Vehicular Disruption Tolerant Networks

During the alert phase it is of paramount importance to reach the vast majority of the concerned population as fast as possible. Our solution makes use of Disruption, or Delay, Tolerant Network (DTN) techniques to improve the message spreading process. This chapter presents a broad overview of DTNs, particularly focusing on Vehicular DTNs (VDTNs). Even though the proposed techniques could also be implemented in a Packet Switched Network the main characteristics and challenges remain basically the same for both Packet Switched and VDTN.

3.1 Introduction

Traditional networks suppose the existence of some path between end-points, short end-to-end round-trip delay time and small loss ratio. Today, however, new applications, environments and types of devices are challenging these assumptions. In Disruption Tolerant Networks, also called sometimes opportunistic networks, an end-to-end path from source to destination may not exist. In this environment nodes can still connect and exchange information, but in an opportunistic way.

Delay Tolerant Networks have been developed as an approach to building architecture models which are tolerant to long delays and/or disconnected network partitions when delivering data to destinations. This chapter discusses the characteristics of these architectures, and many of the protocols developed to ensure packet delivery in these networks. We henceforth use DTN to refer to Delay Tolerant Networking, Disruption Tolerant Networks and Opportunistic Networks. For Vehicular DTN, the acronym VDTN is used.

The vehicular network research field, and more specifically the VDTN research field, have attracted great attention in the last few years. Initiatives such as the i2010 Intelligent Car Initiative Intelligent Car (2009) aim to decrease the number of accidents and CO_2 emissions in Europe, utilizing sensors and vehicle-to-vehicle (V2V) communication. As part of these projects, cars equipped with wireless devices will exchange traffic and road safety information with nearby cars and/or roadside units. In fact, according to the ETSI 102 638 technical report [41], by 2017 20% of the running vehicles will have wireless communication capabilities. The same report estimates that by 2027 almost 100% of the vehicles will be equipped with communication devices.

The design of the core Internet protocols is based on a number of assumptions: these include the existence of some path between endpoints, short end-to-end round-trip delay time, and the perception of packet switching as the right abstraction for end-to-end communications. Furthermore, the efficiency of these protocols is based on assumptions about the resources available to the nodes and the properties of the links between them. Traditionally nodes are considered to be fixed, energy unconstrained, connected by low loss rate links, and communication occurs through the exchange of data between two or more nodes. Today, however, new applications, environments and types of devices are challenging these assumptions and call for new architectures and modes of node operation. Some of these challenges are: intermittent and/or scheduled links, very large delays, high link error rates, energy-constrained devices, with heterogeneous underlying network architectures and protocols in the protocol stack, and most importantly, the absence of an end-to-end path from a source to a destination. Applications on the following environments may pose such challenges: spacecrafts, planetary/interplanetary, military/tactical, disaster response, mobile sensors, vehicular environments, satellite and various forms of large scale ad hoc networks. The variety of these applications, the impossibility of having a fixed wired Internet infrastructure everywhere, and the inclusion of mobility in most of these applications, make these challenges more difficult to surmount. This leads us to a new approach of designing networks, taking into account several constraints and characteristics, using DTN.

3.2 Background

VDTNs have evolved from DTNs and are formed by cars and supporting fixed nodes. Fall [42] is one of the first authors to define and discuss DTNs' potential. According to his definition, a DTN consists of a sequence of time-dependent opportunistic contacts. During these contacts, messages are forwarded from their source towards their destination. This is illustrated in Figure 3.1. In the first contact the origin sends the message to A in time t_1, then A holds the message until it is delivered to the destination in the contact at time t_2.

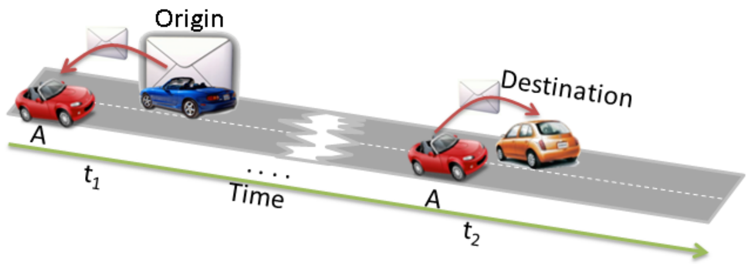

Figure 3.1: A representation of two communication opportunities in a VDTN

Contacts are characterized by their start and end times, capacity, latency, end points and direction. The routing algorithm can use these pieces of information to decide the most appropriate route(s) to deliver a message from its source to its destination. However, routing in a network where the edges among the mobile nodes depend on and vary in time is not a straightforward task. One needs to find an effective route, both in time and space. All nodes along the path should consider the nodes' movement pattern and the possible communication opportunities for message forwarding. Unfortunately, it is not always easy to determine future communication opportunities or even forecast the mobility patterns of the nodes in the network.

Cerf et al. [26] characterizes contacts as:

- Persistent: When they are always available, i.e. a Cable modem.
- On-Demand: When they require an action to start, but after that they work as persistent contacts, for example, a dial-up connection.
- Intermittent scheduled: When the parties involved agree to meet at a

specific location for a determined period of time, i.e., low earth satellite communication window.

- Intermittent opportunistic: Contacts that occur unexpectedly, for example, a car passing by in a non scheduled manner.
- Intermittent predicted: When the contacts are not based on a schedule, but on predictions. A prediction is a contact that is likely to happen, based on the history or other kind of information.

  However, there are no guarantees predicted contacts will actually happen. For example, when people are commuting to work, it is probable that at the same time the same contacts are available, because people normally go at the same time and take the same routes.

Forwarding and routing strategies may vary significantly according to the type of contacts a node, or a network, is expected to encounter. In the case of a DTN with intermittent contact opportunities, the main priority is to maximize the probability of message delivery, and to minimize the end-to-end delay. For networks with more stable and consistent contact opportunities, it is important to discover an efficient path while trying to save as much as possible of the network resources. In a scenario with deterministic message routing and persistent, on-demand or intermittent but scheduled contacts, we may have a chance to achieve optimal performance of the network and manage efficiently the available resources, i.e. spectrum and node energy. However, this is not frequently the case for mobile networks.

Under unpredictable intermittent network conditions, where the mobility obscures present and future topology, nodes can only forward packets randomly based on the likelihood they will eventually arrive to their destination. Then, the problem of delivering messages to their final destination is paramount and dominates that of resource utilization. In this case, flooding and epidemic message-forwarding are popular approaches. Between the two extremes, deterministic contacts and fully opportunistic ones, a broad range of strategies may be used to balance message delivery and resource optimization. Another issue to keep in mind is the duration and bandwidth of contacts. In a vehicular network, the number of contacts may be high, but the duration of each one can often be expected to last only seconds, especially between cars moving in opposite lanes. This significantly limits the amount of information exchanged between nodes.

In 2002 the Internet Research Task Force (IRTF) [63], started a new group called Delay-Tolerant Networking Research Group (DTNRG) [40]. The group was first linked to the Interplanetary Internet Research Group

(IPNRG) [62], however, it soon became clear that the main characteristics of DTNs, i.e. non-interactive, asynchronous communication, would be useful in a broader range of situations. The main aim of DTNRG is to provide architectural and protocol solutions to enable interoperation among nodes in extreme and performance-challenged environments where the end-to-end connectivity may not exist. E.g. public safety, underwater, sensor and ad-hoc networks and extremely degraded connectivity, such as country side networks, to name a few.

3.3 Challenges and techniques

The conditions of DTN operation lead to an architecture that challenges the traditional concepts of most of the network layers. In this section we present some of the major challenges faced by DTN protocols at different network layers. Standard network modeling techniques are also challenged and new ways to model nodes and connections should be created to evaluate the considered protocols. Therefore this section also discusses the different network modeling, traffic modeling, transport layer issues, routing and data dissemination strategies.

3.3.1 Routing

The challenges that DTNs need to overcome have lead the research to heavily focus on routing issues. Routing is considered to be the problem of choosing forwarding strategies that enable messages to pass from the source to the destination. The issues presented in this section pertain to most of the network layers and techniques developed for DTNs. For the case of VDTNs, *store-and-forward*, or *store-carry-and-forward* techniques are used [99]. This means that the nodes which receive a message, store it for some time, possibly carry it to another location, and afterwards forward it to other nodes. This is not new, in the Internet nodes also often momentarily buffer packets as well. However, in this case, nodes try to decrease as much as possible these time intervals. On the other hand, in DTN the storing is used to overcome absence of end-to-end connectivity, and to enable waiting until efficient connections are present. In the *store-and-forward* mechanism each intermediate node is in charge of verifying the integrity of the message before forwarding it. In general, this technique helps us cope with intermittent connectivity, especially in the wilderness or environments requiring high mobility, and may be preferable in situations of long delays in transmission and variable or high error rates.

Mundur and Seligman [79] identify mainly two classes of routing algorithms for DTNs. The first class is based on epidemic routing, in which nodes use opportunistic contact to infect other nodes with the message to be delivered. For this group, the need of network knowledge is minimal. The routing algorithms have no control of node mobility and the forwarding process occurs in a fortuitous way. The second class of algorithms utilizes topology information and the algorithms may control node mobility. For Mundur and Seligman [79], this case is characterized by "islands" of well-connected nodes with intermittent connectivity with other nodes.

Routing issues

Fall [42] and Jain, Fall and Patra [64] present an interesting list of routing issues for DTNs:

- Routing objective: Although the main objective of a routing algorithm is message delivery and DTNs are, by definition, tolerant to delay, that does not mean we should not try to decrease the delay as much as possible. Algorithms should attempt to find a good tradeoff between decreasing the end-to-end delay and saving network resources.

- Reliability: The protocols should be reliable and provide some form of mechanism to inform the nodes that their messages reached the destination. Acknowledged message delivery is an important enhancement of the offered set of services.

- Security: In all types of networks, security is an important factor. However, in DTNs the packets may cross a diverse path to reach the destination and stay for a relatively long time at each intermediate node. The reliability and intentions of the often numerous intermediate nodes may not be always the best ones. Mechanisms to provide message authentication and privacy of the messages' content are of supreme importance.

- Resource allocation: Normally the main routing objectives of maximizing the message delivery ratio and minimizing resource allocation are conflicting. The easiest way to guarantee the message delivery in the smallest amount of time is flooding the network with the message. However, this means a high use of network bandwidth, nodes memory and processing power. These may lead to other problems such as packet collisions, packet drops because of full message queues and surely the waste of the limited amount of energy of the nodes.

- Buffer space: Considering the disconnection problem, messages may be stored for a long period of time before they can be forwarded. The buffer space must be enough to maintain all the pending messages, i.e. messages that have not reached their final destination yet.
- Contact scheduling: The forwarding waiting time is one of the principal elements on DTNs. It is not always clear how long a node will need to keep a message to enable its forwarding. This period may vary from seconds to days.
- Contact capacity: Not only is it not always possible for the contacts to be predicted, but when they do occur, they may be brief. The protocols should take this into account and try to minimize, as much as possible, the use of the spectrum and time with control messages.
- Energy: Mobile nodes may have limited amount of energy and, possibly, difficult access to power sources. Normally, for VDTNs the energy is a factor to be kept in mind but it is not one of the main factors since the vehicle can normally provide enough energy to maintain the communication system.

Evaluation metrics

To evaluate the routing algorithms for DTNs Jones [67], and Sanchez, Franck and Beylot [90], propose the utilization of:

- Delivery ratio: Jones [67] defines delivery ratio as "the fraction of generated messages that are correctly delivered to the final destination within a given time period".
- Latency: Even though the networks and applications are supposed to endure delays, many applications could take advantage of shorter delays. Even more, some applications have time windows of delay resilience, i.e. messages are valid for a certain amount of time, after that the message loses its validity.
- Transmissions: The number of messages transmitted by the algorithms varies and some, that create multiple copies of the message, may send more messages than others.
- Lifetime: Route lifetime is the time a route can be used to forward packets without the need for re-computation.

- End-to-end delay: This evaluation criterion is the time it takes for one message to go from the origin to the destination.
- Capacity: Capacity is the amount of data that may pass through one route during its lifetime.
- Synchronicity: Even in a delay tolerant network, it is possible that, during some intervals, origin and destination are close and the communication may occur directly, or similarly to communication in traditional wireless networks; Synchronicity measures how long this situation where classical communication is possible.
- Simultaneousness: This criteria measures the contact durations, i.e. the time intermediate nodes are in the same area.
- Higher-order simultaneousness: Simultaneousness is computed hop-by-hop. However, the same concept may be applied to a series of nodes. Higher-order simultaneousness is the application of the simultaneousness criteria to k consecutive nodes that are part of the complete path.
- Discontinuity: Discontinuity is the normalized duration of packet storage through the path.

Routing strategies

Recently, Shen, Moh and Chung [95] presented a compact and interesting list of routing strategies for DTNs. Like Mundur and Seligman [79] Shen, Moh and Chung [95] also divide the routing protocols in two families; flooding and forwarding. Flooding strategies are the ones where nodes create copies of the packet and forward to more than one node. Forwarding strategies use the knowledge of the network to select the best paths. A comparison between the generic behavior of flooding and forwarding strategies is depicted in Figure 3.2. Note that flooding strategies result in a significantly higher number of messages compared to forwarding.

Flooding based strategies

One of the simplest possible forwarding strategies is called Direct Contact. In this strategy the node waits until the source comes in contact with the destination before forwarding the data. Jones [67] considers direct contact as a degenerate case of a forwarding strategy. Even though this strategy does not multiply messages, it is considered flooding. The reason is that it does not make use of any topology information the nodes possess. The strategy is simple and presents low resource consumption; however, if the contact

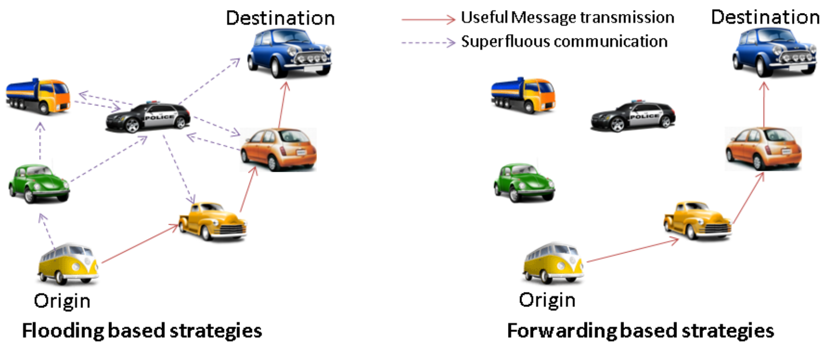

Figure 3.2: Message transmission example comparing flooding and forwarding based strategies

opportunities between source and destination are low, then the delivery rate can also be low.

In the Two hop Relay strategy [67], the source copies the message to the first n nodes that it contacts. These nodes relay the message until they find the destination, it is similar to direct contact, but now not only the source keeps the message, but also n copies of the message are spread among other nodes. With this we increase the required resources, but also the expected delivery ratio. Figure 3.3 illustrates both techniques.

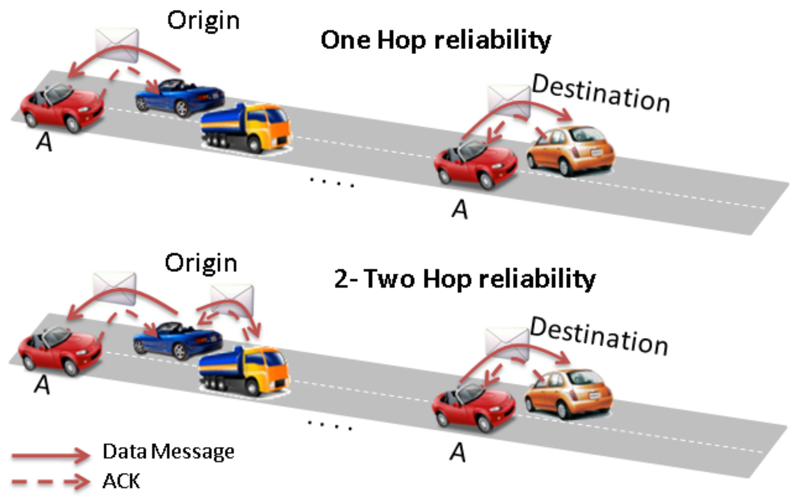

Figure 3.3: One hop reliability and two hop reliability techniques

The tree based flooding strategy of [67] extends the idea of direct contact further in the sense that now all nodes that receive the message may create n copies of it. The message tends to propagate through the network in a controlled flooding that resembles a tree.

Epidemic routing [105] consists of the spreading of the message similarly to that of a virus in an epidemic situation. Each node that receives the message rebroadcasts it to every other node it encounters. The contaminated nodes just keep one copy of the message. This approach is extremely effective, but presents a high resource consumption rate.

Ramanathan et al. [87] present a prioritized version of epidemic routing. This technique imposes a partial ordering on the messages based on costs to destination, source, and expiry time. The costs are derived from link availability information. The technique successfully maintains a gradient of replication density that decreases with increasing distance from the destination. Even though it is also a resource intensive technique it presents lower costs and higher delivery rates than simple epidemic routing.

Forwarding based strategies

Location based routing [67] techniques use geographical information, such as Global Positioning System (GPS) data, to forward data. This strategy is the forwarding one that demands the smallest amount of knowledge of the network structure. With the position information they can estimate the costs and direction to forward the messages.

Source routing strategies calculate the whole path at the origin, prior to sending the packet. This kind of strategy needs to have a fairly consistent view of the network to work properly. On the other hand, in per-hop routing [67], the decisions of which path to take are done on a hop-by-hop basis when the message arrives at each hop. Instead of computing the next hop for each message the per-contact routing technique recomputes its routing table each time a contact arrives and its knowledge of the network increases.

Instead of routing with global contact knowledge, Liu and Wu [71] propose a simplified DTN model and a hierarchical routing algorithm which routes on contact information with three combined methods.

3.3.2 Data dissemination

Data dissemination refers to data-centric communications protocols. The Data Mule project [94] and the Message Ferrying scheme [103], are two of the most well-known data dissemination algorithms for DTNs. They were designed for sensor networks. They propose the use of mobile nodes to collect data from the sensors, buffer it, and deliver the collected data to a sink.

The MULEs (Mobile Ubiquitous LAN Extensions) and ferries utilize nodes navigating through the sensor network to collect data in 'mobile caches'. According to the Data Mule project, all the nodes are fixed and only the cache is mobile. Message Ferrying [103] also considers mobile nodes, but in this approach the nodes are required to follow specific paths and even move in order to help message delivering.

The SPAWN protocol introduced by Das et al. [37] and Nandan et al. [80] discusses how vehicles should interact to accommodate swarming protocols, such as BitTorrent traffic. In SPAWN, the nodes passing through Access Points (APs) collect data that they subsequently exchange with nearby nodes. Nodes are often required to carry traffic useless to them and the BitTorrent protocol is bandwidth intensive, however, swarming protocols is an interesting and effective way for message dissemination among nodes in VDTNs.

3.3.3 Transport issues

The greatest part of the research for DTNs has focused on routing and data dissemination algorithms. However, many other aspects present interesting and valuable challenges. The transport layer is certainly one of the layers that need special attention. Most of the services offered by existing transport layer protocols, such as TCP, have been ignored. For example, end-to-end connections, sequencing, congestion control and reliability are some of the most important features of the TCP protocol. Some of these services may be easily implemented in DTNs while others will require a fair amount of future research. We will focus here on reliability approaches to ensure message delivery on the DTNs.

Hop-by-hop reliability [42] is the most basic and simple reliability strategy to ensure data delivery in DTNs. Each time a node receives a message, it sends an acknowledgement (ACK) of its reception and after that assumes the responsibility for this message across a defined region. For this case an end-to-end ACK is not possible, unless it is a completely new message generated by the destination. The lack of end-to-end reliability of the hop-by-hop approach may be a problem for a series of applications. One way to overcome this problem is the use of Active Receipt [52]. Active receipt is basically an end-to-end acknowledgment created by the destination, addressed to the source of the original message. The receipt is actively sent back through the network. In truth it is a new message that is propagated through the network. Active receipts solve the problem of end-to-end reliability but the price to pay for it may be too high in some situations.

Passive Receipt [52] is another method created to provide end-to-end reliability at a lower cost. The high price of the Active Receipt comes from the generation of two messages in the network instead of just one. To use the terminology of epidemic routing, now we have two messages infecting nodes instead of just one. In this case what Passive Receipt introduces is exactly the concept of an implicit receipt, instead of an active one. The destination, instead of creating a new active receipt, creates an implicit kill message for the first one. The kill message works as a cure for the infected nodes, when they receive this message they know that the message arrived to the destination and that they do not need to rebroadcast the original message anymore. This can be observed in Figure 3.4 - Passive receipt, where the police car does not rebroadcast message 1 after having received message 2. The message is rebroadcast only if the cured nodes meet other node that is re-broadcasting the original message, which is the case of the truck and all previous nodes on that specific path of Figure 3.4. This technique presents a lower flux of messages than the one generated by the active receipt, and the end-to-end reliability is guaranteed, since eventually the source will also receive the passive receipt.

An interesting solution for end-to-end reliability is also proposed by [52] and takes advantage of the number of multiple network infrastructures available nowadays. On the Network-Bridged Receipt approach the nodes may use a different medium access mode to deliver ACKs. For example, while the cell phone network may not present the required data rate for a specific application, or even present a high cost. The cell network may present more than reasonable bandwidth, at a cost effective, to send small ACK messages.

Figure 3.4 presents a schematic description of the reliability approaches presented here. We can see that the number of messages and nodes involved in each one of the techniques vary considerably.

3.4 Modeling techniques for VDTNs

Analytical studies perform an important role in the evaluation and, in consequence, in the development of protocols in every area. Vehicular delay tolerant networks are no different. However, as the constraints of DTNs are somehow particular, compared to traditional wired and wireless networks, the same analytical models and constraints used for the latter may not hold for a DTN environment. An analytical model, or study, "is a proven approach for studying system performance, revealing underlying characteristics, and evaluating communication protocols" [107]. Theoretical works, like the one

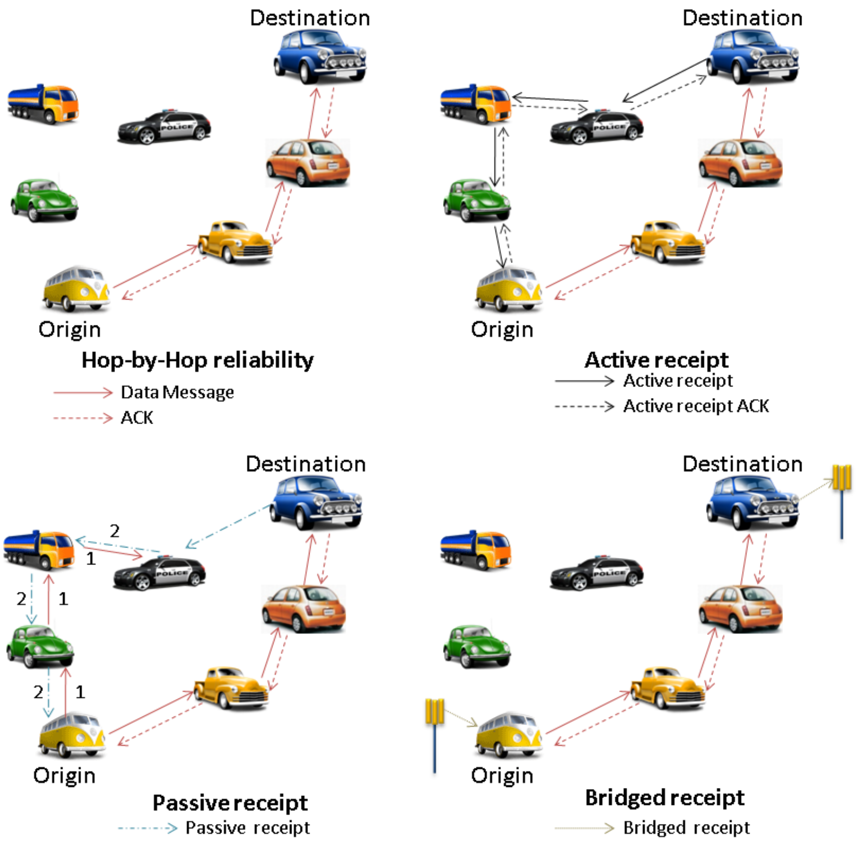

Figure 3.4: Comparison among the reliability approaches messages

of Niyato, Wang, and Teo [83], provide the indication and comparison basis for other simulation or test-beds experiments. Many factors may influence the analytical results of an experiment, e.g. node density, capacity, physical and medium access control characteristics. However the three main factors are: mobility model, data delivery scheme and queue management [107].

3.5 Mobility models

Different DTNs may have different mobility models and mobility directly influences the network structure. The way nodes move, or do not move, may affect: the retransmission delays, frequency of contacts among nodes and energy decay. Mobility can either provide the opportunity for new high quality contact or lead to the breaking of links already established.

Apart from static placement of nodes, probably the simplest mobility model is the Random Walk-based model [117]. In this mobility model nodes choose random points in the area considered and move towards these points at random speeds. The three basic steps for a random way point algorithm, as described by Bettstetter et al. [15] are: first the node chooses randomly a destination, after that it goes towards that destination at a random speed, and finally it waits for a random period of time at the destination point. Some minor variants of this process are also possible, for example Spyropoulos et al. [100] consider random directions instead of positions, but in the end the basic concept is the same.

Some techniques use different well known distributions to control the movement of the nodes. The main advantages of using these distribution based mobility models are that, not only the mathematical model of a well known distribution is easy to implement, but also it is easy to analyze the network behavior afterwards. For example, knowing the nodes distribution makes it easy to calculate the probability of a node crossing a specific network area. Some commonly used distributions are: Normal, poison and exponential.

Markovian mobility models are also a popular choice to model mobility. The main goal of using Markov chains is to create more realistic movement models [32] [22] with real drivers actions, such as motion in the same direction and in adjacent directions, acceleration, stops and sharp turns.

Another model designed to provide realistic mobility patterns, introduced by Haerri, Bonnet and Filali [51] is Kinetic Graphs. This method tries to capture the dynamics of mobile structures and accordingly develop an efficient maintenance for them. Unlike static graphs, kinetic graphs are assumed to be continuously changing and edges are represented by time-varying weights. Kinetic graphs are a natural extension of static graphs and provide solutions to similar problems, such as convex hulls, spanning trees or connected dominating sets, but for continuously mobile networks. This mobility model is implemented in a tool called VanetMobiSim [45], that can generate realistic mobility patterns.

3.6 Delivery schemes

Direct transmission and flooding [106] are two of the most simple delivery schemes possible. In direct transmssion a node simply transmits the message directly to the destination. In flooding schemes it transmits the message to all other nodes it may encounter. The analysis of both schemes is simple since the node behavior is straightforward to predict.

Epidemic dissemination schemes are also extremely popular for VDTNs. For example, the Shared Wireless Infostation Model (SWIM) presents an epidemic Markov dissemination scheme [98]. The scheme is further analyzed and refined in [99]. Wang et al. [107] present a more diverse description of dissemination models.

3.7 Queue management

The way nodes manage their queues is also a determining component in the performance of algorithms for VDTNs. The way one models the queues determines, among others, the way nodes will discard old messages and this in consequence will, possibly, affect the network delivery ratio. The generic queuing analytic framework introduced by Wang et al. [107] is a good starting point for a simple queue model for VDTNs. The models described by Wang have either infinite or finite buffer space. For the infinite buffer space the node's queue is considered to have infinite length. For the finite buffer space it is assumed that each node may hold at most k messages in the queue.

Niyato et al. [83] present an analytical queuing model based on discrete time Markov chains. This work also proposes models for queue performance measures for VDTNs. The proposed performance measures are: Average Number of Packets in Queue of a Mobile Router, Throughput and Average Packet Delivery Delay.

3.8 Applications

One of the main focuses of research in VDTNs in the last few years has been the use of VDTNs in road safety applications. Research such as Xu et al. [111] evaluates the feasibility of using dedicated short range communication to warn vehicles about road accidents. Yang et al. [113] propose the use of V2V to warn vehicles about road conditions and demonstrate the potential of DTNs for real life applications.

3.9 Conclusions

DTN is a young and expanding field. VDTN has also a huge potential because of the imminent appearance of vehicular devices capable of wireless communications. These will operate in a very demanding environment, with intermittent connectivity, where an end-to-end path may not always be present. Even though routing and data dissemination have been the focus of research, areas such as security, topology management, transport layer issues, and higher protocol level concerns are equally important. They present problems that will need to be addressed in the near future. DTN and in particular VDTN is an attractive research field exactly because, in order to achieve the envisioned future of ubiquitous connectivity, we need a solution for these open problems.