Efficient Routing for Safety Applications in Vehicular Networks

METRANS Project DTRS98-G0019

March 2009

Konstantinos Psounis

Electrical Engineering

University of Southern California

Los Angeles, CA 90089

METRANS
Transportation Center
USC | CSULB
10 YEARS OF EXCELLENCE
1998 – 2008

1 Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, and California Department of Transportation in the interest of information exchange. The U.S. Government and California Department of Transportation assume no liability for the contents or use thereof. The contents do not necessarily reflect the official views or policies of the State of California or the Department of Transportation. This report does not constitute a standard, specification, or regulation.

2　Introduction

Vehicular ad hoc networks have received a lot of attention in recent years. This attention is due to two reasons. First and foremost, there are a number of real-life applications that become possible in the presence of such an ad-hoc infrastructure. Examples include increasing road safety by reducing the number of accidents as well as reducing their impact in case of non-avoidable accidents, improving local traffic ow and efficiency of road traffic, and offering comfort and business applications to driver and passengers. Second, it is now technically possible to build such a network. Recent developments in radios, coupled with signicant research work in the area of mobile ad-hoc networks, make it likely to build such applications within five to ten years.

While there has been significant effort to define applications, see for example, the Car to Car Communication Consortium [1], the Vehicle Safety communications Project of the Department of Transportation [4], and the PReVENT project [5], there are still some hard technical challenges that need to be resolved. Perhaps the hardest of them all is how to achieve communication in an environment where network nodes (vehicles) move so fast that the very concept of a wireless link between two nodes is meaningless for time scales larger than a few seconds, and where the density of the nodes can vary drastically in space and time, making the network intermittently connected. The fast mobility renders any proactive routing protocols, that establish end-to-end paths between sources and destinations, useless. The intermittent connectivity renders reactive protocols, that establish end-to-end paths upon demand, non-applicable either.

To address this challenge, we propose using a new approach of routing that is tailored to the needs of vehicular ad hoc networks and is termed as mobility-assisted routing. Mobility-assisted routing departs drastically from the traditional view of networking: When a node (moving vehicle or a static roadside station) wants to send a message to one or more nodes (vehicles), it may transmit a number of copies of the message to one or more distinct relay nodes. Each relay will carry the message further, and may transmit it to a new, better relay or directly to a destination.

The first routing protocol of that type that comes to mind is flooding, according to which whenever two vehicles are within range, they exchange all messages that they dont have in common [41]. The main argument for such an approach is that while flooding clearly wastes some network resources, the majority of VANET applications require the messages to reach a large number of vehicles anyway. Further, since the network can be disconnected, sending the data to everybody should reduce delivery delays. However, recent studies have shown that flooding creates so much contention for the wireless channel, that its performance is, in practice, quite bad. There have a been a number of attempts to alleviate this problem. In [33] the authors examine a number of different schemes to suppress redundant transmissions after a message has been delivered by flooding. In [40, 43] a message is forwarded to another node with some probability smaller than one, i.e. data is gossiped instead of flooded. In [14, 27–29] simple methods to take advantage of the history of past encounters are implemented in order to make fewer and more informed forwarding decisions than flooding. Finally, it has also been proposed that ideas from network coding could be useful to reduce the number of bytes transmitted by flooding [42]. Although all these schemes, if carefully tuned, can improve to an extent the performance of flooding, they are still flooding-based in nature, and thus often exhibit the same shortcomings as flooding [38, 39].

We propose a different approach than flooding that signicantly reduces its overhead, while achieving good performance. The idea is to distribute only a bounded number of copies to a number of relay-vehicles, each of which can then deliver it to the destination or to a new, better relay-vehicle. We refer to these schemes as spraying-based schemes. Spraying schemes keep the number of transmissions small while exploiting the speed of flooding.

To design the optimal spraying scheme, we address the following important questions:

(i) **How many copies should a scheme spray:** We analyze how to choose the number of copies sprayed by a scheme to achieve a given delay. We derive analytical expressions expressing the expected delay in terms of the network parameters, and discuss how to solve for the number of copies. In vehicular

networks, most of the times its not possible to know the network parameters like the number of cars on the highway. So we also describe an online algorithm to estimate the network parameters. Finally, to show that spraying schemes scale, we show that as the number of nodes in the network increases, the percentage of nodes that need to become relays in spraying schemes, in order to achieve the same relative performance, is actually decreasing.

(ii) **How to route each copy:** Once the copies have been sprayed, how does each relay route this copy towards the destination. We propose the use of the single-copy utility-based scheme from [37] for this purpose. Each node maintains a timer for every other node in the network, which records the time elapsed since the two nodes last encountered each other. These timers are similar to the *age of last encounter* in [17], and are useful, because they contain indirect (relative) location information. We show that using these timers or other similar utility functions to route each copy leads to significant performance improvement in the context of vehicular networks. We also discuss how to modify the utility functions to incorporate the presence of roadside stations which have been installed specifically to help delivery in vehicular networks.

(iii) **How to distribute copies:** The choice of spraying method directly affects the expected delay of spraying phase. Further, this delay is independent of the particular single-copy routing scheme that is used to route each copy in the second phase. We first show that if node movements are independent and identically distributed (IID), then allowing each relay to give away half of its copies till it has only one remaining is the optimal strategy. We label this strategy binary spraying. We then show that if node movements are not IID, but instead, each node has an utility associated for each destination, then if this utility function is also used to route each copy through a single copy utility-based scheme, then binary spraying still remains the optimal strategy.

Up till now, we ignore contention in the analysis. Incorporating wireless contention complicates the analysis significantly. This is because contention manifests itself in a number of ways, including (i) finite bandwidth which limits the number of packets two nodes can exchange while they are within range, (ii) scheduling of transmissions between nearby nodes which is needed to avoid excessive interference, and (iii) interference from transmissions outside the scheduling area, which may be significant due to multipath fading [8]. To analyze how do the answers to the previous three questions change if we incorporate contention, we first propose a general framework to incorporate contention in the performance analysis of mobility-assisted routing schemes for ICMNs while keeping the analysis tractable. We then use this framework to derive delay expressions for spraying schemes and use these expressions to understand whether and how do our previous results change?

Our objective is to design highly efficient routing schemes for vehicular ad hoc networks (VANETs), that are tailored to supporting real-life safety-related applications. Hence, we want to understand how the proposed routing algorithms work with realistic vehicle mobility. To accomplish this goal, we first propose a new mobility model which captures the essential characteristics of human-driven mobility. The proposed model is a *time-variant community mobility model*, and is referred to as the TVC model. Using empirical traces, we first show that the TVC model captures the statistics observed in vehicular traces. Then we derive delay expressions for spraying based schemes for a specific instantiation of the proposed mobility model. Finally, we use these expressions to show that spraying schemes achieve very good performance with realistic vehicle mobility too.

We also propose a new protocol to enable one-to-many communication while suppressing duplicate transmissions. Finally, we use showcase applications to demonstrate the applicability and efficacy of the proposed protocols. The end-result of this work is a library of protocols, which we label spraying schemes, which offer a reliable and efficient method of routing messages between vehicles and between vehicles and roadside stations, and support a wide range of safety applications.

3 Optimal Design of Spraying Scheme

In this section, we discuss the problem of efficient routing in vehicular networks, and describe our proposed solution, Spray routing. Our problem setup consists of a number of nodes (vehicles) moving inside a bounded area (city) according to a stochastic mobility model. Additionally, we assume that the network is disconnected at most times, and that transmissions are faster than node movement (i.e. it takes less time to transmit a message x meters far - ignoring queueing delay - than to carry it for the same distance)[1]).

Our study of single-copy routing algorithms [37] showed that using only one copy per message is often not enough to deliver a message with high reliability and relatively small delay in a vehicular network. On the other hand, routing too many copies in parallel, as in the case of flooding-based schemes (e.g. epidemic routing or gossiping), can often have disastrous effects on performance [26]. The total transmissions performed by epidemic routing are orders of magnitude higher than those performed by an optimal scheme. So, under low traffic loads epidemic routing achieves close to optimal delays, but as the traffic input increases it begins to suffer severely from contention and its delay very quickly increases.

Based on these observations, we have identified the following desirable design goals for a routing protocol in vehicular networks. Specifically, an efficient routing protocol in this context should:

- perform significantly fewer transmissions than flooding-based routing schemes, under all conditions.

- generate low contention, especially under high traffic loads.

- deliver a message faster than existing single and multi-copy schemes, and exhibit close to optimal delays.

- deliver the majority of the messages generated;

Additionally, we would like this protocol to also be:

- highly scalable, that is, maintain the above performance behavior despite changes in car density.

- simple, and require as little knowledge about the network as possible, in order to facilitate its implementation.

3.1 Spray and Wait

Since too many transmissions are detrimental on performance, especially as the network size increases, the proposed protocol, *Spray and Wait*, distributes only a small number of copies each to a different relay. Each copy is then "carried" all the way to the destination by the designated relay.

Binary Spray and Wait Binary Spray and Wait routing consists of the following two phases:
- *spray phase*: for every message originating at a source node, L message copies are initially spread to L distinct relays. The source of a message initially starts with L copies; any node A that has $n > 1$ message copies (source or relay), and encounters another node B (with no copies), it hands over to B $\lfloor n/2 \rfloor$ of its copies and keeps $\lceil n/2 \rceil$ for itself; when it is left with only one copy, it switches to the wait phase.

- *wait phase*: if the destination is not found in the spraying phase, each of the L nodes carrying a message copy performs "Direct Transmission" [37] (i.e. will forward the message only to its destination).

---

[1]This is reasonable assumption with modern wireless devices. Assume, for example, that a node has a range of $100m$ and a radio of $1Mbps$ rate. Then, it could send a packet of $1KB$ at a distance of $100m$ in only $8ms$. Even if that node is a fast moving car with a speed of say $65mph$, it could carry the same packet at a mere distance of less than $1m$ in the same $8ms$.

Binary Spray and Wait decouples the number of transmissions per message from the total number of nodes. Thus, transmissions can be kept small and essentially fixed for a large range of scenarios. Additionally, its mechanism combines the speed of epidemic routing with the simplicity and thriftiness of direct transmission. Initially, it "jump-starts" spreading message copies quickly in a manner similar to epidemic routing. However, it stops when enough copies have been sprayed to guarantee that at least one of them will reach the destination, with high probability. Since cars move quickly around the network and "cover" a sizeable part of the network area in a given trip, we will show that *only a small number of copies can create enough diversity to achieve close-to-optimal delays*.

As we mentioned earlier, the basic idea behind Binary Spray and Wait (i.e. extending the 2-hop scheme of [20] to introduce more than one relays) is relatively simple and has been identified as beneficial by other researchers also [15, 31, 33]. However, a number of important questions need to be answered first, before the desirable performance can be achieved: (i) How many copies should a scheme spray? (ii) How should these copies be distributed to different vehicles and roadside stations, i.e is it possible to do better than binary spraying? (iii) How should each of these copies be routed, i.e. is waiting for the destination after spraying the best strategy?

3.2   Deciding the Right Number of Copies

In this section, we analyze how to choose the number of copies used (denoted by L) in order to achieve a specific expected delay. Let us assume that there is a specific delivery delay constraint to be met. One reasonable way to express such a constraint would be as a factor a times the optimal delay ED_{opt} ($a > 1$), since this is the best that any routing protocol could do[2].

We first state theorems which express the expected delay of optimal routing and spray and wait in terms of the network parameters. Throughout this section, we will be making the following assumptions:

Network: M nodes move on a $\sqrt{N} \times \sqrt{N}$ 2-dimensional torus. Each node can transmit up to distance $K \geq 0$ meters away, where $K/\sqrt{N}$ is much smaller than the value required for connectivity [22], and each message transmission takes one time unit.

Mobility Models: We assume that all nodes move according to some stochastic mobility model ("MM"). We next define a mobility property. The statistics of this property will be used in the expected delay expressions for different routing scheme.

Meeting Time Let nodes i and j move according to a mobility model 'mm' and start from their stationary distribution at time 0. Let $X_i(t)$ and $X_j(t)$ denote the positions of nodes i and j at time t. The meeting time (M_{mm}) between the two nodes is defined as the time it takes them to first come within range of each other, that is $M_{mm} = min_t\{t : \|X_i(t) - X_j(t)\| \leq K\}$.

We assume that the "meeting times" of the mobility model "mm" is approximately exponentially distributed or has an exponential tail, with expected meeting time equal to EM_{mm}. It has been shown that a number of popular mobility models like Random Walk [9], Random Waypoint and Random Direction [33, 35], as well as more realistic, synthetic models which are suitable to model contacts between moving vehicles [24] exhibit such (approximately) exponential encounter characteristics. Therefore, the subsequent analysis and algorithms of this and the following section apply to all these models.

Contention: Throughout our analysis we assume that bandwidth and buffer space are infinite. In other words, we assume that there is no contention for these resources. Later sections address how do the results presented in this section after incorporating contention in the analysis.

The following theorem states the expected delivery time of the optimal algorithm.

---

[2]By this, we do not assume that ED_{opt} is always known to the user. If ED_{opt} is not known a could still be used as a measure of how "aggressive" the protocol should be.

Theorem 3.1 *The expected message delivery time of the optimal algorithm ED_{opt} is given by*

$$ED_{opt} = \frac{H_{M-1}}{(M-1)} ED_{mm}, \tag{1}$$

where H_k is the kth Harmonic Number, i.e, $H_k = \sum_{i=1}^{k} \frac{1}{i} = \Theta(\log k)$.

We next state the expected end-to-end delay of Binary Spray and Wait. After the L copies have been sprayed, each of the L relays will independently look for the destination to directly deliver the message (if the latter has not been found yet). We first state the delay of the wait phase in the following Lemma.

Lemma 3.1 *Let EW denote the expected duration of the "wait" phase, if needed, and let EM_{mm} denote the expected meeting time under the given mobility model. Then, EW is given by*

$$EW = \frac{EM_{mm}}{L}. \tag{2}$$

The following theorem calculates the expected delivery time of Binary Spray and Wait. It defines a system of recursive equations that calculates the (expected) residual time after i copies have been spread, in terms of the time until the next copy($i+1$) is distributed, plus the remaining time thereafter. It is important to note that the following result is generic. By plugging into the equations the appropriate meeting time value EM_{mm}, we can calculate the expected delay of Spray and Wait for the respective mobility model [35].

Theorem 3.2 *Let $ED_{sw}(L)$ denote the expected delay of the Binary Spray and Wait algorithm, when L copies are spread per message. Let further $ED(i)$ denote the expected remaining delay after i message copies have been spread. Then, $ED(1) \approx ED_{sw}(L)$, where $ED(1)$ can be calculated by the following system of recursive equations:*

$$
\begin{aligned}
ED(i) &= \frac{EM_{mm}}{i(M-i)} + \frac{M-i-1}{M-i} ED(i+1), \, i \in \left[1, \frac{L}{2}\right]; \\
ED(i) &= \frac{ED_{mm}}{i(M-i)} + \frac{M-i-1}{M-i} \left(\frac{2i-L}{i} ED(i) + \frac{L-i}{i} ED(i+1)\right), \textit{for } i \in \left[\frac{L}{2}+1, L-1\right]; \\
ED(L) &= EW = \frac{EM_{mm}}{L}.
\end{aligned}
$$

The above result, albeit quite useful in accurately predicting the performance of Binary Spray and Wait, is not in closed form. This makes it difficult to theoretically compare the performance of Binary Spray and Wait to that of the optimal scheme, or to calculate the number of copies to be used in closed form. For this reason, in the following lemma we also derive an upper bound that is in closed form, by assuming that Source Spray and Wait is performed, that is, only the source can forward a new copy. Note that Source Spray and Wait always has a larger delay than Binary Spray and Wait.

Lemma 3.2 *The following upper bound holds for the expected delay of Binary Spray and Wait:*

$$ED_{sw} \leq (H_{M-1} - H_{M-L}) EM_{mm} + \frac{M-L}{M-1} EW, \tag{3}$$

where H_n is the nth Harmonic Number, i.e, $H_n = \sum_{i=1}^{n} \frac{1}{i} = \Theta(\log n)$.

This bound is tight for a small L/M ratio, but becomes pessimistic as this ratio grows larger. This is because the bound basically includes the full time until all copies are spread, regardless of whether the destination is found in one of the initial steps of the spraying phase. However, when the number of copies is

Table 1: minimum L to achieve expected delay

a	1.5	2	3	4	5	6	7	8	9	10
recursion	21	13	8	6	5	4	3	3	3	2
bound	N.A.	N.A.	11	7	6	5	4	3	3	2
taylor	N.A.	N.A.	10	7	5	4	3	3	3	2

much smaller than the total number of nodes (which is the case of most interest) this bound is very useful when tuning the performance of Spray and Wait.

The following lemma states that the required number of copies only depends on the number of nodes, and is straightforward to prove from Eq.(3) or Theorem 3.2.

Lemma 3.3 *The minimum number of copies L_{min} needed for Binary Spray and Wait to achieve an expected delay at most aED_{opt} is independent of the mobility model, the size of the network N, and transmission range K, and only depends on a and the number of nodes M.*

The required number of copies $L_{min}(M)$ for Binary Spray and Wait to achieve a desired expected delay can be calculated in any of the following three ways: (i) solve the system of equations of Theorem 3.2 for increasing L, until $ED_{sw}(L) < aED_{opt}$, or (ii) solve the upper bound equation Eq.(3) for L, by letting $ED_{sw} = aED_{opt}$, and taking $\lceil L \rceil$, or (iii) approximate the harmonic number H_{M-L} in Eq.(3) with its Taylor Series terms up to second order, and solve the resulting third degree polynomial:

$$(H_M^3 - 1.2)L^3 + (H_M^2 - \frac{\pi^2}{6})L^2 + \left(a + \frac{2M-1}{M(M-1)}\right)L = \frac{M}{M-1}, \tag{4}$$

where $H_n^r = \sum_{i=1}^n \frac{1}{i^r}$ is the nth Harmonic number of order r.

Method (i) is obviously the most accurate one. However, it is also the most cumbersome. Since the upper bound of Eq.(3) is tight for small L/M values, if the delay constraint a is not too tight, we can use method (ii) or (iii) to quickly get a good estimate for L_{min}.

In Table 1 we compare results for L_{min}, as calculated with each of these three methods for different values of a. We assume the number of nodes M equals 100. 'N.A' stand for 'Non Available' and means that such a low delay value is never achievable by the bound. As can be seen in this table the L found through the approximation is quite accurate when the delay constraint is not too stringent.

3.2.1 Estimating L when Network Parameters are Unknown

Throughout the previous analysis we've assumed that network parameters, like the total number of nodes M, are known. This assumption might be valid in networks operated by a single authority (e.g. sensor networks), however, this assumption will not hold for vehicular networks. So, we next describe how to produce and maintain good estimates of necessary network parameters, like M, and adapt L accordingly.

This problem is difficult in general. A straightforward way to estimate M would be to count unique IDs of nodes encountered already. However, this method requires a large database of node IDs to be maintained in large networks, and a lookup operation to be performed every time any node is encountered. Furthermore, although this method converges eventually, its speed depends on network size and could take a very long time in large disconnected vehicular networks. A better alternative is to produce an estimate of M by taking advantage of inter-meeting time statistics. Specifically, let us define T_1 as the time until a node (starting from the stationary distribution) encounters *any* other node. It is easy to see from Lemma 3.2 that T_1 is exponentially distributed with average $T_1 = EM_{mm}/(M-1)$. Furthermore, if we similarly define T_2 as the time until two *different* nodes are encountered, then the expected value of T_2 equals $EM_{mm}\left(\frac{1}{M-1} + \frac{1}{M-2}\right)$. Cancelling EM_{mm} from these two equations we get the following estimate for M:

$$\hat{M} = \frac{2T_2 - 3T_1}{T_2 - 2T_1}. \tag{5}$$

Estimating M by the procedure above presents some challenges in practice, because T_1 and T_2 are ensemble averages. Since hitting times are ergodic [9], a node can collect sample intermeeting times $T_{1,k}$ and $T_{2,k}$ and calculate time averages $\hat{T}_1$ and $\hat{T}_2$ instead. However, the following complication arises: when a node i meets another node j, i and j become *coupled* [18]; in other words, the next intermeeting time of i and j is not anymore exponentially distributed with average EM_{mm}. In order to overcome this problem, each node keeps a record of recently encountered nodes. Every time a new node is encountered, it is stamped as "coupled" for an amount of time equal to the *mixing* or *relaxation* time for that graph, which is the expected time until a node starting from a given position arrives to its stationary distribution [9]. Then, when node i measures the next sample intermeeting time, it ignores all nodes that it's coupled with at the moment, denoted as c_k, and scales the collected sample $T_{1,k}$ by $\frac{M-c_k}{M-1}$. A similar procedure is followed for $\hat{T}_2$. Putting it altogether, after n samples have been collected:

$$
\begin{aligned}
\hat{T}_1 &= \frac{1}{n} \sum_{k=1}^{n} \left(\frac{M - c_k}{M - 1} \right) T_{1,k}, \\
\hat{T}_2 &= \frac{1}{n} \sum_{k=1}^{n} \left[\left(\frac{M - c_{k-1}}{M - 1} \right) T_{1,k-1} + \left(\frac{M - c_k}{M - 2} \right) T_{1,k} \right].
\end{aligned}
$$

Replacing $\hat{T}_1$ and $\hat{T}_2$ in Eq.(5) we get a current estimate of M. As can be seen by Eq.(5), the estimator for M is sensitive to small deviations of T_1 and T_2 from their actual values. Therefore it is useful for a node to also maintain a running average of M. Specifically, the running estimate $\hat{M}$ is updated with every new estimate $\hat{M}_{new}$ as $\hat{M} = \beta\hat{M} + (1-\beta)\hat{M}_{new}$ ($0 < \beta < 1$, with values closer to 1 providing better stability). We could now use this estimate of M to calculate the number of copies using one of the previous methods.

Figure 1 shows how the online estimate $\hat{M}$, calculated with our proposed method, quickly converges to its actual value for a 200×200 network with 200 nodes, for both the random walk and random waypoint models, again validating the generality of our expressions. (Note that even in this small scenario, our method's convergence is more than two times faster than ID-counting.) Finally, both our method and ID-counting could take advantage of indirect information learning, where nodes exchange known unique IDs or independently collected samples to speed up convergence.

We believe that similar estimators could potentially be constructed for other network parameters or statistics, as well, (e.g. approximate network area N, or various moments for encounter times) which could be used to provide users with predictions of the service level available. We intend to look further into this issue in future work.

3.3 Scalability of Spray and Wait

Having shown how to find the minimum number of copies L_{min} to achieve a delay at most a times the optimal, it would be interesting, from a scalability point of view, to see how the percentage L_{min}/M of nodes that need to receive a copy behaves as a function of M. The reason for this is the following: If we assume a large enough TTL (time-to-live) value is used, flooding-based schemes will eventually give a copy to every node and therefore perform at least M transmissions. Increased contention and the resulting retransmissions will obviously increase this value significantly. On the other hand, Spray and Wait performs L transmissions, and produces very little contention compared to flooding-based schemes. Consequently, the number of transmissions that Spray and Wait performs per message is at most a fraction L_{min}/M of the number of transmissions per message epidemic and other flooding-based schemes perform.
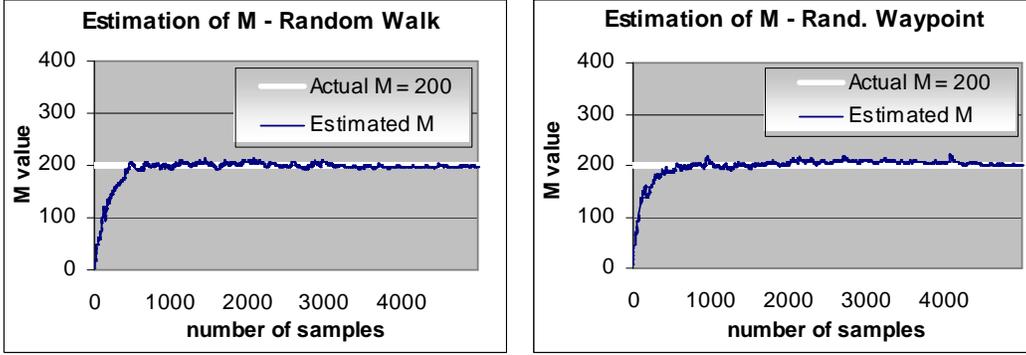
Figure 1: Online estimator of number of nodes (M) — $N = 200 \times 200$, transmission range $= 0$, $\beta = 0.98$, mixing time $= 4000$.

In Figure 2 we depict the behavior of L_{min}/M as a function of M for different values of a. It is important to note there that, *as the number of nodes in the network increases, the percentage of nodes that need to become relays in Spray and Wait, in order to achieve the same relative performance, is actually decreasing*. The intuition behinds this interesting result is the following: when $L \ll M$ the delay of Spray and Wait is dominated by the delay of the wait phase; in that case, if L/M is kept constant, the delay of Spray and Wait decreases roughly as $1/M$ (as $M \to \infty$). On the other hand, the delay of the optimal scheme (and also the spraying delay) decreases more slowly as $\log(M)/M$ [34]. The following Lemma formally states the result.
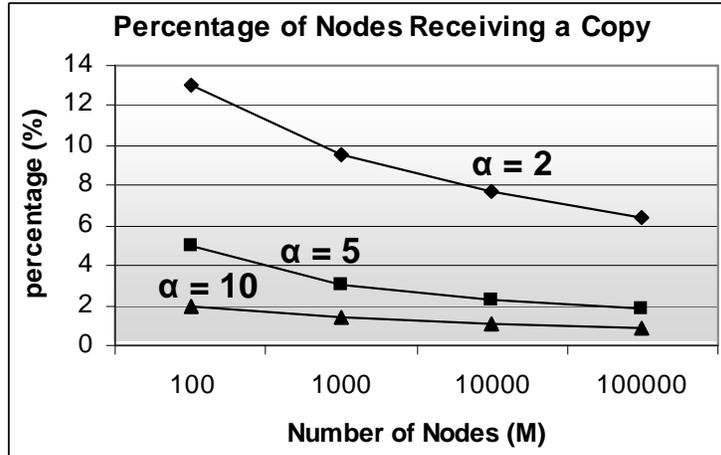

Figure 2: Required percentage of nodes L_{min}/M receiving a copy for spray and wait to achieve an expected delay of aED_{opt}

Lemma 3.4 *Let L/M be constant and let $L \ll M$. Let further $L_{min}(M)$ denote the minimum number of copies needed by Spray and Wait to achieve an expected delay that is at most aED_{opt}, for some a. Then $\frac{L_{min}(M)}{M}$ is a decreasing function of M.*

This behavior of L_{min}/M implies that Spray and Wait is extremely scalable. While, usually, the performance of many schemes (including flooding-based ones, in our case) deteriorates as the number of nodes increase, the relative performance of Spray and Wait improves, making its performance advantage even more

pronounced in large networks. This property is a must for a vehicular network in a large metropolitan area like Los Angeles, where the number of vehicles is expected to be very large.

3.4 Routing Each Copy Separately - "Spray and Focus" Routing

Although Binary Spray and Wait combines simplicity and efficiency, it can be optimized further. Consider a vehicular network in which vehicles move closely within separate, and often sparsely located groups. In such situations, partial paths may exist over which a message copy could be quickly transmitted closer to the destination. Yet, in Spray and Wait a relay with a copy will naively wait until it moves within range of the destination itself. This problem could be solved if some other single-copy scheme is used to route a copy after it's handed over to a relay, a scheme that takes advantage of transmissions (unlike Direct Transmission).

We propose the use of the single-copy utility-based scheme from [37] for this purpose. Each node maintains a timer for every other node in the network, which records the time elapsed since the two nodes last encountered each other[3] (i.e. came within transmission range). These timers are similar to the *age of last encounter* in [17], and are useful, because they contain indirect (relative) location information. Specifically, for a large number of vehicular mobility models, it can be shown that a smaller timer value on average implies a smaller distance from the node in question. Further, we use a "transitivity function" for timer values (see details in [37]), in order to diffuse this indirect location information much faster than regular last encounter based schemes [17]. The basic intuition behind this is the following: in most situations, if node B has a small timer value for node D, and another node A (with no info about D) encounters node B, then A could safely assume that it's also probably close to node D. We assume that these timers are the *only* information available to a node regarding the network (i.e. no location info, etc.).

We have seen in [37] that appropriately designed utility-based schemes, based on these timer values, have very good performance in scenarios were mobility is low and localized. This is the exact situation were Spray and Wait loses its performance advantage. Therefore, we propose a scheme were a fixed number of copies are spread initially exactly as in Spray and Wait, but then each copy is routed independently according to the *single-copy* utility-based scheme which uses a utility function based on these timers. We call our second scheme *Spray and Focus*.

Spray and Focus Spray and Focus routing consists of the following two phases:

- *spray phase*: for every message originating at a source node, L message copies are initially spread – by binary spraying – to L distinct "relays".

- *focus phase*: let $U_X(Y)$ denote the utility of node X for destination Y; a node A, carrying a copy for destination D, forwards its copy to a new node B it encounters, if and only if $U_B(D) > U_A(D) + U_{th}$, where U_{th} (utility threshold) is a parameter of the algorithm.

3.4.1 Evaluation of Spraying Schemes

We have used a custom discrete event-driven simulator to evaluate and compare the performance of different routing protocols under a variety of mobility models and under contention. A slotted collision detection MAC protocol has been implemented in order to arbitrate between nodes contenting for the shared channel. The routing protocols we have implemented and simulated are the following: (1) Epidemic routing ("epidemic"), (2) Randomized flooding with $p = (0.02 - 0.1)$ ("random-flood"), (3) Utility-based flooding ("utility-flood"), (4) Optimal (binary) Spray and Wait ("spray&wait"), (5) Spray and Focus ("spray&focus"), (6) Seek and Focus single-copy routing ("seek&focus") [34], and (7) Oracle-based Optimal routing ("optimal"). (We will use the shorter names in the parentheses to refer to each routing scheme in simulation plots.)

---

[3]In practical situations, each node would actually maintain a cache of the most recent nodes that it has encountered, in order to reduce the overhead involved in a large network.

We choose the number of copies L for Spray and Wait according to the theory of Section 3.1. (Specifically, such that the delay of Spray and Wait would be about $2\times$ that of the Oracle-based Optimal if the nodes were performing random walks.) For Spray and Focus and all other protocols we have tried to tune their parameters in each scenario separately, in order to achieve a good transmissions-delay tradeoff. Finally, in all schemes that use a utility function, including Utility-based flooding, we have used our own utility function proposed in [37], which has been shown to perform better than existing utility functions [29] for most mobility models.

We first evaluate the effect of traffic load on the performance of different routing schemes (Scenario A). We then examine their performance as the level of connectivity changes (Scenario B).

Scenario A - Effect of Traffic Load: 100 nodes move according to the random waypoint model [13] in a 500×500 grid with reflective barriers. The transmission range K of each node is equal to 10. Finally, each node is generating a new message for a randomly selected destination with an increasing rate resulting in average traffic loads (total number of messages generated throughout the simulation) from 200 (low traffic) to 1000 (high traffic).

Fig. 3 depicts the performance of all routing algorithms, in terms of total number of transmissions and average delivery delay. Epidemic routing performed significantly more transmissions than other schemes (from 56000 to 144000), and at least an order of magnitude more than Spray and Wait. Therefore, we do not include it in the transmission plots, in order to better compare the remaining schemes. We also depict two plots for Spray and Wait for two different L values, in order to gain better insight into the transmissions-delay tradeoffs involved. Finally, note that, in this scenario, Spray and Focus had similar performance with Spray and Wait, and thus we don't include results for it. In the next section, we will see in detail scenarios where Spray and Focus can significantly improve the performance of Spray and Wait.

As is evident by Fig. 3, Spray and Wait outperforms all single and multi-copy protocols discussed and achieves its performance goals set at the start of this section. Specifically: (i) under low traffic its delay is similar to Epidemic routing and is $1.4 - 2.2$ times faster than all other multi-copy protocols; it performs an order of magnitude less transmissions than Epidemic routing, and $5 - 6$ times less transmissions than Randomized and Utility-based, and (ii) under high traffic it retains the same advantage in terms of total transmissions, and outperforms *all* other protocols, in terms of delay, by a factor of $1.8 - 3.3$.

As a final note, the delivery ratio of almost all schemes in this scenario was above 90% for all traffic loads, except that of Seek and Focus which was about 70%, and that of Epidemic routing which plummeted to less than 50% for very high traffic, due to severe contention.
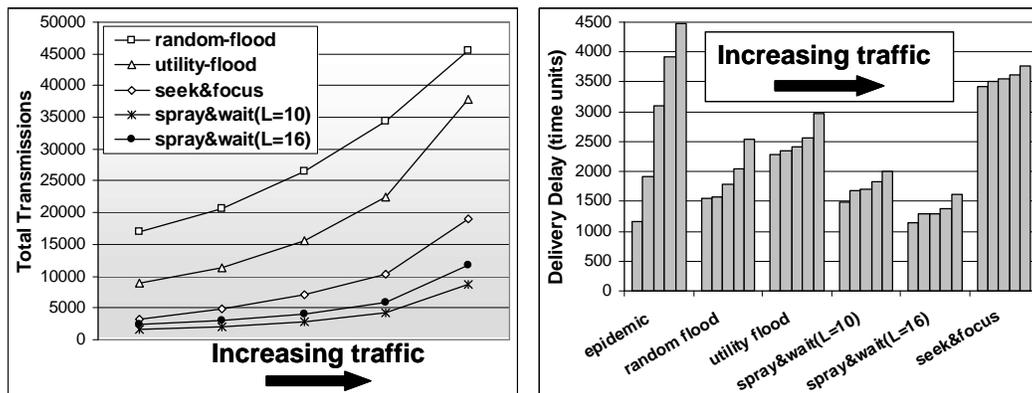

Figure 3: Scenario A - performance comparison of all routing protocols under varying traffic loads.

Scenario B - Effect of Connectivity: In this scenario, the size of the network is 200×200 and the traffic load is medium. We would like to evaluate the performance of all protocols in networks with a large range

of connectivity characteristics, ranging from very sparse, highly disconnected networks, to *almost* connected networks.

Before we proceed, it is necessary to define a meaningful connectivity metric. Although a number of different metrics have been proposed (for example [16]), no widespread agreement exists, especially if one needs to capture both disconnected and connected networks. We believe that a meaningful metric for the networks of interest is the expected *maximum cluster size* defined as the percentage of total nodes in the largest connected component (cluster). This indicates what percentage of nodes have already conglomerated into the connected part of the network, with "one" implying a regular connected network (with high probability).

The above connectivity metric measures "static" connectivity. It indicates how connected a random snapshot of the connectivity graph will be. However, in situations where mobility is exploited to deliver traffic end-to-end, "dynamic" connectivity also plays an important role on performance. Dynamic connectivity can be seen as a measure of how many new nodes are encountered by a given node within some time interval. If nodes move in an IID manner, this is directly tied to the mixing time for the graph representing the network [9]. The larger the mixing time, the more "localized" the node movement, and the longer it will take a node to *carry* a message to a remote part of the network.

In order to evaluate the effect of dynamic connectivity on different protocols, we present two sets of results, one where nodes move according to the random waypoint model and one where nodes perform random walks. The random waypoint has one of the fastest mixing times ($\Theta(\sqrt{N})$), while the random walk has one of the slowest ($\Theta(N)$) [9]. Furthermore, for each mobility model we vary the transmission range K to span the entire static connectivity range.

Figure 4 and Figure 5 depict the number of transmissions and the average delay for the random waypoint and the random walk scenarios, respectively, as a function of transmission range (respective connectivity values are shown in the parentheses).

There are a number of interesting things to notice about these plots. First, although Randomized and Utility Flooding can improve the performance of epidemic routing they still have to perform way too many transmissions to achieve competitive delays. Further, when nodes move according to the random waypoint model, Spray and Wait outperforms all protocols, in terms of both transmissions and delay, for all levels of connectivity. Its performance is close to the optimal, and thus Spray and Focus cannot offer any improvement. On the other hand, when nodes perform random walks, Spray and Wait may exhibit large delays, if the network area is large enough. Here the few copies are spread locally, and then each custodian takes a very long time to traverse the network and reach the destination. Even if the number of copies were increased, it would be the spraying phase that would take a long time, since new nodes are found very slowly. Spray and Focus can overcome these shortcomings and excel (when the network is not too sparse), achieving the smallest delay with only a few extra transmissions. Note though that despite the better utility function used, Utility Flooding is still plagued by its flooding nature and choice of threshold. This problem was even more pronounced when other existing utility functions were used.

Finally, both Spray and Wait and Spray and Focus are quite scalable and robust, compared to other multi-copy or even single-copy options. Epidemic routing and the rest of the schemes manage to achieve a delay that is comparable to the spraying schemes for very few connectivity values only, but perform quite poorly for the vast majority of scenarios. Spray and Wait and Spray and Focus, on the other hand, exhibit great stability. They performs few transmissions across all scenarios, while achieving a delivery delay that decreases as the level of connectivity increases, as one would expect.

3.5   Distributing Copies

In this section, we study how to distribute these L copies. The choice of spraying method directly affects the expected delay of spraying phase. Further, this delay is independent of the particular single-copy routing scheme that is used to route each copy in the second phase.
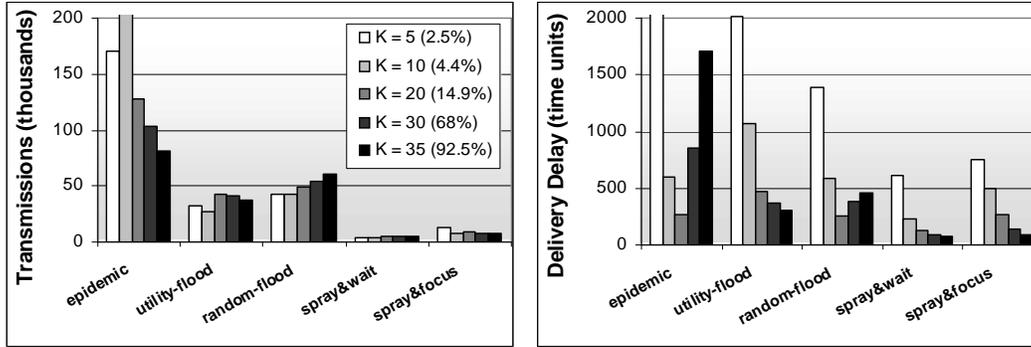
Figure 4: Scenario B - Random Waypoint Mobility: Total transmissions and delay as a function of transmission range K (respective connectivity values are shown in parentheses).
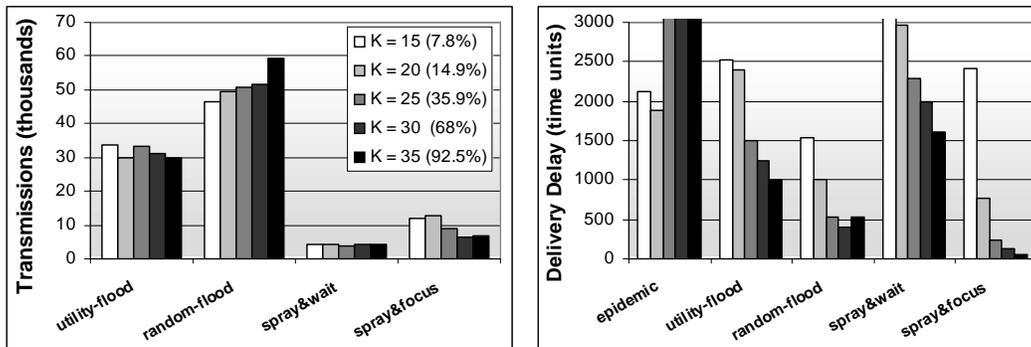

Figure 5: Scenario B - Random Walk Mobility: Total transmissions and delay as a function of transmission range K (respective connectivity values are shown in parentheses).

13

We first state the following theorem which formally shows that binary spraying is optimal when node movement is independent and identically distrbuted (IID).

Theorem 3.3 *When all nodes move in an IID manner, Binary Spraying minimizes the expected time until all copies have been distributed.*

Proof: Let us call a node "active" when it has more than one copies of a message. Let us further define a spraying algorithm in terms of a function $f : N \rightarrow N$ as follows: when an active node with n copies encounters another node, it hands over to it $f(n)$ copies, and keeps the remaining $n - f(n)$. Any spraying algorithm (i.e. any f) can be represented by the following binary tree with the source as its root: assign the root a value of L; if the current node has a value $n > 1$ create a right child with a value of $n - f(n)$ and a left one with a value of $f(n)$; continue until all leaf nodes have a value of 1.

A particular spraying corresponds then to a sequence of visiting all nodes of the tree. This sequence is random. Nevertheless, *on the average*, all tree nodes at the same level are visited in parallel. Further, since only active nodes may hand over additional copies, the higher the number of active nodes when i copies are spread, the smaller the residual expected delay until all copies are spread. Since the total number of tree nodes is fixed ($2^{1+\log L} - 1$) for any spraying function f, it is easy to see that the tree structure that has the maximum number of nodes at every level, also has the maximum number of active nodes (on the average) at every step. This tree is the balanced tree, and corresponds to Binary Spraying. □

Now, if the node movements are not IID, but instead, each node has an utility associated for each destination, which is the most common case in vehicular networks, how does the spraying phase gets modified? We first find the optimal spraying policy under the following set of assumptions, and later discuss what do our results imply for general vehicular networks.

(i) M nodes perform independent random walks on a $\sqrt{N} \times \sqrt{N}$ 2D torus (finite lattice). Each node moves one grid unit in one time unit.

(ii) Each node can transmit up to $K \geq 0$ grid units away, where $\frac{K}{\sqrt{N}}$ is much smaller than the value required for connectivity [22]. We use Manhattan distance $d_{ab} = \|a_x - b_x\| + \|b_x - b_y\|$ to measure proximity between two positions a and b (or between two nodes).

(iii) There is no contention in the network. In other words, the buffer space is infinite, and any communicating pair of nodes do not interfere with any other simultaneous transmission.

(iv) Let the number of copies distributed by the spraying based schemes be denoted by L.

We next state a lemma which will be used in the derivation of the optimal spraying policy.

Lemma 3.5 *Let $E[M(d)]$ denote the expected time until two independent random walks, starting at a distance d from each other, first meet each other. $E[M(d)]$ can be derived by solving the following set of linear equations:*

$$E[M(d)] = \begin{cases} p_{d,d-2}E[M(d-2)] + p_{d,d} \\ E[M(d)] + p_{d,d+2}E[M(d+2)] & d > K \\ 0 & d \leq K \end{cases}, \tag{6}$$

where p_{d_1,d_2} denotes the probability that the two walks are at a distance d_2 from each other in the next time slot given they are at a distance d_1 from each other in the current time slot and, for $d_1 > 3$, it equals

$$\begin{cases} \frac{16d_1-20}{64d_1} & d_2 = d_1 - 2 \\ \frac{16d_1+12}{64d_1} & d_2 = d_1 + 2 \\ \frac{32d_1+8}{64d_1} & d_2 = d_1 \end{cases}, \text{ for } d_1 = 3, \text{ it equals } \begin{cases} \frac{7}{48} & d_2 = 1 \\ \frac{15}{48} & d_2 = 5 \\ \frac{26}{48} & d_2 = 3 \end{cases}, \text{ for } d_1 = 2, \text{ it equals } \begin{cases} \frac{3}{32} & d_2 = 0 \\ \frac{11}{32} & d_2 = 4 \\ \frac{18}{32} & d_2 = 2 \end{cases},$$

for $d_1 = 1$, it equals $\frac{7}{16}$ and $\frac{9}{16}$ for $d_2 = 3$ and $d_2 = 1$ respectively and for $d_1 = 0$, it equals $\frac{4}{16}$ and $\frac{12}{16}$ for $d_2 = 2$ and $d_2 = 0$ respectively.

Now we present an algorithm which will answer the following question: 'Two nodes A and B are within range of each other and A has $l \leq L$ copies of a packet while B has none. The utility of both the nodes is known. Then how many of the l copies should A give to B such that the expected delivery delay is minimized.' Before we proceed, we first specify the utility function we will use. Amongst the different utility functions used in the literature (see [34]), we choose 'the distance to the destination' for our analysis.

Now we derive the algorithm to find the optimal spraying policy. Let a node (label it node A) be a distance d from the destination and has l copies of the packet. Let $D(d, l)$ denote the time this node will take to deliver the packet to the destination. In the future time slots, either one of the following two events can happen first: (i) E_1: Node A meets the destination and delivers the packet. (ii) E_2: Node A meets one of the potential relays. Let the time duration elapsed till event E_i occurs be denoted by T_i, $i = 1, 2$. By definition, T_1 is exponentially distributed with mean $E[M(d)]$. To derive the distribution of T_2, we use the fact that the time it takes to meet one particular relay node is exponentially distributed with mean $E[M]$, where $E[M]$ is the expected meeting time of two random walks. T_2 is the minimum of M [4] such exponentials which is also an exponential with mean $\frac{E[M]}{M}$. Thus, the time duration till one of these two events occur is equal $\min(T_1, T_2)$ and is exponentially distributed with mean $\frac{1}{\frac{1}{E[M(d)]} + \frac{M}{E[M]}}$.

Let node A encounter a potential relay (lets label it node B) before meeting the destination. (The probability of this event is equal to $\frac{\frac{M}{E[M]}}{\frac{1}{E[M(d)]} + \frac{M}{E[M]}}$.) Let node A and B be at a distance d_A and d_B from the destination when they meet. Node A has l copies of the packet while B has none. Let $D_M(d_A, d_B, l)$ denote the minimum additional delay to deliver the packet to the destination. Then,

$$E[D(d, l)] = \frac{1}{\frac{1}{E[M(d)]} + \frac{M}{E[M]}} + \frac{\frac{M}{E[M]}}{\frac{1}{E[M(d)]} + \frac{M}{E[M]}} \sum_{d_A, d_B} P(d_A, d_B) E[D_M(d_A, d_B, l)], \qquad (7)$$

where $P(d_A, d_B)$ is the probability that the two nodes are at a distance d_A and d_B from the destination when they meet.

Node A can give any number from 0 to $l - 1$ copies to the B. If i of the l copies are given to B, then the delivery delay to the destination is the minimum of $D(d_A, l - i)$ and $D(d_B, i)$. Hence,

$$E[D_M(d_A, d_B, l)] = \min_{0 \leq i < l} (E[\min(D(d_A, l - i), D(d_B, i))]) \qquad (8)$$

Note that the solution to Equation (8) gives the optimal spraying policy.

Equations (7) and (8) form a system of non linear equations. Solving these equations will give the optimal spraying policy, but solving a non linear system is not easy. So, we make approximations to simplify these equations. (Note that due to these approximations, the spraying policy obtained is not really the optimal, but it will give an intuition into the structure of the optimal policy.)

First, we assume that the sum of two exponentially distributed random variables is also exponential. With this approximation, the distribution of both $D(d, l)$ and $D_M(d_A, d_B, l)$ can be derived to be exponential. Thus, Equation (7) reduces to the following:

$$E[D(d, l)] = \frac{1}{\frac{1}{E[M(d)]} + \frac{M}{E[M]}} + \frac{\frac{M}{E[M]}}{\frac{1}{E[M(d)]} + \frac{M}{E[M]}} \sum_{d_A, d_B} P(d_A, d_B) \min_{0 \leq i < l} \left(\frac{1}{\frac{1}{E[D(d_A, l-i)]} + \frac{1}{E[D(d_B, i)]}} \right). \qquad (9)$$

Equation (9) is still a system of non linear equations which are not easy to solve. So, we make another approximation by replacing d_A by its expected value. For the random walk mobility model, $E[d_A]$ is equal

---

[4]The number of potential relays is equal to the number of nodes which do not have a copy of the packet. This number is upper bounded by the total number of nodes, M. Since the number of potential relays is unknown at a given time, we use the upper bound on this value.
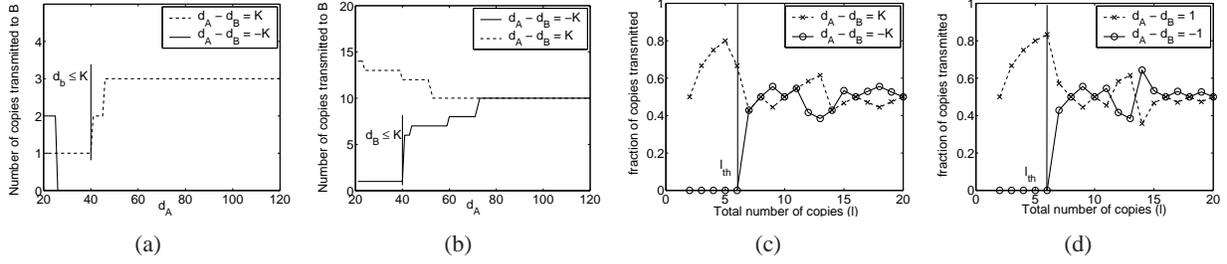
Figure 6: Studying the optimal spraying policy for Spray and Wait. Network Parameters: $N = 150 \times 150$, $M = 40$, $K = 20$. (a) Number of copies given to node B as a function of d_A for $l = 4$. (b) Number of copies given to node B as a function of d_A for $l = 20$. (c) Proportion of copies given to node B as a function of l for $d_A = 75$. (d) Proportion of copies given to node B as a function of l for $d_A = 75$.

to d as the probability of moving in any direction is the same. Replacing d_A by d in Equation (9) yields,

$$E[D(d,l)] = \frac{1}{\frac{1}{E[M(d)]} + \frac{M}{E[M]}} + \frac{\frac{M}{E[M]}}{\frac{1}{E[M(d)]} + \frac{M}{E[M]}} \sum_{d_B} P(d_B \mid d_A = d) \min_{0 \leq i < l} \left(\frac{1}{\frac{1}{E[D(d,l-i)]} + \frac{1}{E[D(d_B,i)]}} \right). (10)$$

In Equation (10), the value of $E[D(d,l)]$ depends only on those $E[D(\hat{d},\hat{l})]$ for which either $\hat{l} < l$ or $\hat{l} = l, \hat{d} \leq d$. So, a dynamic program can be used to solve Equation (10).

The dynamic program will be initialized with the value of $E[D(d,1)]$ which depends on how each copy is routed towards the destination. Section 3.5.1 and 3.5.2 finds its value for Spray and Wait and Spray and Focus.

The only unknown in Equation (10) is $P(d_B \mid d_A = d)$. Since node B is within range of A, d_B will lie within $d - K$ and $d + K$. $P(d_B \mid d_A = d)$ can be derived using elementary combinatorics to be equal to

$$\begin{cases} \frac{K+1}{4K} & d_B = d - K \\ \frac{2}{4K} & d - K + 2 \leq d_B \leq d + k - 2 \\ \frac{K+1}{4K} & d_B = d + K \end{cases}.$$

3.5.1 Spray and Wait

In this section, we first study the optimal spraying policy for spray and wait, then study the spraying policy obtained by solving Equation (10), and finally present a simple heuristic which achieves a expected delay very close to the optimal.

In Spray and Wait, each relay node routes the copy towards the destination using direct transmission. Thus, $E[D(d,1)]$ is the expected time it takes for the relay to meet the destination and is equal to $E[M(d)]$.

Now, we study the spraying policy obtained by solving Equation (10). Let node A which has l copies of the packet meet node B which has none. Let the distance to the destination of both the nodes be denoted by d_A and d_B respectively. Figure 6(a)-6(b) plots the number of copies given to node B versus d_A for different values of l. For $l = 4$, the node which is closer to the destination gets most of the copies while for $l = 20$, most of the times, nearly half of the copies are given away to node B. This observation suggests that the optimal policy behaves differently for different values of l. (Note that node B gets only one copy when it is within the transmission range of the destination because the packet will be delivered at the next transmission opportunity.)

To study the behavior of the optimal policy as l changes, we plot the proportion of copies given to node B as a function of l for different values of $d_A - d_B$ in Figures 6(c)-6(d). In all the cases, there exists a threshold
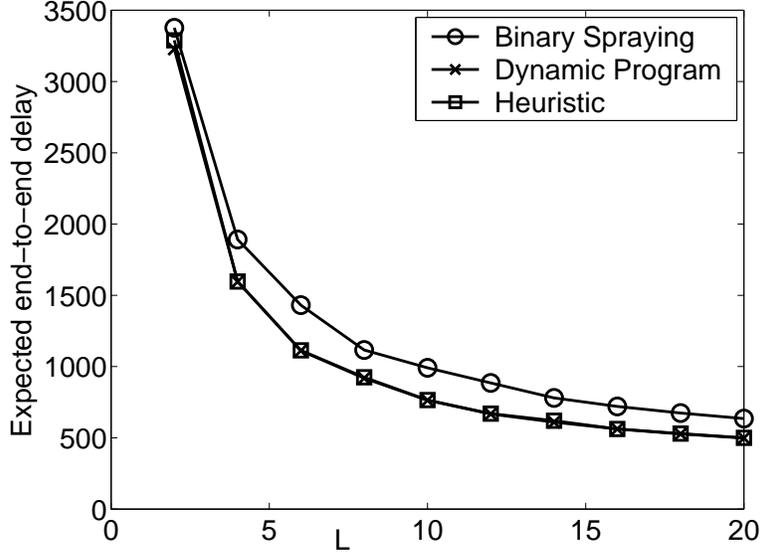
Figure 7: Comparison of the expected end-to-end delay performance of binary spraying, the optimal policy and the proposed heuristic. Network parameters: $N = 150 \times 150, M = 40, K = 20$.

for l below which most of the copies are kept by the node closer to the destination and above which the copy splitting is more or less half and half. We label this threshold as l_{th}.

Based on the above observation, we propose a simple heuristic to distribute copies. (i) If l is less than l_{th} and node A is closer to the destination, then node B is not given any of the copies. (ii) If l is less than l_{th} and node B is closer to the destination, then node B is given $l - 1$ copies. (iii) If l is greater than l_{th}, then node B is given half of the copies. Figures 7-8 compare the performance of the optimal policy, the proposed heuristic and binary spraying for different network parameters. It is easy to see that the proposed heuristic performs very close to the optimal and has a better performance than binary spraying.

3.5.2 Spray and Focus

In this section, we first study the optimal spraying policy for spray and focus, then study the spraying policy obtained by solving Equation (10), and finally present a simple heuristic which achieves a expected delay very close to the optimal.

In Spray and Focus, each relay node performs utility based forwarding towards the destination. First, we derive the value of $E[D(d, 1)]$ to initialize the dynamic program which is used to solve Equation (10).

Lemma 3.6 $E[D(d, 1)]$ can be derived by solving the following set of non linear equations:

$$E[D(d, 1)] = \frac{1}{\frac{1}{E[M(d)]} + \frac{M}{E[M]}} + \frac{\frac{M}{E[M]}}{\frac{1}{E[M(d)]} + \frac{M}{E[M]}} \sum_{d_2} P(d_2 \mid d) E[D(min(d, d_2), 1)]. \tag{11}$$

Proof: In the future time slots either of the following two events can happen first: (i) The node meets the destination and delivers the packet. This time duration is exponentially distributed with mean $E[M(d)]$. (ii) The node meets a potential relay node. This time duration is exponentially distributed with mean $\frac{E[M]}{M}$. Let the relay node be at a distance d_2 from the destination. Then if $d_2 < d$, then the relay node is closer to the destination and it will be given the copy of the packet. The additional time it will take to deliver the packet will be equal to $E[D(d_2, 1)]$. But if $d_2 \geq d$, the original node will retain the copy and the additional time it will take to deliver the packet is still equal to $E[D(d, 1)]$. □
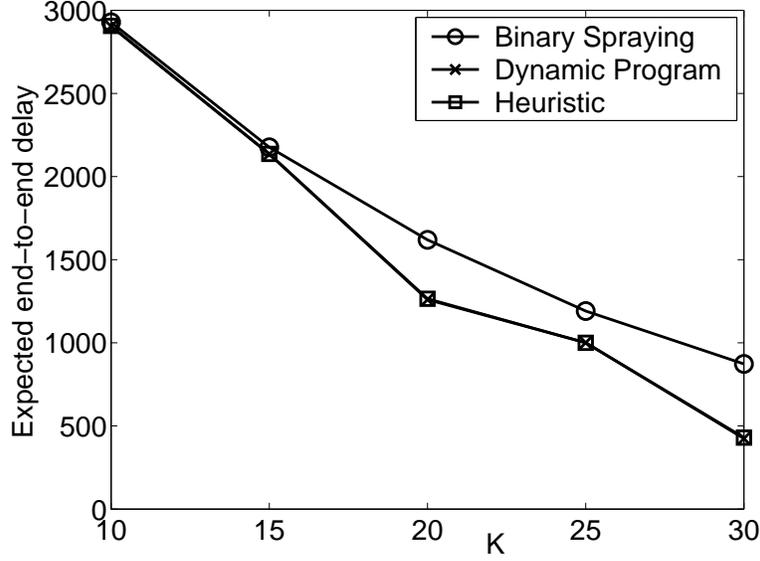
17

Figure 8: Comparison of the expected end-to-end delay performance of binary spraying, the optimal policy and the proposed heuristic. Network parameters: $N = 150 \times 150, M = 40, L = 5$.

(a)   (b)   (c)   (d)

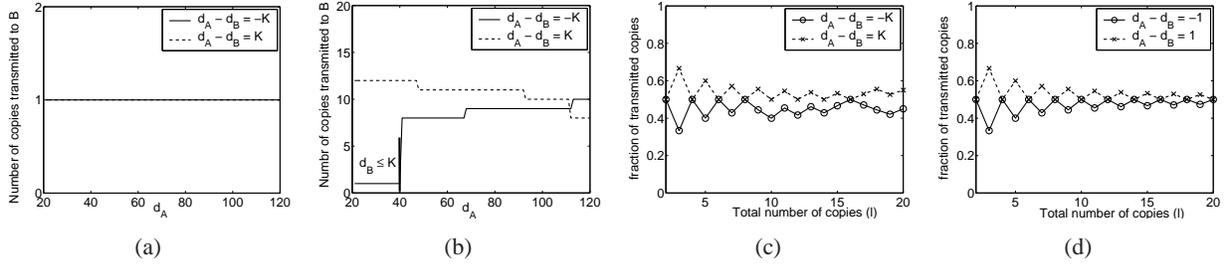Figure 9: Studying the optimal spraying policy for Spray and Focus. Network Parameters: $N = 150 \times 150, M = 40, K = 20$. (a) Number of copies given to node B as a function of d_A for $l = 2$. (b) Number of copies given to node B as a function of d_A for $l = 20$. (c) Proportion of copies given to node B as a function of l for $d_A = 75$. (d) Proportion of copies given to node B as a function of l for $d_A = 75$.

A particular value of $E[D(d, 1)]$ depends only on those values of $E[D(\hat{d}, 1)]$ for which $\hat{d} \leq d$. Hence, a dynamic program can be used to solve Equation (11).

Now, we study the optimal spraying policy obtained by solving Equation (10) after substituting the value of $E[D(d, 1)]$ derived in Lemma 3.6. Figure 9(a)-9(b) plots the number of copies given to node B versus d_A for different values of l. The curves show that most of the times, nearly half of the copies are handed over to node B irrespective of the value of l. To confirm this observation, we plot the proportion of copies given to node B as a function of l for different values of $d_A - d_B$ in Figures 9(c)-9(d). For all the cases, nearly half of the copies are handed over to node B. This suggests that binary spraying should perform close to the optimal policy. Figures 10-11 compare the performance of binary spraying with the optimal policy for different network parameters. These figures show that binary spraying has near optimal performance for Spray and Focus. The near optimal performance of binary spraying is explained by the following two observations: (i) If a node distributes its copies to bad nodes (nodes which have a higher expected delivery delay), it still has its own copy which it can give to a good node whenever it meets one. (ii) Moreover, a bad node will have a chance to give up its copy to good nodes later in the future. Thus, spraying copies as fast as possible will achieve a good

18

Figure 10: Comparison of the expected end-to-end delay performance of binary spraying and the optimal policy. Network parameters: $N = 150 \times 150, M = 40, K = 20$.
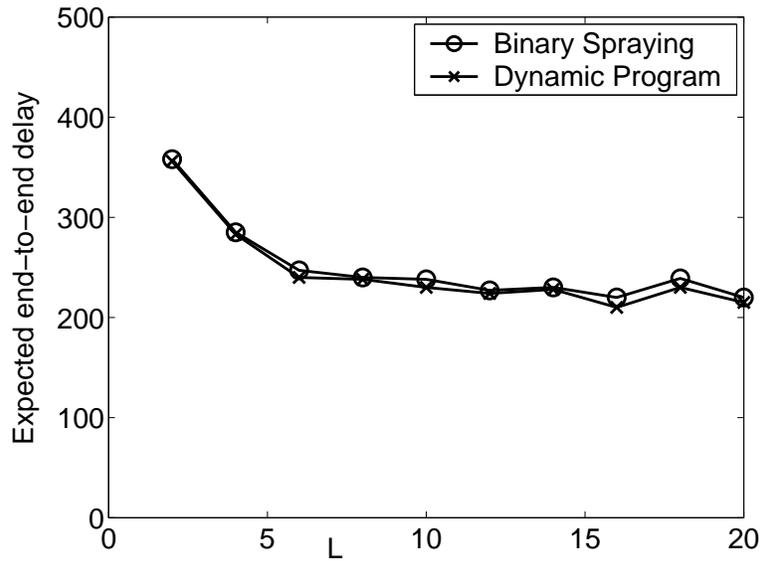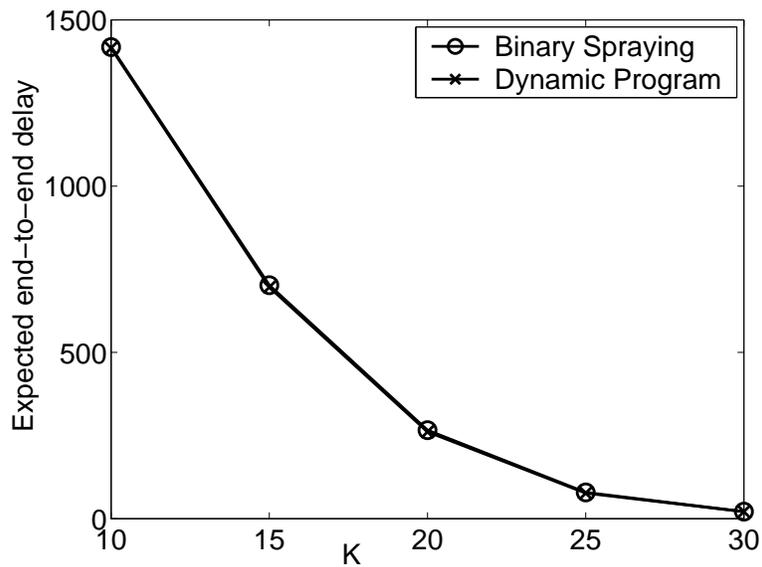

Figure 11: Comparison of the expected end-to-end delay performance of binary spraying and the optimal policy. Network parameters: $N = 150 \times 150, M = 40, L = 5$.

delay performance for Spray and Focus.

3.5.3 Discussion

We now generalize the intuition derived in the previous section to general utility functions. For Spray and Wait, if a smaller utility always means a smaller distance to the destination, there always exists a threshold l_{th} such that the following heuristic performs well: (i) If l is less than l_{th} and node A is closer to the destination, then node B is not given any of the copies. (ii) If l is less than l_{th} and node B is closer to the destination, then node B is given $l-1$ copies. (iii) If l is greater than l_{th}, then node B is given half of the copies. All the utility functions discussed in Section 3.4 satisfy this constraint, hence, the proposed heuristic was found to be very efficient in vehicular networks.

For Spray and Focus, irrespective of the utility function, binary spraying always yields efficient results because the focus phase allows fixing any "wrong" or "bad" decisions made earlier. Hence, for vehicular networks, Binary Spray and Focus was found to be the best spraying protocol.

3.6 Collaboration of communication-capable vehicles and roadside stations

In addition to vehicle to vehicle communication, another form of communication is expected to take place between vehicles and roadside stations along the road. Such stations are envisioned to be installed in intersections, or at regular distances along highways. The correct operation of the binary spray and focus protocols in a vehicular network does not depend on the existence of such infrastructure. Nevertheless, if such stations become available, they can be used to signicantly improve performance.

Spray and focus treats roadside stations similarly to vehicles. However, an important difference is that these stations are assumed to be interconnected, and once a message is received by one of them, it can reach very fast distant locations. So, the utility of these stations is the same for each destination. In other words, if a roadside station comes within range of the destination, then all roadside stations can be assumed to be within the range of the destination. Hence, these stations tend to have a higher utility in general, so it is very likely that vehicles will communicate with each other through roadside stations. We always observed better expected delays and higher delivery probabilities in presence of these roadside stations.

Introducing roadside stations introduces the following change to the analysis. Roadside station is static while the vehicle is moving according to a given mobility model. The duration after which they come within range of each other is no longer one meeting time. This duration is equal to the hitting time which is rigorously defined as follows.

Hitting Time Let a node i move according to mobility model "mm", and start from its stationary distribution at time 0. Let j be a static node with uniformly chosen X_j, then the hitting time (T_{mm}) is defined as the time it takes node i to first come within range of node j, that is $T_{mm} = \min_{t}\{t : \|X_i(t) - X_j\| \leq K\}$.

We next state expressions of the expected hitting time for the two most common mobility models - the Random Direction and the Random Waypoint mobiity models.

We first define the Random Direction mobility model and then state the expression for its expected hitting time.

Random Direction In the Random Direction (RD) model each node moves as follows: (i) choose a direction θ uniformly in $[0, 2\pi)$; (ii) choose a speed according to assumption (d); (iii) choose a duration T of movement from an exponential distribution with average $\frac{\overline{L}}{\overline{v}}$; (iv) move towards θ with the chosen speed for T time units; [5] (v) after T time units pause according to assumption (e) and go to step (i).

---

[5]If the boundary is reached, the node either reflects back or re-enters from the opposite side of the network (torus).

Theorem 3.4 *The expected hitting time ET_{rd} for the Random Direction model is given by:*

$$ET_{rd} = \left(\frac{N}{2K\overline{L}} \right) \left(\frac{\overline{L}}{\overline{v}} + \overline{T}_{stop} \right).$$

(12)

We next define the Random Waypoint mobility model, then state a lemma stating the average distance covered by a node in one epoch, and then state the expression for the expected hitting time Random Waypoint mobility.

Random Waypoint In the Random Waypoint (RWP) model, each node moves as follows [13]: (i) choose a point X in the network uniformly at random, (ii) choose a speed v uniformly in $[v_{min}, v_{max}]$ with $v_{min} > 0$ and $v_{max} < \infty$. Let $\overline{v}$ denote the average speed of a node, (iii) move towards X with speed v along the shortest path to X, (iv) when at X, pause for T_{stop} time units where T_{stop} is chosen from a geometric distribution with mean $\overline{T}_{stop}$, (v) and go to Step (i).

Lemma 3.7 *Let L be the length of an epoch, measured as the distance between the starting and the finishing points of the epoch. Then $EL_{rwp} = 0.3826\sqrt{N}$.*

Theorem 3.5 *The expected hitting time ET_{rwp} for the Random Waypoint model is given by:*

$$ET_{rwp} = \left(\frac{N}{2KEL_{rwp}} \right) \left(\frac{EL_{rwp}}{\overline{v}} + \overline{T}_{stop} \right).$$

(13)

3.7 Incorporating Contention

Up till now, we have ignored contention in the analysis. The assumption of no contention is valid only for very low traffic rates, irrespective of whether the network is sparse or not. For higher traffic rates, contention has a significant impact on the performance, especially of flooding-based routing schemes. Given the small contact durations in vehicular network, contention will have a even more severe affect on performance. To demonstrate the inaccuracies which arise when contention is ignored, we use simulations to compare the delay of three different routing schemes in a sparse network, both with and without contention, in Figure 12. The plot shows that ignoring contention not only grossly underestimates the delay, but also predicts incorrect trends and leads to incorrect conclusions. For example, without contention, the so called spraying scheme has the worst delay, while with contention, it has the best delay.

Incorporating wireless contention complicates the analysis significantly. This is because contention manifests itself in a number of ways, including (i) finite bandwidth which limits the number of packets two nodes can exchange while they are within range, (ii) scheduling of transmissions between nearby nodes which is needed to avoid excessive interference, and (iii) interference from transmissions outside the scheduling area, which may be significant due to multipath fading [8]. So, we first propose a general framework to incorporate contention in the performance analysis of mobility-assisted routing schemes for ICMNs while keeping the analysis tractable. This framework incorporates all the three manifestations of contention, and can be used with any mobility and channel model. The framework is based on the well-known physical layer model [21]. Prior work has used the physical layer model to derive capacity results, see, for example, [19, 21, 32], and has assumed an idealized perfect scheduler. We are interested in calculating the expected delay of various mobility-assisted routing schemes under realistic scenarios, and for this reason we assume a random access scheduler.
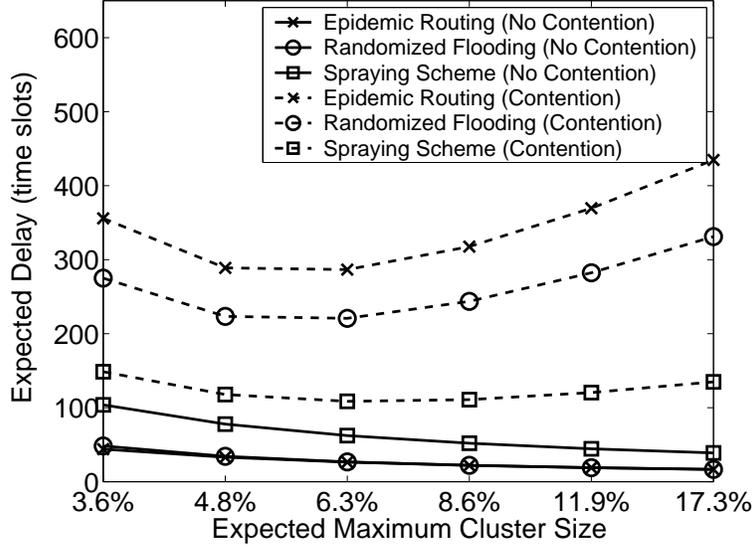
Figure 12: Comparison of delay with and without contention for three different routing schemes in sparse networks. The simulations with contention use the scheduling mechanism and interference model described in Section 3.7.1. The expected maximum cluster size (x-axis) is defined as the percentage of total nodes in the largest connected component (cluster) and is a metric to measure connectivity in sparse networks [38]. The routing schemes compared are: epidemic routing [41], randomized flooding [40] and spraying based routing [39].

3.7.1 The Framework

We assume that there are M nodes moving in a two dimensional torus of area N. We also assume that each node acts as a source sending packets to a randomly selected destination. Finally, we assume the following radio model.

Radio Model: An analytical model for the radio has to define the following two properties: (i) when will two nodes be within each other's range, (ii) and when is a transmission between two nodes successful. (Note that we define two nodes to be within range if the packets they send to each other are received successfully with a non-zero probability.) If one assumes a simple distance-based attenuation model without any channel fading or interference from other nodes, then two nodes can successfully exchange packets without any loss only if the distance between them is less than a deterministic value K (also referred to as the transmission range), else they cannot exchange any packet at all. The value of K depends on the transmission power and the distance attenuation parameter. However, in presence of a fading channel and interference from other nodes, even though two nodes can potentially exchange packets if the distance between them is less than K, a transmission between them might not go through. A transmission is successful only when the signal to interference ratio (SIR) is greater than some desired threshold.

We assume the following radio model: (i) Two nodes are within each other's range if the distance between them is less than K, and (ii) any transmission between the two is successful only if the SIR is greater than a desired threshold Θ. Note that this model is not equivalent to a circular disk model because any transmission between two nodes with a distance less than K is successful with a certain probability that depends on the fading channel model and the amount of interference from other nodes.

We now present the framework for a mobility model with a uniform node location distribution. Commonly used mobility models like random direction and random waypoint on a torus satisfy this assumption [12, 35]. The proposed framework can be easily extended to any other mobility model [26] in which the process governing the mobility of nodes is stationary and the movement of each node is independent of each other.

We first identify the three manifestations of contention and describe how do they affect message exchange.

Finite Bandwidth: When two nodes meet, they might have more than one packet to exchange. Say two nodes can exchange s_{BW} packets during a unit of time. If they move out of the range of each other, they will have to wait until they meet again to transfer more packets. The number of packets which can be exchanged in a unit of time is a function of the packet size and the bandwidth of the links. We assume the packet size and the bandwidth of the links to be given, hence s_{BW} is assumed to be a given network parameter. We also assume that the s_{BW} packets to be exchanged are randomly selected from amongst the packets the two nodes want to exchange[6].

Scheduling: We assume an ideal CSMA-CA scheduling mechanism is in place which avoids any simultaneous transmission within one hop from the transmitter and the receiver. Nodes within range of each other and having at least one packet to exchange are assumed to contend for the channel. For ease of analysis, we also assume that time is slotted. At the start of the time slot, all node pairs contend for the channel and once a node pair captures the medium, it retains the medium for the entire time slot.

Interference: Even though the scheduling mechanism is ensuring that no simultaneous transmissions are taking place within one hop from the transmitter and the receiver, there is no restriction on simultaneous transmissions taking place outside the scheduling area. These transmissions act as noise for each other and hence can lead to packet corruption.

In the absence of contention, two nodes would exchange all the packets they want to exchange whenever they come within range of each other. Contention will result in a loss of such transmission opportunities. This loss can be caused by either of the three manifestations of contention. In general, these three manifestations are not independent of each other. We now propose a framework which uses conditioning to separate their effect and analyze each of them independently.

Main Idea: Lets look at a particular packet, label it packet A. Suppose two nodes i and j are within range of each other at the start of a time slot and they want to exchange this packet. Let p_{txS} denote the probability that they will successfully exchange the packet during that time slot. First, we look at how the three manifestations of contention can cause the loss of this transmission opportunity.

Finite Bandwidth: Let E_{bw} denote the event that the finite link bandwidth allows nodes i and j to exchange packet A. The probability of this event depends on the total number of packets which nodes i and j want to exchange. Let there be a total of S distinct packets in the system at the given time (label this event E_S). Let there be $s, 0 \leq s \leq S-1$, other packets (other than packet A) which nodes i and j want to exchange (label this event E_s^S). If $s \geq s_{BW}$, then the s_{BW} packets exchanged are randomly selected from amongst these $s+1$ packets. Thus, $P(E_{bw}) = \sum_S P(E_S) \left(\sum_{s=0}^{s_{BW}-1} P(E_s^S) + \sum_{s=s_{BW}}^{S-1} \frac{s_{BW}}{s+1} P(E_s^S) \right)$. To simplify the analysis, we make our first approximation here by replacing the random variable S by its expected value in the expression for $P(E_{bw})$[7] (see Equation (14) for the final expression for $P(E_{bw})$). Note that simulations results presented in [26] verify that this approximation does not have a drastic effect on the accuracy of the analysis.

Scheduling: Let E_{sch} denote the event that the scheduling mechanism allows nodes i and j to exchange packets. The scheduling mechanism prohibits any other transmission within one hop from the transmitter and the receiver. Hence, to find $P(E_{sch})$, we have to determine the number of transmitter-receiver pairs which

---

[6]Note that assuming a random queueing discipline yields the same results as FIFO in our setting (yet simplifies analysis). This is so because a work conserving queue yields the same queueing delay for constant size packets irrespective of whether the queue service discipline is FIFO or random queueing. In addition, due to packet homogeneity (all packets are treated the same) the expected end-to-end delay will also be the same. Of course, if packet homogeneity is lost, for example by assigning higher priority to packets that are closer to their destination, the expected end-to-end delay will decrease as packets with a smaller end-to-end service requirement will be serviced first.

[7]We incorporate the arrival process through $E[S]$ in the analysis. $E[S]$ depends on the arrival rate through Little's Theorem. Thus, after deriving the expected end-to-end delay for a routing scheme in terms of $E[S]$, Little's Theorem can be used to express the delay in terms of only the arrival rate.

have at least one packet to exchange and are contending with the i-j pair. Let there be a nodes within one hop from the transmitter and the receiver (label it event E_a) and let there be c nodes within two hops but not within one hop from the transmitter and the receiver (label it event E_c). These c nodes have to be accounted for because a node at the edge of the scheduling area can be within the transmission range of one of these c nodes and will contend with the desired transmitter/receiver pair. Let $t(a, c)$ denote the expected number of possible transmissions contending with the i-j pair. By symmetry, all the contending nodes are equally likely to capture the channel. So, $P(E_{sch} \mid E_a, E_c) = 1/t(a, c)$.

Interference: Let E_{inter} denote the event that the transmission of packet A is not corrupted due to interference given that nodes i and j exchanged this packet. Let there be $M - a$ nodes outside the transmitter's scheduling area (this is equivalent to event E_a). If two of these nodes are within the transmission range of each other, then they can exchange packets which will increase the interference for the transmission between i and j. Lets label the event that packet A is successfully exchanged inspite of the interference caused by these $M - a$ nodes as I_{M-a}. Then, $P(E_{inter} \mid E_a) = P(I_{M-a})$.

Packet A will be successfully exchanged by nodes i and j only if the following three events occur: (i) the scheduling mechanism allows these nodes to exchange packets, (ii) nodes i and j decide to exchange packet A from amongst the other packets they want to exchange, and (iii) this transmission does not get corrupted due to interference from transmissions outside the scheduling area. Thus,

$$
p_{txS} = P(E_{bw}) \sum_{a,c} P(E_a, E_c) P(E_{sch} \mid E_a, E_c) P(E_{inter} \mid E_a)
$$

$$
= \left(\sum_{s=0}^{s_{BW}-1} P(E_s^{E[S]}) + \sum_{s=s_{BW}}^{E[S]-1} \frac{s_{BW} P(E_s^{E[S]})}{s+1} \right) \times \sum_{a,c} \frac{P(E_a) P(E_c \mid E_a) P(I_{M-a})}{t(a, c)}. \tag{14}
$$

Expressions for the unknown values on Equation (14) can be easily derived using geometric arguments. Please refer to [26] for details.

We next study how does the optimal spraying scheme change after incorporating contention in the analysis. We first state a sequence of lemmas which state the expected delay expressions for source spray and wait (spraying scheme in which only source is allowed to spray copies) and fast spray and wait [26] (which yields a lower bound on binary spray and wait). We will then use these delay expressions to illustrate if and how the conclusions drawn in the previous sections change.

Before stating the lemmas, we define two additional mobility properties. The delay expressions will be stated in terms of these two.

Inter-Meeting Time Let nodes i and j start from within range of each other at time 0 and then move out of range of each other at time t_1, that is $t_1 = min_t\{t : \|X_i(t) - X_j(t)\| > K\}$. The inter meeting time (M_{mm}^+) of the two nodes is defined as the time it takes them to first come within range of each other again, that is $M_{mm}^+ = min_t\{t - t_1 : t > t_1, \|X_i(t) - X_j(t)\| \le K\}$.

Contact Time Assume that nodes i and j come within range of each other at time 0. The contact time τ_{mm} is defined as the time they remain in contact with each other before moving out of the range of each other, that is $\tau_{mm} = min_t\{t - 1 : \|X_i(t) - X_j(t)\| > K\}$.

Now we state the delay expressions for the two spraying schemes. Let $E[D_{ssw}^{mm}(m)]$ denote the expected time it takes for the number of nodes having a copy of the packet to increase from m to $m+1$ for source spray and wait routing. First, we state the value $E[D_{ssw}^{mm}(m)]$, and then state the expected end-to-end delay for source spray and wait (denoted by $E[D_{ssw}^{mm}]$) in terms of $E[D_{ssw}^{mm}(m)]$.

Lemma 3.8 $E[D_{ssw}^{mm}(m)] = \begin{cases} \frac{E[M_{mm}]}{(M-1)p_{success}^{ssw}} & 1 \le m < L \\ \frac{E[M_{mm}]}{L p_{success}^{ssw}} & m = L \end{cases}$ where $p_{success}^{ssw} = 1 - (1 - p_{txS}^{ssw})^{E[\tau_{mm}]}$.

Theorem 3.6 $E[D_{ssw}^{mm}] = \sum_{i=1}^{L} p_{dest}^{ssw}(i) \sum_{m=1}^{i} E[D_{ssw}^{mm}(m)].$

Similarly, let $E[D_{fsw}^{mm}(m)]$ denote the expected time it takes for the number of nodes having a copy of the packet to increase from m to $m+1$ for fast spray and wait routing. Again, first we state the value $E[D_{fsw}^{mm}(m)]$, and then state the expected end-to-end delay for fast spray and wait (denoted by $E[D_{fsw}^{mm}]$).

Lemma 3.9 $E[D_{fsw}^{mm}(m)] = \begin{cases} \frac{E[M_{mm}]}{m(M-m)p_{success}^{fsw}} & 1 \leq m < L \\ \frac{E[M_{mm}]}{Lp_{success}^{fsw}} & m = L \end{cases}$ where $p_{success}^{fsw} = 1 - \left(1 - p_{txS}^{fsw}\right)^{E[\tau_{mm}]}.$

Theorem 3.7 $E[D_{fsw}^{mm}] = \sum_{i=1}^{L} p_{dest}^{fsw}(i) \sum_{m=1}^{i} E[D_{fsw}^{mm}(m)].$

We now re-visit the three fundamental questions related to spraying-based schemes and comment on how do the conclusions drawn without considering contention change after incorporating contention in the analysis.
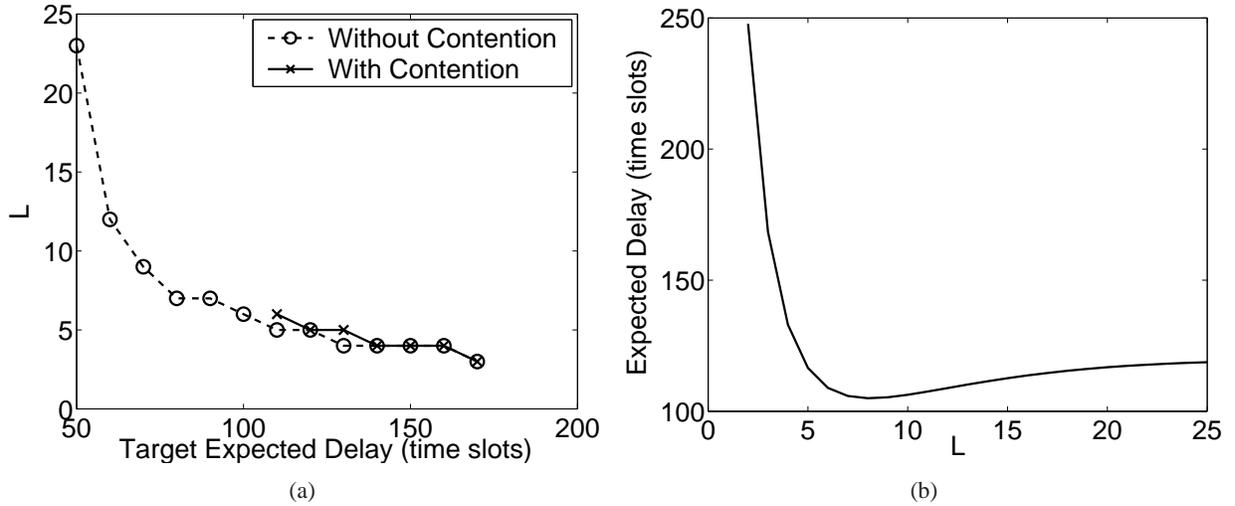

Figure 13: (a) Minimum value of L which achieves the target expected delay for source spray and wait. (b) L against expected delay (with contention). Network parameters: $N = 100 \times 100, K = 8, M = 150, \Theta = 5, E[S] = 70, \overline{T}_{stop} = 0, \overline{v} = 1, s_{BW} = 1.$

3.7.2 Deciding the Right Number of Copies

This section studies the error introduced by ignoring contention when one has to find the minimum value of L (the number of copies sprayed) in order for a spraying-based scheme to achieve a specific expected delay. (Note that we want the minimum value of L which achieves the target delay as bigger values of L consume more resources.) We choose the source spray and wait scheme with the random waypoint mobility model as the case study in this section. We numerically solve the expression for $E[D_{ssw}^{rwp}]$ in Theorem 3.6 to find the minimum value of L which achieves a target delay and plot it in Figure 13(a) both with and without contention for a sparse network. (For the expected delay of source spray and wait without contention, we use the expression derived in [39].) This figure shows that an analysis without contention would be accurate for smaller values of L (smaller values of L generate lower contention in the network), however it would predict that one can use a large number of copies to achieve a target expected delay which actually will not be achievable in practice due to contention. For example, the analysis without contention indicates that a delay of 50 time units is achievable with $L = 23$ while the contention-aware analysis indicates that it is not achievable. Figure 13(b)

shows that $L = 23$ results in an expected delay of more than 118 time units, which is also achievable by $L = 5$. Thus choosing a value of L based on predictions from a contention-ignorant analysis led to a value of delay which is not only much higher than expected but also would have been achieved by nearly four times fewer copies. Thus, we conclude that the analysis without contention will give accurate results only for smaller values of L, and larger values of L should not be chosen as they merely create more contention without reducing the expected delay.
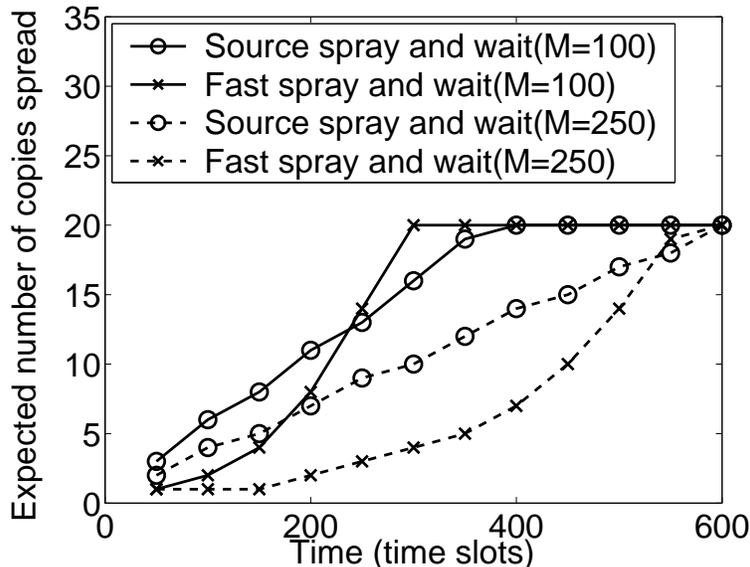

Figure 14: Comparison of fast spray and wait and source spray and wait: Expected number of copies spread vs time elapsed since the packet was generated. Network parameters: $N = 100 \times 100$ square units, $K = 5, \Theta = 5, s_{BW} = 1$ packet/time slot, $L = 20$. Expected maximum cluster size (metric to measure connectivity) for these network parameters is equal to 4.6% for $M = 100$ and 5.2% for $M = 250$.

3.7.3 Routing Each Copy Separately

Utility-based forwarding reduces the number of copies required (L) to achieve a given delay. Thus it reduces contention in the spraying phase. However, after copies have been distributed, it requires multiple message exchanges in the focus phase which increases contention. Amongst the two, the contention reduction in the spraying phase dominates, hence, the conclusions drawn without incorporating contention in the analysis still hold. We give a numerical example to support this claim in Section 4.4.3.

3.7.4 Distributing Copies

As shown in Section 3.5, spraying copies as fast as possible is the best way to spread copies if all the relay nodes are equal/homogeneous. To answer whether spraying the copies as fast as possible is optimal with contention, we compare fast source spray and wait and source spray and wait for the random waypoint mobility model. Since fast spray and wait spreads copies whenever there is any opportunity to do so, it has the minimum spraying time when there is no contention in the network. On the other hand, since source spray and wait does not use relays to forward copies, it is one of the slower spraying mechanisms when there is no contention in the network.

Now we study how fast the two schemes spread copies of a packet when there is contention in the network. Figure 14 plots the number of copies spread as a function of the time elapsed since the packet was generated. Somewhat surprisingly, depending on the density of the network, source spray and wait

can spray copies faster than fast spray and wait. This occurs because fast spray and wait generates more contention around the source as it tries to transmit at every possible transmission opportunity. Such a behavior is expected for dense networks, but these results show that increased contention can deteriorate fast spray and wait's performance even in sparse networks. This issue is more aggravated in vehicular networks as contact durations are small. In general, unless the network is very sparse, strategies which spray copies slower yield better performance than more aggressive schemes thanks to reducing contention.

4   Analysis with Realistic Mobility Models - "Community-based Mobility Model"

To understand the performance of spray and focus routing with realistic vehicle mobility, we propose a new mobility model. Like a good mobility model, the proposed model has the following three characteristics: (i) it captures *realistic* vehicular mobility patterns of scenarios in which one wants to eventually operate the network; (ii) at the same time the proposed model is *mathematically tractable*; this is very important to allow the derivation of performance bounds and to understand the limitations of various protocols under the given scenario; (iii) finally, it is *flexible* enough to provide qualitatively and quantitatively different mobility characteristics by changing some parameters of the model, yet in a repeatable and scalable manner as designing a new mobility model for each existing or new scenario is undesirable.

The proposed model is a *time-variant community mobility model*, and is referred to as the TVC model. One salient characteristic in the TVC model is *location preference*. Another important characteristic is the *time-dependent, periodical behavior* of nodes. To our best knowledge, this is the first *synthetic mobility model* that captures non-homogeneous behavior in both *space* and *time*.

To establish the flexibility of our TVC model we show that we can match its two prominent properties, *location visiting preferences* and *periodical re-appearance*, with a vehicle mobility trace[2].

Finally, in addition to the improved realism, the TVC model can be mathematically treated to derive analytical expressions for important mobility properties of interest, such as the meeting time, the inter-meeting time etc. We illustrate how to derive the statistics of these quantities, and then use them to derive expressions for spray and focus routing for a particular instantiation of the model.

4.1   Time-variant Community Mobility Model

After analyzing a large number of traces [24], we observed two important properties that are common in all of them: (a)*skewed location visiting preferences* and (b)*time-dependent mobility behavior* [23]. Spcifically, the *location visiting preference* refers to the percentage of time a node spends at a given location and the *time-dependent mobility behavior* refers to the observation that nodes visit different locations, depending on the time of the day. We believe that these two properties are prevalent in any human-driven mobility. This belief is supported by typical daily activities of humans: most of us tend to spend most time at a handful of frequently visited locations, and a recurrent daily or weekly schedule is an inseparable part of our lives. It is essential to design a model that captures such spatial-temporal preferences of human mobility in many contexts.

We next present the design of our *time-variant community (TVC) mobility model*. We illustrate the model with an example in Fig. 15 and use this example to introduce the notations we use (see Table 2) in the rest of the paper.

First, to induce *skewed location visiting preferences*, we define some *communities* (or heavily-visited geographic areas). Take time period 1 (TP1) in Fig. 15 as an example, the communities are denoted as $Comm_j^1$ and each of them is a square geographical area with edge length C_j^1.[8] A node visits these communities with

---

[8]For all parameters used in the paper, we follow the convention that the subscript of a quantity represents its community index,

Table 2: Parameters of the time-variant community mobility model[1]

N	Edge length of simulation area
V	Number of time periods
T^t	Duration of t-th time period
S^t	Number of communities in time period t
C_j^t	Edge length of community j in time period t
$Comm_j^t$	The j-th community during time period t
$p_{i,j}^t$	The probability to choose community j when the previous community is i, during time period t
π_j^t	Stationary probability of an epoch in community j during time period t
$v_{min}, v_{max}, \overline{v}$	Minimum, maximum, and average speed[1]
$D_{max,j}, \overline{D_j}$	Maximum and average pause time after each epoch[1]
$\overline{L_j}$	Average epoch length for community j
$P_{move,j}^t | P_{pause,j}^t$	Probability that a node is moving	pausing when being in community j during period t
P_j^t	Fraction of time the node is in state j ($P_j^t = P_{move,j}^t + P_{pause,j}^t$)
K	Transmission range of nodes
$A(a_j^t, b_k^t)$	The overlapped area between $Comm_j^t$ of node a and $Comm_k^t$ of node b
w^t	A specific relationship between a target coordinate and the communities in time period t
Ω^t	The set of all possible relationships between a target coordinate and the communities in time period t
$P_h(w^t)$	Unit-time hitting probability under the specific scenario w^t
$P_H(w^t)$	Hitting probability for a time period t under specific scenario w^t
P_m^t	Unit-time meeting probability in time period t
P_M^t	Meeting probability for a time period t

different *probabilities* (details are given later) to capture its spatial preference in mobility. In the TVC model, the mobility process of a node consists of *epochs* in these communities. When the node chooses to have an *epoch* in community j (we say that *the node is in state j* during this epoch), it starts from the end point of the previous epoch within $Comm_j^1$ and the epoch length (movement distance) is drawn from an exponential distribution with average $\overline{L_j}$, in the same order of the community edge length. The node then picks a random speed uniformly in $[v_{min}, v_{max}]$, and a direction (angle) uniformly in $[0, 2\pi]$, and performs a random direction movement within the chosen community with the chosen epoch length[9]. The first difference between the TVC model and the standard Random Direction model is hence the spatial preference and location-dependent behavior. Note that, a node can still roam around the whole simulation area during some epochs, by assigning an additional community that corresponds to the whole simulation field (e.g. $Comm_3^1$). We refer to such epochs as *roaming epochs*.

We next explain how a node selects the next community for a sequence of epochs. At the completion of an epoch, the node remains stationary for a pause time uniformly chosen in $[0, D_{max,j}]$. Then, depending on its current state i and time period t, the node chooses the next epoch to be in community j with probability $p_{i,j}^t$. This community selection process is essentially a time-variant Markov chain that captures the spatial and temporal dependencies in nodal mobility and thus makes the community selection process in the TVC model non-*i.i.d.*, an important feature absent in many synthetic mobility models even if they consider non-uniform mobility features. Now, if the end point of the previous epoch is in $Comm_j^t$ (this can be the case when the node has two consecutive epochs in $Comm_j^t$, or $Comm_j^t$ contains $Comm_i^t$), the node starts the next epoch directly. If, on the other hand, the node is currently not in $Comm_j^t$, a *transitional epoch* is inserted to bridge the two epochs in disjoint communities. The node selects a random coordinate point in the next community, moves directly towards this point on the shortest straight path with a random speed drawn from $[v_{min}, v_{max}]$, and then continues with an epoch in the next community. Hence the movement trajectory of a node is always continuous in space.

We next introduce the structure in time. To capture time-dependent behavior, one creates multiple *time periods* with different community and parameter settings. As an example, there are $V = 3$ time periods with duration T^1, T^2, and T^3 in Fig. 15. These time periods follow a *periodic structure* (e.g., a simple recurrent structure in Fig. 15 or the weekly schedule in Fig. 16). This setup naturally captures the *temporal preferences* (e.g., go to work during the days and home during the nights) and *periodicity* in human mobility. On the time boundaries between time periods, each node continues with its ongoing epoch, and decides the next epoch according to the new parameter settings in the new time period when it finishes the current epoch.

As a final note, we choose to construct the TVC model with simple building blocks introduced above due to its amenability to theoretical analysis [35] and flexibility. To further explain the flexibility of our TVC model, we note that the number of communities in each time period (denoted as S^t) can be different, and the communities can overlap (as in TP1 in Fig. 15) or contain each other (as in TP2 in Fig. 15). Finally, the time period structure, communities, and all other parameters could be assigned differently for *each node* to capture node-dependent mobility (e.g., people following different schedules, with different working places, etc.), while nodes can share some communities (i.e., the popular locations) as well. This construction allows for maximum flexibility when setting up the simulations for nodes with heterogeneous behaviors[10].

---

and the superscript represents the time period index.

[9]To avoid boundary effects, if the node hits the community boundary it is re-inserted from the other end of the area (i.e., "torus" boundaries). Note that we could also choose random waypoint or random walk models for the type of movement during each epoch.

[10]When necessary, we use a pair of parentheses to include the node ID for a particular parameter, e.g., $C_j^t(i)$ denotes the edge length of the j-th community during time period t for node i.
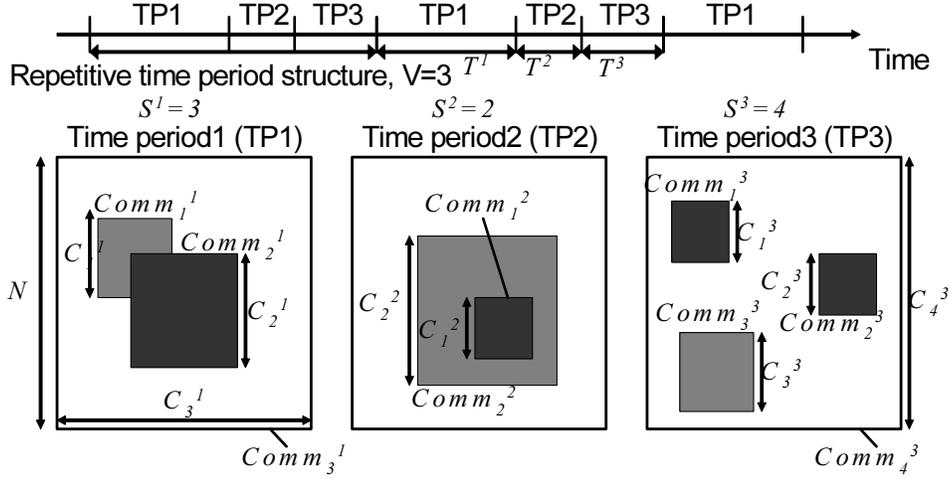
Figure 15: Illustration of a generic scenario of time-variant mobility model, with three time periods and different numbers of communities in each time period.
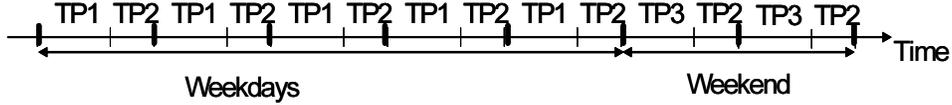

Figure 16: An illustration of a simple weekly schedule, where we use time period 1 (TP1) to capture weekday working hour, TP2 to capture night time, and TP3 to capture weekend day time.

4.2 Model Validation

The TVC model described in the previous section provides a general framework to model a wide range of mobility scenarios. In this section, our aim is to demonstrate the model's flexibility and validate its realism by generating synthetic traces from the model, with matching mobility characteristics to a well-known, publicly-available VANET trace. However, it is important to note that the use of such a model is not merely to match it with any specific trace instance available; this is only done for validation and calibration purposes. Rather, the goal is to be able to reproduce a much larger range of realistic mobility instances than a single trace can provide[11].

We first outline a general 3-step systematic process to construct specific mobility scenarios. Then, we demonstrate our success to generate matching mobility characteristics with three qualitatively different traces. All the parameter values we use in this section are also available in [6].

STEP 1: Determine the Structure in Space and Time

• (1.1 Number of communities) Each community in the TVC model corresponds to a location visited frequently by nodes. The number of communities needed is thus determined by how closely one wants the mobility characteristics to match with the curves. Due to the nature of skewed location visiting preference, in our experience, only two or three communities are needed to capture up to 85% of the user online time spent at the most popular locations. Such a simple spatial structure yields simple theoretical expressions. However, if one wants the model to capture more details (e.g., for detailed simulation), the user can instantiate as many communities as needed to explicitly represent the less visited locations.

---

[11]We have made our mobility trace generator available at [6]. The tool provides mobility traces in both ns-2 compatible format and time-location (i.e., (t, x, y)) format.

• (1.2 Location of communities) If the map of the target environment is available, one should observe the map and identify the points of attraction in the given environment to assign the communities accordingly. Alternatively, if the map is not available, one can instantiate communities at *random locations*.

• (1.3 Time period structure) Typically, human activities are bounded by daily and weekly schedules so a time period structure shown in Fig. 16 would suffice for most applications. If capturing finer behavior based on time-of-day is necessary, one could additionally split the day into time periods with different mobile node behavior.

STEP 2: Assign Other Parameters After the space/time structure is determined, one has to determine the remaining parameters for each community and time period. This includes π_j^t, $\overline{D_j^t}$, and $\overline{L_j^t}$, which represent the stationary probability (which is calculated after selecting proper $p_{i,j}^t$'s that lead to a desired stationary distribution using simple Markov chain theory), average pause time, and average epoch length, respectively, at community j during time period t.

• The average epoch length in each community, $\overline{L_j^t}$, should be at least in the same order as the edge length of the community, C_j^t. This is to ensure that the end point of the epoch becomes almost independent of its starting point, since the mixing time of the corresponding process becomes quite small. (The motivation for this requirement is to keep the theoretical analysis tractable.)

• The average duration the node stays in community j is given by $\pi_j^t(\overline{D_j^t} + \overline{L_j^t}/\overline{v})$. The ratio between the durations the node stays in each community shapes the location visiting preference curve.

• The highest peak of the re-appearance probability curve is determined by the weighted average probability of the node appearing in the same community during the same type of time period. This value is $\sum_{t=1}^{V} \frac{T^t}{\sum_{k=1}^{V} T^k} \sum_{j=1}^{S^t} (P_j^t)^2$, where P_j^t denotes the fraction of time the node spends in community j.

STEP 3: Adjust User On-off Pattern (Optional) The mobility trace generated by the TVC model is an "always-on" mobility trajectory (i.e., the mobile nodes are always present somewhere in the simulation field). However, in some situations some nodes might be absent occasionally. Thus one may need to make *optional* adjustments to turn nodes off in the generated trace, depending on the actual environment to match with. To address this we assign a probability $P_{on,j}$ as the probability for the node to be "on" in community j.

We now show that *skewed location visiting preferences* and *periodical re-appearance* are also prominent mobility properties in vehicle mobility traces. We obtain a vehicle movement trace from [2], a website that tracks participating taxis in the greater San Francisco area. We process a 40-day trace obtained between Sep. 22, 2006 and Nov. 1, 2006 for 549 taxis to obtain their mobility characteristics. The results are shown in Figures 17(a) and 17(b) with the label *Vehicle-trace*.

We use 30 communities and the weekly time schedule in (STEP1). We need more communities for this trace as the taxis are more mobile and visit more places than people on university campuses. From the actual trace, we discover that the taxis are offline (i.e., not reporting their locations) when not in operation. Hence we assume that the nodes are "on" only when they are moving. The pause times between epochs are considered as breaks in taxi operation. Therefore in (STEP3), $P_{on,j}^t = (\overline{L_j^t}/\overline{v})/(\overline{D_j^t} + \overline{L_j^t}/\overline{v})$, and we adjust the parameters in a similar way as described in the previous section. The curves in Figures 17(a) and 17(b) with label *Model* match with the curves with *Vehicle-trace* label well. As a final note, although vehicular movements are generally constrained by streets and our TVC model does not capture such microscopic behaviors, designated paths and other constraints could still be added in the model's map (for vehicular or human mobility) without losing its basic properties. We defer this for future work.

4.3 Derivation of Meeting and Inter-meeting Times

One of the biggest advantages of our model is that, in addition to the realism, it is also analytically tractable with respect to the important mobility properties required to analyze protocol performance. In this section, we demonstrate this property by deriving the meeting and the inter-meeting times for a specific instantiation
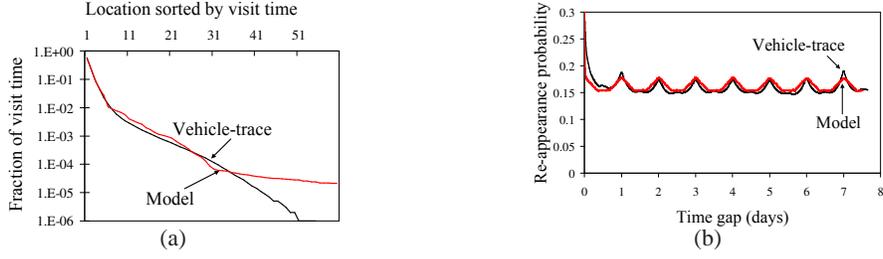
Figure 17: Matching mobility characteristics of the synthetic trace to the vehicle mobility trace. (a) Location visiting preferences. (b) Periodical re-appearance.

of the model. We refer to this instantiation as the community-based mobility model as it ignores the time-dependencies.

Community-based Model Nodes move inside the network as follows:

- each node i has a local community C_i of size $\|C_i\| = c^2 N, c \in (0, 1]$; a node's movement consists of a sequence of local and roaming epochs.

- a **local epoch** is a Random Direction movement[12] restricted inside area C_i with average epoch length $\overline{L}_c$ equal to the expected distance between two points uniformly chosen in C_i.

- a **roaming epoch** is a Random Direction movement in the entire network with expected length $\overline{L}$.

- *(local state L)* if the previous epoch of node i was a local one, the next epoch is a local one with probability $p_l^{(i)}$, or a roaming epoch with probability $1 - p_l^{(i)}$.

- *(roaming state R)* if the previous epoch of node i was a roaming one, the next epoch is a roaming one with probability $p_r^{(i)}$, or a local one with probability $1 - p_r^{(i)}$.

Lemma 4.1 calculates some useful probabilities, and follows easily from elementary probability theory.

Lemma 4.1 *Let us denote as $\pi_l^{(i)}$ and $\pi_r^{(i)}$ the probability that a given epoch of node i is a local or a roaming one, respectively. Let us further denote the probability that, at any time, the node is: (a) moving in local epoch as $p_{ml}^{(i)}$, (b) moving in roaming epoch as $p_{mr}^{(i)}$, (c) pausing after a local epoch as $p_{pl}^{(i)}$, (d) pausing after a roaming epoch as $p_{pr}^{(i)}$. Then:*

$$\pi_l^{(i)} = \frac{1 - p_r^{(i)}}{2 - p_l^{(i)} - p_r^{(i)}}, \qquad\qquad \pi_r^{(i)} = \frac{1 - p_l^{(i)}}{2 - p_l^{(i)} - p_r^{(i)}}$$

$$p_{ml}^{(i)} = \frac{\pi_l^{(i)} \frac{\overline{L}_c}{\overline{v}}}{\pi_l^{(i)} \overline{T}_l + \pi_r^{(i)} \overline{T}_r}, \; p_{mr}^{(i)} = \frac{\pi_r^{(i)} \frac{\overline{L}}{\overline{v}}}{\pi_l^{(i)} \overline{T}_l + \pi_r^{(i)} \overline{T}_r}, \quad p_{pl}^{(i)} = \frac{\pi_l^{(i)} \overline{T}_{stop}^l}{\pi_l^{(i)} \overline{T}_l + \pi_r^{(i)} \overline{T}_r}, \; p_{pr}^{(i)} = \frac{\pi_r^{(i)} \overline{T}_{stop}}{\pi_l^{(i)} \overline{T}_l + \pi_r^{(i)} \overline{T}_r}.$$

Table 3 summarizes the new notation specific to the community model described above. We will focus here on the case where each node i has its own community C_i, but all nodes have the same mobility characteristics, that is, $p_l^{(i)} = p_l$ and $p_r^{(i)} = p_r, \forall i$ (i.e. drop the (i) from all probabilities). The heterogeneous case is only a straightforward extension of this [35].

Meeting Time: To compute the expected meeting time, we break the problem into the following two cases: (i) *non-overlapped communities*, which refers to the case where the communities of the two nodes under study are disjoint, and (ii) *overlapped communities*, which refers to the case where the communities of the two nodes are the same. (We ignore partial overlap to simplify analysis.) Each of these two cases are analyzed separately, and then we take a weighted average over the two cases. The following theorem states the result.

---

[12]Note that each node could also perform Random Waypoint movement in each epoch, instead of Random Direction.

Table 3: Notation for the Community-based mobility model

C_i	community of node i: $\|C_i\| = c^2 N, c \in (0, 1]$
p_l	probability that next epoch is local, given that previous epoch was local
p_r	probability that next epoch is roaming, given that previous epoch was roaming
π_l	probability that a given epoch is a local one
π_r	probability that a given epoch is a roaming one
p_{mr}	probability that a node is in roaming state and moving
p_{ml}	probability that a node is in local state and moving
p_{pr}	probability that a node is in roaming state and pausing
p_{pl}	probability that a node is in local state and pausing
$\overline{L}_c$	expected length of local epoch
$\overline{T}^l_{stop}$	expected pause time for a local epoch
$\overline{T}_l$	expected local epoch duration $(\overline{L}_c/\overline{v} + \overline{T}^l_{stop})$
$\overline{T}_r$	expected roaming epoch duration $(\overline{L}/\overline{v} + \overline{T}_{stop})$

Theorem 4.1 *The probability distribution of the meeting time M_{comm} under the Community-based mobility model can be approximated by the weighted sum of two exponential distributions, with expected v alue:*

$$EM_{comm} = (1 - c^2)EM_{comm,diff} + c^2 EM_{comm,same}. \tag{15}$$

where,

$$EM_{comm,diff} = \left[\frac{2K\overline{v}}{N} \left(\hat{v}_{rd}((p_{mr} + p_{ml})^2 - p_{ml}^2) + (2p_{mr}(p_{pr} + p_{pl})) + (2p_{ml}p_{pr}) \right) \right]^{-1}$$

$$EM_{comm,same} = \left[\frac{2K\overline{v}}{N} \left(\frac{\hat{v}_{rd}p_{ml}^2}{c^2} + \frac{2p_{ml}p_{pl}}{c^2} + \hat{v}_{rd}\left((p_{mr} + p_{ml})^2 - p_{ml}^2\right) + 2p_{mr}(p_{pr} + p_{pl}) + 2p_{ml}p_{pr} \right) \right]^{-1}$$

are the expected meeting time for nodes with non-overlapping and overlapping communities, respectively.

As a special case, in some real-life situations each node tends to move most of the time in a very small area that is different for each node (e.g. at home), and that could be entirely covered by the node's antenna, while the network might be much larger (e.g. a city-wide wireless Metropolitan Area Network). In this case, the probability distribution for the meeting time can be again approximated by a single exponential, simplifying some derivative results.

Corollary 4.1 (Small Community) *When the community size of nodes is much smaller than the network area $(\|C_i\| \ll N)$, the meeting time $EM_{comm}^{(small)}$ under the Community-based Random Direction model is exponentially distributed with mean value:*

$$EM_{comm}^{(small)} = \frac{ET_{rd} + \frac{1-p_r}{1-p_l}\frac{N}{2K\overline{L}}\overline{T}^l_{stop}}{p_m^c \hat{v}_{rd} + 2(1 - p_m^c)}, \tag{16}$$

where $p_m^c = \frac{(1-p_l)\overline{L}/\overline{v}}{(1-p_r)\overline{T}^l_{stop} + (1-p_l)\overline{T}_r}$.

Inter-meeting Time: To calculate the inter-meeting times, we again condition on the two subcases of overlapping and non-overlapping communities. We first state the result for the simpler case of non-overlapping communities.
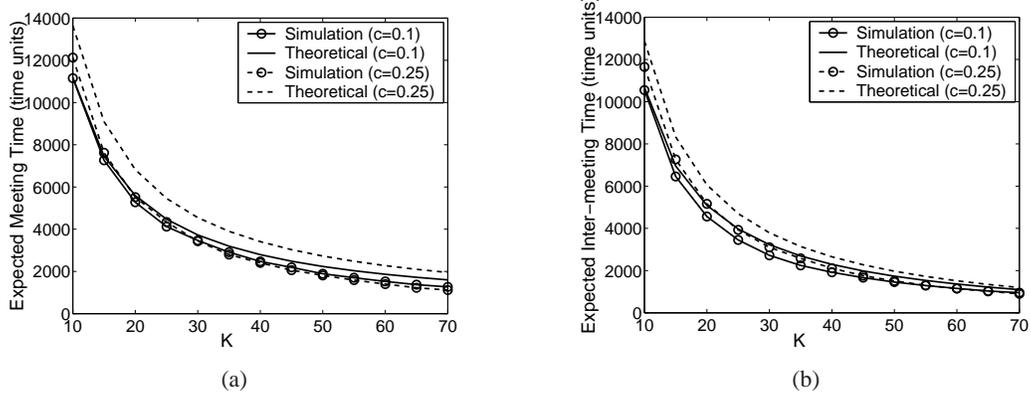
Figure 18: Simulation and analytical results for the Community-based mobility model. (a) Meeting time. (b) Inter-meeting time. Network parameters: $N = 500 \times 500, \overline{L} = 150, p_l = 0.9, p_r = 0.5, \overline{v} = 1.0, \overline{T}_{stop} = \overline{T}^l_{stop} = 0$.

Lemma 4.2 *The expected inter-meeting time for nodes with non-overlapped communities is $EM^+_{comm,diff} = EM_{comm,diff}$.*

When the communities of the two nodes overlap, then the situation becomes slightly more complicated. Specifically, if the two nodes meet within their community, there is a high probability that they will meet again quite fast. The following lemma states the result.

Lemma 4.3 *The expected inter-meeting time for nodes with overlapping communities is*

$$EM^+_{comm,same} = p^+_1 E[M^+_1] + p^+_2 E[M^+_2] + (1 - p^+_1 - p^+_2)EM_{comm,same}, \tag{17}$$

where (i) p^+_1 is the probability that when the two nodes met, both were in their local states and only one of the nodes was moving, and $E[M^+_1]$ is the expected inter-meeting time for this case, (iii) p^+_2 is the probability that when the two nodes met, both were in their local states and moving, and $E[M^+_2]$ is the expected inter-meeting time for this latter case.

We next state the value of the expected inter-meeting time, EM^+_{comm}, in terms of $EM^+_{comm,same}$ and $EM^+_{comm,diff}$ in the following theorem.

Theorem 4.2 *The expected inter-meeting time of the Community-based mobility model is*

$$EM^+_{comm} = (1 - c^2)EM^+_{comm,diff} + c^2 EM^+_{comm,same}. \tag{18}$$

More results as well as the derivation of all the results presented in this section and the expressions for $p^+_1, E[M^+_1], p^+_2$ and $E[M^+_2]$ can be found in [36].

Accuracy of the Analysis: Figures 18(a) and 18(b) compare the analytical and simulation results for the expected meeting and inter-meeting times under the Community-based mobility model. As can be seen, theory matches simulations quite closely.

4.4 Analyzing Spraying-based Routing Schemes

In this section, we state the expected delay values for epidemic routing, fast spray and wait and fast spray and focus. Please refer to [26] for the derivation of these values. We study epidemic routing as it forms

the basic building block of fast spraying. To simplify the presentation in this section, we assume that there are r small communities, and these communities are assumed to be small enough such that all nodes within a community are within each other's range. We also assume that the nodes spend most of their time within their respective communities. Finally, we assume that the number of nodes sharing a community is equal across all r communities, that is the number of nodes sharing a community is equal to $\frac{M}{r}$.

4.4.1 Epidemic Routing

This section derives the expected delay of epidemic routing for the community-based mobility model. Since each node spends most of its time within its community (which implies $E[M_{comm,diff}] >> E[M_{comm,same}]$), we make an approximation to simplify the exposition by assuming that with high probability, a node starting from its stationary location distribution will first meet a node within its own community than a node belonging to a different community. This implies that once a node gets a copy of a packet, with high probability, all members of its community will get the copy before any node outside its community. A simple outcome of this is that the first $\frac{M}{r} - 1$ nodes to get a copy of the packet belong to the source's community.

We first state how much time it takes for all nodes within the source's community to get a copy of the packet. This derivation is different from all the derivations in previous sections because $E[M_{comm,same}] \neq E[M_{comm,same}^+]$. Thus, we need to keep track of which pair of nodes have met in the past but were unable to successfully exchange the packet. We model the system using the following state space: (m, m_p) where $1 \le m \le \frac{M}{r}$ is the number of nodes which have a copy of the packet and $0 \le m_p \le m\left(\frac{M}{r} - m\right)$ is the number of node pairs such that only one node of the pair has a copy of the packet, they have met at least once after the node (which has the copy) received its copy, and they were unable to successfully exchange this packet in their past meetings. Let $E[D_{in}(m)]$ denote the expected time it takes for the number of nodes having a copy of the packet to increase from m to $m+1$ given $m < \frac{M}{r}$ (which implies that all nodes within the source's community have not yet received a copy of the packet).

Lemma 4.4 $E[D_{in}(m)] = \sum_{m_p=0}^{m\left(\frac{M}{r} - m\right)} p_{m,m_p} \frac{E[T_{m,m_p}]}{1 - p_{m,m_p}^{self}}$, where $E[T_{m,m_p}]$ is the expected time elapsed till one of the nodes not having a copy meets a node having a copy of the packet given that the system is in state (m, m_p), p_{m,m_p}^{self} is the probability that the system remains in the state (m, m_p) after these nodes (which met after $E[T_{m,m_p}]$) are unable to successfully exchange the packet, and p_{m,m_p} is the probability that the system visits state (m, m_p).

Please refer to [26] for the expressions of $E[T_{m,m_p}]$, p_{m,m_p}^{self} and p_{m,m_p}.

Next, we state the value of $E[D_{epidemic}^{comm}(m)]$ which is the expected time it takes for the number of nodes having a copy of the packet to increase from m to $m + 1$.

Lemma 4.5 $E[D_{epidemic}^{comm}(m)] = \begin{cases} E[D_{in}\left(rem\left(m, \frac{M}{r}\right)\right)] & \text{if } rem\left(m, \frac{M}{r}\right) \neq 0 \\ \frac{E[M_{comm,diff}]}{m(M-m)p_{success2}^{epidemic}} & \text{if } rem\left(m, \frac{M}{r}\right) = 0 \end{cases}$ where $p_{success2}^{epidemic} = 1 - \left(1 - p_{txS2}^{epidemic}\right)^{E[\tau_{comm,diff}]}$ and $rem(x, y)$ is the remainder left after dividing x by y.

Finally, we derive the expected delay of epidemic routing for the community based mobility model (denoted by $E[D_{epidemic}^{comm}]$) in terms of $E[D_{epidemic}^{comm}(m)]$.

Theorem 4.3 $E[D_{epidemic}^{comm}] = \sum_{i=1}^{M-1} \frac{1}{M-1} \sum_{m=1}^{i} E[D_{epidemic}^{comm}(m)]$.

4.4.2 Fast Spray and Wait

This section derives the expected delay of fast spray and wait routing scheme for the community-based mobility model. As before, first we derive the value of $E[D^{comm}_{fsw}(m)]$. For $m < L$ (in the spray phase), the value of $E[D^{comm}_{fsw}(m)]$ is derived in a manner similar to the derivation of $E[D^{comm}_{epidemic}(m)]$ as flooding is used to spread the L copies in the spray phase. Next, we state the value of $E[D^{comm}_{fsw}(L)]$ which is the expected time to find the destination in the wait phase.

Lemma 4.6 $E[D^{comm}_{fsw}(L)] = \frac{\frac{M}{r} - \hat{l}}{M - L} \left(\sum_{m_p=0}^{\hat{l}\left(\frac{M}{r} - \hat{l}\right)} p_{\hat{l},m_p} E[T_{\frac{m_p}{\frac{M}{r} - \hat{l}}}] \right) + \left(1 - \frac{\frac{M}{r} - \hat{l}}{M - L} \right) \frac{E[M_{comm,diff}]}{L p^{fsw}_{success2}}$, where $\hat{l} = rem\left(L, \frac{M}{r}\right)$, $E[T_s]$ is the expected time till the destination receives a copy of the packet given there are s nodes belonging to the destination's community which were unable to successfully exchange the packet with the destination in the past, and $p^{fsw}_{success2} = 1 - \left(1 - p^{fsw}_{txS2}\right)^{E[\tau_{comm,diff}]}$.

Finally, we derive the expected delay of fast spray and wait for the community based mobility model (denoted by $E[D^{comm}_{fsw}]$) in terms of $E[D^{comm}_{fsw}(m)]$.

Theorem 4.4 $E[D^{comm}_{fsw}] = \sum_{i=1}^{L} p^{fsw}_{dest}(i) \sum_{m=1}^{i} E[D^{comm}_{fsw}(m)]$.
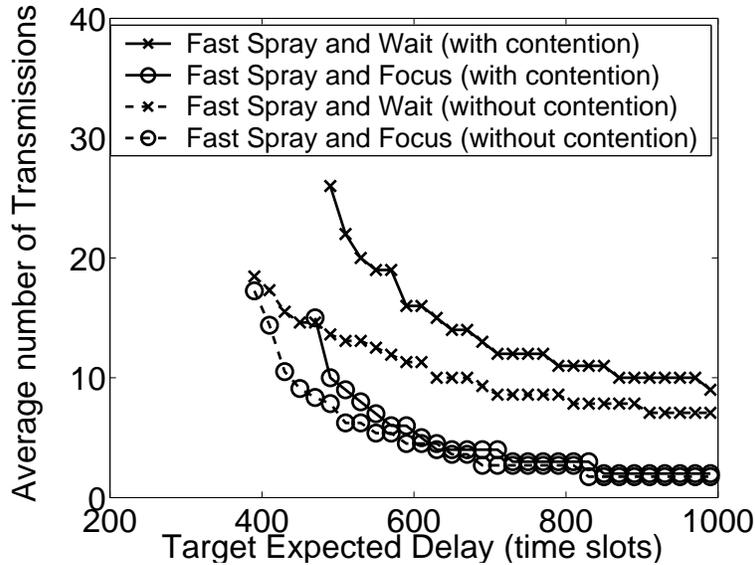

Figure 19: Comparison of fast spray and wait and fast spray and focus. Average number of transmissions required to deliver the packet to the destination vs target expected delay. Network parameters: $N = 500 \times 500$ square units, $M = 40, K = 20, \Theta = 5, s_{BW} = 1$ packet/time slot, $p_l = 0.8, p_r = 0.15, r = 4$.

4.4.3 Fast Spray and Focus

For community-based mobility models, [25] proposed the use of a simpler function as a utility function for their 'Label' scheme: If a relay meets a node which belongs to the same community as the destination, the relay hands over its copy to the new node. We use this simple and effective utility function to route copies of the packet in the focus phase. For example, handing over the copy of the packet to a vehicle which shares the parking lot with the destination will get the message delivered faster and reliably.

This section derives the expected delay of fast spray and focus for the community-based mobility model. As before, first we derive $E[D^{comm}_{fsf}(m)]$. Since flooding is used to spread the copies in the spray phase,

$E[D_{fsf}^{comm}(m)]$ for $m < L$ can be derived in a manner similar to the derivation of $E[D_{epidemic}^{comm}(m)]$. The next lemma derives the value of $E[D_{fsf}^{comm}(L)]$ which is the expected time it takes for the packet to get delivered to the destination in the focus phase.

Lemma 4.7 $E[D_{fsf}^{comm}(L)] = \frac{\frac{M}{r}-\hat{\imath}}{M-L}\left(\sum_{m_p=0}^{\hat{\imath}(\frac{M}{r}-\hat{\imath})} p_{\hat{\imath},m_p} E[T_{\frac{m_p}{\frac{M}{r}-\hat{\imath}}}]\right) + \left(1 - \frac{\frac{M}{r}-\hat{\imath}}{M-L}\right)\left(\frac{E[M_{comm,diff}]}{L\frac{M}{r}p_{success2}^{fsf}} + \frac{\frac{M}{r}-1}{\frac{M}{r}}\left(E[M_{comm,same}]\right.\right.$

$+\left.\left.\frac{(1-p_{successs1}^{fsf})E[M_{comm,same}^+]}{p_{success1}^{fsf}}\right)\right)$, where $\hat{\imath} = rem\left(L, \frac{M}{r}\right)$, $p_{success1}^{fsf} = 1 - \left(1 - p_{txS1}^{fsw}\right)^{E[\tau_{comm,diff}]}$ and $p_{success2}^{fsf} = 1 - \left(1 - p_{txS2}^{fsw}\right)^{E[\tau_{comm,same}]}$.

Now we derive the expected delay of fast spray and focus for the community based mobility model (denoted by $E[D_{fsf}^{comm}]$) in terms of $E[D_{fsf}^{comm}(m)]$.

Theorem 4.5 $E[D_{fsf}^{comm}] = \sum_{i=1}^{L} p_{dest}^{fsf}(i) \sum_{m=1}^{i} E[D_{fsf}^{comm}(m)]$, where $p_{dest}^{fsf}(i) = \begin{cases} \frac{1}{M-1} & i < L \\ \frac{M-L}{M-1} & i = L \end{cases}$.

We now use the analysis presented in this section to validate the claims made in Section 3.7.3 through a numerical example. We study how much performance gains are achieved by spray and focus over spray and wait (for the community-based mobility model) both with and without contention in the network by plotting the minimum value of the average number of transmissions it takes to achieve a given target expected delay for both the schemes in Figure 19. We first find the minimum value of L which achieves the given target expected delay for both the schemes and then find the average number of transmissions which is equal to $\sum_{i=1}^{L} i p_{dest}^{R}(i)$. (The minimum value of L is computed using the analytical expressions derived in this section. The value of $p_{dest}^{R}(i)$ for both the schemes was derived in Theorems 4.4 and 4.5.) We observe that fast spray and focus outperforms fast spray and wait even with contention in the network, with gains being larger with contention. Since $E[M_{comm,diff}] >> E[M_{comm,same}]$, forwarding a copy to any node in the destination's community in the focus phase significantly reduces the delay for the same L without significantly increasing the contention as it requires only one extra message per copy. Hence, fast spray and focus shows more performance gains over fast spray and wait after incorporating contention.

5 Support for Multicasting

While there are safety applications that involve two vehicles only, for example, applications that prevent accidents resulting from changing lanes when a vehicle is in the blind spot of another, the majority of safety applications, or example, pre- and post-crash warnings, involve a large number of vehicles.

In such one-to-many communication scenarios it is important to avoid duplicate transmissions. To understand what this means, consider a scenario where two vehicles 0_1 and 0_2 have a collision on a highway which results in blocking a number of lanes. These vehicles would broadcast a warning message, that would be received by vehicles close to the collision, say vehicles $1_i, i = 1, 2, 3, \ldots$. These vehicles, in turn, would forward the warning message to vehicles further away from the collision, say vehicles $2_i, i = 1, 2, 3, \ldots$. A duplicate transmission would result if 0_i would send two messages to 1_i, one for itself and one to be forwarded to 2_i. A duplicate transmission would also result if 2_i would directly receive a second warning message from 0_i once it is closer to the collision. Simple rules, translated into utility values for each potential receiver can be used to suppress such duplicate transmissions.

To suppress duplicate transmissions, we use the following idea. Let 0_1 and 0_2 broadcast the message. Let $1_i, i = 1, 2, 3 \ldots$ receive this broadcast. Then amongst these nodes, only the nodes which are furthest away from the origin of the message (in terms of distance) are allowed to re-broadcast. Let the set of nodes which receive this next broadcast be denoted by 2_i. Again the same rule applies. Only nodes furthest away from the origin are allowed to re-broadcast. Nodes which lie between the origin of the message and the furthest away

nodes do not transmit that message. This suppresses duplicate transmissions. The idea is similar to the EXOR protocol [11].

The main challenge in the design of the protocol is to define a metric which reflects the distance from the origin of the message. We could use a metric similar to the one used in [11], however, fast-moving vehicles do not allow the collection of link losses. So, we experimented with the following idea. Each node j which receives a broadcast starts a random timer proportional to the received signal strength. If j receives the same message from another node, then it imples that another node further away also received the same message. So j cancels the timer. On the other hand, if j does not receive any message before the timer expires, it broadcasts the message. We found the protocol to suppress most duplicate messages while ensuring delivery to all nodes.

6 Applications and Experiments

6.1 Evaluating Applications using Realistic Vehicle Mobility Models

In this section, we evaluated two safety applications in simulations either using the binary spray and focus routing or the proposed multicasting algorithm. To evaluate these applications, we use realistic vehicle mobility models. Note that the research community has used in the past a number of unrealistic mobility models. We used a better model, the so-called TVC model, to design and evaluate routing schemes. However, this model is still not realistic enough. In the absence of real ample data, we use synthetic mobility traces generated using the following two mobility models.

Freeway mobility model [10]: This model emulates the behaviour of of vehicles on a freeway. In this model, maps are used. There are several freeways on the map and each freeway has lanes in both directions. The dierences between Random Waypoint and Freeway are the following: (a) Each mobile node is restricted to its lane on the freeway. (b) The velocity of mobile node is temporally dependent on its previous velocity. (c) If two mobile nodes on the same freeway lane are within the safety distance (SD), the velocity of the following node cannot exceed the velocity of preceding node.

Due to the above relationships, the Freeway mobility pattern is expected to have spatial dependence and high temporal dependence. It also imposes strict geographic restrictions on the node movement by not allowing a node to change its lane.

Manhattan mobility model [10]: This model emulates the behaviour of vehicles on streets defined by maps. The map is composed of a number of horizontal and vertical streets. Each street has two lanes for each direction (north and south direction for vertical streets, east and west for horizontal streets). The mobile node is allowed to move along the grid of horizontal and vertical streets on the map. At an intersection of a horizontal and a vertical street, the mobile node can turn left, right or go straight. This choice is probabilistic: the probability of moving on the same street is 0.5, the probability of turning left is 0.25 and the probability of turning right is 0.25. The velocity of a mobile node at a time slot is dependent on its velocity at the previous time slot. Also, a node's velocity is restricted by the velocity of the node preceding it on the same lane of the street. The inter-node and intra-node relationships involved are the same as in the Freeway model. Thus, the Manhattan mobility model is also expected to have high spatial dependence and high temporal dependence. It too imposes geographic restrictions on node mobility. However, it diers from the Freeway model in giving a node some freedom to change its direction.

We evaluate the following two applications with the Freeway mobility model and the Manhattan mobility model.

- **Post-crash Warning:** Once an accident occurs, approaching vehicles should be warned to prevent subsequent accidents, inform drivers to use alternate routes, etc. This is a standard multicasting scenario, where the warning message should propagate to a large number of vehicles, using multi-hop

paths formed by vehicles and possibly some roadside stations. We used the multicasting technique described in Section 5 to distribute the warning message. The message includes the information of the direction of travel in which collision occured and whether the collision occured on a surface street or on a highway. If the collision is on a highway, message is propagated only in the reverse direction while on a surface street, the message is propagated in all directions.

We used IEEE 802.11(b) protocol at the MAC layer in simulations, set the data rate to 11Mbps, and assumed the message size to be 128 bytes. We switched off RTS/CTS. We found that the message spread throughout the network within a few seconds even without any roadside stations. Assuming zero propagation delay for communication between different roadside stations, using them reduced the delivery delay to less than couple of seconds.

- **Pre-crash Warning:** This application refers to a situation where a number of vehicles communicate to each other to warn their drivers that there is high possibility of a collision. First, a two-way point to point communication takes place between the two cars that might collide. Then, a warning message is multicasted to a number of trailing vehicles.

  To implement pre-crash warning, we use multicasting with a constant TTL value of 6. So, the message is not forwarded beyond 6 hops. This avoids flooding the message to cars to whom the pre-crash message is not intended. We again use IEEE 802.11(b) MAC, set the data rate to 11Mbps and assume the message size to be 64 bytes. The message size is smaller as the information to be delivered is less. We observed that the message spread to all cars within 6 hops within 0.5s of its origination.

- **SOS Services:** In this application a vehicle is periodically broadcasting a SOS message. The objective is to route this message to the nearest roadside station, which is then expected to forward the message to an emergency center, e.g. a police station. We use spray and focus routing scheme to forward this message to the nearest roadside station. All roadside stations are assumed to be a common node, that is the address of all roadside stations is the same. So, if a vehicle sees any roadside station, it updates its utility function.

  We again use IEEE 802.11(b) protocol without RTS/CTS, set the data rate to 11Mbps and assume the message size to be 64 bytes. The message size is smaller because the information to be delivered is less. Expected delay was always less than 10s even in a sparse network with very fast moving cars with short contact durations.

- **Curve Speed Warning:** This application requires cooperation between vehicles and roadside stations. The stations monitor the speed of the cars and inform the drivers if they are approaching the curve with too much speed. When the approaching speed is beyond some limit, the roadside stations sends a warning message to the errant vehicle, and this message is then transformed to a pre-crash warning by the errant vehicle and multicasted to all vehicles within 6 hops.

The applications we study show the efficacy of the proposed routing schemes proposed in delivering essential data within a reasonable time frame. Simulating and evaluating other safety applications is left as future work.

6.2   Experiments

Simulations and analysis are very important tools in designing routing solutions. However, they both have their limitations. Motivated by this, we decided to go beyond simulations and evaluate the spray and focus routing scheme on an actual testbed. So, we acquired a number of ICOP eBox-3854 which are a mini-PC running Click [30] and Linux 2.6.20. Each node is equipped with a Senao NMP-8602 wireless card running the madwifi driver [3] and an omni-directional antenna. These wireless cards will be operated in IEEE 802.11(b) promiscuous mode in a manner similar to the simulations.

We are currently in the process of implementing spray and focus over the click router. We also plan to talk with the USC Transportation Office [7] to install these devices in their fleet of trams and "campus cruisers" that connect the two main campuses, and, in addition, cover any route inside a 10-mile radius around the two main campuses.

References

[1] Car to car communication consortium. http://www.car-to-car.org/index.php?id=128.

[2] Cab spotting, a project that tracks taxi mobility in the san francisco bay area. URL `http://cabspotting.org/api`.

[3] MadWifi. http://madwifi.org/.

[4] Vehicle safety communications pro ject (nhtsa). http://www-nrd.nhtsa.dot.gov/pdf/nrd-12/060419-0843.

[5] Prevent. http://www.prevent-ip.org/en/home.htm.

[6] Simulation codes for the tvc model. URL `http://nile.cise.ufl.edu/\~{}weijenhs/TVC\_model`.

[7] Trojan Transportation. http://transnet.usc.edu.

[8] D. Aguuayo, J. Bicket, S. Biswas, G. Judd, and R. Morris. Link-level measurements from an 802.11b mesh network. In *Proceedings of ACM SIGCOMM*, 2004.

[9] Dave Aldous and James Fill. Reversible markov chains and random walks on graphs. (monograph in preparation.). *http://stat-www.berkeley.edu/users/aldous/RWG/book.html*.

[10] F. Bai, N. Sadagopan, and A. Helmy. The important framework for analyzing the impact of mobility on performance of routing for ad hoc networks. *Elsevier Ad Hoc Networks Journal*, 1, November 2003.

[11] S. Biswas and R. Morris. Opportunistic routing in multi-hop wireless networks. In *Proceedings of ACM SIGCOMM*, 2005.

[12] J. Y. Le Boudec and Milan Vojnovic. Perfect simulation and stationarity of a class of mobility models. In *Proceedings of IEEE INFOCOM*, 2005.

[13] Josh Broch, David A. Maltz, David B. Johnson, Yih-Chun Hu, and Jorjeta Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocols. In *Mobile Computing and Networking*, 1998.

[14] X. Chen and A. L. Murphy. Enabling disconnected transitive communication in mobile ad hoc networks. In *Proceedings of ACM POMC*, August 2001.

[15] R. M. de Moraes, H. R. Sadjadpour, and J.J. Garcia-Luna-Aceves. Throughput-delay analysis of mobile ad-hoc networks with a multi-copy relaying strategy. In *Proceedings of IEEE SECON*, 2004.

[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] Richard Durrett. *Probability: Theory and Examples*. Duxbury Press, second edition, 1995.

[19] A. Gamal, J. Mammen, B. Prabhakar, and D. Shah. Throughput-delay tradeoff in wireless networks. In *Proceedings of IEEE INFOCOM*, 2004.

[20] Matthias Grossglauser and David N. C. Tse. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Transactions on Networking*, 10(4), 2002.

[21] P. Gupta and P. Kumar. Capacity of wireless networks. *IEEE Transactions on Information Theory*, 46 (2), 2000.

[22] P. Gupta and P. R. Kumar. Critical power for asymptotic connectivity in wireless networks. In *in Stochastic Analysis, Control, Optimization and Applications*. Birkhauser, Boston, 1998.

[23] W. Hsu and A. Helmy. On important aspects of modeling user associations in wireless lan traces. In *The Second International Workshop On Wireless Network Measurement*, 2006.

[24] W. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy. Modelling spatial and temporal dependencies of user mobility in wireless mobile networks. *to appear at IEEE/ACM Transactions on Networking*, 2009.

[25] P. Hui and J. Crowcroft. How small labels create big improvements. In *Proceedings of IEEE PerCom Workshop on Intermittently Connected Mobile Ad Hoc Networks*, 2007.

[26] A. Jindal and K. Psounis. Contention-aware performance analysis of mobility-assisted routing. *IEEE Transactions on Mobile Computing*, February 2009.

[27] 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.

[28] J. Leguay, T. Friedman, and V. Conan. Dtn routing in a mobility pattern space. In *Proceedings of ACM SIGCOMM workshop on Delay Tolerant Networking (WDTN)*, 2005.

[29] Anders Lindgren, Avri Doria, and Olov Schelen. Probabilistic routing in intermittently connected networks. *SIGMOBILE Mobile Computing and Communication Review*, 7(3), 2003.

[30] Robert Morris, Eddie Kohler, John Jannotti, and M. Frans Kaashoek. The Click modular router. *SIGOPS Oper. Syst. Rev.*, 1999.

[31] G. Sharma and R.R. Mazumdar. On achievable delay/capacity trade-offs in mobile ad hoc networks. In *Workshop on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (Wi'Opt'04)*, 2004.

[32] G. Sharma, R. R. Mazumdar, and N. B. Shroff. Delay and capacity trade-offs in mobile ad hoc networks: A global perspective. In *Proceedings of IEEE INFOCOM*, 2006.

[33] T. Small and Z.J. Haas. Resource and performance tradeoffs in delay-tolerant wireless networks. In *Proceedings of ACM SIGCOMM workshop on Delay Tolerant Networking (WDTN)*, 2005.

[34] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Single-copy routing in intermittently connected mobile networks. In *Proceedings of IEEE SECON*, 2004.

[35] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Performance analysis of mobility-assisted routing. In *Proceedings of ACM MOBIHOC*, 2006.

[36] T. Spyropoulos, A. Jindal, and K. Psounis. An analytical study of fundamental mobility properties for encounter-based protocols. *International Journal of Autonomous and Adaptive Communications Systems (IJAACS)*, 2008.

[37] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Efficient routing in intermittently connected mobile networks: The single-copy case. *IEEE/ACM Transactions on Networking*, 2008.

[38] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Efficient routing in intermittently connected mobile networks: The multi-copy case. *IEEE/ACM Transactions on Networking*, 2008.

[39] Thrasyvoulos Spyropoulos, Konstantinos Psounis, and Cauligi S. Raghavendra. Spray and wait: Efficient routing in intermittently connected mobile networks. In *Proceedings of ACM SIGCOMM workshop on Delay Tolerant Networking (WDTN)*, 2005.

[40] Yu-Chee Tseng, Sze-Yao Ni, Yuh-Shyan Chen, and Jang-Ping Sheu. The broadcast storm problem in a mobile ad hoc network. *Wireless Networks*, 8(2/3), 2002.

[41] Amin Vahdat and David Becker. Epidemic routing for partially connected ad hoc networks. Technical Report CS-200006, Duke University, April 2000.

[42] J. Widmer and J.-Y. Le Boudec. Network coding for ecient communication in extreme networks. In *Proceedings of ACM SIGCOMM workshop on Delay Tolerant Networking (WDTN)*, 2005.

[43] X. Zhang, G. Neglia, J. Kurose, and D. Towsley. Performance modeling of epidemic routing. In *Networking*, 2005.