RESEARCH ARTICLE

Movement Activity Estimation and Forwarding Effects for Opportunistic Networking Based on Urban Mobility Traces

Karin Anna Hummel[1][*] and Andrea Hess[2]

[1] ETH Zurich, Computer Engineering and Networks Laboratory, Communication Systems Group, Gloriastrasse 35, 8092 Zurich, Switzerland

[2] University of Vienna, Faculty of Computer Science, Research Group Entertainment Computing, Lenaugasse 2/8, 1080 Vienna, Austria

ABSTRACT

Opportunistic, mobility-assisted, or encounter networking is a method based on ad-hoc networking and introduced to disseminate data in a store-and-forward manner by means of spontaneously connecting mobile devices. While in many networked systems mobility is treated as a challenge requiring additional management, in opportunistic networks movement facilitates networking as it creates additional contacts between devices. These new networking opportunities can be exploited in addition to traditional wireless infrastructure networks or in absence of these networks. Hereby, algorithms for opportunistic data dissemination make use of information about social ties, regularities in movement, and the future path of mobile entities. The availability of this information is reasonable for areas such as campuses or conference venues, where social or professional ties are strong or when traveling by, e.g., public transport lines or vehicles following a navigation system. Other movement activities of humans in larger areas often lack this information and new techniques are required to derive similar useful movement information.

By observing movement characteristics of network users such as average velocities or revisiting patterns, estimates about the likelihood of getting in contact with other devices can be estimated. Our approach goes one step further by introducing users' movement activities derived from movement patterns typical in, e.g., tourist movement, shopping activities, or evening activities. Movement activities are notions summarizing a particular movement situation which is meaningful to users and can be used to further estimate user needs and user generated network traffic. In case movement patterns are uncertain or fragmentary, knowledge about activities may help to faster estimate average movement characteristics. The main objective of this paper is to detail the approach of relating activities to observed multi-variate mobility characteristics based on the Naïve Bayes classifier. The approach is applied to four typical urban movement use case activities including pedestrian and vehicular movement. Results are presented based on two different experimental training sets consisting of GPS outdoor traces: first, a training set of emulated movement activities and, second, a training set consisting of labeled real-world daily activities over one month tracked by volunteers. The results of the classification study confirm that movement can be characterized as proposed. By using mobility activities and corresponding distributions of movement characteristics, the impact of activities on opportunistic forwarding performance in terms of contact and inter-contact time, forwarding distance and coverage of an area, and predictability of the future path of a moving device is investigated.

[*] **Correspondence**

ETH Zurich, Computer Engineering and Networks Laboratory, Communication Systems Group, Gloriastrasse 35, 8092 Zurich, Switzerland

Wirel. Commun. Mob. Comput. 2012; **xx**:2–21

Prepared using wcmauth.cls

1

WIRELESS COMMUNICATIONS AND MOBILE COMPUTING

Wirel. Commun. Mob. Comput. 2012; **xx**:2–21

1. INTRODUCTION

In the early years of mobile wireless networking research, network protocols were often evaluated by using simple random mobility models, like the Random Waypoint mobility model, which do not describe real-life movement sufficiently and can lead to non-uniform node distributions within an area as shown in [1]. When optimizing network protocols under mobility in wireless infrastructure networks and Mobile Ad-hoc Networks (MANETs), a wrong assumption about the mobility hinders the success of protocols in real-world settings.

Using appropriate mobility models for opportunistic and delay-tolerant networks is even more important and challenging because mobility is exploited for connectivity opportunities. Here, accurate mobility models are required by the networking algorithms to show their full potential. The requirement of more realistic mobility models, i.e., models that are representative for the observation goal and include realistic assumptions about pathways, obstacles, velocities, etc., has led to collecting real-life traces as found, e.g., in the CRAWDAD database [2]. As *movement* is seen as being an important property of mobile networks, mobility models have moved to the center of investigation. Section 2 gives an overview of related work on mobility modeling.

Opportunistic networking relies on mobile nodes getting in range and in contact with one another. The quality of the contact can be measured in terms of contact-time and inter-contact time. The quality of the dissemination protocol is linked to the dissemination purpose. For example, sending a message to a particular receiver node requires other routing and forwarding mechanisms than spreading information within a certain area irrespective of the receiving nodes' identities. In the latter case, the availability of information in the area of interest is a measure of the dissemination quality. The movement of nodes, i.e., the primary cause for contacts, influences these forwarding characteristics of opportunistic networks. Hence, it should be possible to improve data dissemination in the opportunistic network by selecting devices due to their movement characteristics best fitting for forwarding, e.g., because they show a large mobility range, tend to revisit positions frequently, or travel with moderate velocities.

Our research questions focus on whether we can describe observed movement characteristics in terms of movement activities, and whether these movement activities show different impacts on opportunistic networking. In [3], we introduced and defined major mobility characteristics, gave a brief description of the possible effects of mobility characteristics on opportunistic networking, and presented the approach of relating mobility characteristics and movement activities exemplified for emulated mobility traces. In [4], a long-term real-world study on daily activities tracked by GPS devices is described. In this paper, we summarize the results briefly and extend them by presenting an approximation of the empirical distribution functions of observed characteristics by well-known distribution functions. Further, we investigate the impacts of the empirical distribution functions on forwarding capabilities of mobile devices using agent-based simulation.

In Section 3, we derive a set of mobility characteristics related to the time and space domain based on a comprehensive literature survey and own extensions. The classification approach relating movement characteristics to movement activities is presented in Section 4. A Naïve Bayes classifier applied to a multi-variate data set is used to infer movement activities. The classification approach is evaluated by a set of labeled GPS traces of emulated and real-life urban movement activities of test participants in Section 5. The selected movement activities are: *way to work*, going out for daily *shopping*, *evening activity*, and being a *tourist* in a city.

Finally, we relate opportunistic networking quality to movement activity. A discussion of possible impacts of mobility on opportunistic networking is given in Section 3 and investigated in Section 6. Based on simulation, an opportunistic network of mobile nodes is analyzed in terms of contact time and inter-contact time, forwarding distance and coverage, and predictability of movement. The differences in networking quality are discussed along the movement activities.

2. RELATED WORK ON REALISTIC MOBILITY MODELING FOR OPPORTUNISTIC NETWORKS

One way to model mobility realistically and useful for networking protocols is given by integrating geographic and topological restrictions [5]. Simple mobility models, for instance, are based on grid topologies while more complex models use detailed city maps. Another approach is to use traces captured from human movement as a source to derive realistic mobility models. Mobility models extracted from traces try to capture existing patterns by statistical analysis and pattern recognition techniques. In [6], the resulting mobility model is based on known statistical models, such as log-normal or Gaussian distributions, that can be assigned to the properties speed, pause time, and to the placement of hotspot regions on the simulation area. In [7], a time-variant community mobility model is proposed based on location visiting preferences and periodical re-appearances at the same location observed in WLAN traces. These movement characteristics are referred to and extended in the following sections of this paper.

Recent studies of human movement traces have shown that human movement behavior can be characterized by using truncated power-law flights. In [8], the authors discuss the mapping of human trajectories to the power-law flights, called Lévy flights. In principle, the observation showed that people tend to travel most of the time over short distances, while some people travel occasionally over hundreds of kilometers. The SLAW (Self-similar Least Action Walk) mobility model is based on similar statistical patterns combined with a least action trip planning algorithm that is based on the observation that people attempt to minimize the total traveling distance by visiting nearby destinations first [9].

Based on surveys, in [10], the Working Day Movement Model is proposed, which pictures the everyday life of average people. Similar to our approach, activities are modeled and considered. This mobility model consists of three sub-models for home-, office-, and evening activities and shortest path transitions between these activities. The shortest paths on the city map to the next destination depends on the selected transport modality such as walking, going by car or by bus which defines, e.g., the velocity. Derived from urban traffic planning research, a definition of a mobility model considering aspects of urban mobility is detailed in [11]. Here, the mobility of pedestrians during a workday is modeled in detail by using an agent model describing dynamics between moving nodes and parameters describing activities and tasks determined for each pedestrian based on time use studies.

Since our approach describes also the relation between mobility characteristics and activities, it contributes to the field of activity modeling based on real-world mobility data. In location-based activity recognition, high-level activities are derived by detecting significant places and inferring the types of the places (such as work, home, at a friend, and parking lot) [12]. Additionally, activities at these places are detected by combining GPS traces, temporal information (e.g., time of day, stay time, day of week), and semantic information from geographic databases (e.g., location of bus stops, restaurants, stores). Similarly, daily activity patterns are generated in [13] based on GSM roaming and handover data and Point of Interest (POI) information characterizing the major activity in each area on the map. In [14, 15], the Reality Mining dataset [16] is analyzed, where an activity type (home, work, other) is derived from a participant questionnaire for each GSM cell association. Particularly in [14], the authors investigate the usefulness of certain time points for predicting future activities. Similarly, we use labeled data in our approach that are derived from user log diaries.

Apart from mobility and activity modeling, approaches to opportunistic routing and forwarding based on mobility patterns are related to our work. Whether and how mobility influences the performance of opportunistic networking is a major question. In [17], the sensitivity of connectivity metrics to changes of mobility parameters is investigated. In this study, pedestrian mobility models are included based on an advanced civil engineering mobility simulator capable to capture realistic interactions of pedestrians with the physical environment. In [18], forwarding nodes are selected by predicting the future contacts of a node based on a 'delivery predictability' metric assuming that a node is suitable for forwarding a message to another node if the two nodes meet one another frequently. Every time a node encounters another node, the 'delivery predictability' is increased while it is decreased if they do not encounter each other for a while. A similar method based on the mean inter-meeting time known for each node pair in the network is used in [19] to enhance forwarding protocols for delay-tolerant networks. In [20, 21], routing decisions are grounded on similarities in patterns like

meeting frequency and visited locations of destination and forwarding nodes. In [22], location visiting and velocity patterns are utilized by a mechanism for message replica deletion in delay-tolerant networks. A controlled epidemic forwarding approach where relay nodes are chosen depending on their path history is proposed in [23]. Apart from the domain of ad-hoc networking, knowledge about mobility patterns has been utilized, for example, in cellular networks for paging schemes based on mobility profiles considering most (recently) visited cells and temporal patterns [24, 25]. Although our work is focused on opportunistic networks, the study of movement activities based on GPS traces is a contribution to a wider range of mobile networks, e.g., also to cellular networks and their services where micro mobility and user movement context are of interest.

3. MOBILITY CHARACTERISTICS

Since to the best of our knowledge there is no comprehensive summary of mobility characteristics and patterns, we present an overview and definition of commonly used movement characteristics followed by a discussion about how these characteristics can be considered by opportunistic data forwarding algorithms.

3.1. Mobility Characteristics

Movement can be defined as the process of changing location (over time). Consequently, the dimensions *space* and *time* are used to classify movement characteristics. Additionally, *spatio-temporal characteristics* are introduced correlating spatial and temporal features of movement. The summary presented here is based on an extensive literature survey. Specific emphasize is put on the characteristics further used in this paper for individual trips.

Spatial characteristics are used to capture movement in the space domain of single users as well as of groups. The characteristics commonly used are:

- **Prevalence.** Prevalence is a metric giving the fraction of time a user spends at a given location [26, 27, 28]. This metric is often used as a spatial metric as it determines the spatial distribution of users. Spatial characteristics that are derived from the basic prevalence metric are, for instance, *popular locations* in terms of 'location visiting preferences' [7] or 'hotspot regions' [6]. In this paper, the metric will not be used since the individual micro-mobility traces of our empirical study do not provide a sufficient number of samples for the whole area they cover.
- **Flight length.** The flight length has been defined in [29] as the distance between two end points of a path where no movement pause occurred and the distance to a straight line connecting the end points from each sampled position in between is below a certain threshold. In this paper, we calculate the flight length as the length of a path traveled between two consecutive pause phases.
- **Activity range.** The activity range of a node can be defined as the area covering all locations that have been visited, as defined in [28]. In this paper, we use the term **mobility range** similarly to activity range and define it in terms of distance to a center for arbitrary trips, where the whole trip is encapsulated by a rectangle and the "center" is defined as the center of the rectangle.

Temporal characteristics commonly used are:

- **Pause time.** The pause time is the duration of a phase between two consecutive movement phases. Pause times are extracted, for example, based on WLAN data in [30] or based on GPS data in [9].
- **Persistence.** Persistence can be defined as the amount of time a user stays continuously connected to an access point (AP). In [26, 27], persistence is used to infer the session length distribution. This metric is not used further in this paper, since we do not consider user sessions, but only dwelling periods already described by pause times.

4   *Wirel. Commun. Mob. Comput.* 2012; **xx**:2–21

Prepared using wcmauth.cls

- **Start time.** The start (and end) time of a trip allows to capture day time dependent behavior. In relation to our approach is the use of starting and ending time in [6], where the time of the first and the last appearance of a mobile device on a campus during a workday is considered as mobility characteristic.

Spatio-temporal characteristics are based on correlations between spatial and temporal movement properties:

- **Revisits.** In [7], the authors define the characteristic 'periodical re-appearances', i.e., the re-appearance at the same AP after a certain time gap. This metric allows to capture time-dependent behavior, such as recurrent daily and weekly schedules. Similarly, the time until a node returns to an area after exiting this area ('return time') [31] and the remaining time for visiting a location ('hitting time') [7] have been investigated. In this paper, both the **number of revisited positions** and the **time between revisits** are used to capture revisits of locations.
- **Meetings.** Frequency and duration of contacts between nodes are of particular interest for opportunistic forwarding mechanisms. The inter-contact time is a common meeting characteristic giving the time elapsed between two successive contact periods for a given node pair, which is investigated in [32, 33, 34, 31, 35] for various data sets. Additional meeting characteristics studied in related literature are contact time [33, 34], meeting time [34], (inter-)any-contact time [33], and time distance [35]. Since we analyze individual movements in larger areas, meetings are rare and only considered as forwarding metric when analyzing the impact of mobility on the network in the simulation study (Section 6).

In addition to these characteristics, metrics derived directly from the position traces used in this paper are: **velocity**, i.e., the speed between two consecutive positions in m/s, and intensity of **change of direction** in degrees between two track-points used to describe the smoothness of movement.

It is worth mentioning that mobility characteristic extraction from real life traces is constrained by the data capturing method determining the granularity of time and location observations. For example, using cellular or WLAN based mobility data, some fine granular movements can not be detected and only estimated in the models. Some indoor positioning systems, GPS, and Differential GPS (D-GPS) provide trace data of finer granularity, but are often not available for higher numbers of mobile nodes.

3.2. Mobility Effects for Opportunistic Forwarding

Whether opportunistic networks can be formed between mobile devices carried by their owners while meeting one another on the street, in buildings, or on public transport strongly depends on the users' movement patterns. Here, we assume individual movement without looking at group movement caused by social relations and ties or determined, e.g., by public transport such as joining on the same bus. Table I briefly summarizes how the metrics defined in Section 3 may affect data forwarding in opportunistic networks along the selected forwarding metrics: connection duration and frequency of connection establishments (connection opportunities), forwarding distance and area covered, and predictability of future locations. The forwarding distance captures the amount of different areas reached per node while the area covered captures the availability of information in the area as a dissemination result of all nodes. The possible correlations described are intended to give direction to further investigations.

The connection duration is crucial to determine whether networked nodes stay in contact long enough for the required network load to be transmitted. The frequency of connection establishment and forwarding distance of a node affect its contribution to dissemination. Finally, the predictability of future locations is of importance in case a forwarding protocol makes use of estimated future routes of nodes. Note, that the *meeting* metrics used to describe meetings of devices or humans can be directly mapped to the forwarding metrics that describe connections of networked devices.

To illustrate the relations by an example, we assume people on the way to their workplace who dwell just for some minutes at bus stations, whereas during evening activities, they pause at various locations such as theaters or bars for longer periods. With respect to forwarding, nodes with a fitting move/pause pattern can be selected by opportunistic protocols as preferred forwarding nodes. For example, a heuristic forwarding protocol aware of the pause times of a node may select nodes with shorter pause times for distributing small information items quickly.

Table I introduces some potential impacts of single mobility characteristics on forwarding characteristics using simplifying assumptions, e.g., assuming that changes of direction do not follow any (periodic) pattern which would increase predictability. Further, a forwarding metric is affected by a combination of characteristics simultaneously and also by the combined behavior of mobile device under observation (device density, presence or absence of group behavior, etc.). Our focus is set on investigating nodes along their individual movement characteristics to derive estimates for a node's networking opportunities. Hence, we focus on contacts between (mostly) strangers without considering group mobility. In Section 6, we present a simulation experiment that investigates the forwarding capabilities of different movement activities.

Mobility characteristic		Forwarding metric
Prevalence, pause time, persistence		
Pause/dwell time	short	short connection duration
	long	long connection duration
Activity range, flight length, velocity		
Mobility (activity) range	small	short forwarding distance
	large	long forwarding distance
Flight length	short	short forwarding distance
	long	long forwarding distance
Meetings		
Contact frequency	low	few connection establishments
	high	many connection establishments
Contact time	short	short connection duration
	long	long connection duration
Revisits, changes of direction		
Revisited locations	few	bad predictability of future locations
	many	good predictability of future locations
Changes of direction	few	good predictability of future locations
	many	bad predictability of future locations

Table I. Potential impacts of individual mobility characteristics on forwarding.

4. DERIVING ACTIVITIES FROM CHARACTERISTICS

Micro mobility characteristics can be aggregated and used to directly infer probabilities of connectivity opportunities, but there is still a lack of a semantic description of movement pattern combinations. The notion of a *movement activity* is introduced to improve the user's understanding of movement, which is important for experimental tests with users. Additionally, knowing the movement activity can help to derive mobility characteristics in case mobility patterns are fragmented and not fully known. A movement activity is here seen as a composition of mobility patterns that can be summarized by a name corresponding to the purpose of the trip of a mobile user or entity. Beyond the networking view, the knowledge about activities of mobile users can be utilized by applications to provide activity-aware services and service adaptations.

The set of mobility characteristics is defined as $C = \{C_1, C_2, ... C_K\}$, where K is the number of different characteristics and may vary depending on the modeling purpose. To derive the probability of the occurrence of a specific activity, we follow a *Naïve Bayes* classification approach to detect an activity given a specific mobility characteristic vector $V_j = (v_{j1}, v_{j2}, ... v_{jK})$ where each v_{jk} describes a movement characteristic C_k (movement feature), e.g., in terms of an interval, a binary value, etc. In case only one movement characteristic is considered, the vector is reduced to a scalar.

The calculation is based on Bayes' Theorem ($1 \leq i \leq n$ and $1 \leq j \leq m$, where n is the number of different activities and m the number of different characteristic vectors):

$$P(A_i|V_j) = \frac{P(V_j|A_i)P(A_i)}{P(V_j)}, \text{ where}$$

6   *Wirel. Commun. Mob. Comput.* 2012; **xx**:2–21

Prepared using wcmauth.cls

A_i is a movement activity or trip and V_j is the movement characteristic vector observed. Once this probability is calculated for all different activities A_i, the activity with highest probability is selected as classification output. *Naïve Bayes* classification has been selected as it provides good results while being simple to use. However, the activity and movement vector classes can be used by other classification approaches as well.

In our work, we investigate individual movement captured by GPS devices. Based on position traces the empirical distribution of each characteristic C_k is derived and used during classification. The considered characteristics of our first study are (see Section 3): Velocity, change of direction, flight length, mobility range (i.e., distance to center), pause time, number of revisited positions, and time between revisits. In a second step, we introduce also start time as an important characteristic.

5. MOVEMENT CHARACTERIZATION EXPERIMENTS

As we focus on investigating the differences in human movement behavior in terms of movement characteristics, we perform a two-step experimental study. In the first study, we emulate mobility activities in urban areas and generate 'artificial' GPS mobility traces of selected mobility activities for training a Naïve Bayes classifier; while in the second study, we use labeled every-day GPS mobility traces collected during real-world movement for training. This approach of first considering emulated movement allows to describe movement activities without disturbances and to detail the potential of the classification approach for the real-world data set.

After describing the movement use cases and the data sets used, movement characteristic extraction and the empirical cumulated density functions are presented for the emulated training set. Then, we present classification results using the emulated movement activity and parts of the real-world data sets as training data and parts of the real-world data set as test data.

5.1. Movement Activity Use Cases

Four different typical movement activities are defined as use cases. The activities have been selected due to their applicability in urban areas. In the real-world data set, these scenarios were the most frequent scenarios reported by the volunteers. Additionally, the movement characteristics are expected to differ in these scenarios:

- **Way to Work.** In this movement activity, the straightest way from home to office is taken using various means of transport (public transport, pedestrian movement, car, etc.).
- **Shopping Activity.** These trips are characterized by a situation where a person starts the trip at a location such as at home or at the office, visits several shops, and returns.
- **Evening Activity.** These trips describe an activity where public places, like theaters, restaurants, bars, are visited after, e.g., leaving the office in the evening.
- **Tourist Activity.** This activity describes a typical (mainly pedestrian) movement of tourists in a city, e.g., pausing at sights to take pictures and visiting multiple sights in a city area.

5.2. Mobility Data Sets Used

The *artificial data set* consists of GPS trips generated by emulating outdoor activities. Test persons acted as if they were performing specific movement activities in Vienna (four trips corresponding to the four movement activities considered). While in this controlled experiment the data set showed no artifacts, it is clear that it only approximates daily activities and lacks variations.

The *real-world data set* consists of daily observation of 13 volunteering test persons during a long-term study over one month. In addition to the GPS traces collected, the test persons also took notes by identifying the type of movement activity, e.g., that one trip was a regular morning trip to work, a shopping trip, etc. During preprocessing of the GPS position

	Evening	Shopping	Tourist	Work	All
Pedestrian	35	18	25	82	160
Vehicular	6	11	6	69	92
All	41	29	31	151	252

Table II. Number of trips in the real-world data set split into pedestrian (including public means of transport) and vehicular trips.
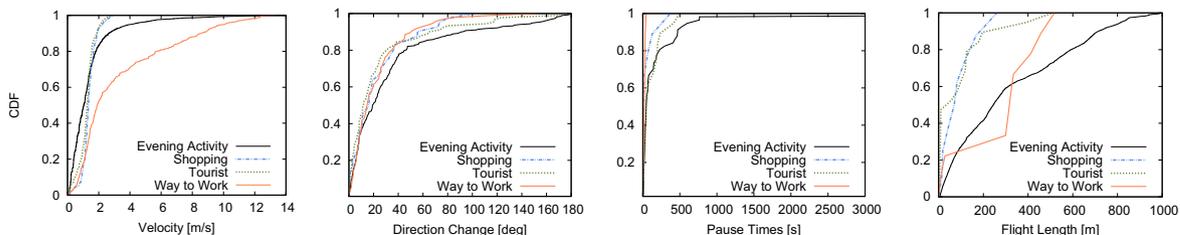

Figure 1. Empirical CDFs of (a) velocity, (b) change of direction, (c) pause time, and (d) flight length occurrences.

data, measurement outliers that occurred due to shielding by obstacles such as houses were eliminated by analyzing the velocity values using the Local Outlier Factor (LOF) algorithm [36]. The real-world data set comprises a greater variation of movement, local specific biases such as the influence of certain area topographies are reduced, and real-life behavior can be observed.

During both studies, GPS positions were logged approximately once per second by small GPS receivers carried by the test persons. After filtering and selecting the most frequently reported trips, the real-world data set used consists of 252 trips in total – namely 151 Way to work, 29 Shopping, 41 Evening, and 31 Tourist trips (see Table II). Trips of the first three categories were collected in Vienna, while Tourist trips were collected in a couple of European cities and vacation spots.

5.3. Investigating Movement Characteristics

In this section, the measured mobility characteristics in the emulated data set are analyzed. Additionally, specifics about aggregating the movement characteristics valid for both the artificial and the real-world data sets are presented.

Velocity, Change of Direction, Pause Time, and Flight Length

Velocity, flight length, and pause times have been calculated directly along the definitions previously introduced. For changes of direction, we used a minimal distance value of 20m, meaning that a change of direction is investigated after a distance of 20m has been reached to the previous position. The calculation compares the current direction with the direction measured previously. Changes of direction range from $0°$ to $180°$ and summarize both right and left turns.

The empirical CDF of **velocity** for each type of trip is depicted in Figure 1(a). The lower velocities in the mostly pedestrian activities *Shopping* and *Tourist* are visible; *Shopping* yields a lower occurrence for velocities below 1m/s. As *Way to work* and *Evening* consists of walking and public transport parts, higher velocities are more likely.

The occurrences of intensity of **change of direction** show a more homogeneous picture among all the activities (Figure 1(b)). The highest number of occurrences is given in the *Way to work* movement activity for changes of direction below $90°$, which is plausible due to the more limited turning when driving.

The **pause time** (Figure 1(c)) occurrences show different patterns for the activities. While for the *Way to work* activity most occurrences lie below 40s, the other activities show different accumulation points visualized by steeper curves in
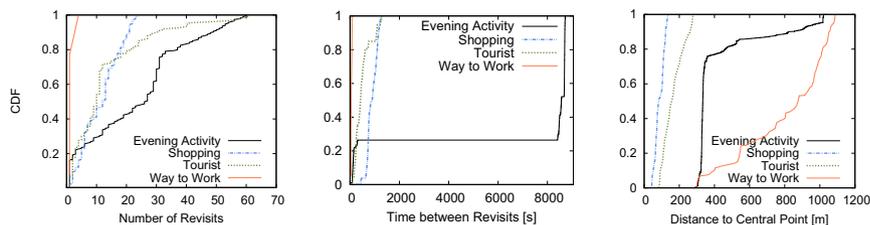
8

Wirel. Commun. Mob. Comput. 2012; **xx**:2–21

Prepared using wcmauth.cls

Figure 2. Empirical CDFs of (a) number of revisits per position, (b) times between revisits, and (c) mobility range (distance to center).

certain pause intervals such as between 175s and 200s for the *Tourist* activity. The longest pause times (up to 10 000s) are calculated for *Evening* activities.[*]

The empirical CDFs of **flight length** occurrences (Figure 1(d)) show a high fraction of short 'flights' amounting to less than 100m in the activities *Shopping* and *Tourist*, while at the same time, the flight length in the latter shows a higher variation (ranging from 3m to 510m). In case of the two activities including public transport mobility, i.e., *Evening* activity and *Way to work*, short 'flights' occur less frequently and occurrences strongly depend on the means of transport.

Revisiting Positions and Mobility Range

The frequency of position revisits allows to describe loops in pathways. Positions are treated as the same position, if they lie within a configurable position radius. We show here the results for one selected radius $r = 20$m. A revisit is only assumed if the moving person has first moved away from the position, here, the configurable distance is set to 50m.

Figure 2(a) shows the occurrences of the **number of position revisits**. The number of revisits per position is generally low for the movement activity *Way to work* (all occurrences below 4 revisits). In the *Tourist* movement activity, an accumulation point lies in the interval from 8 to 11 revisits and a moderate occurrence also at higher numbers of revisits, which corresponds to the behavior of tourists walking around sights.

In Figure 2(b), the occurrences of **time between revisits** are depicted. In the *Way to work* activity, the times between revisits are mostly below 65s as only very short periods may occur when, e.g., walking back on a street section after getting off a bus. In the *Tourist* activity, a high number of occurrences is given in the interval between 150s and 550s while the *Shopping* activity clearly shows the path back from the shop revisiting locations after approximately 600s. The *Evening* activity shows the longest times between revisits, i.e., more than 8 000s.

Finally, Figure 2(c) illustrates the cumulative occurrences of the **mobility range**. The small movement areas for the activities *Shopping* and *Tourist* result in short distances. In the *Way to work* activity, the movements nearly show a steady-going curve between 300m and 1 200m, whereas the line for the *Evening* activity exhibits a high density of position measurements between 290m and 340m although the maximum lies above 1 000m.

5.4. Activity Detection

The Naïve Bayes classifier of the open source machine learning library WEKA [37] has been applied to the movement characteristic vectors. Here, the results are presented for two experiments: first, using the artificial data set for training, and, second, using the real-life data set for training.

Results of the Artificial Training Set

Previous to classification into the four activities, we defined two to three meaningful categories for each characteristic (e.g., for velocity v we used two classes for $v \leq 2$m/s and $v > 2$m/s to distinguish between pedestrian and faster movements) and calculated the occurrences both for the training and a test set (both sets normalized). For training

---

[*]For readability reasons, the x-axis in Figure 1(c) has been limited to 3000s.

Wirel. Commun. Mob. Comput. 2012; **xx**:2–21

Prepared using wcmauth.cls   **9**

Training data set	Velocity	Direction change	Pause time	Flight length	Number revisits	Time betw. revisits	Mobility range	Start time	Multi-variate
(a) Artificial	0.35	0.48	0.38	0.46	0.39	0.43	0.38	–	0.61
(b) Real-world	0.57	0.65	0.66	0.64	0.66	0.64	0.58	0.59	0.81

Table III. Classification success rate for (a) an example test set of four real-world traces after training the classifier with the artificial data set, and (b) for the real-world test set after training with the real-world training set.

and classification, we separated the trip into sequences of mobility characteristic combinations observed in the trip and classified these sequences. For performing the classification, we selected four real-world traces arbitrarily.

As we used only few categories to differentiate between four different activities (i.e., two to three), the amount of correctly classified data items based on single characteristics yields expectedly bad results as summarized in Table III. The use of multi-variate characteristic data achieved a higher success rate for classification of 0.61, which is an encouraging outcome for only four arbitrarily selected trips.

It is expected that in particular an enriched training data set will yield better results due to increased representativeness and diversity. Further, the analysis of the classification process also showed that *Evening* activities exhibited many similarities to *Tourist* and *Way to work* activities resulting in false classifications.

Results of the Real-world Training Set

The real-world data set has been divided into a training and a test data set by randomly selecting half of the traces for training and a distinctive half of the remaining traces for testing. In comparison to the previous classification based on the artificial training set, the training and test data sets were structured differently. Every trip was classified in a single classification step. In addition to the characteristics already defined in the first study, we added the *start time* of a movement as an important discriminator to the multi-variate data set. Adding this characteristics allows to differentiate between 'early morning' trips, 'forenoon', 'afternoon', and 'evening/night' and, thus, to reflect the time-dependent aspect of trip purposes. This is useful, e.g., to differentiate between Evening and Way to work movement activities.

Previous to classification into activities – similar to the artificial data set – meaningful interval categories were defined for each mobility characteristic. Here, we increased the number of categories based on the insights of the first results for the emulated training set; now four to five categories are used. During experiments based on different interval settings, i.e., equally sized intervals, the achievable results showed lower true positive rates. Thus, we selected the interval settings corresponding to meaningful interval limits for use in our further study. For example, the value range of the characteristic *velocity* is separated into categories of pedestrian moving (velocity v of about $0 - 2$m/s), fast walking or running motion (v of about $2 - 8$m/s), moderate driving on city streets (v of about $8 - 20$m/s), and fast vehicle speed ($v > 20$m/s), such as observed for trains, or cars on a highway. For the other categories, similar meaningful intervals are used.

After training of the Naïve Bayes classifier, the classifier was tested against the test set. The classification of the real data test set led to an overall *classification success rate*, or *true positive rate*, of 0.81. In Table IV the numbers of trips assigned correctly and incorrectly during classification are given. The numbers of incorrect labels assigned expose similarities between different movement activities. To exemplify these similarities we identified explanations for incorrect classifications by investigating the experimental data set. Shorter pedestrian *Tourist* trips are classified as *Shopping* movement activity for instance, while *Way to work* trips are classified as *Evening* activity if starting at a late hour and containing pause times.

We observed that for *Way to work* and *Evening* activity a true positive rate of 0.91 and 0.80 was achieved, respectively. These two movement activities are well represented in the real-world dataset, while for the two other activity types, fewer data were available. To have a closer look at the false positives in the movement activities with smaller data set and lower classification success rate, we separated the data sets for *Shopping* and *Tourist* movement activities into pedestrian and vehicular trips. The resulting second classification matrix presented in Table IV (right hand side) allows a more detailed analysis of the cause for incorrect classifications in the categories *Shopping* and *Tourist*. For example, these numbers show that most *vehicular Shopping* trips are labeled incorrectly due to significant similarities to *Way to work* trips.

	Assigned label						Assigned label					
	E	S	T	W			E	S-p	T-p	W	S-v	T-v
Evening (E)	16	0	1	3		**Evening (E)**	14	0	2	3	1	0
Shopping (S)	0	8	2	4		**Shopping ped. (S-p)**	0	8	0	1	0	0
Tourist (T)	0	6	8	1		**Tourist ped. (T-p)**	1	2	8	1	0	0
Way to work (W)	4	0	3	68		**Way to work (W)**	3	1	1	67	3	0
						Shopping veh. (S-v)	0	0	1	3	1	0
						Tourist veh. (T-v)	0	0	0	0	0	3

Table IV. Classification matrix for (i) the four different movement activities and (ii) for six different movement activity types including a distinction between pedestrian and vehicular scenarios for movement activities with low classification success rates.

Similar to the first evaluation step based on the artificial training data set, we also classified the real-word test set using each single movement characteristic in addition to the multi-variate approach (the resulting classification success rates are included in Table III). For reducing the set of features again and as none of the single characteristics yields an acceptable result, we recommend to use subsets of multiple features. The subsets, however, should be investigated further in terms of classification success.

6. OPPORTUNISTIC FORWARDING EXPERIMENTS

The purpose of these investigations is to demonstrate to which extent mobility characteristics may determine the forwarding behavior of opportunistic networks. Mobility characteristics can be observed on single devices without further knowledge about contacts and are, therefore, easy to use in real-world applications. To investigate the potential of trip activity awareness, the activities introduced previously are used for determining a specific probabilistic mobility behavior. Then, a discussion of the influences of mobility on forwarding characteristics is presented.

6.1. Forwarding Metrics

Reflecting the descriptive considerations of Section 3.2, we evaluate an opportunistic forwarding scenario along the following metrics and their definitions:

Connection Establishment and Duration
- **Connection duration or contact time.** The contact time is measured as the time two nodes stay in communication range. For independent node mobility as used in the simulation, it is expected that the contact time is higher for activities containing longer pause times.
- **Inter-connection time or inter-contact time.** For reasons of completeness, we include the inter-contact time in addition to the contact time in our observations. The inter-contact time gives the time a node is continuously without any contact opportunities to other nodes. The lower the inter-contact time, the more frequently a node connects to another node. For independent nodes, the inter-contact time depends on the node density in an area.

Forwarding Distance and Coverage
- **Forwarding distance.** The forwarding distance of a mobile node is defined by the number of different sub-areas of the simulation area visited by the mobile node, or, in other words the capabilities of the node to reach different remote areas during its movement. In the simulation model, each sub-area is represented by a stationary node tracing contacts. This metric is expected to be affected by the mobility characteristics flight length and mobility range.
- **Coverage (availability of information).** In addition to the forwarding distance, the spatial distribution of information is investigated by determining the fraction of sub-areas that have received an information item, termed

coverage. In the simulation, the availability of information gives the fraction of stationary nodes representing a sub-area that have received the data. Similarly to the forwarding distance, it is expected that flight length and mobility range influence this metric. Additionally, as data is persistently kept at the stationary nodes, it is also influenced by the trip start time distribution.

Predictability of Future Locations of a Linear Predictor

The **prediction accuracy** rates the predicted next location based on the actual mobility path. The prediction accuracy depends first of all on the quality of the predictor, i.e., how well the predictor fits the movement and its regularities. In our simulation, a prediction is done every ten minutes by extrapolating the movement of a node linearly, taking current position and direction into account. In case, a node is within a predicted traveling corridor given by the area surrounding the predicted straight line (here, width of 500m), the prediction is counted as being a correct prediction. The distance to the predicted line is calculated every 10s. For linear extrapolation, the velocity is an important characteristic when a fixed time interval is used for predicting the next location. For more complex predictors considering more past visited locations and pathway regularities, also frequent revisits of past locations may result in better predictability due to the pattern generated by revisits. Further, a low number of irregular changes of direction together with a low intensity of changes of direction is expected to improve predictability.

6.2. Experimental Simulation Setup

To investigate the forwarding behavior of mobile nodes, agent-based simulation is used. As we are more interested in the general capabilities of the distributed system, the simulation setup considers information dissemination while not detailing the networking layers in this study (e.g., network congestions is not included in the model). Movement activities are simulated according to the mobility characteristics extracted from the real-world data set of GPS traces, i.e., the empirical probability functions described in Section 6.3. The simulation is based on AnyLogic, a Java based simulation environment [38].

Each movement activity has been simulated separately and we performed five simulation runs per activity. The mobile nodes are inserted uniformly distributed on the simulation area according to the observed start time distribution. As the start time determines whether nodes start their trip at all, we select the simulation runtime in order to compensate for start time differences and to assure fairness between, e.g., evening and daytime activities. The simulation area of 6km in width and length represents the size of, e.g., an inner-city area of a European city.

Few mobile nodes carry information right from the start of their trip (configurable fraction that is currently set to 0.1), whereas the rest of the mobile and stationary nodes may or may not receive the information from mobile nodes met during the simulation time. A trip ends when a mobile node reaches its mobility range.

A grid of stationary infrastructure nodes with a fixed distance in between is placed on the simulation area to determine data dissemination (coverage) and forwarding distance. These nodes are only receiving data from mobile nodes and do not disseminate data themselves. Figure 3 visualizes an excerpt of the stationary infrastructure grid at two time steps. There are basically two types of possible communication links of mobile nodes: connections to stationary nodes to infect a grid cell with a message and connections to other mobile nodes located within transmission range. In the current parameter setting, the simulation includes 169 infrastructure nodes as a consequence of the inter-spacing width of 500m. Table V summarizes the configuration parameters of the simulation.

6.3. Mobility Modeling Approach

The mobility model used in the simulation is based on the empirical probability density functions of the real-world data sample, i.e., the simulation uses the EPDFs of the characteristics velocity, change of direction, pause time, flight length, mobility range, and start time. The characteristics number of revisits and time between revisits are not used as input parameters but emerge during the simulation runs. To make the results comparable to other work, we approximate the EPDFs by fitting appropriate probability functions showing similar behavior for each of the mobility characteristics:
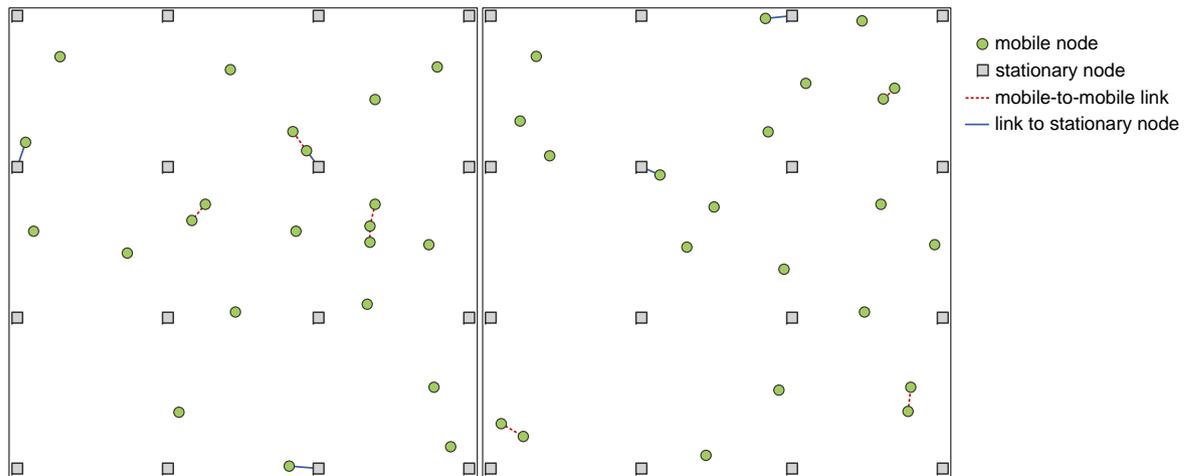
12   *Wirel. Commun. Mob. Comput.* 2012; **xx**:2–21

Prepared using wcmauth.cls

Figure 3. Visualization of infrastructure grid section (4x4 stationary nodes) with mobile nodes at simulation time step t_i and $t_i + \Delta$.

Area size	6km x 6km
Simulation time (per activity)	17 hours – Evening: 17:00-10:00; other activities: 7:00-24:00
Number of mobile nodes	200
Fraction of nodes carrying data at start	0.1
Transmission range	100m
Infrastructure node inter-spacing	500m

Table V. Parameter settings of the simulation runs in the open simulation system.

velocity, change of direction, pause time, and flight length. Fitting, i.e., parameter estimation of the distribution functions is based on Maximum-likelihood fitting using the statistic tool R [39].

Velocity

The mapping of known probability distributions to the empirical cumulative density functions of velocity resulted in the selection of a Lognormal distribution for all activities. A similar approach using the Lognormal distribution to approximate velocity behavior is described in [6] for velocity values extracted from campus traces. The Lognormal distribution is given by Equation 1; the parameter settings are summarized in Table VI.

Empirical and fitted CDFs of the velocity characteristic are depicted in Figure 4(a). All curves show significant inclinations (significant number of trips) in both pedestrian and vehicular speeds corresponding to the mixed movement nature of the real-world observation. However, the pedestrian speeds of about $1 - 1.5$m/s are more frequent in the *Evening* (blue line), *Tourist* (cyan line), and *Shopping* (red line) activity. *Way to work* (green line) inhibited more vehicular sequences while in *Tourist* activities also travels over highways were captured resulting in higher speeds in vehicular cases.

$$f(x) = \frac{1}{x\sigma\sqrt{(2\pi)}} e^{-\frac{(log(x)-\mu)^2}{2\sigma^2}} \tag{1}$$

Change of Direction

Changes of direction intensities are approximated by an exponential distribution showing similar curve characteristics as the empirical distribution with the parameter setting given in Table VII detailing parameter λ of Equation 2. Note, that changes of direction to the right (0 to 180 degrees) and to the left (0 to -180 degrees) are mapped to a positive rotation value (0 to 180 degrees).
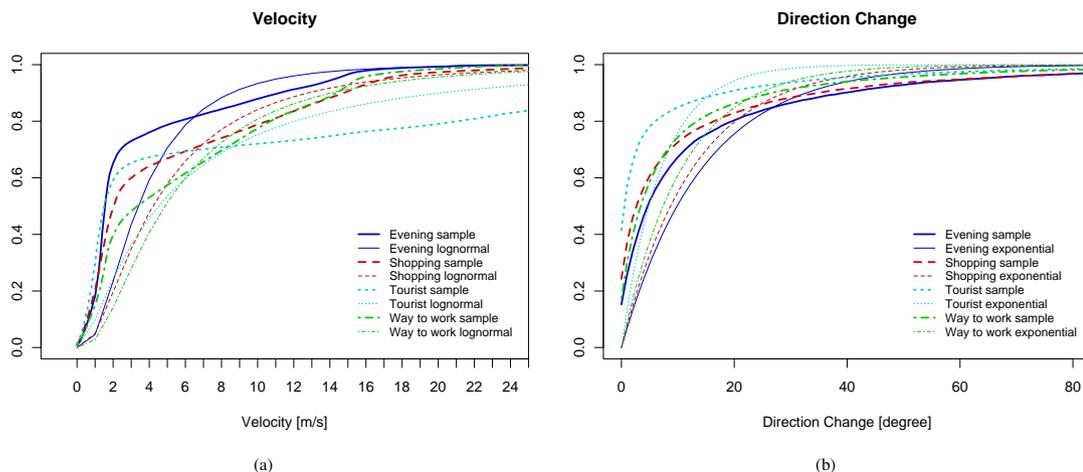
Wirel. Commun. Mob. Comput. 2012; **xx**:2–21

Prepared using wcmauth.cls   **13**

Figure 4. (a) Velocity: ECDFs and fitted lognormal CDFs and (b) Change of Direction: ECDFs and fitted exponential CDFs.

Movement activity	Distribution function	Parameter settings
Evening	Lognormal	$\mu = 1.216950991,\ \sigma = 0.722702962$
Shopping	Lognormal	$\mu = 1.434877973,\ \sigma = 0.868246447$
Tourist	Lognormal	$\mu = 1.506114822,\ \sigma = 1.166707076$
Way to work	Lognormal	$\mu = 1.582293212,\ \sigma = 0.828569221$

Table VI. Parameter setting for velocity distribution function.

$$f(x) = \lambda e^{-\lambda x} \tag{2}$$

Movement activity	Distribution function	Parameter settings
Evening	Exponential	$\lambda = 0.0703799465$
Shopping	Exponential	$\lambda = 0.0797647278$
Tourist	Exponential	$\lambda = 0.1433312136$
Way to work	Exponential	$\lambda = 0.094152855$

Table VII. Parameter setting for the distribution function of change of direction.

The empirical and fitted CDFs for change of direction are shown in Figure 4(b). The ECDF curves show similar behavior, only in the *Tourist* (cyan line) and *Way to work* (green line) activities, higher relative numbers of cases show smaller changes of direction (about 0.85 to 0.90 smaller than 20 degrees). While for *Way to work* movement, this result is expected, the unexpected long highway trips of *Tourist* activities are the reason for the results of these trips. For *Shopping* and *Evening* activities, pedestrian movement caused a higher variability in changes of direction.

Pause Time

Pause time experimental distribution results were approximated by Pareto and Lognormal distributions that showed similar results in terms of appropriateness for approximation. In compliance to related work, Pareto distribution fitting was chosen and is presented here (parameter settings given in Table VIII for Equation 3 as used in R [40]). In [9], a Pareto distribution is introduced to describe GPS traces investigated for the SLAW mobility model and in [6], a Lognormal distribution was mapped to the pause times found in WLAN data estimated based on extracted velocity. Both resulting insights are in-line with our observations.
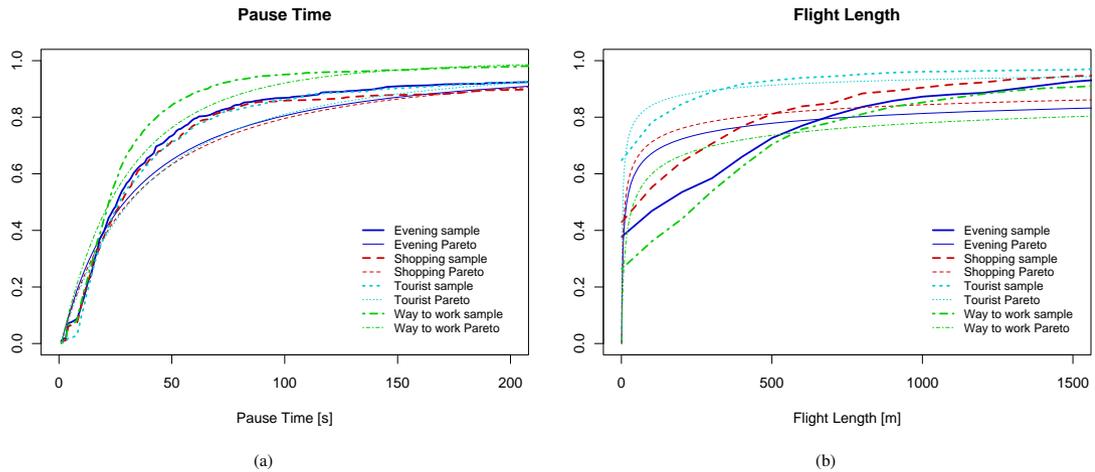
14

Wirel. Commun. Mob. Comput. 2012; **xx**:2–21

Prepared using wcmauth.cls

Figure 5. (a) Pause Time: EDCFs and fitted Pareto CDFs and (b) Flight Length: EDCFs and fitted Pareto CDFs.

$$f(x) = \alpha \frac{s^{\alpha}}{(x+s)^{\alpha+1}}, \tag{3}$$

where $x > s$. The fitted Pareto CDFs are visually compared to the empirical CDFs for pause time in Figure 5(a) (limited to a pause time of 200s to enhance readability). The *Way to work* activity showed a relative high number of shorter pause times (green line) while the other three activities showed similar pause times containing a significantly high fraction of about 0.1 to 0.2 longer than one minute.

Movement activity	Distribution function	Parameter settings
Evening	Pareto	$s = 41.3474534, \alpha = 1.3399386$
Shopping	Pareto	$s = 46.5458099, \alpha = 1.3921989$
Tourist	Pareto	$s = 64.5206558, \alpha = 1.7890021$
Way to work	Pareto	$s = 122.8639776, \alpha = 4.2903128$

Table VIII. Parameter setting for pause time distribution function.

Flight Length

In related work, flight length is often assumed to follow a power-law distribution [8]. Our empirical PDFs have been well approximated by both the Pareto and Lognormal distribution. Here, the fitted parameter settings of the Pareto distribution (Equation 3) are described by Table IX.

The ECDFs for flight length along with the fitted Pareto CDFs are depicted in Figure 5(b). The more pedestrian-like activities *Tourist* (cyan line) and *Shopping* (red line) showed higher occurrences of shorter flight lengths than the *Way to work* (green line) and *Evening* (blue line) activities, which showed quite similar behavior (slightly higher fraction of short flight lengths for *Evening* activities).

Movement activity	Distribution function	Parameter settings
Evening	Pareto	$s = 1.06479645, \alpha = 0.24506560$
Shopping	Pareto	$s = 0.94243061, \alpha = 0.26666162$
Tourist	Pareto	$s = 0.71566740, \alpha = 0.37313322$
Way to work	Pareto	$s = 3.18195439, \alpha = 0.26296255$

Table IX. Parameter setting for flight length distribution function.

Wirel. Commun. Mob. Comput. 2012; **xx**:2–21   **15**

Prepared using wcmauth.cls

Start Time and Mobility Range

The characteristics start time and mobility range show very particular distributions. Instead of trying to find appropriate mixture distributions with limited use in other studies, we depict only the empirical PDFs in Figure 6(a) and Figure 6(b). These EPFDs are further used in our simulation.

In Figure 6(a), the curves start when the first occurrences are encountered, e.g., at 5p.m. for *Evening* activities (blue line), in the morning for *Way to work* (green line) when the highest relative numbers have been observed, and considerably later – as expected – in the cases of *Shopping* (red line) and *Tourist* (cyan) showing also the lunch break when hardly any trips start. In particular for distinguishing daily routines that show temporal regularities, the start time is an important mobility characteristic.

The mobility range as visualized in Figure 6(b) (limited to 6000m for readability reasons) shows peaks at smaller ranges in case of *Tourist* or *Evening* activities (at about 1000m for *Tourist* and 3000m for *Evening*) but also larger ranges. The *Shopping* and *Way to work* activities resulted in multiple peaks at smaller and larger ranges reflecting the different distances traveled from home to work place or to a shop or mall. The large mobility ranges where unexpected in the *Shopping* activity case at first sight, but could be explained by traveling frequently to shopping malls when living in the outer districts of a city. Further, the long travels of tourists visiting sights outside the urban area are causing larger ranges.
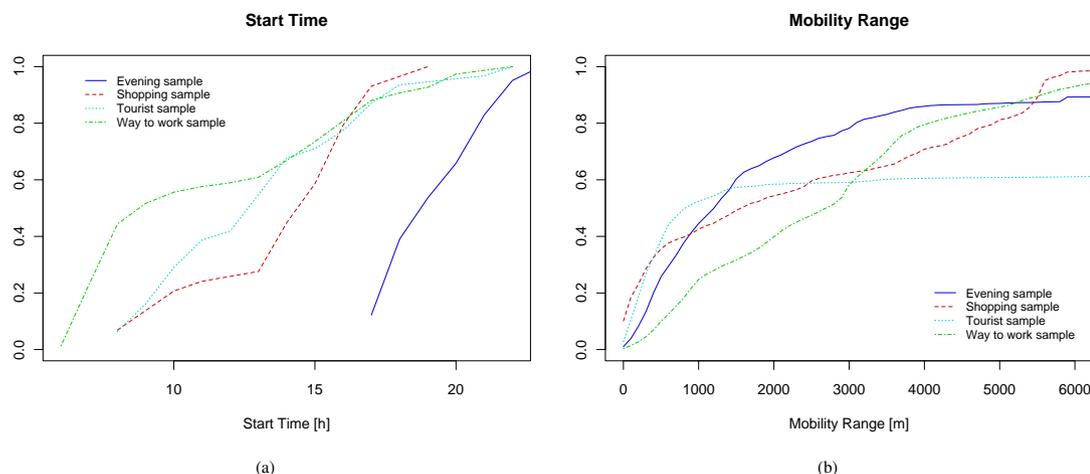

Figure 6. (a) Start Time ECDFs and (b) Mobility Range ECDFs.

6.4. Discussion of Simulation Results

We conducted five simulation runs for every movement activity. For every metric defined in Section 6.1, we discuss the median and visualize the first and third quartiles of the distribution of the five runs as box plots including 'whiskers', i.e., the extension of the box to the most extreme data points which are no more than 1.5 times the interquartile range from the box (minimum and maximum values in the plots). The interquartile range is a measure of the statistical dispersion (note, that the percentiles and the extreme values may collapse in the plots if the ranges are small). The relative frequencies of twelve equally sized intervals are summarized – except the last interval, which is an open interval containing the sum of all occurrences greater or equal the lower limit of the interval. The results should give first insights in potential correlations between movement activities and forwarding capabilities.

Contact Time and Inter-contact Time

The activities *Evening*, *Shopping*, and *Tourist* show quite similar distributions of contact times, while the *Way to work* activity shows more occurrences of contact times < 100s (median and mean at 0.69). Figure 7(a) visualizes the median of
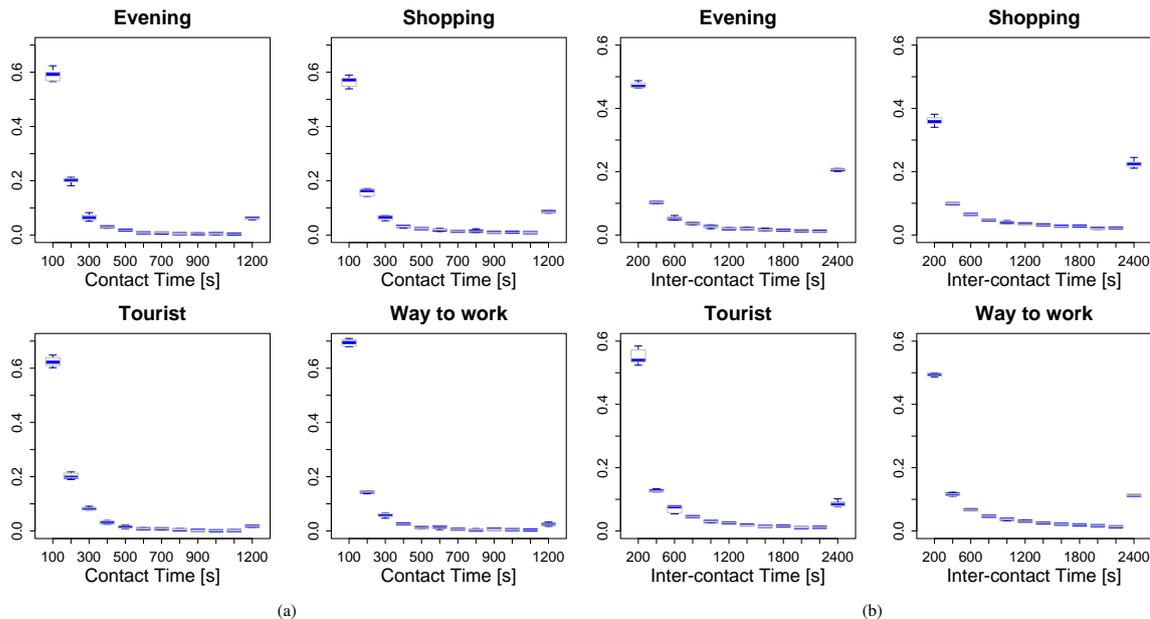
Figure 7. Median (blue horizontal line), first and third quartile (lower and upper bound of the box), and 'whiskers' (extensions of the box) of the empirical distributions of (a) contact time and (b) inter-contact time.

the five simulation runs. The high number of occurrences of short contact times for the *Way to work* activity corresponds to the high number of occurrences of short pause times while the other activities show quite similar pause time distributions (see Figure 5(a)).

The results of inter-contact time observations are depicted in Figure 7(b). This metric highly depends on the density of nodes, which is influenced by the mobility of the nodes, e.g., nodes stop moving when reaching their mobility range, and nodes move outside the observation area due to a large mobility range (in particular in combination with high velocities). All activities show high occurrence of short inter-contact times, e.g., the fraction of inter-contact times < 200s ranges from 0.36 (*Shopping*) to 0.54 (*Tourist*). In case of *Shopping*, also a higher number of occurrences of longer inter-contact times is observed (inter-contact time $\geq 2\,400$s, 0.23 of occurrences).

Forwarding Distance and Coverage (Availability of Information)

The forwarding distance and the coverage (availability of information in the observation area) describe the efficiency of dissemination via mobile nodes. Figure 8(a) shows the median of the forwarding distance and Figure 8(b) the median of the coverage over time.

The highest fraction of nodes that achieve a short forwarding distance of less than four stationary nodes is exhibited by *Tourist* trips. Here, a considerable number of nodes does not even reach one infrastructure node (forwarding distance is about 0 for a fraction of about 0.12; forwarding distance < 2 for a fraction of 0.36). The highest fraction of nodes showing higher number of contacts with stationary nodes can be found in the *Way to work* activity. The extreme cases *Tourist* and *Way to work* show highest numbers of occurrences of short flight lengths and long flight lengths respectively (see Figure 5(b)). While *Way to work* has a high fraction of moderate mobility ranges (see Figure 6(b)), the other activities show higher fractions of smaller and larger ranges (due to vehicular movement), the latter leads to moving out of the observed simulation area.

The mean and the median of the maximum coverage achieved by the different movement activities in the five different runs are: *Way to work* mean: 0.97 (median: 0.96), *Tourist* mean: 0.78 (median: 0.8), *Shopping* mean: 0.77 (median: 0.76), and *Evening* mean: 0.7 (median: 0.7). The low coverage values of *Evening* activities is due to a high proportion of longer
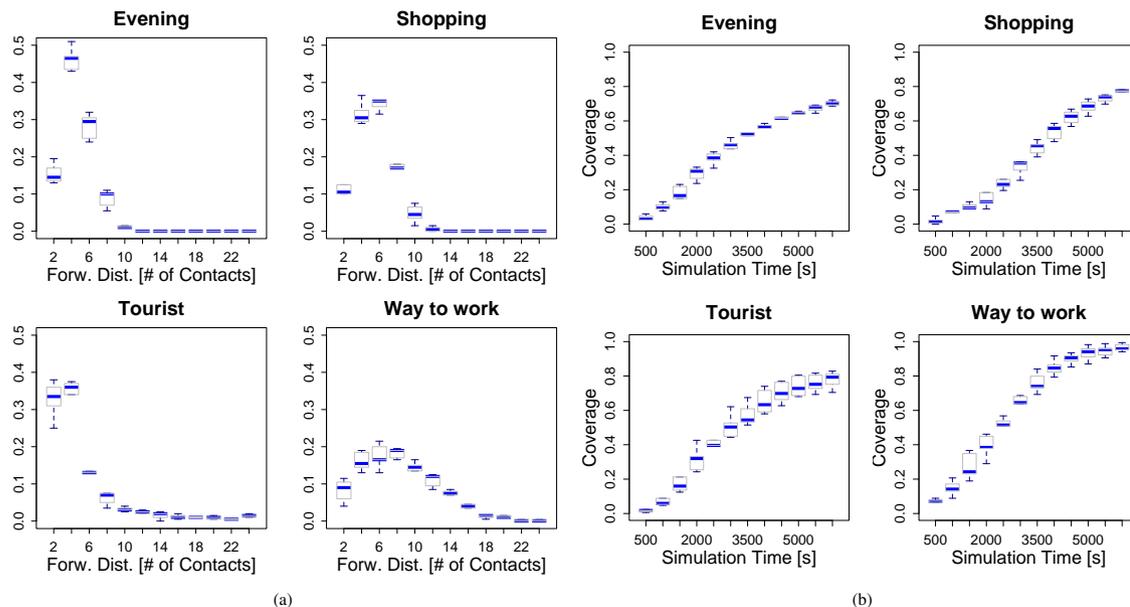
Figure 8. Median (blue horizontal line), first and third quartile (lower and upper bound of the box), and 'whiskers' (extensions of the box) of (a) the distribution of forwarding distances and (b) the coverage (availability of information) along the simulation time.

pause times (and shorter flight lengths). Further, the nodes' range allows them to move outside the observation area soon. Similar to the results in terms of the forwarding distance, *Way to work* activities achieve the highest coverage.

Predictability of Future Locations of a Linear Predictor

The predictability of future locations is of interest if route information is required by a forwarding algorithm. Figure 9 depicts the median of the prediction accuracy in the range 0.8 to 1.0 accuracy (truncated for readability reasons). It has to be mentioned, that the linear extrapolation shows a high prediction accuracy in all the simulation runs, which is a consequence of the prediction parameter settings. High velocities show effects when using the linear extrapolation prediction with fixed prediction time intervals. As a consequence, a lower fraction of prediction accuracies ≥ 0.95 is observed for the *Way to work* activity. The high fraction of long pause times of the *Evening* activity causes the high accuracies in this scenario.

As a concluding remark it has to be mentioned, that the predictor's logic and parameter settings should ideally be adapted to each movement activity (movement characteristics) to achieve a sufficiently high prediction accuracy.

7. CONCLUSION

We discussed and detailed commonly used movement characteristics based on a comprehensive literature survey and presented classification and opportunistic forwarding results. Primarily, the motivations for the experiments were to evaluate to which extent movement activities can be recognized based on the mobility characteristics defined and to investigate the impact of movement activities on forwarding metrics in opportunistic networks.

By using a Naïve Bayes classifier, daily movement activity use cases were classified based on an artificial training set consisting of emulated movement activities and on a real-world training and test set consisting of 252 trips tracked by volunteers using GPS devices for one month. An overall promising classification success rate of 0.81 could be achieved for the real-world data set.
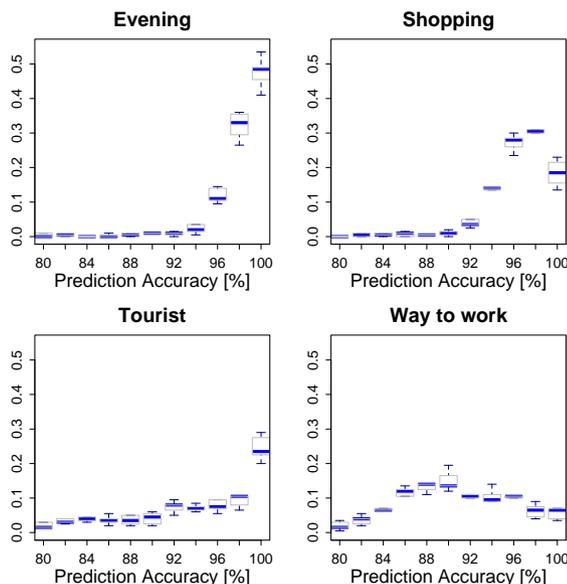
18   *Wirel. Commun. Mob. Comput.* 2012; **xx**:2–21

Prepared using wcmauth.cls

Figure 9. Median (blue horizontal line), first and third quartile (lower and upper bound of the box), and 'whiskers' (extensions of the box) of prediction accuracy distribution.

The investigation of forwarding effects of nodes moving along the different daily activities introduced was performed using agent-based simulation. The simulation results showed that the activity resulting in longer pause time fractions also resulted in higher fractions of longer contact times. The additional metric inter-contact time was influenced by the node density. The forwarding distance, i.e., the number of areas visited by the mobile node and the resulting coverage of the area in terms of availability of information are higher for movement activities showing longer flight length and larger mobility range occurrences. When using mobility ranges to introduce end points of trips, higher ranges led to the situation that nodes moved outside the area of interest which effected forwarding metrics as well.

Finally, when a route prediction of a node is required by the opportunistic forwarding algorithm general predictability is an important factor. We investigated the accuracy of simple linear extrapolation based prediction of future locations and observed differences between the movement activities. As fixed prediction intervals were used in our prediction setting, higher velocities were causes for lower accuracies.

Based on the insights of our study on mobility activities and forwarding characteristics, we plan to extend our analysis and, finally, to provide opportunistic forwarding strategies based on movement activities and characteristics of mobile nodes.

ACKNOWLEDGEMENTS

REFERENCES

1. Bettstetter C, Resta G, Santi P. The Node Distribution of the Random Waypoint Mobility Model for Wireless Ad Hoc Networks. *IEEE Transactions on Mobile Computing* 2003; **2**(3):257–269, doi:10.1109/TMC.2003.1233531.

2. CRAWDAD: A Community Resource for Archiving Wireless Data At Dartmouth. http://crawdad.cs.dartmouth.edu/.

3. Hummel K, Hess A. Movement Activity Estimation for Opportunistic Networking Based on Urban Mobility Traces. *IFIP Wireless Days 2010*, 2010; 303–316, doi:10.1109/WD.2010.5657723.

4. Hummel K, Hess A. Estimating Human Movement Activities for Opportunistic Networking: A Study of Movement Features. *WOWMOM 2011: 2011 IEEE Int. Symp. on a World of Wireless, Mobile and Multimedia Networks*, 2011; 1–7, doi:10.1109/WoWMoM.2011.5986468.

5. Jardosh A, Belding-Royer EM, Almeroth KC, Suri S. Towards Realistic Mobility Models for Mobile Ad Hoc Networks. *MobiCom '03: Int. Conf. on Mobile Computing and Networking*, ACM: New York, NY, USA, 2003; 217–229, doi:10.1145/938985.939008.

6. Kim M, Kotz D, Kim S. Extracting a Mobility Model from Real User Traces. *INFOCOM '06: IEEE Int. Conf. on Computer Communications*, 2006; 1–13, doi:10.1109/INFOCOM.2006.173.

7. Hsu W, Spyropoulos T, Psounis K, Helmy A. Modeling Time-Variant User Mobility in Wireless Mobile Networks. *INFOCOM '07: IEEE Int. Conf. on Computer Communications*, 2007; 758–766, doi:10.1109/INFCOM.2007.94.

8. Gonzalez MC, Hidalgo CA, Barabasi AL. Understanding Individual Human Mobility Patterns. *Nature* 2008; **453**(7196):779–782, doi:10.1038/nature06958.

9. Lee K, Hong S, Kim SJ, Rhee I, Chong S. SLAW: A Mobility Model for Human Walks. *INFOCOM '09: IEEE Int. Conf. on Computer Communications*, 2009; 855–863, doi:10.1109/INFCOM.2009.5061995.

10. Ekman F, Keränen A, Karvo J, Ott J. Working Day Movement Model. *MobilityModels '08: ACM SIGMOBILE Workshop on Mobility Models*, ACM: New York, NY, USA, 2008; 33–40, doi:10.1145/1374688.1374695.

11. Kim J, Sridhara V, Bohacek S. Realistic Mobility Simulation of Urban Mesh Networks. *Ad Hoc Networks* 2009; **7**(2):411–430, doi:10.1016/j.adhoc.2008.04.008.

12. Liao L, Fox D, Kautz H. Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields. *International Journal of Robotics Research* 2007; **26**(1):119–134, doi:10.1177/0278364907073775.

13. Phithakkitnukoon S, Horanont T, Lorenzo GD, Shibasaki R, Ratti C. Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data. *Int. Conf. on Human Behavior Understanding (HBU'10)*, 2010; 14–25.

14. Choujaa D, Dulay N. Predicting Human Behaviour from Selected Mobile Phone Data Points. *12th ACM Int. Conf. on Ubiquitous Computing*, Ubicomp '10, ACM: New York, NY, USA, 2010; 105–108, doi:10.1145/1864349.1864368.

15. Farrahi K, Gatica-Perez D. What Did You Do Today? Discovering Daily Routines from Large-Scale Mobile Data. *16th ACM International Conference on Multimedia (MM '08)*, ACM: New York, NY, USA, 2008; 849–852, doi: 10.1145/1459359.1459503.

16. Reality Mining Dataset. http://reality.media.mit.edu/download.php.

17. Helgason O, Kouyoumdjieva ST, Karlsson G. Does Mobility Matter? *WONS'10: 7th International Conference on Wireless On-demand Network Systems and Services*, 2010; 9–16.

18. Lindgren A, Doria A, Schelén O. Probabilistic Routing in Intermittently Connected Networks. *SIGMOBILE Mobile Computing and Communications Review* 2003; **7**(3):19–20, doi:10.1145/961268.961272.

19. Liu C, Wu J. An Optimal Probabilistic Forwarding Protocol in Delay Tolerant Networks. *10th ACM Int. Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '09)*, ACM: New York, NY, USA, 2009; 105–114, doi: 10.1145/1530748.1530763.

20. Leguay J, Friedman T, Conan V. Evaluating Mobility Pattern Space Routing for DTNs. *INFOCOM '06: IEEE Int. Conf. on Computer Communications*, 2006; 1–10, doi:10.1109/INFOCOM.2006.299.

21. Burns B, Brock O, Levine B. MV Routing and Capacity Building in Disruption Tolerant Networks. *INFOCOM '05. IEEE Int. Conf. on Computer Communications*, 2005; 398–408, doi:10.1109/INFCOM.2005.1497909.

20

Wirel. Commun. Mob. Comput. 2012; **xx**:2–21

Prepared using wcmauth.cls

22. Kaveevivitchai S, Ochiai H, Esaki H. Message Deletion and Mobility Patterns for Efficient Message Delivery in DTNs. *IEEE Int. Conf. on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2010; 760 –763, doi:10.1109/PERCOMW.2010.5470535.

23. Bromage MK, Koshimoto JT, Obraczka K. TAROT: Trajectory-Assisted Routing for Intermittently Connected Networks. *CHANTS '09: 4th ACM Workshop on Challenged Networks*, ACM: New York, NY, USA, 2009; 9–18, doi:10.1145/1614222.1614225.

24. Zang H, Bolot JC. Mining Call and Mobility Data to Improve Paging Efficiency in Cellular Networks. *ACM Int. Conf. on Mobile Computing and Networking (MobiCom '07)*, ACM: New York, NY, USA, 2007; 123–134, doi: 10.1145/1287853.1287868.

25. Wu HK, Jin MH, Horng JT, Ke CY. Personal Paging Area Design Based on Mobile's Moving Behaviors. *Conf. of the IEEE Computer and Communications Societies (INFOCOM '01)*, 2001; 21–30, doi:10.1109/INFCOM.2001.916683.

26. Balazinska M, Castro P. Characterizing Mobility and Network Usage in a Corporate Wireless Local-area Network. *MobiSys '03: International Conference on Mobile Systems, Applications, and Services*, 2003; 303–316, doi: 10.1145/1066116.1066127.

27. Tuduce C, Gross T. A Mobility Model Based on WLAN Traces and its Validation. *INFOCOM '05: IEEE Int. Conf. on Computer Communications*, 2005; 664–674, doi:10.1109/INFCOM.2005.1497932.

28. Kim J, Helmy A. The Challenges of Accurate Mobility Prediction for Ultra Mobile Users. *SIGMOBILE Mobile Computing and Communications Review* 2009; **13**(3):58–61, doi:10.1145/1710130.1710143.

29. Rhee I, Shin M, Hong S, Lee K, Chong S. On the Levy-Walk Nature of Human Mobility. *INFOCOM '08: IEEE Int. Conf. on Computer Communications*, 2008; 924–932, doi:10.1109/INFOCOM.2008.145.

30. Yoon J, Noble BD, Liu M, Kim M. Building Realistic Mobility Models from Coarse-Grained Traces. *MobiSys '06: Int. Conf. on Mobile Systems, Applications and Services*, ACM: New York, NY, USA, 2006; 177–190, doi: 10.1145/1134680.1134699.

31. Karagiannis T, Le Boudec JY, Vojnović M. Power Law and Exponential Decay of Inter Contact Times Between Mobile Devices. *MobiCom '07: ACM Int. Conf. on Mobile Computing and Networking*, 2007; 183–194, doi: 10.1145/1287853.1287875.

32. Chaintreau A, Hui P, Crowcroft J, Diot C, Gass R, Scott J. Impact of Human Mobility on Opportunistic Forwarding Algorithms. *IEEE Transactions on Mobile Computing* 2007; **6**:606–620, doi:10.1109/TMC.2007.1060.

33. Hui P, Chaintreau A, Scott J, Gass R, Crowcroft J, Diot C. Pocket Switched Networks and Human Mobility in Conference Environments. *ACM SIGCOMM Workshop on Delay-tolerant Networking*, WDTN '05, ACM: New York, NY, USA, 2005; 244–251, doi:10.1145/1080139.1080142.

34. Jindal A, Psounis K. Fundamental Mobility Properties for Realistic Performance Analysis of Intermittently Connected Mobile Networks. *PerCom Workshops '07: IEEE Int. Conf. on Pervasive Computing and Communications Workshops*, 2007; 59–64, doi:10.1109/PERCOMW.2007.52.

35. Srinivasan V, Motani M, Ooi WT. Analysis and Implications of Student Contact Patterns Derived from Campus Schedules. *MobiCom '06: Int. Conf. on Mobile Computing and Networking*, ACM: New York, NY, USA, 2006; 86–97, doi:10.1145/1161089.1161100.

36. Breunig MM, Kriegel HP, Ng RT, Sander J. LOF: Identifying Density-based Local Outliers. *SIGMOD Record* 2000; **29**(2):93–104, doi:10.1145/335191.335388.

37. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. The WEKA Data Mining Software: An Update. *SIGKDD Explorations* November 2009; **11**(1):10–18, doi:10.1145/1656274.1656278.

38. XJ Technologies. AnyLogic Simulation Software. http://www.xjtek.com/anylogic/.

39. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 2010. URL http://www.R-project.org, ISBN 3-900051-07-0.

40. R Package Actuar. http://cran.r-project.org/web/packages/actuar/actuar.pdf.

Prepared using wcmauth.cls