Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institut für
Technische Informatik und
Kommunikationsnetze

Zurinet

Using public transportation as a DTN backbone

Leo den Hartog

Master Thesis
Advisors: Dr. Thrasyvoulos Spyropoulos, Franck Legendre
Supervisor: Prof. Dr. Bernhard Plattner

Abstract

This thesis focuses on the mobility properties of route- and schedule-based networks, such as public transportation networks. We are interested in the behaviour of nodes in such networks and the feasibility of using them as a DTN backbone. We model the Zurich and Amsterdam tram networks and analyse the distribution of the inter-contact times. Our first contribution is that we find these inter-contact times to be exponentially distributed. We then adapt simple analytical models based on random mixing and exponential inter-contact times for the expected delivery delay in delay tolerant networks to our network and compare them to experimental results. The analytical model proves to be an accurate predictor of real performance over a variety of densities. Finally we perform a feasibility study to determine the potential of enabling stops to relay messages and analyse the congestion for epidemic routing.

Contents

List of Figures

Chapter 1

Introduction

The density of smart-phones in the population is growing fast. People are used to carrying their phone, which is nowadays more like a small computer, everywhere. This provides an existing deployment of potential delay-tolerant network nodes on a great scale. Therefore, there also exists a good opportunity to benefit from opportunistic networking approaches, provided that the quality of service is sufficient for specific applications.

A lot of research in the field of mobility in delay tolerant networks has been performed. Most experimental results are based on human mobility in specific surroundings [2], and suffer from some known measurement issues. On the other hand, synthetic traces are derived from simulations based on random way-point mobility or simple adaptations thereof. Existing traces focus mainly on the contact and inter-contact properties to describe mobility. For the inter-contact times in these traces, some seem to be modelled best by a power law distribution, others by an exponential one. No conclusive evidence has been found to prove either one generally correct.

We study the Zurich tram network using real schedules and simulations, trying to find an approximation of the distribution of inter-meeting times between the trams to determine whether route- and schedule-based networks (RSBNs) can also be analytically modelled according to existing models and whether such a network could act as a backbone for a bigger DTN. Little previous work exists on RSBNs [1] and usually not on a scale as big and dense as the Zurich network. In addition to the real schedules, we derive synthetic schedules based on the real ones to vary the density of nodes in the network. We also add random delays to individual nodes to create a generic simulation on which we can verify our results within different schedules.

Our paper has two major contributions. First, we show that the inter-meeting times in the Zurich network, as well as in the Amsterdam tram network, are approximately exponentially distributed. Independent of the exact schedule, inter-meeting times in route-based networks seem to be exponentially distributed on a wide range of tram intervals. Secondly, we propose a simple analytical model to describe the expected delivery delay for routing messages in such a RSBN using epidemic routing. The analytical model found is based on well known Markov chain models for epidemic routing [9] and holds well for all schedules and network topologies considered. This shows that the expected delivery delay in a route- and schedule-based network can be calculated through a simplistic model derived from the general formula for epidemic message spread in other types of networks, despite the evidently more ”structured” mobility in RSBNs.

In the last section, we look at the feasibility of using the Zurich tram network as a DTN backbone. We study the effect of enabling stops to store and forward messages as well as the influence of bandwidth constraints on congestion in the network and try to give some estimates for the throughput of such a network.

Chapter 2

Related work

Little previous work exists that studies the mobility properties of RSBNs. DieselNet is a real-world DTN test bed over the bus network that connects the several sites of the campus. Zhang et al. [1] analyse the inter-contact times between the buses and study the performance of epidemic routing in that network. They could not show conclusive evidence that the inter-contact times can be modelled generally and focus on individual lines rather than the aggregate. Hui et al. [2, 8] created the well-known haggle traces, consisting of six separate datasets of human mobility at conferences, on campuses and throughout the city of Cambridge. They calculated the inter-contact times and show that it can be modelled by a power law distribution. These traces also have common measurement issues because nodes fail or the contacts are too short and are not recognized due to the granularity of the measurement. Furthermore, mobility of people at a conference is certainly very different than the structured and scheduled one in a RSBN. A paper by Gaito et al. [3] examines the possibility of using the bus network of Milano as a DTN backbone. They propose and evaluate a routing algorithm for their network, but contrary to our approach they are not particularly interested in modelling the mobility. Karagiannis et al. [4] focus on fitting the inter-contact times of mobility in delay tolerant networks. They do not include RSBNs in their work, but show for several settings using human mobility and one for vehicular movement, that neither the exponential nor the power law distribution is able to fit all of them. Depending on density and size of a network, different distributions are used to model the aggregated inter-contact times.

In this work, we analyse the mobility properties of RSBNs in more detail. We study the inter-contact times of real schedules and find them to be approx. exponentially distributed. Furthermore, using simulations we find that this holds for a large range of schedule densities and networks. Finally we propose a simple analytical model to predict the performance of epidemic spread over RSBNs.

Chapter 3

Modelling inter-contact times

To get a better understanding of RSBNs, we first wanted to look at the mobility of nodes. RSBN have a different kind of mobility than other delay tolerant networks, because (i) nodes run on predefined, known schedules and (ii) nodes share a set of paths. Many papers analyse the inter-meeting times of nodes in various networks [8], because inter-meeting times are considered a basic parameter on which the performance of routing in delay tolerant networks depends. They conclude that inter-meeting times have either an exponential or a power-law distribution. For RSBNs, only little work has been done [1, 3] and no conclusive evidence for either distribution has been published.

As our primary dataset, we choose the Zurich tram and bus network, because it is dense, rather big and keeps its schedule quite precisely. Including buses, it consists of 16 lines on 230 stops that go through the city center with an average schedule interval of approximately 6 minutes. It has a star like topology with almost every line passing one of the three main stations, which is shown in Figure 3.1. We simulated the network based on the real schedule for the times from 9 am to 3 pm, because this is the most regular schedule without interference from lines that stop early or start late. We assumed that nodes stop for 30 seconds at any given tram stop and we made sure that they are at each stop according to the schedule, travelling at a constant speed.

Contrary to other papers [2, 4], we focus on the any-inter-contact time rather than the pair-wise inter-contact time, because measuring contacts between single nodes in a tram network proves difficult because (i) trams might switch identities at final stops and (ii) not all trams meet each other. So we assume similar to the work of Gaito et al. [3] that meetings between specific trams are less important than the fact of meeting another tram at all. Meetings between individual trams also matter less in a tram network, because several nodes travel on the same route and exchange messages in a predefined interval. Individual trams may also be exchanged, the important factor being to which line a tram belongs, not the individual node id.

The first question that arose from this reasoning is the importance of inter-meeting times between nodes of the same line. These intra-line meetings happen at fixed intervals based on the frequency of trams running on the line. Because of that, the any-inter-meeting times distribution would be upper bounded by half of the interval with which trams start their routes. That would mask the tail of the any-inter-meeting time distribution and cut off an important part of mobility related information. We assume that nodes which only meet nodes of the same line for a long time (i.e. on their way to a final stop they do not share), have a different, smaller effect on (epidemic) routing and those meetings should therefore be excluded. So we first compared the any-inter-contact time between all trams to the any-inter-contact time excluding meetings between trams of the same line in

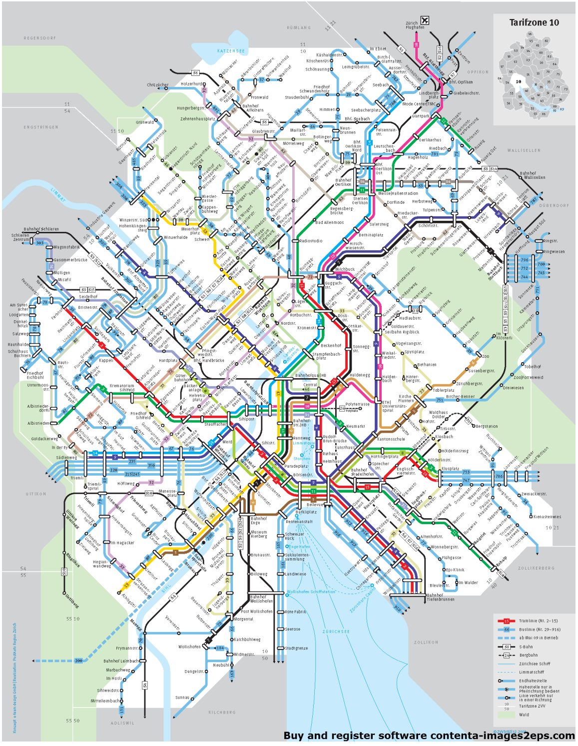

Figure 3.1: Zurich tram network

Figure 3.2.

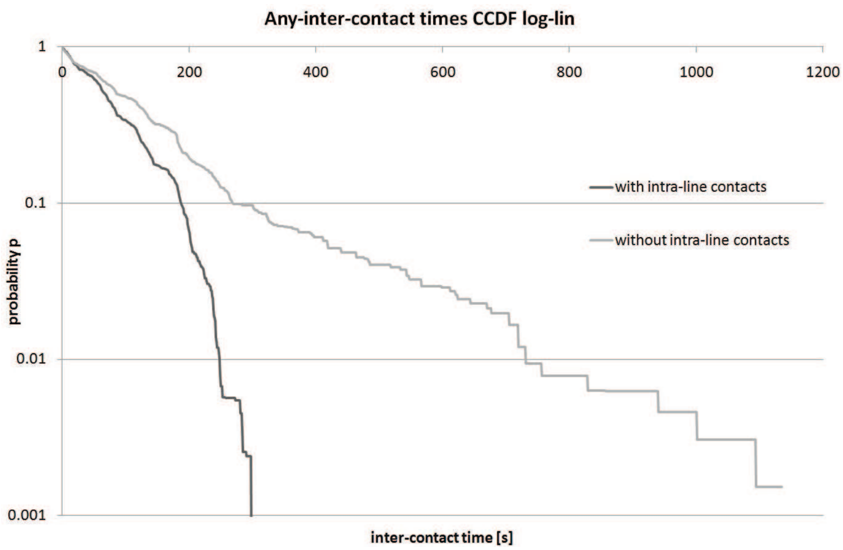

Figure 3.2: Any-inter-contact times with and without intra-group contacts in log-linear scale in Zurich on the real schedule

It is clearly seen that the distribution of the any-inter-contact times with intra-line contacts in Figure 3.2 is cut off at around 300 seconds, which is the average interval between two nodes of the same line. The distribution without intra-line contacts in Figure 3.2 shows nodes having inter-meeting times up to 1100 seconds, so the effect of the upper bound is clearly shown in these graphs. With the infection model for message spread, intra-line contacts benefit only from the first node of a line in both directions. Other nodes just meet already infected nodes, because they just met a carrier a short time ago. So we assume that the model without intra-line contacts is more appropriate routing.

The any-inter-meeting distribution excluding intra-group contacts in log-linear scale looks approximately linear, which hints for an exponential distribution. We tried to fit the distribution with several common models like log-normal and Weibull, but the exponential model delivered the most accurate fit for the distribution. Weibull actually performs as good as the exponential distribution, but with β approximately 1 it is essentially exponential. The comparison of the different distributions is shown in Table 3.1, we used the sum of squared error as the measurement for accuracy. The exponential distribution fitted for the any-inter-meeting times distribution is shown in Figure 3.4. According to our assumptions from the plots, the exponential model fits the distribution well. The exponential parameter λ does not vary too much between different lines, and can therefore be aggregated over the whole network as is seen in Figure 3.3.

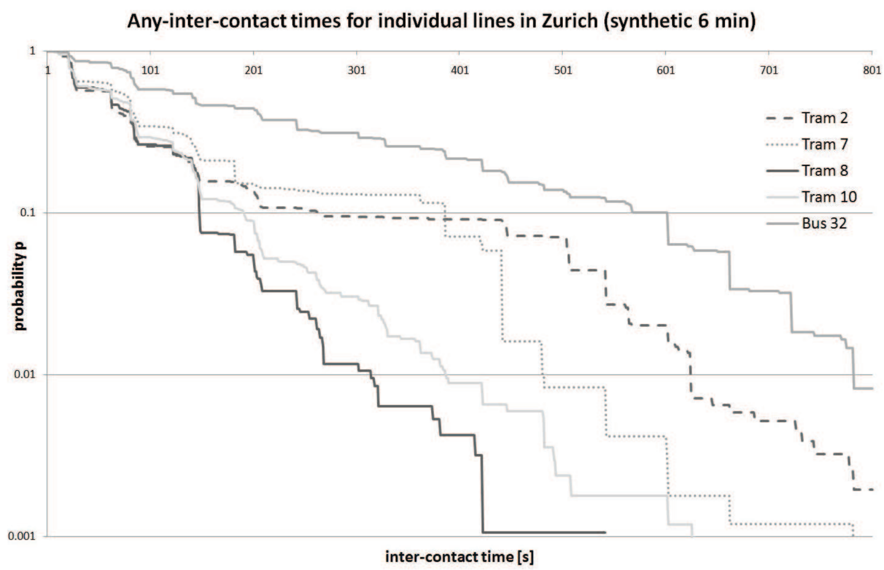

Figure 3.3: Any-inter-contact times of individual lines in Zurich

Distribution	α	β	SSE	R-Square	RMSE
Exponential	0.008		0.278	0.994	0.015
Power Law	5.567	0.481	17.070	0.639	0.119
Weibull	129.500	0.944	0.240	0.995	0.014
Log-normal	1.048	4.438	0.645	0.986	0.023

Table 3.1: Comparison of different distributions for the any-inter-meeting times in Zurich

The formulas for the distributions in a ccdf plot are given as:

$$y_{exp}(x) = e^{-\alpha * x}$$

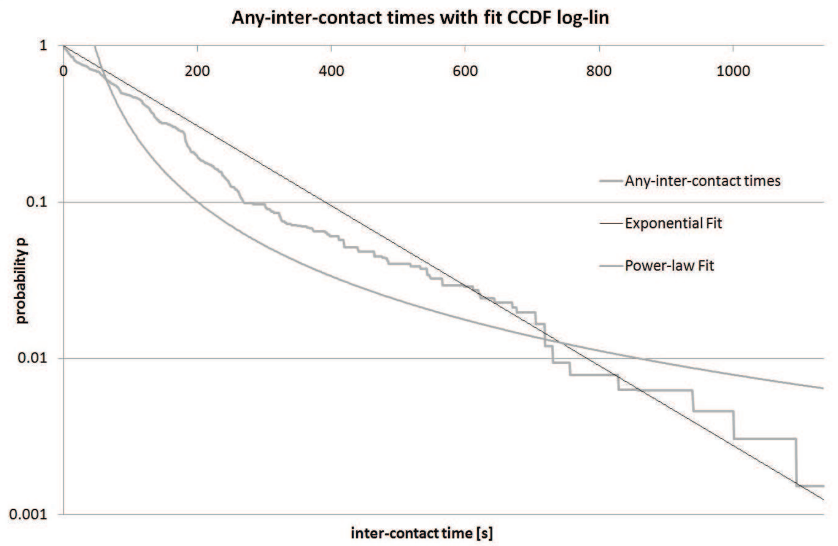

Figure 3.4: Any-inter-contact times without intra-group contacts with exponential and power law fitting in Zurich

$$y_{power}(x) = (\frac{\alpha}{x})^{\beta}$$

$$y_{weibull}(x) = e^{-(\frac{x}{\alpha})^{\beta}}$$

$$y_{log-normal}(x) = \frac{1}{2} - \frac{1}{2} * erf(\frac{ln(x) - \beta}{\alpha * \sqrt{2}})$$

As scales for the exactness of a certain distribution we chose three well-known statistical measurements.

- SSE: Sum of squared error
- R-Square: The coefficient of determination, $R^2 = 1 - \frac{SSE}{SST}$ with SST as the total sum of squares.
- RMSE: Mean square error

We have thus validated that the exponential distribution is a good fit for the real tram schedule running at an interval of approximately 6 minutes. To further test and validate our conclusions for different schedules, we implemented a synthetic model for the tram schedule. We kept the original topology and routes, but replaced the schedule with a synthetic model. Instead of the real starting times, we defined an average interval I with which trams start their run at a final stop. With this we can vary the number of nodes in the network to check whether the exponential fit also performs best for different node densities.

We ran simulations with tram intervals between 3 and 30 minute. For 30 minutes, the network becomes very sparse, corresponding to a late night schedule for example, whereas routes are almost completely full at the 3 minutes interval. The simulations should therefore cover most realistic scenarios of a RSBN. The any-inter-contact times for the synthetic scenarios are plotted in Figure 3.5.

Table 3.2 shows the R-square values for different intervals. The exponential model delivered a very good fit for all intervals considered. Even for a very dense network, the exponential model did not break. To take the variations from the fixed interval into account, we distributed the exact starting times randomly over the time $[-\frac{I}{2}; \frac{I}{2}]$, but the results were similar.

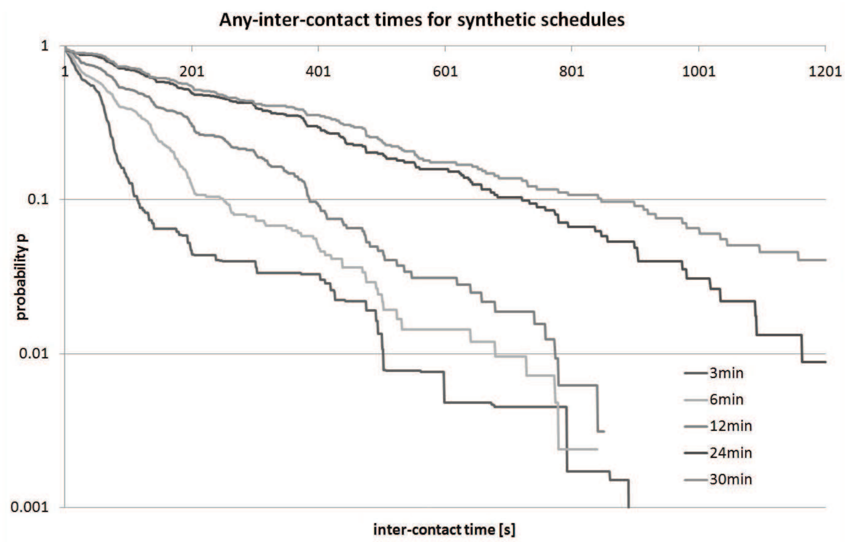

Figure 3.5: Synthetic any-inter-contact times for intervals 3 to 30 minutes in Zurich on synthetic schedules

Interval[s]	3 min	6 min	12 min	24 min	30 min
Exponential	0.979	0.989	0.995	0.996	0.995
Power Law	0.673	0.663	0.628	0.617	0.635
Weibull	0.979	0.995	0.997	0.996	0.996
Log-normal	0.977	0.986	0.983	0.982	0.986

Table 3.2: Comparison of R-square for different synthetic schedules in Zurich

Counter-intuitively these results confirm that the any-inter-meeting times are in fact distributed exponentially over various densities in the Zurich network. The exponential model seems to be independent of the density of the network, so we also wanted to analyse the importance of the network structure. For that reason, we simulated the Amsterdam network as a second benchmark for the exponential model. The Amsterdam tram-network has a similar density (16 lines, 230 stops) but a quite different topology as is shown in Figure 3.6. Whereas Zurich has a star like topology, Amsterdam focuses around a center where half the lines have one of their final stops and the other half of the lines circle around tangentially to the center. In the case of the Amsterdam network, the any-inter-contact time is also best fitted with an exponential model compared to other models (Table 3.3). The any-inter-meeting time distribution for a synthetic 6 minutes interval with the exponential fit is shown in Figure 6 in log-lin.

Our simulations render considerable evidence that any-inter-meeting times in RS-BNs are distributed exponentially. That the mobility in a rather complex network behaves similarly to a random way-point model is interesting, because it has many implications on routing. Why the any-inter-meeting times are exponentially distributed should be analysed in further work. Preliminary studies will implement a synthetic topology to generalize our conclusions for any randomly created network.

Distribution	α	β	SSE	R-Square	RMSE
Exponential	0.008		0.834	0.982	0.026
Power-law	5.272	0.487	16.81	0.629	0.118
Weibull	120.6	0.999	0.834	0.981	0.026
Log-normal	1.008	4.389	1.152	0.975	0.031

Table 3.3: Comparison of different distributions for the any-inter-meeting times in the Amsterdam network

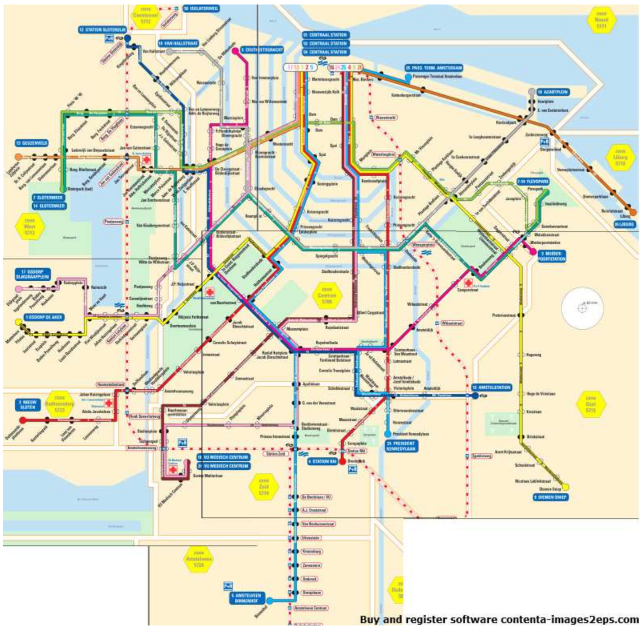

Figure 3.6: Amsterdam tram network

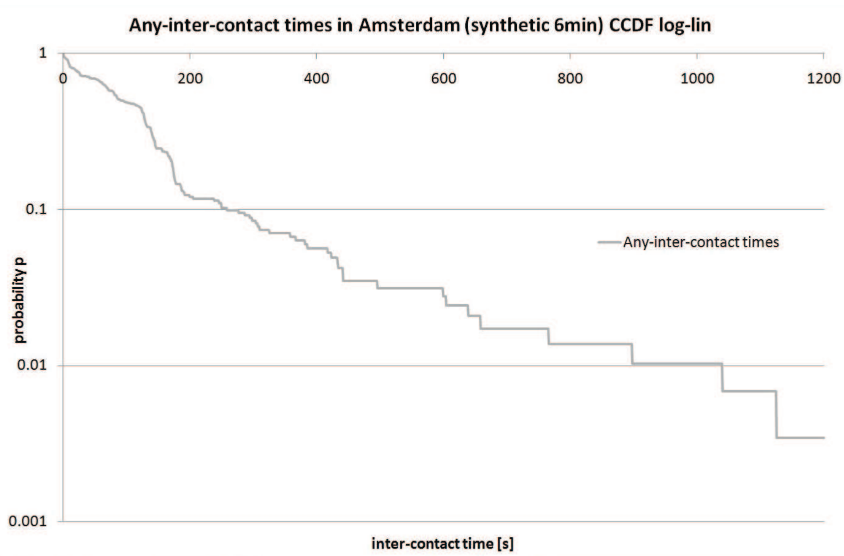

Figure 3.7: Any-inter-contact time without intra-group contacts in Amsterdam

Chapter 4

Analysis of epidemic message spread

Inter-meeting times are a core indication for the performance of epidemic message spread in DTNs. We found that the any-inter-meeting times in RSBNs are distributed exponentially, so standard models for delay tolerant networks should render reasonable results for RSBNs as well. Therefore we want to try to adapt a simple analytical model for the performance of epidemic routing to analyse routing in RSBNs. It would be interesting if a RSBN, which has predefined paths and schedules, could be modelled similarly to a random way-point scenario where all nodes move arbitrarily and are able to reach any destination. If estimates for the performance of epidemic routing deliver approximate results, it would give a new perspective to the concept of routing in RSBNs. Many well-studied ideas of delay-tolerant networks could be adapted for RSBNs and simple routing protocols might be able to deliver good performance.

4.1 Analytical model

In this section we try to predict the performance of routing in a RSBN based on existing concepts for delay tolerant networks. There are several models for epidemic message spread in networks with exponentially distributed inter-meeting times. For our network we assume:

- Message sources and destinations are randomly chosen stops.
- A node that can deliver a message to the destination always delivers.
- Bandwidth and buffer constraints are not taken into account because we are primarily interested in the optimal expected delivery delay without constraints due to congestion.

For the RSBNs, we need to additionally make the following assumptions:

- Any-inter-contact times distribution of all nodes is exponentially distributed over the same parameter λ. (We saw that this holds for the Zurich and Amsterdam networks)
- Routes are interconnected in such a way that all nodes are theoretically able to see almost every other node. (This is only an approximation, as in practice there is a bias as to which lines a given line sees more often)

In our simulations, messages are created at stops, so the overall delivery delay E_{dd} needs to include the time E_{pickup} to initially pick up the message. The time between when the first node picks up the message until the node which delivers the message is infected is E_{inter}. Finally, the last node takes time to deliver the message to the destination stop approximated by E_{drop}.

E_{pickup} is dependent on the interval I between trams and the average number of lines M_{lps} that visit a certain stop. So the average time until a message is picked up equals the interval between trams divided by two (for both directions), M_{lps} (because more than one line might visit a stop) and two (half for the average distance of the closest tram at the moment of message creation). This renders E_{pickup} to:

$$E_{pickup} = \frac{I}{2 * 2 * M_{lps}} \tag{4.1}$$

To calculate E_{inter} we use an approach by Daley and Gani [9]. They use the distribution of pairwise inter-contact times to calculate the expected delivery delay of a message from any node to another. It is based on a markov chain using identical, independent and exponentially distributed inter-contact times. If the mobility of a RSBN behaves similarly to the random waypoint networks they considered, we can adapt that model for any-inter-contact times, using the markov chain in figure 4.1 with node 1 being the first node to pick up the message.

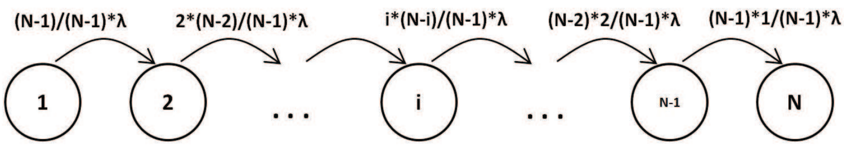

Figure 4.1: Markov Chain for any-inter-contact times

We calculate E_{inter} analogue to Daley and Gali. E_{inter} equals the average of the estimated times to reach states E_k with k nodes infected:

$$E_{inter} = \sum_{k=1}^{N-1} E_k \frac{1}{N-1} \tag{4.2}$$

The meeting rate for a single node meeting an uninfected node, with i nodes infected is: $i * \frac{(N-i)}{N-1} * \lambda$. This renders E_k to:

$$E_k = \sum_{i=1}^{k} \frac{N-1}{i * (N-i) * \lambda} \tag{4.3}$$

Inserting Equation 3 into Equation 2 leads to the final E_{inter}:

$$E_{inter} = \frac{1}{\lambda} * \sum_{k=1}^{N-1} \sum_{i=1}^{k} \frac{1}{i * (N-i)} \approx \frac{ln(N)}{\lambda} \tag{4.4}$$

using
$N =$ the number of Nodes
$\lambda =$ the distribution rate of the any-inter-contact times

E_{drop} is the time it takes the final node to go the destination stop. It is calculated as the average length of a line L_{line} ($L_{line} \approx 34min$ for the Zurich network) divided by a factor x. This factor should be between 2 and 4 for realistic schedules. Based on experimental results, we scale the factor from 4 to 2 according to the interval that trams start in the range of 3 to 30 minutes. The denser the network is, the more likely is it that the final node gets the message closer to the destination. This leads to:

$$E_{drop} = \frac{L_{line}}{2} * \frac{I + 1800}{3600} \tag{4.5}$$

Combining Equations 1, 4 and 5, the formula for E_{dd} becomes:

$$E_{dd} = E_{pickup} + E_{inter} + E_{drop} \tag{4.6}$$

$$E_{dd} = \frac{I}{2 * 2 * M_{lps}} + \frac{ln(N)}{\lambda} + \frac{L_{line}}{2} * \frac{I + 1800}{3600} \tag{4.7}$$

With:

$$N = \frac{L_{line}}{I} * N_{line} * 2 \tag{4.8}$$

This provides a closed form equation for the expected delivery delay for our transportation networks. The essential part for the estimated delivery delay, the time for the interconnection between nodes, is based on well studied work for general delay tolerant networks with exponentially distributed inter-meeting times. We made two key assumptions for the RSBNs that we studied. Our first assumptions was that nodes are able to meet almost any other node, which is true for the Zurich as well as the Amsterdam network although the probability of meeting with individual nodes is not the same. This assumption will not hold true for all transportation networks, so further studies should go into generalizing the argument for less interconnected networks. Secondly we assumed that the inter-meeting times distribution is the same for all nodes. In realistic networks, some lines are of course more interconnected than others, but we believe that this does not have a large influence, because the different λ_i for the individual lines vary only slightly.

4.2 Experimental results

Now we want to validate the accuracy of our analytical forms by comparing them to experimental results. Preliminary simulations are performed on the DTN simulator ONE (ref. ONE). To be able to implement synthetic schedules and run them in reasonable time, all results are calculated by a proprietary simulator. Existing simulators such as ONE were too slow and did not offer enough design flexibility to render all our simulations.

For Zurich, the specific variables for the topology are the average number of lines per stop M_{lps}:

$$M_{lps} = \frac{385}{228} = 1.6812$$

and the average number of nodes N in the network at any given time according to Equation 8:

$$N = \frac{L_{line}}{I} * N_{line} * 2 = \frac{2040}{I} * 16 * 2 = \frac{65280}{I}$$

Hereby, we get the following values for E_{inter} in Table 4.1 and E_{drop} in Table 4.2 for the intervals between 6 and 24 minutes. Both formulas hold for these intervals with an error of less than 14 percent. From this we calculate E_{dd}, as is shown in Table 4.3.

Interval[s]	Analyt. E_{inter} [s]	Sim. E_{inter} [s]	Diff. [s] ([%])
360	623	692	-69 ([-11])
540	767	870	-103 ([-13])
720	854	909	-55 ([-6])
900	1036	1114	-78 ([-8])
1080	1144	1160	-16 ([-1])
1260	1170	1245	-75 ([-6])
1440	1216	1379	-163 ([-13])

Table 4.1: Comparison of E_{inter} for the ZH network

Interval[s]	Analyt. E_{drop} [s]	Sim. E_{drop} [s]	Diff. [s] ([%])
360	612	592	20 ([-3])
540	663	677	-14 ([-2])
720	714	737	-23 ([-3])
900	765	790	-25 ([-3])
1080	816	812	4 ([-0])
1260	867	820	47 ([-5])
1440	918	854	64 ([-7])

Table 4.2: Comparison of E_{drop} for the ZH network

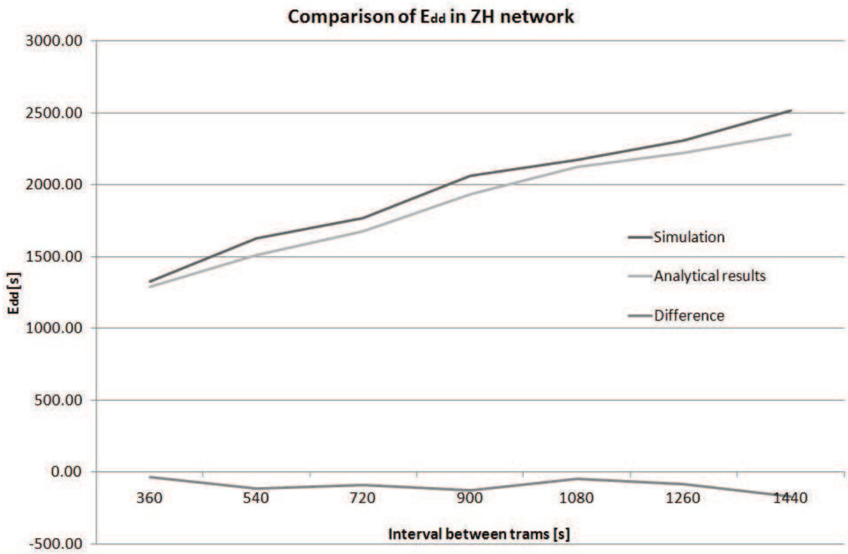

Figure 4.2: Comparison of experimental and analytical results in the ZH network

Table 4.3 shows, that the analytical results match the experimental results with an error of less than 8 percent for all intervals between 6 and 24 minutes. From this we conclude that the analytical model for the E_{dd} holds for the Zurich network.

For Amsterdam, the specific variables for the topology are again the average number of lines per stop M_{lps}:

$$M_{lps} = \frac{386}{224} = 1.7232$$

and the average number of nodes N in the network at any given time according

Interval[s]	λ	Analyt. E_{dd} [s]	Sim. E_{dd} [s]	Diff. [s] ([%])
360	0.00834	1288	1323	-35 (-3)
540	0.00626	1510	1624	-114 (-7)
720	0.00528	1675	1767	-92 (-5)
900	0.00414	1934	2059	-125 (-6)
1080	0.00359	2120	2170	-50 (-2)
1260	0.00337	2224	2307	-83 (-4)
1440	0.00314	2347	2517	-170 (-7)

Table 4.3: Comparison of E_{dd} for the ZH network

to Equation 8:

$$N = \frac{L_{line}}{I} * N_{line} * 2 = \frac{1939}{I} * 16 * 2 = \frac{62040}{I}$$

So for intervals ranging from 6 to 24 minutes we get Table 4.4.

Interval[s]	λ	Analyt. E_{dd} [s]	Sim. E_{dd} [s]	Diff. [s]
360	0.00836	1250	1570	-320 (-20)
540	0.00562	1552	1994	-442 (-22)
720	0.00457	1736	2265	-501 (-22)
900	0.00363	2023	2488	-465 (-19)
1080	0.00321	2194	2502	-245 (-12)
1260	0.00306	2280	2538	-258 (-10)
1440	0.00269	2481	2850	-369 (-13)

Table 4.4: Comparison of E_{dd} for the ADAM network

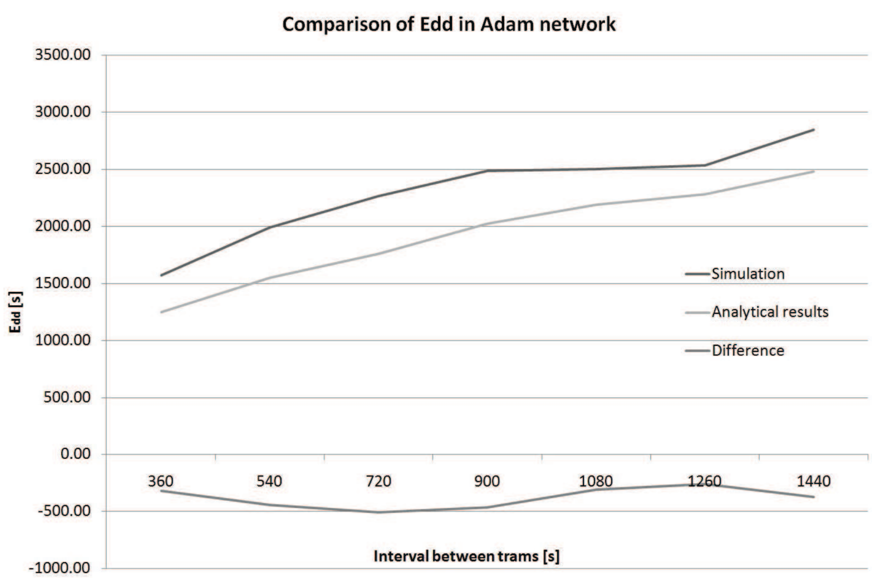

Figure 4.3: Comparison of experimental and analytical results in the Adam network

The results in the Amsterdam network are a bit worse than in the ZH network. But the difference between analytical and experimental results stays within a 22 percent range for intervals of 6 to 24 minutes. The different topology changes the outcome especially on very dense networks. The results improve significantly if E_{drop} is changed to a constant value $E_{drop} = \frac{L_{line}}{2}$ as is shown in Table 4.5. Due to the different topology, the time to deliver the message rises and the balance between E_{inter} and E_{drop} changes. What specific factor of the topology causes this will be researched in future work.

Interval[s]	λ	Analyt. E_{dd} [s]	Sim. E_{dd} [s]	Diff. [s]
360	0.00836	1637	1570	67 (4)
540	0.00562	1892	1994	-102 (-5)
720	0.00457	2055	2265	-210 (-10)
900	0.00363	2266	2488	-222 (-9)
1080	0.00321	2387	2502	-115 (-5)
1260	0.00306	2425	2538	-113 (-4)
1440	0.00269	2578	2850	-272 (-10)

Table 4.5: Comparison of E_{dd} for the ADAM network with adapted E_{drop}

The simulations give us enough evidence to believe that the performance of epidemic routing in RSBNs can in fact be modelled with sufficient accuracy, using simple models for epidemic message spread in delay tolerant networks. We believe that the main factor for this are the exponentially distributed inter-contact times. In future work, we plan to investigate whether our conclusions hold for arbitrary synthetic RSBN topologies and to identify the parameters that make an RSBN topology amenable to such simple epidemic model analysis.

Chapter 5

Feasibility study for routing

In this last section, we will look into our public transportation network from a different angle. We want to perform a simple feasibility study for routing in RSBNs. The previous section confirmed that RSBNs in fact behave similarly to other DTNs, and that epidemic routing performs as the standard model assumes. We primarily looked at epidemic routing in order to test our assumptions rather than to implement a specific routing protocol. To study the routing process in more detail, we identified two main issues. First we want to analyse the effects of including routers at tram stops to improve performance. A tram network is built with the idea of people getting on and off at stops, but includes the possibility of waiting there. In a similar way, enabling stops to relay messages should enhance the performance of the network, because no paths are lost, but contact opportunities would increase especially in sparse networks. Secondly, we are interested in congestion, due to the immense overhead of epidemic routing, to analyse the need for implementing more sophisticated protocols and the benefits these could offer.

5.1 Effects of stop relays on performance

Enabling stops to relay messages is bound to have a beneficial effect on the E_{dd}. Including stops does not destroy any paths, but is able to create new contact opportunities. Therefore the sparser a network is, the bigger the effect should be. We categorized stops by the number of lines that cross them, to analyse if specific stops have a bigger benefit than others. We enabled relaying in stops with a number of lines equal or above a certain threshold T_h and compared the results for different thresholds. We compared the results for intervals between 3 and 24 minutes in Tables 5.1 and 5.2.

Interval[s]	No Forwarding	T_h 6	T_h 5	T_h 4	T_h 3	T_h 2
180	1150	1150	1145	1142	1132	1118
360	1415	1413	1396	1395	1370	1358
540	1660	1658	1649	1647	1573	1562
720	1970	1898	1875	1868	1833	1824
900	2200	2135	2108	2101	1967	1922
1080	2335	2310	2246	2236	2121	2119
1260	2311	2284	2236	2224	2140	2124
1440	2693	2618	2540	2528	2410	2394
N_{stops}	0	3	7	17	41	101

Table 5.1: Forwarding at stops with Threshold T_h based on lines [s]

Interval[s]	No Forwarding	T_h 6	T_h 5	T_h 4	T_h 3	T_h 2
180	100	100	100	99	98	97
360	100	100	99	99	97	96
540	100	100	99	99	95	94
720	100	96	95	95	93	93
900	100	97	96	96	89	87
1080	100	99	96	96	91	91
1260	100	99	97	96	93	92
1440	100	97	94	94	89	89
N_{stops}[%]	0	1	3	7	18	44

Table 5.2: Forwarding at stops with Threshold T_h based on lines [%]

As expected, there is a beneficial effect of enabling stops to relay. For small intervals, the effect is rather small (only 6 percent with more than 100 stops included). The bigger the interval and therefore the sparser the network is, the more the observed benefits. But even for a pretty sparse network at a 24 minutes interval, the benefit is only 11 percent. We believe this is due to the robustness of the epidemic routing protocol. Even in a sparse network, epidemic routing creates enough messages to ensure a reasonable spread, although a lot of contact opportunities are missed. Many lines share, especially in the center, a few stops in a row, so even without enabling stops to relay messages, enough contact opportunities seem to arise. Consequently, if the amount of resource usage for epidemic routing is acceptable for a given network, (e.g. enough bandwidth and memory at nodes) installing additional wireless relays at stops would not offer significant benefits. To get a better understanding of this point, we will next investigate the incurred level of congestion and its effect.

5.2 Congestion

Congestion is a natural problem for any delay tolerant network. The tram network has the advantage over other DTNs, that buffer space and battery power are not big constraints. An average sized router with a big hard drive can easily be fitted into a tram and connect to the main power line. The constraint that matters most is bandwidth for message transmission. We assume that messages between trams are forwarded through standard IEEE-802.11 (WLAN) communication. That should suffice for several messages at a station, but some connections appear when two tram pass by each other, so they are only in range for about 4 seconds. Based on an average speed of 10 m/s per tram, that would imply a distance of 80 meter in movement. According to a study performed with cars in an urban environment [6] at a distance of about 75 meters, the signal to noise ratio sinks below 15dB.

For the feasibility of routing over such a network, we want to study the effect of congestion for epidemic routing. Epidemic routing has the biggest overhead compared to any other (reasonable) protocol, so if epidemic routing is feasible, other protocols should perform good as well, in terms of congestion. For the bandwidth constraints, we assume:

- Bandwidth: 4Mbit/s, according to Singh et al. for crossing vehicles [6]
- Message size: 2 Mbyte = 16 Mbit, because that is just enough to apply a wide range of services such as email or multimedia messaging.
- Buffer uses FIFO

- No buffer constraints
- No energy constraints

With these, we tested the Zurich network for different message loads at a tram-interval of 6 minutes. We compared two different scenarios. One has no prior knowledge of E_{dd} whatsoever and performs epidemic routing with a TTL of 30 minutes. The second simulation also runs epidemic routing, but it uses a TTL-oracle [5] , which means that it knows the exact time to delivery up front and can set its TTL accordingly. This is a reasonable assumption in a RSBN, because the message path and delivery time could be calculated or at least approximated in a real environment through a time-variant Dijkstra-algorithm [7].

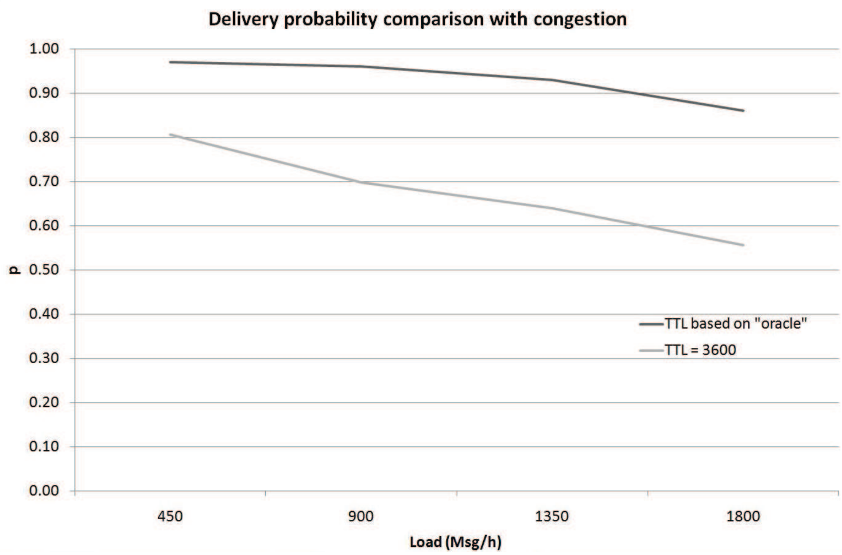

Figure 5.1: Comparison of congestion with and without TTL-oracle

The comparison shows a clear loss due to congestion in epidemic routing without an oracle. We think this is because the E_{dd} varies a lot between messages as is seen in Figure 5.2, thus many messages keep being sent through the network uselessly and block the useful transmissions in the FIFO buffers. The epidemic routing with a TTL-oracle performs well, even with big message loads (88 percent at 3.6 GB per hour).

This shows that with implementation of an oracle, routing is realistically feasible even with epidemic routing. Creating a TTL oracle in a route- and schedule-based network is simple, because the E_{dd} as well as the optimal path in a network without delays can be calculated easily with a time-variant Dijkstra-algorithm [7]. A Dijkstra algorithm should also be able to calculate reasonable approximations for the individual delivery delays, even if delays are taken into account. However, enough slack for queuing delays, missed contacts, etc. should be allowed to ensure a high delivery probability. Otherwise the protocol could simply save previous delivery times and adapt the TTL for future messages accordingly. Different protocols could take further advantage of the inherent properties of route- and schedule-based networks, so it should be possible to decrease the overhead more without losing performance. One such proposal, albeit with increased delays, can be found in [3]. Depending on the application, other message sizes could be considered. The congestion should not vary much, as long as the message size is small enough to be transmitted when passing another tram. So with packets small enough, epidemic routing with a TTL

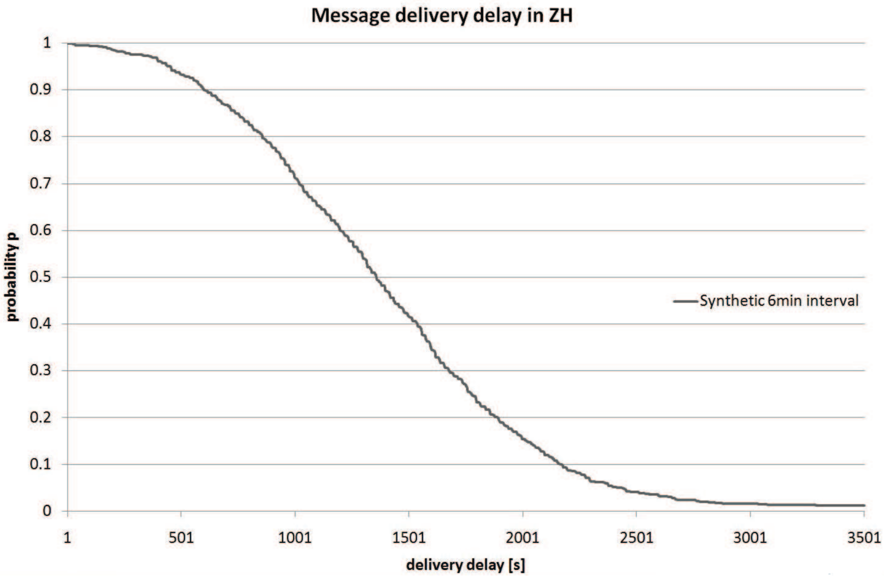

Figure 5.2: Distribution of delivery delay for individual messages

oracle can handle more than 2000 Mbyte per hour with a delivery probability of about 95 percent, as is seen in Figure 5.1.

Chapter 6

Conclusion

This paper has two main conclusions. First we show that any-inter-contact times in RSBNs are approximately exponential. This model holds for two different topologies with real schedules as well as a wide range of different densities on synthetic schedules. We believe this finding enhances our understanding of route-based networks and enables us to create analytical models for the performance of routing in such networks.

From the model for the distribution of inter-meeting times in a RSBN, we secondly create an analytical model for the expected delivery delay. For epidemic message spread, it is derived from a simple Markov-chain model for exponentially distributed inter-meeting times. The analytical results are compared to experimental values in the Zurich as well as the Amsterdam network over a wide range of densities. The model proves to be accurate within a 10 percent margin for all realistic schedules in Zurich and 22 percent in Amsterdam. This shows that the epidemic message spread in RSBNs can indeed be closely approximated with a simple model based on random mobility, although the Amsterdam example hints that the individual topology of a network has a big influence on the behaviour of epidemic routing.

Finally we did a feasibility study for routing to analyse the possibility of using a tram network as a DTN backbone. Our preliminary studies have shown, that routing should be feasible with standard protocols. The effect of enabling stops to relay messages was rather small, but could prove to be an important factor for different routing protocols. Congestion is naturally an issue for epidemic message spread, but on medium message loads even epidemic routing performs well within realistic bandwidth constraints.

Further work should go into the direction of confirming our results on more topologies and entirely different network sizes. We focussed mainly on the analysis of the network rather than taking advantage of the unique properties which RSBNs offer. The next step would be to go deeper into routing and study the performance of different routing protocols and the benefits of implementing oracles such as calculating the optimal path based on the schedule up front. Scheduled networks have the advantage, that all meetings can be predicted or, with delays taken into consideration, at least approximated closely, which should be reflected in a routing protocol.

Bibliography

[1] X. Zhang, J. Kurose, B. N. Levine, D. Towsley and H.Zhang, ”Study of a Bus-Based Disruption Tolerant Network: Mobility Modelling and Impact on Routing”. in *Proc. ACM Intl. Conf. on Mobile Computing and Networking (MOBICOM)*, September 2007.

[2] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, ”Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms”. *IEEE INFOCOM*, 2006.

[3] S. Gaito, D. Maggiorini, E. Pagani, G. P. Rossi, *Probabilistic Routing over a Bus Switched Network*. Universita degli Studi di Milano, Milano, Italy, 2009. (http://cartoon.nptlab.dico.unimi.it).

[4] T. Karagiannis, J. Y. Le Boudec, M. Vojnovic, ”Power Law and Exponential Decay of Inter Contact Times between Mobile Devices”. *ACM MobiCom 2007*, September 9 to 14, 2007, Montreal, Quebec, Canada.

[5] S. Jain, K. Fall, and R. Patra, ”Routing in a Delay Tolerant Network”. *ACM SIGCOMM*, 2004, Aug. 30 to Sept. 3, 2004, Portland, Oregon, USA.

[6] J. P. Singh, N. Bambos, B. Srinivasan, and D. Clawin ”Wireless LAN Performance Under Varied Stress Conditions in Vehicular Traffic Scenarios”. Stanford, California, USA, *IEEE INFOCOM*, 2002.

[7] D. C. Lee. ”Proof of a Modified Dijkstras Algorithm for Computing Shortest Bundle Delay in Networks with Deterministically Time-Varying Links”. *IEEE Communications Letters*, Vol. 10, No. 10, October 2006.

[8] P.Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft and C. Diot, ”Pocket Switched Networks and Human Mobility in Conference Environments”. *Proceedings of the ACM SIGCOMM 2005 Workshop on Delay-Tolerant Networking (WDTN)*, Philadelphia, Pennsylvania, USA, August 2005.

[9] D. Daley and J. Gani, *Epidemic Modelling: An Introduction*. Cambridge, United Kingdom: Cambridge University Press, 1999.