Opportunistic Communication and Human Mobility

Ólafur Helgason, Sylvia T. Kouyoumdjieva, and Gunnar Karlsson

Abstract—Many mobility models currently used for evaluating wireless communication systems have weak resemblance to reality and there is a lack of understanding on which characteristics of human mobility affect system performance. In particular, most current mobility models assume a free flow of nodes and do not consider how mobility is affected by interactions with other persons and with the physical environment. They also assume a closed system, not considering the effect of node arrival and departure. The structure of space in which the mobility occurs is either not considered at all, or only in a limited way. In this work, we address human pedestrian mobility for evaluation of wireless communication to determine which of the aforementioned aspects need to be captured and to what level of detail. We focus on opportunistic communication in the form of ad-hoc and delay-tolerant networks. For the evaluation, we use mobility models from the field of transportation and urban planning that are used for designing and dimensioning public spaces for comfort and safety of pedestrians in rush hour and emergency evacuation. The models capture micro-mobility of pedestrians better than most mobility models used in mobile networking since the application domain requires realistic representation of node interactions with the physical environment and with other nodes. Our results show that the free flow assumption used in most models does not have a significant performance impact. We also conclude that performance is not very sensitive to accurate estimation of the probability distributions of mobility parameters such as speed and arrival process. Our results, however, suggest that it is important to capture the scenario and space in which mobility occurs since these may affect performance significantly.

Index Terms—Mobility models, opportunistic networking, wireless systems

1 Introduction

It is known that mobility significantly affects the performance of wireless communication systems [1]–[3]. On one hand, mobility may disrupt ad hoc networks when connected nodes move away from each other. On the other hand, mobility may contribute to data dissemination when nodes store, carry and forward messages. Thus, the effects of mobility can lead either to improvement, or to deterioration of wireless system performance, depending on the mode of communication. This is the motivation for us to study realistic mobility.

Our focus is on common modes of mobile communication, namely where mobile devices are carried by pedestrians in built environments. There have been performance evaluations of such systems that use synthetic mobility models, where nodes move randomly in a closed area. Due to the randomness and the infinite sojourn time, nodes in such models might not represent human mobility patterns. Models have been presented to address the shortcomings. In [4] the authors develop a detailed analytical model to study the connectivity properties of pedestrian mobility along a street. The pedestrian mobility model presented in [5] incorporates obstacles to restrict node movement and wireless transmission. The authors of [6] devise a pedestrian mobility model based on social network theory, while [7] presents and analyzes a movement model based on inter-contact times and contact time distributions that follow closely those of real-life traces. Finally, in [8] the authors present SLAW, a model that is able to produce synthetic walk traces based on social context information. Despite the steps taken towards more realistic modeling of pedestrian mobility, there are still a number of issues that are left out of the modeling scope. For example, pedestrian mobility models usually do not consider node-to-node interactions. The space where mobility occurs is either not considered at all, or only in a very limited way. Cultural aspects, such as differences in personal space requirements and walking speed, are usually neglected. Moreover, most models are closed systems, and the effect of arrival processes and sojourn times in the observed area is not considered. Studies based on random mobility, without consideration of physical interaction, and without arrival and departures of nodes might give misleading conclusions on the performance of systems.

Measurements of mobility might provide a means to build realistic mobility models and the measurement traces may be used for simulation studies. The scale of most measurement campaigns has however been limited to a few tens of nodes, and are often limited to university campuses and theme parks (rather than urban settings). Most measurement studies also employ coarse temporal and spatial sampling of node positions and therefore only capture human mobility at time-scales of tens of seconds, minutes or even hours [3], [9]. Hence the measurements

• *The authors are with the KTH ACCESS Linnaeus Centre and Electrical Engineering, Stockholm 10044, Sweden.*
E-mail: {olafur.helgason, stkou, gk}@ee.kth.se.

Manuscript received 23 Nov. 2012; revised 10 Oct. 2013; accepted 28 Nov. 2013. Date of publication 15 Dec. 2013; date of current version 2 July 2014. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, *and reference the Digital Object Identifier below.*
Digital Object Identifier 10.1109/TMC.2013.160

capture only snapshots of nodal movements, and not exact paths. Moreover, the under-sampling leads to missed node contacts which in turn might affect the system performance [10].

Advanced mobility models for flows of pedestrians as well as for individual pedestrians in a flow have been devised in the area of urban planning and transportation research [11]. These models are primarily used for designing and dimensioning large-scale public spaces in order to optimize the flow of people as well as for drawing emergency and evacuation strategies. It is thus of primary importance for them to *realistically* capture the structure of the physical space, how nodes interact with it and how they interact with each other. These models therefore capture micro-mobility of pedestrians better than mobility models used in mobile networking. We believe that these models can be of great use for evaluating the performance of mobile communication and we are convinced that they can give insight into which elements of mobility affect performance.

In this paper we study how mobility affects the performance of wireless communication systems. Our approach is based on a detailed and realistic micro-level mobility model implemented in Legion Studio; a commercial pedestrian simulator package commonly used by architects and civil engineers for designing and dimensioning large public spaces. The mobility in Legion Studio is based on advanced analytical models [12] that have been calibrated and validated by measurement studies [13]. We investigate a system where nodes (pedestrians) move in structured real-life spaces such as urban outdoor areas and subway stations. We further allow for a small number of stationary nodes to take part in the communication process.

Our work seeks answers to the following questions:

- How do micro-level mobility parameters (such as speed, arrival process, and personal space requirements) affect the connectivity parameters (such as link duration, contact rate, and path duration) of a mobile wireless system?
- Is connectivity highly sensitive to even small changes in input mobility parameters?
- How does the scenario in which mobility occurs affect the connectivity parameters?
- How sensitive is the performance of access points with respect to location?
- Are some input parameters more important than others? If so, which ones?

Furthermore, we address whether mobility can be captured by relatively simple analytic or simulation models to give meaningful performance results, or whether other more advanced measures are needed. In other words, what are acceptable simplifications and abstractions when modelling mobility for the sake of evaluating wireless systems? For instance, is a rough estimate of input parameters sufficient for performance evaluation or are accurate estimates of the empirical distributions needed?

This paper is a follow-up of our previous work [10]. In this work we present full details of the analysis and extend the previous work with new and additional results, where we analyse a richer set of mobility and connectivity metrics, leading to more definitive conclusions. We also expand on the previous work by studying multi-hop connectivity and connectivity between mobile and stationary nodes.

The rest of this paper is organized as follows. In Section 2 we present a classification for human mobility and discuss background and related work in the context of this classification. In Section 4 we describe our simulation settings and performance metrics and in Section 5 we present the statistical analysis and our results. In Section 6 we conclude and discuss directions for future work.

2 Background and Related Work

Human mobility can be classified in three mobility levels: strategic, tactical and operational mobility [11]. The strategic level describes daily movement patterns of individuals, for example going to work or shopping. Based on the strategic decisions, the tactical level is focused on choosing the travel path (which can be the shortest path for getting to certain destination) depending on the environmental factors, e.g. obstacles or congestion. In the end, the operational level represents the physical process of human movement, focusing on walking speed, physical size of the nodes and interaction with others due to queuing or to avoid traffic.

Each of these levels is likely to affect the performance of wireless communication networks differently. On one side, the decisions taken on strategic and tactical level can affect the inter-meeting times between specific nodes, which is of high importance for delay-tolerant networks. On the other side, operational level decisions are likely to affect node connectivity, as well as durations of different contacts both between nodes that are in each other's communication range, as well as for nodes which communicate over one (or more) relay nodes. This determines the amount of data that can be transferred over a contact which is of great importance for forwarding contents between nodes, or for disseminating data for content distribution applications.

Recently, a line of works has focused on human mobility at the strategic and tactical levels. Some of these works analyse and characterize the mobility effects due to temporal clustering of nodes that belong to the same social network [6], [8], or diurnal regularity in human movements [7], [14]. Unlike these works, our work focuses on micro-level mobility at the operational level and how individual contacts are affected by mobility details. Modelling mobility at this level for the sake of evaluating wireless communication has received less attention but [5] and [15] are two examples. The mobility model in [5] captures movements of nodes where obstacles block or restrict both their freedom to roam and their signal propagation. It is however quite restrictive in the types of obstacles that can be modelled and it does not capture node-to-node interactions. The work in [15] is perhaps the most closely related work to our study. The authors find that mobility impacts the connectivity graph of nodes and that different mobility models impact the connectivity graph in different ways. By simulating various ad-hoc routing protocols under the different mobility models they confirm that the choice of mobility model matters and that performance ranking of protocols may vary with the mobility model used. In particular, they find that there is a strong correlation between the link duration and path duration metrics of the connectivity graph [16] and performance of ad-hoc routing protocols. Their mobility models however do not have the same operational-level details as the mobility model we use and they only consider closed systems with relatively few nodes while we study open systems and collect statistics from a large sample population.

The importance of correct capturing of pedestrian and crowd mobility has received plenty of interest in the last decade. A number of analytical models have been developed, most of them in the area of social science [17] and urban planning [18], as well as gaming [19]. Those studies have until now been separated from the field of wireless communications. A step towards coupling pedestrian mobility to the field of wireless communication has been taken in [4].

3 Mobility Model: Legion Studio

The tool we use in this work for capturing human mobility is Legion Studio [20], a commercial software package initially developed for designing and dimensioning large-scale spaces via multi-agent simulation of pedestrian behaviours. It is used mainly by architects and civil engineers. Legion Studio allows import of AutoCAD drawings of built structures such as subway stations and a grid of streets in urban areas.

For a simulated scenario, the arrival distribution and target walking speed distribution of pedestrians can be set, as well as the mobility objectives (e.g. decisions taken at crossroads) and target destinations for the pedestrians. Legion Studio then provides a multi-agent pedestrian model for simulating the pedestrian movements and collision avoidance. The multi-agent pedestrian model is based on advanced analytical and empirical models and although Legion Studio is a commercial product (such as Matlab and the Opnet simulator), a thorough description of the models is available [12] and a short summary is provided in the following sub-section.

The mobility model in Legion Studio has been validated and calibrated by measuring and monitoring pedestrian behaviour, both in controlled settings and with real pedestrian crowds in physical environments [13]. One of the main contributions of our work is thus the introduction of agent-based simulation for evaluating mobility with respect to communication network performance.

Despite an extensive search (such as in the Crawdad database and other sources) we have not been able to find any mobility trace with a fine enough granularity for a meaningful comparison with our results. The Legion Studio traces we use in this study are available at CRAWDAD (http://crawdad.org/kth/walkers/)

In a simulation each pedestrian is represented by a two-dimensional circular entity of a size that approximates an actual person. The navigation decisions of each entity are mostly based on the interaction with other nodes, as well as on the interaction with the surrounding environment, although overriding certain choices is possible. Like in real life, the movement patterns follow a *least effort* principle where each entity tries to maintain its personal space and minimize its dissatisfaction before choosing its next move. Legion Studio provides different aspects of the dissatisfaction factor: inconvenience, frustration and spatial discomfort. Moreover, Legion Studio incorporates all three levels of human mobility (strategic, tactical and operational) for each entity in the simulation, thus allowing correct capturing of events like queuing behind slower nodes at bottlenecks. To resemble reality even more, it allows the use of open systems, where entities enter and leave the system according to a predefined pattern. The ability to record both the spatial and temporal positions of each node during its lifetime in the system provides extensive information that can be used for evaluating mobile communication systems based on thorough examination of user behaviours.

4 Simulation Setup

Each simulation run conducted in Legion Studio results in a mobility trace file, where the position of each node is captured every 0.6 s. For obtaining the connectivity metrics from the mobility traces we feed the traces into a system simulator [21] and study both ad-hoc connectivity directly between mobile nodes and connectivity with stationary nodes (such as access points).

We assume that the mobile nodes can communicate over a short-range radio (e.g. 802.11 or Bluetooth). The radio model we employ is simple: if two nodes are within a fixed range r of each other they can communicate. Unless otherwise noted we assume that $r = 10$ m. The motivation for using this relatively short default range is that a longer range gives problems when the range extends through the physical infrastructure and hence the disk model is too simple to represent real propagation (we remark that ray tracing is infeasible since we only have 2D renderings of spaces and it is computationally expensive). Another motivation is the attenuation of human bodies when the density of people increases. Our experience from experimenting with WiFi on smartphones [22] is that the practical outdoor WiFi range is typically at most 50-60 m with no obstacles and line of sight communication. In indoor and in constrained environments the range is shorter.

Physical and link-layer issues such as interference, shadowing, fading or MAC-layer contention are not considered. We realize that these factors affect the performance of wireless systems and that the interplay between them and mobility can be complex. Moreover, the effect of these factors can depend strongly on the particular radio being used. Here we however aim at examining the properties of an ideal mobile wireless system, isolating the effect of mobility.

4.1 Mobility Scenarios

Our evaluation mainly considers the following two scenarios: an outdoor urban scenario, modeling the Östermalm area of central Stockholm, and an indoor scenario, recreating a two-level subway station (Fig. 1).

The urban outdoor scenario consists of a grid of interconnected streets with lengths varying between 20 m and 200 m. Each street has a width of 2 m which is representative of a sidewalk. The observed area is connected to the outside by 12 passages, and we assume that nodes enter the area through each of those passages with equal arrival rates

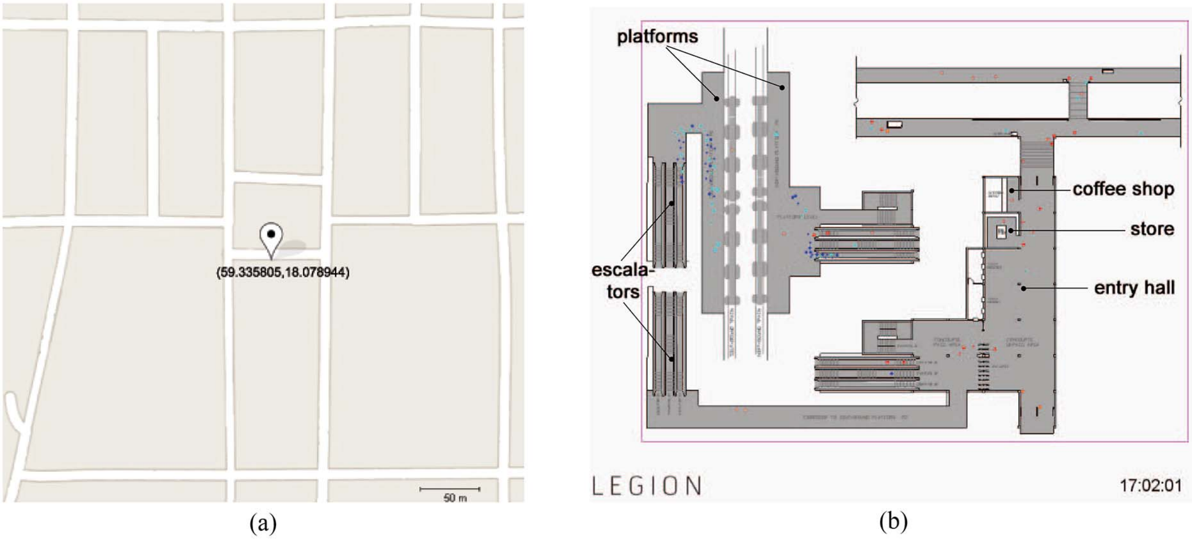


Fig. 1. Simulation scenarios. (a) Ostermalm area in down-town Stockholm. (b) Two-level subway station.

denoted by λ. Upon arrival at an intersection, nodes continue moving along the same street with probability 0.5 or change their direction and turn in an adjoining street with equal probabilities. Nodes roam the streets in this manner until an exit passage is selected and the node leaves the area. The active area of the outdoor scenario is 5872 m^2. The scenario can be characterized as a high mobility scenario, since nodes move constantly throughout their lifetime in the observed areas. We note that we have experimented with different street selection probabilities and replaced the center streets with a wide square and found that this does not significantly affect our results and does not alter our conclusion.

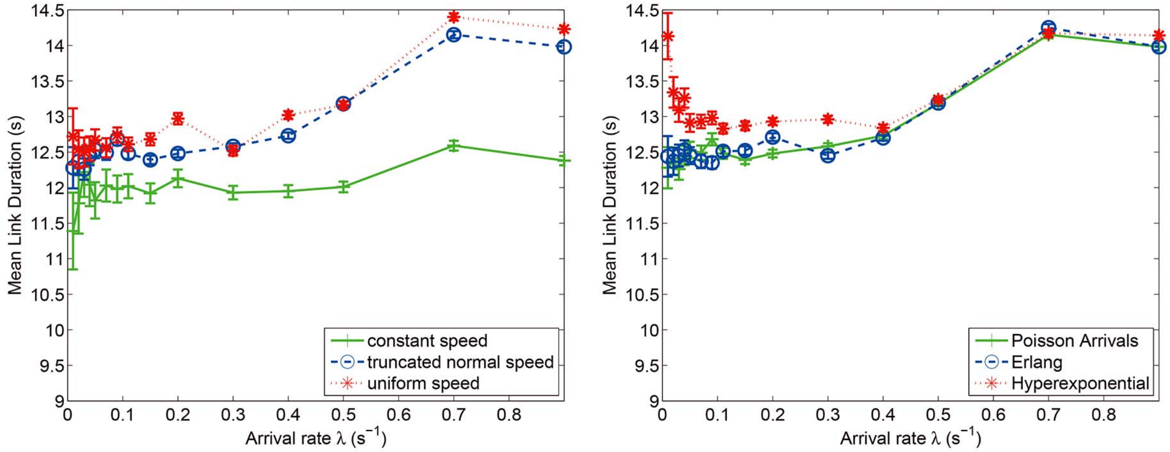
The indoor scenario defines a train platform connected via escalators to the upper entry-level. Nodes can arrive on foot from any of five entry points of the subway station, or when trains arrive at the platforms. The train arrivals contribute to the burstiness of the node arrivals and departures. Nodes congregate while waiting for a train to arrive at one of the platforms, or while taking a break in the store or the coffee shop at the entry level. Since the station is relatively densely populated, and its structure severely constrains mobility, the physical interaction of nodes is high and leads to queuing. The main bottleneck where queuing is observed are the escalators. Nodes can leave the area either by boarding a train or by walking out through one of the exits on the upper level. This subway station model is provided with the Legion Studio simulator and the arrivals and departures are determined by the model which we cannot modify. The active area of the scenario is 1921 m^2.

4.2 Performance Metrics

The main focus of our evaluation is to explore how mobility affects the wireless *connectivity graph*. Various studies have shown that for mobile wireless systems there is a strong correlation between metrics defined on the connectivity graph and the protocol or system level metrics [2], [3], [15]. The connectivity graph at time t is an undirected graph $G(t) = (V(t), E(t))$ where $V(t)$ is the set of mobile nodes, and the edge set $E(t)$ consists of the wireless radio links between nodes. Thus we have that link $(n_i, n_j) \in E(t)$ if $D(n_i, n_j) \leq r$ at time t where $D(n_i, n_j)$ is the euclidean distance between nodes n_i and n_j. Our evaluation studies the following metrics.

- *Link duration:* This is the time when two nodes are physically within direct communication range. More formally, if there exists an $\epsilon > 0$ such that $(n_i, n_j) \in E(\tau)$ for all $\tau \in [t, t+d]$ and $(n_i, n_j) \notin E(t-\epsilon)$ and $(n_i, n_j) \notin E(t+d+\epsilon)$, then d is defined as the link duration.
- *Contact rate:* We define the contact rate as the number of non-zero link durations per node, divided by the lifetime of the node in the simulation. Since we study open systems, the contact rate is a more suitable metric than the number of contacts, since the rate is (by definition) normalized by the simulation sojourn time of a node.
- *Inter-contact time:* We define the inter-contact time as the elapsed time from the beginning of one contact to the beginning of the next. We note that in our model, nodes can have multiple simultaneous contacts.
- *Path duration:* A path $P = \{n_1, n_2, \ldots, n_k\}$ exists between nodes n_1 and n_k at time t if $\{(n_1, n_2), (n_2, n_3), \ldots, (n_{k-1}, n_k)\} \subset E(t)$. The path duration is the time during which two remote nodes are physically connected via the same set of relay nodes. It is defined in a similar manner as the link duration. If there exists an $\epsilon > 0$ such that $\{(n_1, n_2), (n_2, n_3), \ldots, (n_{k-1}, n_k)\} \subset E(\tau)$ for all $\tau \in [t, t+d]$ and $\{(n_1, n_2), (n_2, n_3), \ldots, (n_{k-1}, n_k)\} \not\subset E(t-\epsilon)$ and $\{(n_1, n_2), (n_2, n_3), \ldots, (n_{k-1}, n_k)\} \not\subset E(t+d+\epsilon)$ then we define d as the duration of path P.

The above metrics are important for the performance of most mobile wireless systems. Our definition of link duration includes the node discovery and contact setup time, which can differ from one technology to another. The amount of data that can be transferred over a contact



(a) (b)

Fig. 2. Impact of speed distribution (a) and arrival process (b) on mean link duration for the Östermalm scenario.

thus depends on the channel bit-rate and the remaining link duration after connection setup. The contact rate is a measure of the contact opportunities available to nodes and the contact rate between any pair of nodes has been shown to be a metric that highly affects the performance of routing protocols in delay tolerant networks [3]. For multi-hop networks the path duration is an important metric that affects the feasibility of ad-hoc routing protocols [2]. The path duration is a measure of the time that two nodes can communicate over a particular path.

4.3 Measuring Performance Metrics

In our evaluation we study both mean values and full distributions of the observed performance metrics where the distributions are estimated using statistical analysis of simulation data. The time resolution of the mobility traces exported from Legion Studio is 0.6 s and therefore we update the connectivity graph with the same resolution and update the measured metrics. The beginning/ending of a contact is assumed to be uniformly distributed in the sampling interval before/after the contact is first/last observed in our simulations. During a simulation run we compute the connectivity graph in every round and update the measured metrics. In all simulations we start collecting data after a warm-up period, i.e. the system starts empty and after the warm-up period we collect statistics from 1000 nodes. The length of the warm-up period depends on the scenario and arrival rate; to determine its length we follow Welch's procedure outlined in [23].

We estimate the path duration metric with the duration of the shortest path (i.e. the path with minimum number of links) since calculating all possible paths is not feasible. Finding the shortest path is usually the goal of ad-hoc routing protocols and therefore it serves as a good measure of performance. When a path is found between two previously unconnected nodes, the path is stored. In the next round the duration of those paths that still exist is incremented. If a new shorter path is found we keep track of this new path as well as all previous paths between the node pair that still exist.

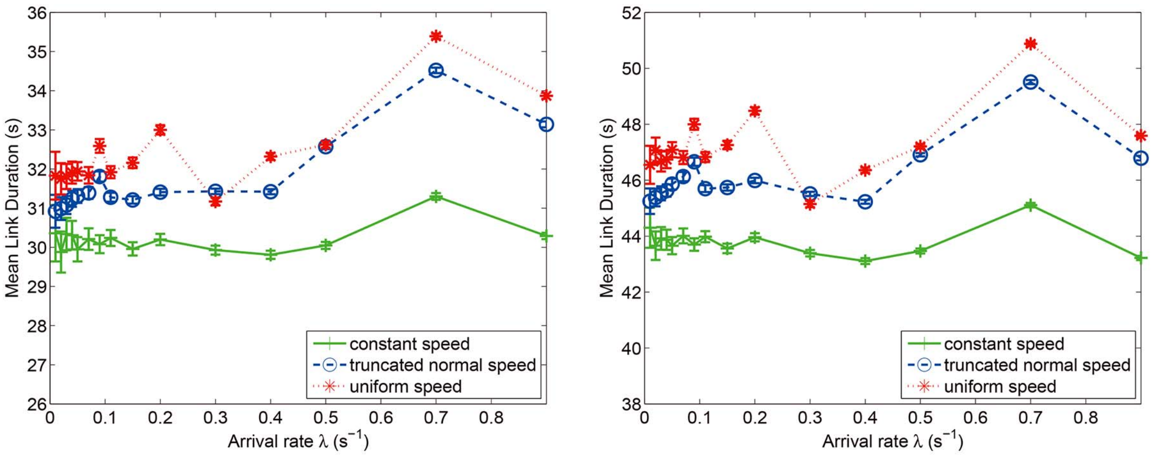
5 Results

5.1 Effect of Speed and Arrival Process

We use the Östermalm scenario to examine how the input speed distribution and arrival process affect the connectivity metrics and their distributions. Legion Studio allows us to configure the *target speed* of nodes as an input parameter. Due to queuing behind slow nodes or speeding up when overtaking others, nodes tend to change their actual speed during the simulation runs. Legion Studio allows one to also configure different (pre-defined) cultural profiles that affect node parameters such as level of dissatisfaction, physical size and desired inter-node distance. We have examined those parameters, and we found that the connectivity metrics are insensitive towards them. Therefore we will not discuss them further.

Fig. 2 shows the effect of target speed distribution and arrival process on the mean link duration. All mean values in this paper (Fig. 2, Fig. 3 and Fig. 5) are plotted showing 95% confidence intervals. Note that the confidence intervals might be invisible due to their narrow width. In Fig. 2(a) we study three different speed distributions, all with the same mean speed of 1.3 m/s; (*i*) constant speed, (*ii*) uniform with minimum speed 0.6 m/s and maximum of 2.0 m/s and (*iii*) truncated normal distribution with minimum speed 0.6 m/s and maximum of 2.0 m/s. The reason for choosing 1.3 m/s as a mean speed is the study in [24]. Moreover, we conducted experiments with both higher and lower mean speed values, and observed only a shift by a constant factor in the connectivity curves, but no other performance change. In order to facilitate the comparison, we assume a Poisson arrival process for all cases. However, due to different factors such as cultural profile or age, people do not move at constant speed. That is the reason to believe that the uniform and the truncated normal distributions resemble reality to a higher extent. It can be seen that the uniform distribution suggests slightly longer link durations because of the higher amount of slower nodes in the scenario.

In Fig. 2(b) we show the effect of different arrival processes on the mean link duration. We have chosen the processes such that they have the same mean λ: (*i*) a Poisson



(a) (b)

Fig. 3. Impact of speed distribution on mean link duration for a communication range of 30 m (a) and 50 m (b) for the Östermalm scenario.

process with rate λ, (*ii*) a 4-stage Erlang where each stage has rate 4λ, and (*iii*) a two-phase hyper-exponential inter-arrival time distribution with arrival rates 0.35λ and 5.7λ in the first, respectively second phase, and selection probabilities 0.31 and 0.69. The variation coefficient of the Erlang and the hyper-exponential distribution is 1/2 and 2 respectively. The speed distribution is truncated normal (0.6, 2.0) with mean 1.3 m/s. At lower rates the hyper-exponential arrival process results in longer link durations; we believe the reason for such behaviour is the bursty nature. Due to the burstiness, nodes entering the scenario have higher chances to be connected upon arrival and stay connected as they move along in a group. Such burstiness is often observed in urban life where nodal clustering is usually caused by traffic lights, train arrivals and departures, etc. The clustering of nodes in the scenario leads to queuing, forcing faster nodes to slow down behind those with leisurely pace, thus increasing the link durations with the surrounding nodes.

Fig. 3 shows that increasing the transmission range only changes the scale of the link durations but otherwise the conclusions are the same.

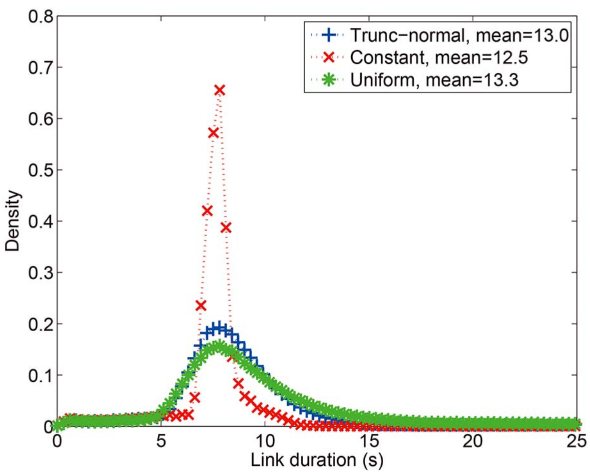


Fig. 4. Link duration histograms for the Östermalm scenario with different speed distributions.

In addition to studying the mean values we have also looked at how the distribution of the link duration is affected by the speed and arrivals. Fig. 4 shows the histograms of link durations for the three speed distributions under a Poisson arrival process with a per-entry arrival rate of $\lambda = 0.15\ s^{-1}$. To facilitate a comparison between the histograms they all have equal fixed-width bins and each data point in the plot is the centre of a bin. All the histograms are characterized by a sharp peak in the link durations at approximately 8 s. This is due to contacts with nodes moving in opposite directions. The mean duration of this type of contacts is given by $d = 2r/|v_a - v_b|$ where r is the communication range and $|v_a - v_b|$ is the relative speed of the nodes. For $r = 10$ m and a mean speed of 1.3 m/s the average duration of these contacts is 7.7 s. The figure illustrates how speed affects both the height and the width of the spike. The basic shape of the empirical contact distribution however remains the same in all cases and this shape is determined by the mobility scenario. The histograms for the Erlang and hyper-exponential arrivals have high resemblance to those shown and are therefore omitted from the plot.

Fig. 5 shows the impact of speed and arrival processes on the mean contact rate. The contact rate increases linearly at low arrival rates but at higher rate the node density increases, the linearity is violated, and the contact rate grows slower. This is due to node interactions. We also see that the contact rate is insensitive to both arrival process and speed distribution of nodes. We have also found that the same applies to the empirical distribution of the inter-contact time, i.e. it is highly insensitive to both the target speed distribution and node arrival process. Here we examine only a range of arrival rates ($\lambda \in [0.01s^{-1}, 0.90s^{-1}]$). For arrival rates higher than 0.90 s^{-1}, the physical limitations of the space restrict node movements to a point where all nodes become static. For arrival rates lower than 0.01 s^{-1} the contact parameters tend towards 0. We refer the reader to the analytical model in [4] for more details on modeling the lower extreme. We note that these numbers are valid for the particular scenarios examined in the paper and may differ when the mobility occurs in a different physical environment.

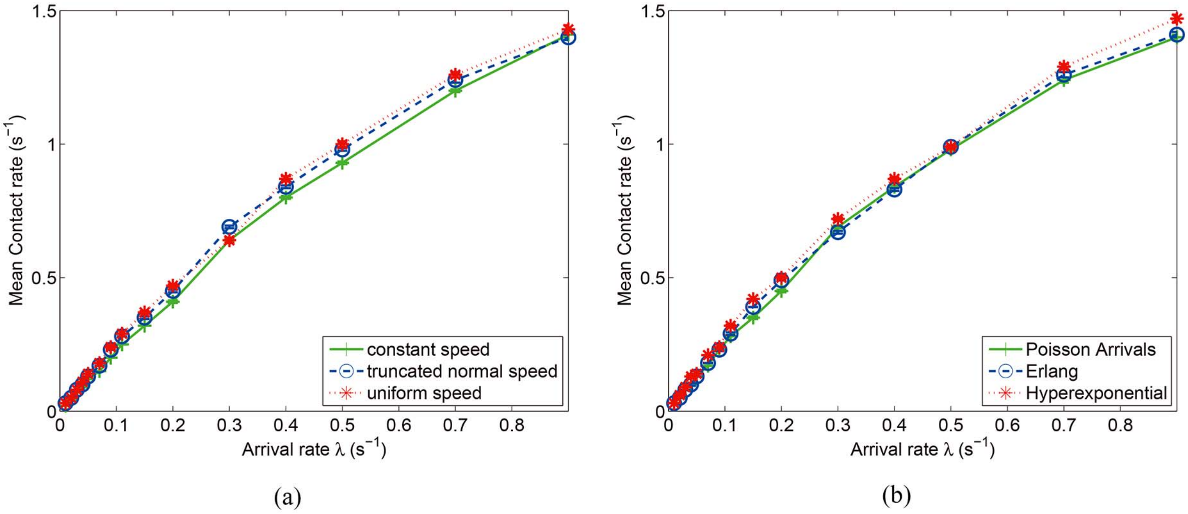


Fig. 5. Impact of speed distribution (a) and arrival process (b) on mean contact rate for the Östermalm scenario.

Fig. 6 presents the sojourn time density for the Östermalm scenario under different speed distributions and different arrival processes. For both experiments we choose to compare scenarios with mean arrival rate $\lambda = 0.15\ s^{-1}$. We see that changes in the speed distribution account for variations in the sojourn time distribution (Fig. 6(b)) while changes in the arrival process (Fig. 6(a)) do not affect the sojourn time distribution. It is interesting to note that the constant speed distribution creates largest deviations whilst providing lowest mean sojourn time durations. A closer investigation of the graphs in Fig. 6(b) however shows that the sojourn time density has the same behaviour over all speed distributions. The reason why this behaviour is more distinct in the constant speed distribution case lies in the fact that since pedestrians move at constant speed, traversing the same path by different users results in the same sojourn time. In the other two cases (a uniform and a truncated normal distribution) the curves are smoothed because of the variations in the sojourn times introduced by users moving at different speeds while traversing the same route.

Fig. 7 presents a comparison between the sojourn time densities for the Östermalm and the Subway scenario. We can see that the sojourn times in the Östermalm scenario are much longer (given the larger space in which nodes move), thus the tail of the distribution is thicker. The Subway scenario is characterized by shorter sojourn times and a thinner tail. Fig. 7 also shows the fitting of a Gamma distribution to capture the characteristics of sojourn time distributions for both scenarios. We see that the Gamma distribution underestimates the maximum in both distributions however it does describe the tails of the distributions well.

5.2 Effect of Scenario

As previously mentioned, the Östermalm scenario is characterized by high mobility. In the Subway scenario, however, nodes can take a break in the coffee shop, or cluster while waiting at the platforms for a train arrival. In Fig. 8(a) we investigate how the link duration histogram is affected by the differences in the scenarios. In order to make the comparison easier we select a configuration of the

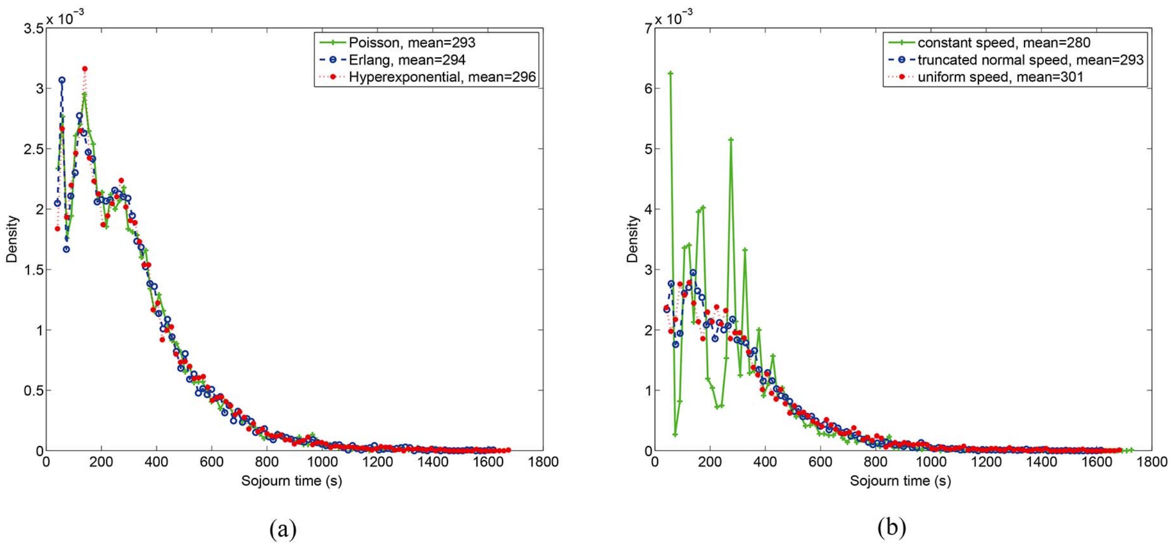
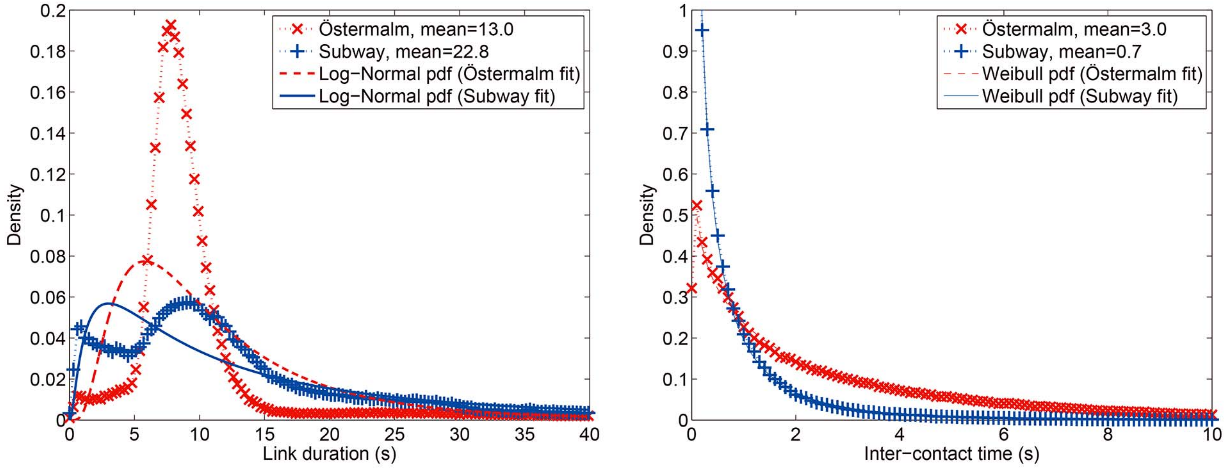


Fig. 6. Density of sojourn times for the Östermalm scenario under (a) different arrival processes and truncated normal speed distribution, and (b) different speed distributions and a Poisson arrival process with $\lambda = 0.15\ s^{-1}$.



(a) (bm)

Fig. 8. Link duration (a) and inter-contact time (b) comparison between the Östermalm and Subway scenarios.

Östermalm scenario that gives comparable node density (ρ) to the Subway scenario. If we define $\bar{N}$ to be the mean number of nodes in the scenario, and A to be the effective size of the area (i.e. the area in which nodes can move), we can calculate $\rho = \bar{N}/A$. The mean number of nodes $\bar{N}$ can be obtained from Little's law as $\bar{N} = \lambda_{tot}\bar{T}$ where λ_{tot} is the mean total arrival rate, and $\bar{T}$ is the mean sojourn time of nodes in the scenario. We measured the sojourn time from our simulations, and use this value for calculating the node density. Thus, we have found that the mean density for the Subway scenario is 0.09 $nodes/m^2$ and that a per-entry arrival rate of $\lambda = 0.15\ s^{-1}$ gives approximately the same node density for the Östermalm scenario.

When comparing the link duration histograms for the two scenarios we see a significant difference. The histogram for the Subway scenario is more complex with at least two visible peaks, one small peak at very short contacts and another one at about 8 s. It also has a thicker tail due to clustering and waiting, and this is reflected in a significantly longer mean link duration: 22.8 s for the Subway scenario and 13.0 s for the Östermalm scenario.

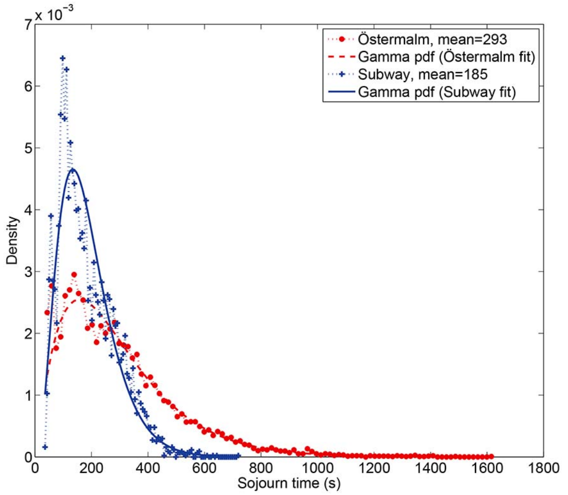


Fig. 7. Comparison between the sojourn time densities for the Östermalm and the Subway scenarios.

We use the Kolmogorov-Smirnov (K-S) statistical test to compare the empirical link duration distribution with standard probability distributions: exponential, gamma, Weibull, log-normal, Rayleigh and Pareto. The parametrization of the reference distributions is based on maximum likelihood estimation and the results are shown in Table 1. None of the reference probability distributions give a particularly good fit for the link duration but the log-normal distribution gives the lowest K-S statistics (which indicates a better fit) for both scenarios. Fig. 8(a) however shows that the log-normal fit underestimates the spike at 8 s in the Östermalm scenario and that the log-normal fit does not capture the multi-modality of the link duration in the Subway scenario. This suggests that the link duration cannot be accurately modelled by a single standard distribution but should rather be approximated by a mixture of distributions.

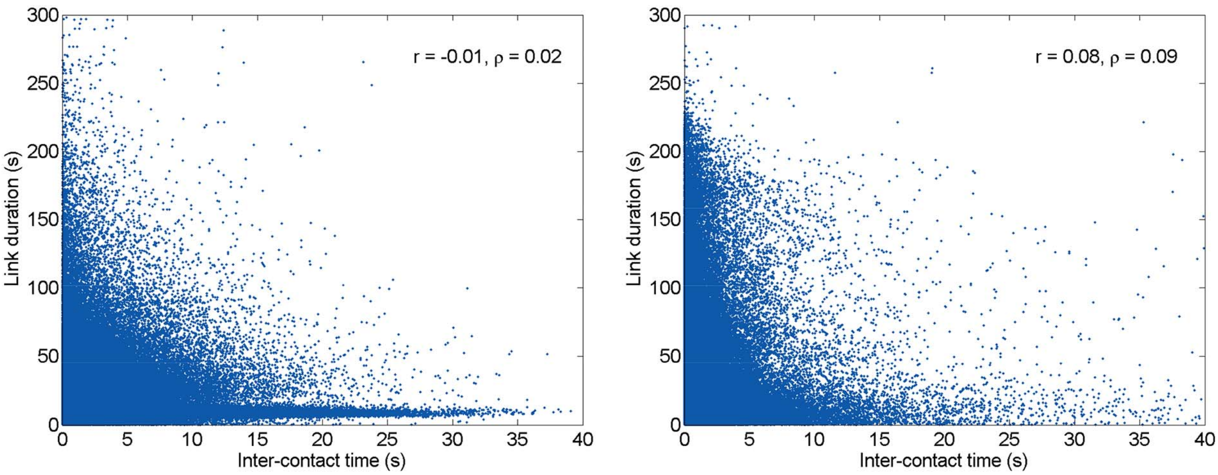
We have also studied the distribution of the inter-contact times in the two scenarios and found that they can be well approximated with a Weibull distribution (see Fig. 8(b) and Table 1). Recent experiments using Bluetooth devices to log contact statistics of humans have suggested that both link durations and inter-contact times (often referred to as inter-any-contact times in those works) are approximated by power-law distributions [25]. This contradicts our results and in [10] we show that the coarse neighbour discovery process of Bluetooth filters out most of the short contact opportunities and much of the effects of operational level mobility that we consider in this work and thus skews the distribution. Due to the long neighbour discovery time, Bluetooth is not well suited for opportunistic mobile systems.

Identifying the underlying empirical distributions for the inter-contact time and link duration is important for understanding whether complex mobility models like Legion Studio are necessary to evaluate communication systems or whether some simplifications can be made. If, for a given scenario, the sequence of link durations and inter-contact times of a node are uncorrelated, they could be modelled as a renewal process where the time between contacts and the duration of each contact are sampled independently from known probability distributions. Fig. 9 shows the scatter

TABLE 1
K-S Statistics for Distributions Fitted with Simulation Results

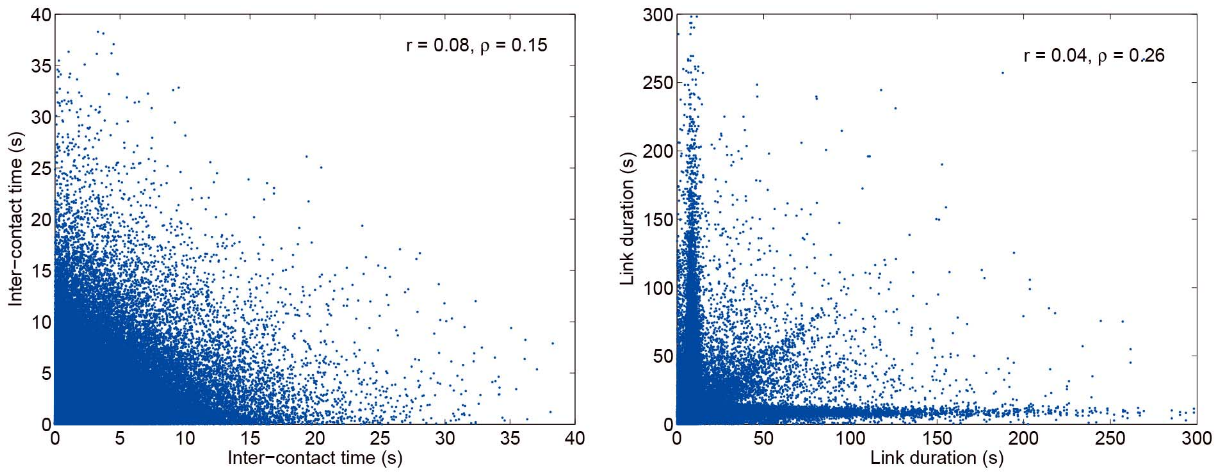
	range (m)	Exponential	Log-Normal	Gamma	Weibull	Pareto	Rayleigh
Östermalm Link Duration	10	0.27	0.20	0.27	0.25	0.30	0.57
Östermalm Inter-Contact Time	10	0.08	0.06	0.02	0.01	0.04	0.37
Subway Link Duration	10	0.15	0.07	0.13	0.11	0.09	0.49
Subway Inter-Contact Time	10	0.21	0.07	0.04	0.01	0.08	0.56
Östermalm Path Duration, h=2	10	0.23	0.15	0.14	0.18	0.24	0.42
Östermalm Path Duration, h=3	10	0.24	0.18	0.15	0.18	0.24	0.38
Subway Path Duration, h=2	10	0.20	0.11	0.16	0.19	0.17	0.66
Subway Path Duration, h=3	10	0.27	0.15	0.20	0.23	0.22	0.70
Östermalm Path Duration, h=2	30	0.10	0.08	0.08	0.09	0.11	0.36
Östermalm Path Duration, h=3	30	0.12	0.11	0.11	0.13	0.15	0.43
Subway Path Duration, h=2	30	0.10	0.09	0.09	0.10	0.12	0.38
Subway Path Duration, h=3	30	0.12	0.11	0.10	0.12	0.14	0.39

Lower value indicates a better fit.



(a) (b)

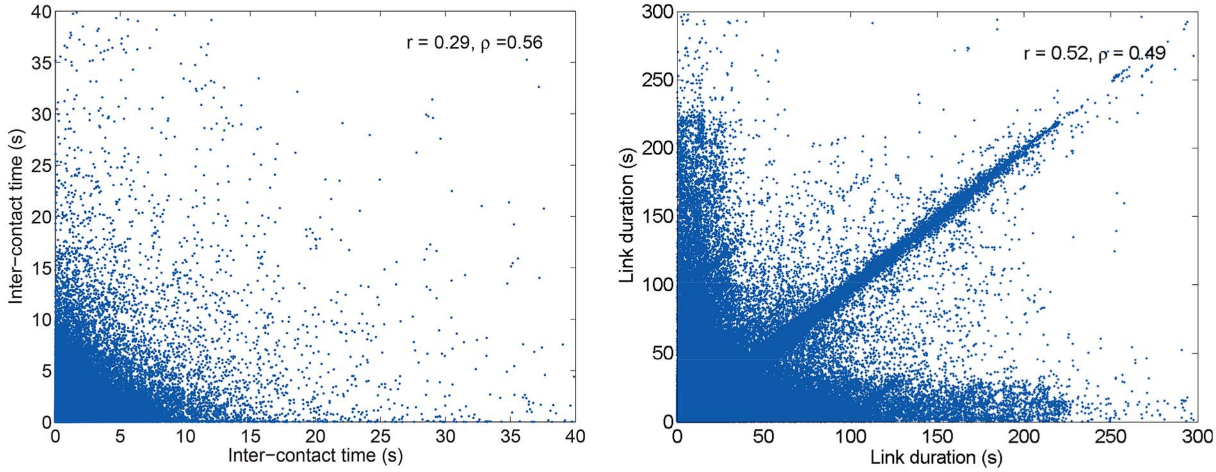
Fig. 9. Scatterplots of (inter-contact time, link duration) pairs for the Östermalm (a) and Subway (b) scenarios. r and ρ denote the Pearson and Spearman rank correlation coefficients respectively.



(a) (b)

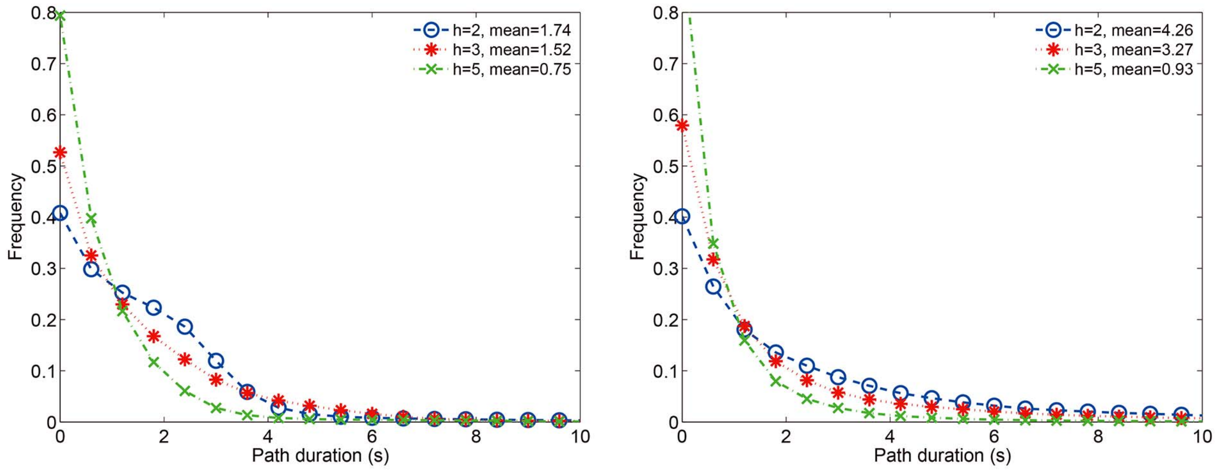
Fig. 10. Correlation of two subsequent inter-contact times (a) and link-durations (b) for the Östermalm scenario.

plots of inter-contact times versus link durations for the Östermalm (Fig. 9(a)) and Subway (Fig. 9(b)) scenarios. Figs. 10 and 11 further show the correlation between two subsequent inter-contact times and link-durations for both scenarios. We also show the Pearson correlation coefficient (r) and the Spearman rank correlation coefficient (ρ) for the data sets. The Pearson correlation coefficient is a measure of the linear dependence between two random variables and Spearman's coefficient is a measure of whether an increase or decrease in one of the random variables results in an increase or decrease in the other. Although there is low correlation between inter-contact time and link duration for



(a) (b)

Fig. 11. Correlation of two subsequent inter-contact times (a) and link-durations (b) for the Subway scenario.



(a) (b)

Fig. 12. Path duration histograms for the Östermalm (a) and Subway (b) scenarios for a 10 m communication range.

both scenarios, visual inspection shows a clear correlation between two successive link-durations for both scenarios, as indicated by the clustering of data points around the vertical and horizontal streaks in the figures (although this correlation is non-linear as indicated by a relatively low Pearson coefficient). Therefore the connectivity process cannot be accurately modelled as a renewal process.

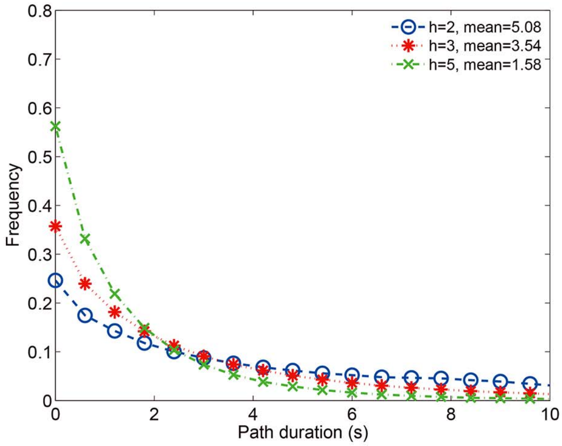
It has been previously shown that there is a strong correlation between path duration at the connectivity graph level and at the routing protocol level [15]. Moreover, the path duration at the connectivity graph level is a good approximation of the duration of a path as seen by an ad-hoc routing protocol.

Figs. 12 and 13 show the path duration histograms for the Östermalm and Subway scenarios for paths with 2, 3 and 5 links and for communication range of 10 m and 30 m. We see that there is a significant difference in the path duration between the scenarios, particularly with a 10 m range. For the Östermalm scenario, the mean duration of a two-hop path is only 1.7 s while it is 4.3 s for the Subway scenario. A longer communication range leads to longer path durations and for a 30 m range the mean duration of a two-hop path is 5.1 s and 4.7 s for the Östermalm and Subway scenarios respectively. In general, paths are short-lived and therefore it is unlikely that ad-hoc routing will perform well, particularly not in the high mobility Östermalm scenario.

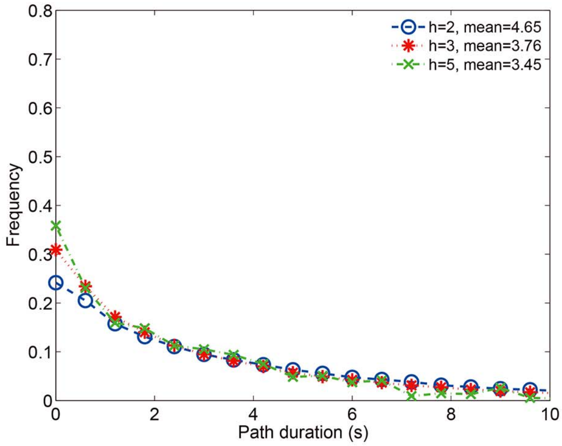
It has been proposed that the path duration distribution can be approximated with an exponential distribution [15], [16]. Table 1, however, shows that a log-normal distribution generally gives the best fit among the reference distributions we use. A gamma distribution also gives a good fit in the scenarios we study except for the Subway scenario with a 10 m range where the log-normal distribution gives a clearly better fit. An exponential distribution only gives a comparable goodness of fit for a 30 m communication range when compared to the log-normal and gamma distributions.

5.3 Effect of Node Interactions

Two of the major strengths of the Legion Studio mobility simulator are the accurate modelling of the physical environment and of node dynamics and interactions. The previous subsection studied how mobility is affected by the



(a)



(b)

Fig. 13. Path duration histograms for the Östermalm (a) and Subway (b) scenarios for a 30 m communication range.

scenario and in this section we study the effect of node interactions.

When evaluating mobile networks of human carried devices, it is commonly assumed that nodes flow freely and that their physical movements are not restricted by other nodes. Moreover, it is also often assumed that all the nodes in the area under inspection are participating in the mobile network. In reality we know that this is not the case. In dense urban areas, people are constrained by the movements of other pedestrians regardless of whether they are part of the mobile network or not. To study the effects of node interactions we consider a simple scenario that allows us to focus only on this aspect of mobility. The scenario we study is a 100 m long and 2 m wide sidewalk with pedestrians arriving at both endpoints. Nodes arrive according to a Poisson process with rate λ and they traverse the whole length of the sidewalk with a target speed, drawn from a uniform distribution over the interval (0.6, 2.0) m/s. For evaluating the effect of node interactions we examine the following configurations:

- No node interaction. Each node maintains its target speed throughout the sidewalk and is not affected by other nodes.
- Small-scale node interaction. We model the street segment in Legion Studio and the target speed of nodes is therefore affected by others. In particular, nodes may slow down due to congestion or due to a slower node in front of them.
- Node interaction with background traffic. We define two types of mobile users in Legion Studio: active nodes that carry mobile devices and can participate in the mobile network, and background nodes who are moving along the sidewalk but do not participate in the mobile network. We denote the arrival rate of background nodes by λ_{bg}.

In our evaluation we set the arrival rate of active nodes at each street endpoint to $\lambda = 0.15\ s^{-1}$ and the Legion Studio simulations are conducted with a background arrival rate λ_{bg} of 0, 0.1 and 0.2 s^{-1}. Our results suggest that the inter-contact time distribution is not sensitive to the node interactions but the link duration histograms show a mild sensitivity to node interactions as seen in Fig. 14. When the node density increases, nodes

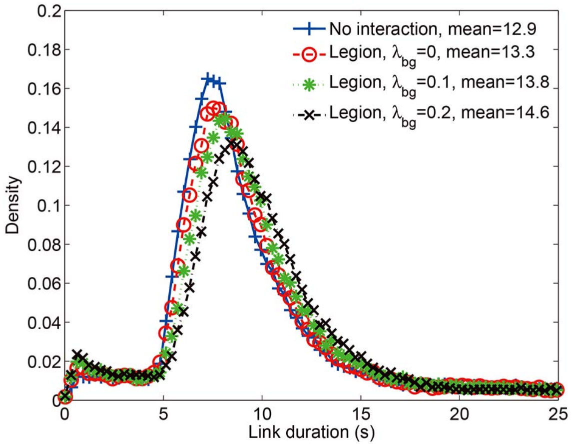


Fig. 14. Link duration histograms for nodes moving in a street with different node interactions.

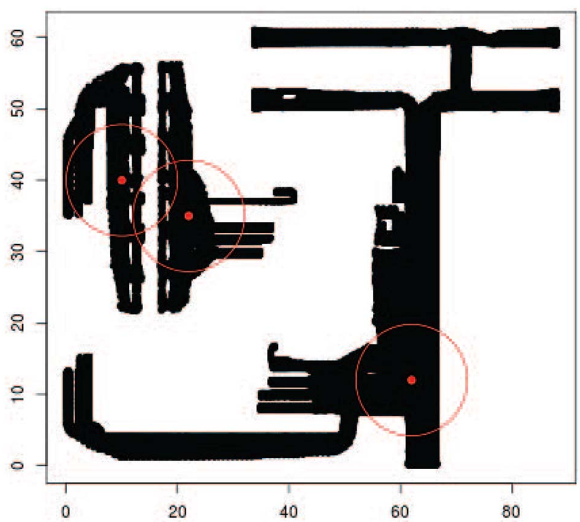


Fig. 15. Access Points and a trace of all node positions in the Subway scenario (x and y positions are given in meters).

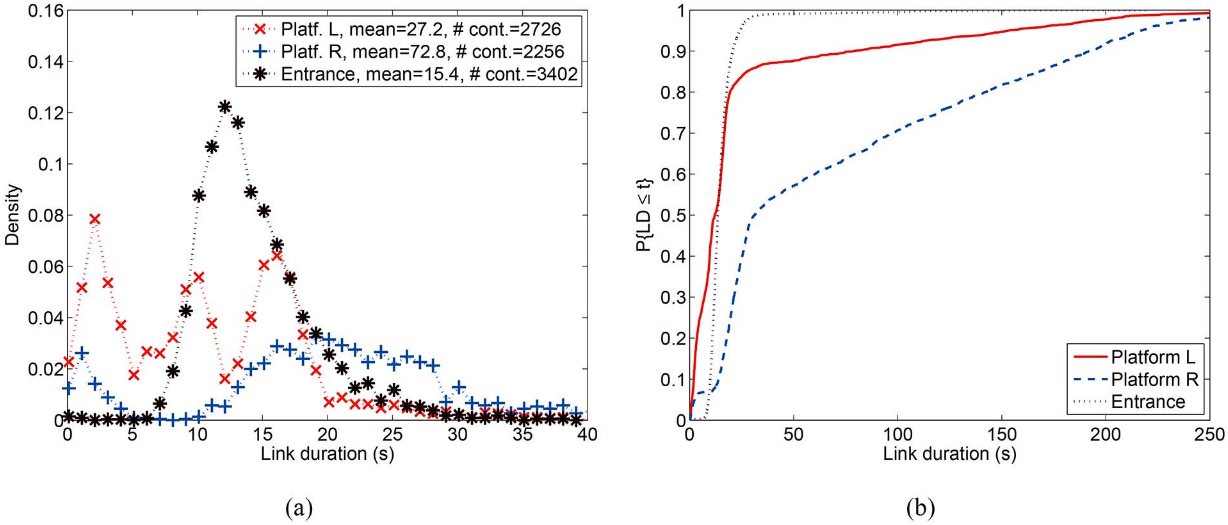


Fig. 16. Empirical distribution of link duration for the different APs in the Subway scenario. PDF (left) and CDF (right).

are forced to slow down due to the interactions with others, which again leads to a slight increase in link duration. For most wireless systems longer link durations are correlated with better performance. Assuming a free-flow of nodes would therefore lead to a pessimistic evaluation of performance.

5.4 Stationary Nodes

So far we have studied the effect of mobility on connectivity in an ad-hoc scenario where all nodes are mobile. Studying connectivity between mobile and stationary nodes is however also of interest since content can be seeded into an opportunistic network by stationary nodes such as 802.11 APs or DTN throwboxes [26]. In addition to studying sensitivity with respect to mobility metrics and scenario, we are also interested in assessing the importance of access point location.

Fig. 15 shows a trace of all node positions in the subway scenario and the position and range of three APs that we have placed in our Legion Studio simulations. The AP positions were selected to be representative of different movement patterns within the subway station. The two APs on the platform capture bulk arrivals and departures of train passengers as well as people waiting for a train on the platforms. The AP at the entrance captures almost all nodes that go through the station but we expect nodes to quickly pass through its range except when the station is crowded and passengers queue at the tolls.

Fig. 16 shows the empirical link duration distribution (histogram and CDF) for the different access points along with the mean link duration and number of contacts per AP. We clearly see that the link duration distributions are quite different for the different APs and their positions. Waiting for the train on the platforms leads to longer link durations. While the entrance AP sees many more contacts than the other APs, almost all of these are relatively short, falling within the sharp spike between 6 s and 30 s, and there is virtually no tail of longer durations in the distribution. Fig. 17 shows that the link duration for the entrance AP is not very sensitive to the walking speed distribution of the

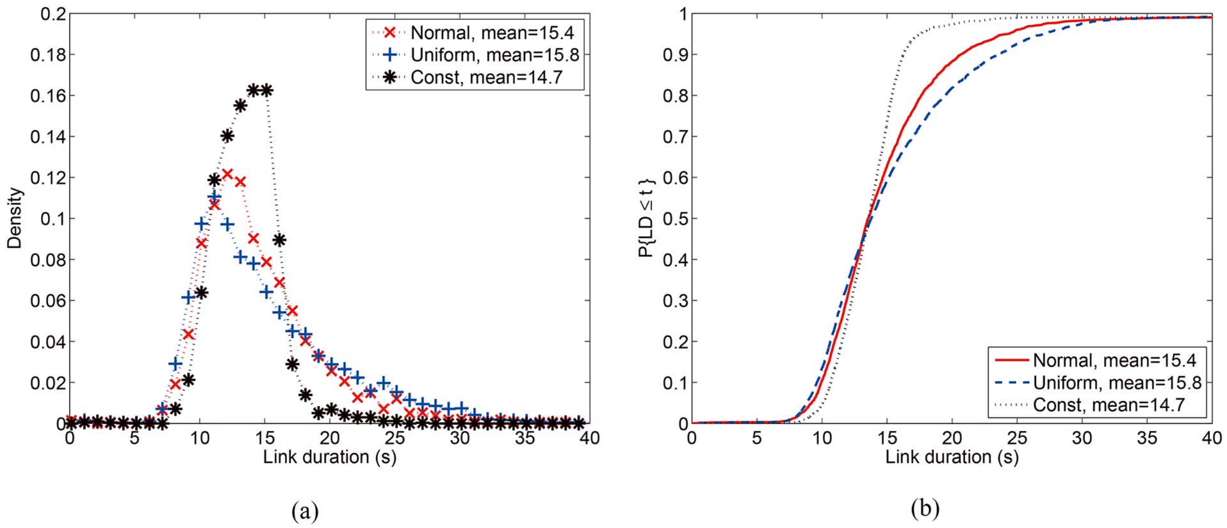
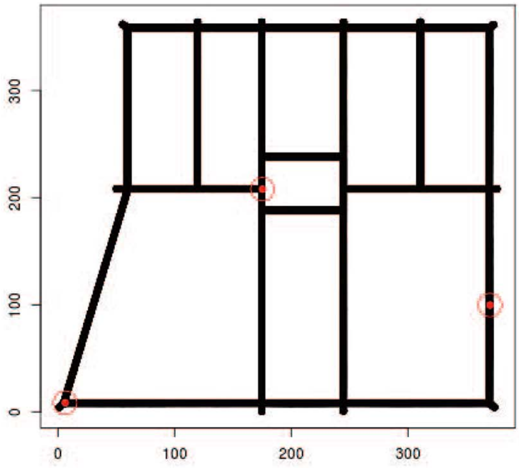
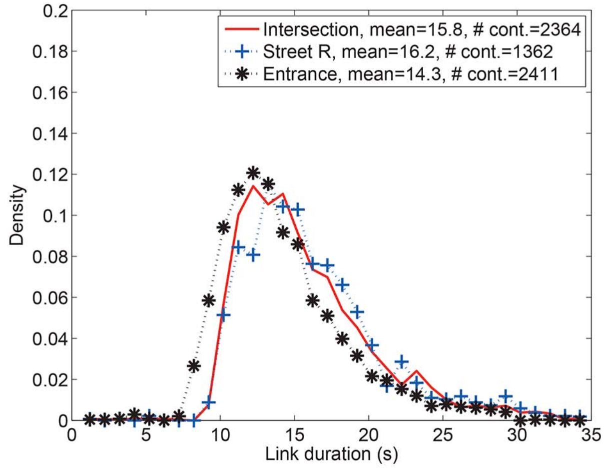


Fig. 17. The effect of walking speed distribution (normal, uniform and constant) on link duration distribution for the entrance AP in the Subway scenario.



(a)



(b)

Fig. 18. AP positions in the Östermalm scenario (left) and link duration histogram for the different APs (right). x and y positions are given in meters.

mobile nodes and it confirms our previous observation that increased walking speed variance increases link durations (same results apply to the other APs).

We also studied the effects of mobility on AP connectivity in the Östermalm scenario and Fig. 18(a) shows the AP positions: SW-Entrance, Intersection near centre and in the middle of a lower left street. Mobility in the Östermalm scenario is much more homogenous than in the Subway and this is reflected in the link duration histograms for the different APs (Fig. 18(b)). Althought the number of contacts is different for different AP positions, the link duration distributions are still very similar. In conclusion, our observations show that in scenarios with non-homogeneous mobility, access point position can significantly affect node connectivity.

6 Conclusion

In this work we have studied how pedestrian mobility affects the connectivity of wireless systems. We focus on identifying effects due to details in the operational-level mobility process, such as distribution of target speed, node arrival process, node-to-node interactions and specifics of the scenario in which mobility occurs.

Our approach is simulation based and we use Legion Studio, an agent-based simulator that uses state-of-the art pedestrian mobility models that realistically capture node interactions and the structure of the space in which mobility occurs. To the best of our knowledge, Legion Studio is the most advanced and realistic simulation model available for micro-level pedestrian mobility and it has, as far as we know, not been used by others for evaluating mobile communication systems.

We have statistically analyzed the empirical distributions for the connectivity metrics and how sensitive they are to changes in mobility input parameters. Our findings show that changes in micro-level mobility input metrics do not affect the connectivity significantly. The empirical distributions of the connectivity metrics are relatively insensitive to changes in input parameters, showing only a modest change in mean values and retaining their basic shape. This is positive since *if* the statistics were highly sensitive to modest changes in input parameters, capturing them by a model would be hard and it would be difficult to identify valid parameters for the model. Comparing connectivity metrics across scenarios however shows more difference and this is particularly reflected in the link duration and path duration metrics. We also studied how the performance of stationary nodes, such as access points, is affected by mobility and AP position. We find that the connectivity of stationary nodes in the Subway scenario depends strongly on position since in this setting mobility of the nodes differs significantly between different areas within the scenario. All of the above suggests that when modelling mobility, accurately capturing the scenario and its structure is much more important than a detailed estimation of input mobility parameters.

We show that node-to-node interactions in dense scenarios only have a small effect on connectivity and they lead to an increase in link durations which is likely to improve performance for most applications. Mobility models that assume a free flow of nodes are therefore likely to give slightly pessimistic performance results when compared to similar models that capture node interactions.

One of the main contributions of our work is the introduction of agent-based simulation as a tool for evaluating the effects of mobility on wireless communication. This allows us to study mobility on a finer timescale and for larger populations than most other works and address and evaluate the effect of assumptions (such as free flow of nodes and insensitivity to walking speed distribution) that are commonly made without justification when studying mobility and wireless communication.

A question that still remains open is whether modeling of pedestrian mobility is necessary in the first place. After examining two distinct mobility scenarios with comparable input parameters, we found a common mobility model hard to devise. We see this finding as a sign for the need of defining an extensive suit of benchmark traces for evaluating mobility in different environments rather than attempting to derive a one-size-fits-all analytical model for pedestrian mobility.

REFERENCES

[1] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

[2] F. Bai, N. Sadagopan, and A. Helmy, "The important framework for analyzing the impact of mobility on performance of routing for ad hoc networks," *Ad Hoc Netw. J.*, vol. 1, no. 4, pp. 383–403, 2003.

[3] A. Chaintreau *et al.*, "Impact of human mobility on the design of opportunistic forwarding algorithms," in *Proc. IEEE INFOCOM*, Barcelona, Spain, 2006.

[4] V. Vukadinović, Ó. R. Helgason, and G. Karlsson, "An analytical model for pedestrian content distribution in a grid of streets," *Math. Comput. Modell.*, vol. 57, no. 11–12, pp. 2933–2944, 2013.

[5] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri, "Towards realistic mobility models for mobile ad hoc networks," in *Proc. MobiCom*, San Diego, CA, USA, 2003, pp. 217–229.

[6] M. Musolesi and C. Mascolo, "A community based mobility model for ad hoc network research," in *Proc. REALMAN*, Florence, Italy, 2006.

[7] F. Ekman, A. Keränen, J. Karvo, and J. Ott, "Working day movement model," in *Proc. MobilityModels*, Hong Kong, China, 2008, pp. 33–40.

[8] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "SLAW: A mobility model for human walks," in *Proc. INFOCOM*, Rio de Janeiro, Brazil, 2009.

[9] N. Eagle and A. Pentland, "Reality mining: Sensing complex social systems," *Pers. Ubiquit. Comput.*, vol. 10, no. 4, pp. 255–268, 2005.

[10] Ó. Helgason, S. Kouyoumdjieva, and G. Karlsson, "Does mobility matter?" in *Proc. IEEE/IFIP WONS*, Kranjska-Gora, Slovenia, 2010.

[11] S. P. Hoogendoorn and P. H. L. Bovy, "Pedestrian route-choice and activity scheduling theory and models," *Transp. Res. B Methodol.*, vol. 38, no. 2, pp. 169–190, 2004.

[12] M. Kagarlis, "Method and apparatus of simulating movement of an autonomous entity through an environment," United States Patent No. U.S. 7,188,056 B2, Mar. 2007.

[13] J. L. Berrou, J. Beecham, P. Quaglia, M. A. Kagarlis, and A. Gerodimos, "Calibration and validation of the legion simulation model using empirical data," in *Pedestrian and Evacuation Dynamics*. Berlin, Germany: Springer, 2007, ch. 3, pp. 167–181.

[14] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, Jun. 2008.

[15] F. Bai, N. Sadagopan, B. Krishnamachari, and A. Helmy, "Modeling path duration distributions in MANETs and their impact on reactive routing protocols," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 7, pp. 1357–1373, Sept. 2004.

[16] Y. Han, R. J. La, and H. Zhang, "Path selection in mobile ad-hoc networks and distribution of path duration," in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006.

[17] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Phys. Rev. E*, vol. 51, no. 5, pp. 4282–4286, May 1995.

[18] I. Benenson and P. Torrens, *Geosimulation: Automata-Based Modeling of Urban Phenomena*. Chichester, U.K.: Wiley, 2004.

[19] *Pedsim Pedestrian Crowd Simulator* [Online]. Available: http://pedsim.silmaril.org/

[20] *Legion Studio* [Online]. Available: http://www.legion.com/

[21] Ó. Helgason and K. Jónsson, "Opportunistic networking in OMNeT++," in *Proc. Simutools, OMNeT++ Workshop*, Marseille, France, 2008.

[22] Ó. Helgason, E. A. Yavuz, S. T. Kouyoumdjieva, L. Pajevic, and G. Karlsson, "A mobile peer-to-peer system for opportunistic content-centric networking," in *Proc. ACM SIGCOMM Mobiheld 2010 Workshop*, New Delhi, India.

[23] A. Law and D. Kelton, *Simulation Modeling and Analysis, 3/e*. New York, NY, USA: McGraw-Hill, 2000.

[24] *Pace Of Life* [Online]. Available: http://www.paceoflife.co.uk/

[25] P. Hui *et al.*, "Pocket switched networks and human mobility in conference environments," in *Proc. WDTN*, Philadelphia, PA, USA, 2005.

[26] N. Banerjee, M. D. Corner, and B. N. Levine, "An energy-efficient architecture for dtn throwboxes," in *Proc. IEEE INFOCOM*, 2007.

Ólafur Helgason received the M.Sc degree in computer science from the University of Iceland in 2002, and the Ph.D. degree in electrical engineering from KTH in 2011. He is a Post-Doctoral Researcher at the Laboratory for Communication Networks, KTH, Stockholm, Sweden. Previously, he was a Researcher at the Systems and Services Laboratory, Reykjavik University, Reykjavik, Iceland, and as a Professional Programmer. In 2001, he was an intern at AT&T Shannon Laboratories, Florham Park, NJ, USA. His current research interests include mobility modelling for wireless communication, opportunistic networking, and mobile peer-to-peer systems.

Gunnar Karlsson received the M.Sc. degree from Chalmers University of Technology, Gothenburg, Sweden, in 1983, and the Ph.D. degree from Columbia University, New York, NY, USA, in 1989. He is a Professor of KTH, the Royal Institute of Technology, Stockholm, Sweden, since 1998. He is the Director of the Laboratory for Communication Networks. Previously, he was with IBM Zurich Research Laboratory and the Swedish Institute of Computer Science (SICS). He was a Visiting Professor at EPFL, Switzerland, and the Helsinki University of Technology in Finland, and ETH Zurich in Switzerland. His current research interests include quality of service, wireless LAN developments, and wireless content distribution.

Sylvia T. Kouyoumdjieva is currently pursuing the Ph.D. degree at the Laboratory for Communication Networks at KTH, the Royal Institute of Technology, Stockholm, Sweden. She received both the B.Sc. and M.Sc. degrees in telecommunications from the Technical University of Sofia, Bulgaria, in 2006, and from the Royal Institute of Technology, Stockholm, Sweden in 2009, respectively. Her current research interests include opportunistic networking and mobile peer-to-peer systems.