-encounter intervals, some stored unexpired bundles may not have enough residual lifetimes for a contact to occur. In particular scenarios (e. g: deep space) bundles may even expire while propagating to their intended receivers. Early discarding of those bundles may significantly reduce buffer occupancy and appears as an interesting congestion control mechanism.

Finally, it is impossible to determine what approach is the right one when there are no real delay-tolerant networks. Perhaps the most important future work is to actually build DTNs and applications that use them, and then to see what routing problems occur in the real world. At the moment, there are a number of prototype DTNs that are being used to measure connectivity properties [11, 12, and 34]. These projects are an extremely important first step, however these networks are not yet being used for any real applications. It is difficult to predict the requirements for DTN routing strategies and to evaluate their performance, without any information about what traffic patterns would be relevant. Building more experimental deployments and applications is critical for being able to determine where DTN routing strategies need improvement. DTNs must be able to integrate multiple types of networks together. This means that techniques will be required that allow messages to be exchanged between DTNs that have different properties and possibly use different routing protocols. None of the work presented here addresses this issue. Additionally, there is an extremely important network that most DTNs will want to communicate with: The Internet. One proposal that enables communication between DTNs and the Internet is the Tether less Communication Architecture [36]. Any widely adopted DTN routing protocol will need to address these issues.

CONCLUSION

Delay-tolerant networks (DTNs) are a promising new development in network research, that offer the hope of connecting people and devices that hitherto were either unable to communicate, or could do so only at great cost. DTN is a network of regional networks i.e. it is an overlay on top of regional networks, including the Internet. It is an attempt to extend the reach of networks. This paper surveyed various existing techniques; protocols for routing in this type of networks. And also discovered a wide variety of methods that address the routing problem, which explained into two key properties: knowledge and replication.

This survey and classification provide following observations i.e. first, in cases where message volume is low, simple epidemic routing works extremely well. This suggests that small experimental deployments could be rapidly developed based on epidemic routing, allowing researchers to have real network topology and traffic data, which could be used to design new routing strategies. And second, to achieve a high delivery ratio with low resource consumption, hybrid techniques that rely on both knowledge about the topology and replication will be required. This has been implicitly noted by several of the researchers in the field, though the challenge is to determine the correct balance between redundancy and resource consumption, and to find manageable solutions for using network topology information.

REFERENCES

1. Hasson, D. R. Fletcher, and D. A. Pentland, “A road to universal broadband connectivity,” in Proceedings of Development by Design (dyd02), December 2002.
2. J. Ott and D. Kutscher, “Drive-thru internet: IEEE 802.11b for “automobile” users,” in Proceedings of IEEE INFOCOM, pp. 362–373, March 2004.
3. E. Brewer, M. Demmer, B. Du, M. Ho, M. Kam,S. Nedevschi, J. Pal, R. Patra, S. Surana, and K. Fall,“The case for technology in developing regions,” IEEEComputer, vol. 38, pp. 25–38, May 2005.
4. T. Small and Z. J. Haas, “The shared wirelessinfostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way),” inProceedings of ACM MobiHoc, pp. 233–244, June 2003.
5. S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf,B. Durst, K. Scott, and H. Weiss, “Delay-tolerant networking: an approach to interplanetary internet,”IEEE Communications Magazine, vol. 41,pp. 128–136, June 2003.
6. R. Y. Wang, S. Sobti, N. Garg, E. Ziskind, J. Lai, and A. Krishnamurthy, “Turning the postal system into a generic digital communication mechanism,” in Proceedings of ACM SIGCOMM, pp. 159–166, 2004.
7. T. Spyropoulos, K. Psounis, and C. S. Raghavendra, “Spray and wait: an efficient routing scheme for intermittently connected mobile networks,” in Proceedings of the ACM SIGCOMM Workshop on Delay-Tolerant Networking (WDTN’05), pp. 252–259, August 2005.
8. K. Fall, “A delay–tolerant network architecture for challenged internets,” in Proceedings of ACM SIGCOMM, pp. 27–34, August 2003.
9. A. Rabagliati, “Wizzy digital courier – how it works,” April 2004. http://www.wizzy.org.za/article/ article print/19/1/2/.
10. W. Zhao, M. Ammar, and E. Zegura, “Controlling themobility of multiple data transport ferries in adelay-tolerant network,” in Proceedings of IEEEINFOCOM, vol. 2, pp. 1407–1418 vol. 2, April 2005.
11. J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine,“Maxprop: Routing for vehicle-baseddisruption-tolerant networking,” in Proceedings ofIEEE INFOCOM, April 2006.
12. P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft,and C. Diot, “Pocket switched networks and humanmobility in conference environments,” in Proceedingsof the ACM SIGCOMM Workshop on Delay-TolerantNetworking (WDTN’05), pp. 244–251, August 2005.
13. J. N. Al-Karaki and A. E. Kamal, “Routing techniquesin wireless sensor networks: a survey,” IEEE WirelessCommunications, vol. 11, no. 6, pp. 6–28, 2004.
14. [14] S. Jain, M. Demmer, R. Patra, and K. Fall, “Using redundancy to cope with failures in a delay tolerant network,” in Proceedings of ACM SIGCOMM, pp. 109–120, October 2005.
15. Y. Wang, S. Jain, M. Martonosi, and K. Fall, “Erasure-coding based routing for opportunistic networks,” in Proceedings of the ACM SIGCOMM Workshop on Delay-Tolerant Networking (WDTN’05), pp. 229–236, August 2005.
16. D. Nain, N. Petigara, and H. Balakrishnan, “Integrated routing and storage for messaging applications in mobile ad hoc networks,” Mobile Networks and Applications (MONET), vol. 9, pp. 595–604, December 2004.
17. R. C. Shah, S. Roy, S. Jain, and W. Brunette, “Data mules: modeling a three-tier architecture for sparse sensor networks,” in Proceedings of Sensor Network Protocols and Applications, pp. 30–41, May 2003.
18. H. Jun, M. H. Ammar, and E. W. Zegura, “Powermanagement in delay tolerant networks: A framework and knowledge based mechanisms,” in Proceedings ofIEEE SECON, September 2005.

19. S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," in Proceedings of ACM SIGCOMM, vol. 34, pp. 145–158, ACM Press, October 2004.
20. M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," IEEE/ACM Transactions on Networking, vol. 10, pp. 477–486, August 2002.
21. R. H. Frenkiel, B. R. Badrinath, J. Borres, and R. D. Yates, "The infostations challenge: balancing cost and ubiquity in delivering wireless data," IEEE Personal Communications, vol. 7, no. 2, pp. 66–71, 2000.
22. A. Vahdat and D. Becker, "Epidemic routing for partially-connected ad hoc networks," Tech. Rep. CS-2000-06, Duke University, July 2000.
23. W. Zhao, M. Ammar, and E. Zegura, "A message ferrying approach for data delivery in sparse mobile ad hoc networks," in Proceedings of ACM Mobi Hoc, (New York, NY, USA), pp. 187–198, ACM Press, May 2004.
24. D. Marasigan and P. Rommel, "MV routing and capacity building in disruption tolerant networks," in Proceedings of IEEE INFOCOM, vol. 1, pp. 398–408, March 2005.
25. Q. Li and D. Rus, "Sending messages to mobile users in disconnected ad-hoc wireless networks," in Proceedings of ACM MobiCom, pp. 44–55, August 2000.
26. J. Widmer and J.-Y. Le Boudec, "Network coding forefficient communication in extreme networks," inProceedings of the ACM SIGCOMM Workshop onDelay-Tolerant Networking (WDTN'05), pp. 284–291,August 2005.
27. T. Small and Z. J. Haas, "Resource and performancetrade-offs in delay-tolerant wireless networks," inProceedings of the ACM SIGCOMM Workshop onDelay-Tolerant Networking (WDTN'05), pp. 260–267,August 2005.
28. A. Lindgren, A. Doria, and O. Schel´en, "Probabilistic routing in intermittently connected networks," Lecture Notes in Computer Science, vol. 3126, pp. 239–254, January 2004.
29. A. Demers, D. Greene, C. Houser, W. Irish, J. Larson,S. Shenker, H. Sturgis, D. Swinehart, and D. Terry,"Epidemic algorithms for replicated databasemaintenance," SIGOPS Operating Systems Review,vol. 22, pp. 8–32, January 1988.
30. J. A. Davis, A. H. Fagg, and B. N. Levine, "Wearablecomputers as packet transport mechanisms inhighly-partitioned ad-hoc networks," in Proceedings ofInt'l Symp. on Wearable Computers (ISWC'01),pp. 141–148, October 2001.
31. R. Handorean, C. Gill, and G.-C. Roman, "Accommodating transient connectivity in ad hoc and mobile settings," Lecture Notes in Computer Science, vol. 3001, pp. 305–322, January 2004.
32. K. Tan, Q. Zhang, and W. Zhu, "Shortest pathrouting in partially connected ad hoc networks," inProceedings of Global Telecommunications Conference(GLOBECOM), vol. 2, pp. 1038–1042, December 2003.
33. P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh,and D. Rubenstein, "Energy-efficient computing forwildlife tracking: design tradeoffs and earlyexperiences with zebranet," in Proceedings ofASPLOS-X, pp. 96–107, October 2002.
34. Dartmouth College, The Dartmouth Wireless TraceArchive. http://crawdad.cs.dartmouth.edu/.
35. K. A. Harras and K. C. Almeroth, "Transport layer issues in delay tolerant mobile networks," in Proceedings of IFIP-TC6 Networking, May 2006.
36. A. Seth, P. Darragh, S. Liang, and S. Keshav, "Anarchitecture for tetherless communication." University of Waterloo, unpublished manuscript,http://blizzard.cs.uwaterloo.ca/keshav/home/Papers/data/05/tca.pdf, June 2005.
37. K. A. Harras, K. C. Almeroth, and E. M. Belding-Royer, "Delay tolerant mobile networks (DTMNs): Controlled flooding Networking, vol. 3462, pp. 1180–1192, May 2005.
38. M. Musolesi, S. Hailes, and C. Mascolo, "Adaptive routing for intermittently connected mobile ad hoc networks," in Proceedings of IEEE WoWMoM, pp. 183–189, June 2005.
39. J. Lebrun, C.-N. Chuah, D. Ghosal, and M. Zhang, "Knowledge-based opportunistic forwarding in vehicular wireless ad hoc networks," in Proceedings of IEEE Vehicular Technology Conference (VTC), vol. 4, pp. 2289–2293, May 2005.
40. E. P. C. Jones, L. Li, and P. A. S. Ward, "Practical routing in delay-tolerant networks," in Proceedings of the ACM SIGCOMM Workshop on Delay-Tolerant Networking (WDTN'05), pp. 237–243, August 2005.
41. [41] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in Proceedings of ACM MobiCom, pp. 243–254, August 2000.
42. M. Mauve, A. Widmer, and H. Hartenstein, "A survey on position-based routing in mobile ad hoc networks," IEEE Network, vol. 15, no. 6, pp. 30–39, 2001.
43. T. S. E. Ng and H. Zhang, "Predicting internet network distance with coordinates-based approaches," in Proceedings of IEEE INFOCOM, vol. 1, pp. 170–179 vol.1, June 2002.
44. J. Leguay, T. Friedman, and V. Conan, "Evaluating mobility pattern space routing for DTNs," in Proceedings of IEEE INFOCOM, April 2006.
45. T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Single-copy routing in intermittently connected mobile networks," in Proceedings of Sensor and Ad Hoc Communications and Networks (SECON), pp. 235–244, October 2004.
46. R. D. Poor, "Gradient routing in ad hoc networks." MIT Media Laboratory, unpublished manuscript http://www.media.mit.edu/pia/Research/ESP/texts/poorieeepaper.pdf, 2000.