An Analysis of Human Mobility using Real Traces

Tiago S. Azevedo, Rafael L. Bezerra, Carlos A. V. Campos and Luís F. M. de Moraes
Department of Systems Engineering and Computer Science, Federal University of Rio de Janeiro
P.O. Box: 68.511, CEP: 21.941-972, Rio de Janeiro, RJ, Brazil
E-mails: {tiago, bezerra, beto, moraes}@ravel.ufrj.br

***Abstract*—We present a new analysis for human mobility through real traces captured in a recreational environment. The goal of this analysis is to investigate the motion components, in both a qualitative and quantitative way, and thus, to get a better knowledge of pedestrian mobility behavior. Moreover, some results based upon human mobility captured in a real scenario with GPS equipment are presented. In these captured data, we verified that the speed and acceleration components follow a Normal distribution, while the direction angle variation component and the pause time measure follow a Lognormal distribution. Finally, we show that velocity and direction angle change components of captured scenario have a temporal dependence unlike the random mobility models.[1]**

***Index Terms*—Mobile networks, human mobility and measurements.**

I. Introduction

Applying mobility traces generated by real measurements is becoming increasingly necessary in the performance evaluation of Mobile Ad hoc Networks (MANETs). In the last few years, real experiments and testbeds have been proposed in order to obtain more accurate and representative results for mobility in wireless networks than those obtained by simulations and analytical modeling – [1], [2], [3], [4]. However, those studies have been conducted without the goal of capturing the real movement of devices in non-structured networks. The acquisition of real data is important because it allows investigation of movement components, in order to answer the following questions:

1) How is the detailed behavior of human mobility in real scenarios?
2) Are there specific characteristics in a particular behavior? What are they?

Answering the above questions is an important step to gain insight about human mobility, so that we may be able to validate models associated to real data and to allow the development of new, more realistic models related to specific scenarios. In addition, these are open problems in this context, which have been discussed by the scientific community, as it can be seen for example in the CRAWDAD [5] and Haggle [6] projects.

In this context, this work aims to investigate the behavior of pedestrian movement through real measurements. The main motivation for this investigation is to have detailed knowledge of human mobility behavior in real scenarios, for the validation of existing models and also to allow the development of new models. In this context, data from real movement is captured and used in an investigation aiming to present a picture of human mobility characterization. To investigate the mobility a methodology will be proposed, through which the movement components will be obtained and measures and statistics of interest about these components will be calculated. These measures will be presented for a set of real acquired data.

The rest of this paper is organized as follows. In Section II, we describe the main published works related with characterization of mobility for MANETs. Our proposal to analyze the movement components is presented in Section III. Section IV presents a case of study of an application of our proposal and some obtained results are described. Finally, Section V concludes the paper and presents some directions for future works.

II. Related Work

The effort applied by the scientific community on the development of new mobility models is increasing, since mobile technologies in wireless networks are constantly changing. Due to these developments, several mobility models were proposed and they will be presented below.

The most used mobility models are: random-walk, Brownian and random waypoint (RWP) models. The random-walk and Brownian models have been proposed to represent the movement of matter and the living beings, and later was used for evaluation of mobile networks and MANETs, as described in [7]. The RWP model, on the other hand, is a variation of random-walk model. This model is the most used today, however, some years ago some undesirable characteristics were discovered in these models, demonstrating a non-realistic behavior as shown in [8], [9], [10]. Other *random models* were proposed, as the smooth random mobility model [11], a Markovian model presented in [12] and the Levy-walk model [13], [14]. These models try to reduce the non-realistic characteristics of the RWP model, as random choice of velocity or movement direction. In addition, some obstacle-based models were proposed in [15], [16].

Recently, several realistic models based on infrastructure-based networks and on real scenarios have been proposed. A mobility model extracted from real user traces of WLANs is presented in [1]. This model estimates the physical location of users from traces of device association with access points (APs) in a WLAN. Besides, it was discovered that the velocity and pause time of movement follow a Lognormal distribution.

[1]This work has been supported by CAPES, CNPq and FAPERJ. Authors in alphabetical order of last names.

In [17], it is proposed a realistic mobility model combining coarse-grained wireless traces with a map of the measured space and association data between WiFi clients and APs. The obtained movement through association data from clients with APs has some undesirable characteristics, as ”ping-pong effect”, ”location problems of APs”, ”physical micro-variations” and ”erroneous reproducibility”, as described in [18]. Furthermore, depending on APs density in the scenario, the accuracy of movements and positions obtained can be affected. In such case, to obtain fine-grained data is essential to decrease the effect of these characteristics. This is one goal of our work that will be presented in Section III.

III. Mobility Analysis Proposal

Based on [8], [19] and motivated by [1], [17], [20], we present an alternative proposal for mobility analysis of people and vehicles. This proposal is based in a quantitative and qualitative detailed analysis of movement components and the dependence relationship between these components. This analysis will be applied to the context of the physical space in which people and vehicles are inserted. Thus, the movement components (velocity, acceleration and direction angle change) and pause time measure can be defined as:

- *Velocity* is the scalar velocity of a device (user) at a given time interval, and can be calculated as the ration between the displacement of the device, in relation to a given time interval, and this time interval.
- *Acceleration* is the scalar velocity variation rate of the device (user) at some time interval.
- *Direction angle change* is defined as being the angle of movement direction change of a device at a specific moment. Given the current, last and next position of the device, the angle can be computed using the Law of Cosines.
- *Pause time* is the interval of consecutive instants that a device remains stopped.

After the description of mobility components and measures used in our analysis, we will explain about the methodology used for our analysis of mobility.

A. Methodology

This methodology is composed by a set of procedures arranged in stages of development. Each stage has a high dependence on the previous stages, thus execution of these stages should be in chronological order. However, the accuracy level of the achieved results will depend on the values chosen for parameters related to each stage of such methodology.

The stages of this methodology are: scenario definitions, data capture and processing, data statistical analysis and identification of behavior patterns. These stages will be described in the following.

1) Scenario Definitions: This methodology is of general use and wide range, however, it is important to specify scenarios in which motion will be investigated. Therefore, it must be taken into consideration the following parameters for each scenario: type of scenario, area size, number of devices and network density, which is given by the ratio between the number of devices and area size of the wireless network.

For instance, we have different types of scenarios for wireless mobile networks: pedestrians and vehicles moving into urban centers or in rural areas, traffic of vehicles on a highway, operations of search and rescue, military applications, people walking inside buildings, among others. All these scenarios can be considered, however, restrictions on the use of these scenarios can be imposed. Then, the stage of data collection and processing will be described to treat those restrictions.

2) Data Capture and Processing: After defining the scenario, it is important to explain the procedures for data capture and processing. For this, independent samples must be chosen, in which each sample should represent the movement of different users.

To obtain data for this analysis it was necessary to use a location system. One suitable location system used is the GPS (Global Positioning System), which enables the capture of the device’s real position. However, it is necessary to mention some desirable characteristics that the GPS device must have so that the space-time analysis can be performed correctly. The procedure for calculating the device location coordinates may contain errors generated by variation in accuracy level of the GPS device used. The range of such errors may vary from centimeters to tens of meters, therefore, the lower the error, the more accurate the device location coordinates will be.

Once data is captured, the filtering of samples will be required, which is the choice of data effectively used in the characterization. Thus, the interval time between samples and its components should be considered. Also, the samples inconsistent data should be removed, as recording errors in files and values discrepancies caused by errors in location calculation. This errors may be caused by multi-path problem of the signal received by the GPS receiver when around buildings and constructions, and interference from trees and objects located in the line of sight of the satellite constellation. Thus, we recommend the use of GPS receivers with a good precision. The greater the precision of GPS the fewer errors will occur and therefore more accurate results will be achieved, allowing greater reliability of data.

After data capture and filtering, we have the statistical analysis stage.

3) Data Statistical Analysis: In order to observe and have more quantitative and qualitative knowledge of data, a statistical treatment should be applied in the movement components and measures of interest captured. This treatment includes the calculation several statistical measures, such as: average, standard deviation, variance, coefficient of variation, minimum value, maximum value, autocorrelation of components, empirical probability distribution function (epdf) and empirical cumulative distribution function (ecdf) of these measures and mobility components.

4) Identification of Behavior Patterns: The occurrence of specific behaviors in the captured movement will be investigated in this stage through the analysis of data obtained in the previous stages. This stage aims to detect mobility patterns or important information about mobility through the captured traces. In case that the results obtained in the previous stages are unsatisfactory, that is, they present some different characteristics than expected, then a re-adjustment in the configuration parameters used in stage 3 or even in stage 2 will be needed. These re-adjustments should be made until that the obtained results are as expected. Furthermore, this stage searches answers for the questions presented in Section I and it is important to enable a more adequate representation of human mobility on real scenarios.

B. Considerations

From the description and specification of these stages, we believe that this methodology will enable the quantification and qualification tasks of components and the measure of interest to be taken more accurately and correctly, as it is recommended in [20], thus permitting the analysis of wireless devices mobility to be more reliable.

As searched results with the use of this characterization methodology, we had:

- The possibility of obtaining deep and detailed information about mobility components captured in real scenarios. This can be realized with synthetic mobility models too;
- To have knowledge of important information for the development of new mobility models containing more realistic features.

IV. Applying the Analysis Proposed

How it was described in Section II, there are few papers about human mobility captured in the real scenarios. In this context, an investigation of component and measure of pedestrian mobility through the analysis proposed is presented in this section. Thus, the experiment description, the data treatment and the achieved results will be described as follow.

A. Experiment Description

The experiments took place at Quinta da Boa Vista's Park localized at Rio de Janeiro City. The park has many trees, lakes and caves, and still holds the City Zoo and the National Museum. This park was chosen because it is conducive to collecting GPS readings. The data collection occurred from January up to May 2008. We used the Trimble GPS Geo XM handheld receiver with high precision (sub-metric location error) for data collection. The GPS device works only with line of sight of satellites, so the data collection should be in an open area and a satellite planning should be made. According to the planning, the best time to do the collections was from 9 am to 4 pm. However, it was observed that in certain areas the GPS device showed instability even with many satellites. That is, the stability of the device depends on the positioning of satellites. Additionally, the sky should be with few clouds for a better functioning of GPS, which did not happen some times.

In order to capture the data, approximately 120 people walking by the Park were randomly chosen to participate on the research as volunteers. The time of each trace was from 300 to around 1300 seconds, sampled at each one second. From 120 traces were chosen 100 traces that had at least 600 seconds of time. Moreover, the time remnant from 600 seconds of 100 traces chosen was not considered in our analysis. Therefore, we analyze 100 traces containing up to 600 seconds of time. Figure 1 shows GPS trace samples from the park. In this figure, we can observe that the displacement of people in this park was defined by the trails and streets, but not randomly. Moreover, we can see that people walking through the same places at different times and with a high spatial regularity. This was also recently identified in movement of mobile phone users investigated in [4].

Figure 1. Samples GPS traces from the Quinta da Boa Vista's Park.

B. Data Treatment

After capturing the real traces, a differential correction was applied on the data to improve the accuracy of the GPS measurements. The software GPS Pathfinder Office [21] was used to do these correction. Therefore, the second stage of methodology will be applied for screening and filtering the data. This stage is necessary because the recording errors in files and values discrepancies caused by errors in location calculation by the GPS equipment. Thus, displacements upper to 2.5 meters per second were not considered.

Once the errors are removed from the traces, the statistical analysis can be done in the metrics of interest (described in Section III-A3). The achieved results from the statistical analysis will be presented in what follow.

C. Achieved Results

Some results about the captured movement were obtained through the proposed methodology and will be presented along this section. The value of components: velocity, acceleration, direction angle change and pause time were extracted from experiments through the methodology presented in Section III-A. Thus, the average, variance, standard deviation, coefficient of variation, minimum, maximum and number of samples of these components were calculated and will be shown in Table I. In this table, we can see that the mean value of velocity, direction angle change and pause time movement components was 1.1 m/s, 34.3 degree and 3.6 seconds, respectively. These values are important results in context to representation of human mobility by simulation of this scenario.

Table I
MEASURE FOR THE EACH COMPONENT OF MOBILITY

Measure	Vel. (m/s)	Accel. (m/s^2)	Dir. ang. change $(^\circ)$	Pause time (s)
Average	1.127	0.0004	34.3313	3.6227
Variance	0.2835	0.0473	1,854.5554	37.0564
Std. Deviation	0.5324	0.2175	43.0645	6.0874
Coef. of Variation	0.4724	576.0821	1.2544	1.6803
Minimum	0.007	-1.19796	0.000089	1
Maximum	2.499921	1.185976	179.9868	106
N. of Samples	39,617	39,617	39,617	1,092

As mentioned in [2], [3], [9], [19] it is important to investigate the characteristics of distributions of each mobility component. Thus, the empirical probability distribution function (epdf) and empirical cumulative distribution function (ecdf) for each component of real data were computed through the methodology presented in Section III-A. In Figure 2, we present the epdf of velocity component of real data. Besides, we can observe that the curve of velocity distribution has a likeness with the Normal distribution, except for the initial part of this curve. This part has a concentration of values due to the failure of GPS in capture of pauses on human movement, because instead of GPS to capture velocities equal to zero, it captures positive values. Therefore, a large amount of captured smaller values instead to zero leads to this concentration.

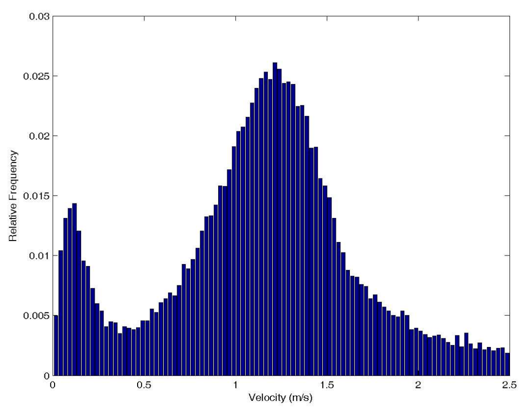

Figure 2. Empirical probability distribution function (epdf) of velocity component.

In order to compare the real data to these known probability distribution functions, we use the distributions Exponential, Gamma, Lognormal, Normal and Weibull. The Maximum Likelihood Estimation (MLE) method, implemented by MATLAB software, was used to choose the parameters of each distribution fitting to the real data. To verify the distribution that best fits the real data two methods were used: (i) - Mean Square Error (MSE) and (ii) - Kolmogorov-Smirnov (K-S) test. Details about these methods can be found in [22]. Thus, in Figure 3 we show the ecdf of velocity component of real data and the Exponential, Gamma, Lognormal, Normal, Weibull distribution functions adjusted through these data. We can observe that the epdf of velocity component is more approximated to Normal distribution. Furthermore, we verify that the Exponential, Gamma and Lognormal cumulative distribution functions adjusted to the real data containing values upper to 2.5, but these values were not shown in Figure 3.

Moreover, the MSE and K-S test were computed from distributions to real data described in Table II, and also the Normal distribution was identified as better fitting to velocity component of real data, because of its smaller MSE and K-S values. Thus, from the obtained results we can conclude that the velocity component of real data follows a Normal distribution function.

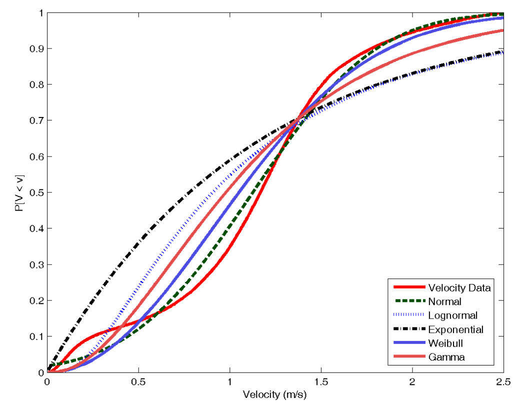

Figure 3. Empirical cumulative distribution function (ecdf) of velocity component and a comparison with Exponential, Lognormal, Normal, Gamma and Weibull cumulative distribution functions.

Table II
MSE AND K-S TEST BETWEEN REAL MOBILITY COMPONENTS AND SOME PROBABILITY DISTRIBUTION FUNCTIONS.

MSE					
Measure	Normal	Lognormal	Exponential	Weibull	Gamma
Velocity	0.00114	0.015743	0.025666	0.004319	0.009347
Acceleration	0.003768				
Dir. ang. change	0.017592	0.000286	0.007422	0.000953	0.001783
Pause time	0.014204	0.004299	0.005969	0.005414	0.006212
K-S test					
Measure	Normal	Lognormal	Exponential	Weibull	Gamma
Velocity	0.059284	0.215795	0.280499	0.118627	0.170827
Acceleration	0.093359				
Dir. ang. change	0.21509	0.045648	0.137508	0.054586	0.075204
Pause time	0.33322	0.278784	0.241215	0.270474	0.254035

With the purpose of analyzing the acceleration (velocity variation) of real data, the epdf behavior of acceleration component is shown in Figure 4.

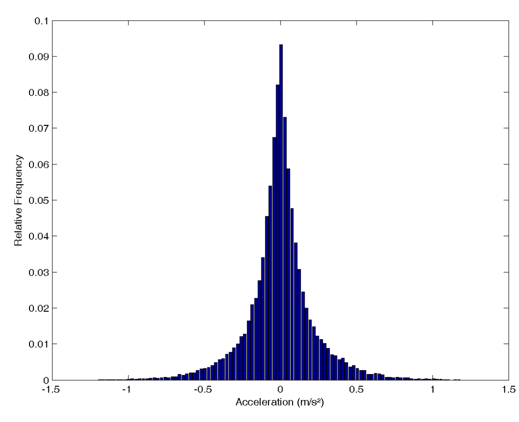

Figure 4. Empirical probability distribution function (epdf) of acceleration component.

Furthermore, we present in Figure 5 an adjustment of acceleration ecdf with the Normal distribution, which is the only distribution that assumes negative values. Through a visual analysis and by small MSE and K-S test achieved, we can verified that this distribution fits to acceleration curve of real data.

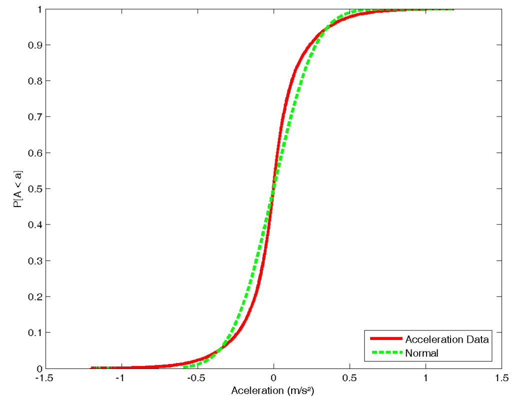

Figure 5. Empirical cumulative distribution function (ecdf) of acceleration component and a comparison with the Normal cumulative distrib. function.

The direction angle change is important to investigate how the movement direction is behaving. Unlike the *zig-zag behavior* of the RWP model, the human movement presents a smooth variation in direction angle change. This can be seen in the tracking of movements shown in Figure 1, as well in the direction angle change distribution of real movement obtained from the experiments and shown on Figure 6. In this figure, we can observe that most of the angle values are between 0 and 45 degrees and with an average value of 34.33° and otherwise the direction angle change of RWP model that have a variation uniform between 0 and 180°.

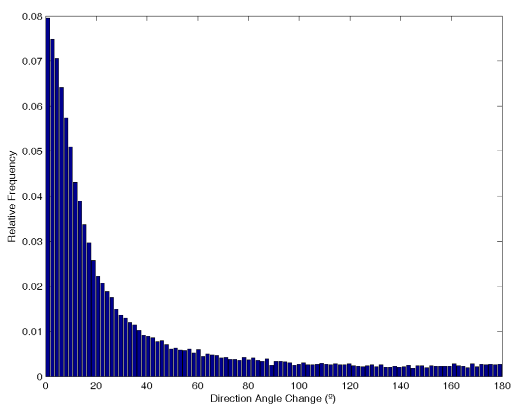

Figure 6. Empirical probability distribution function (epdf) of direction angle change component.

Besides, we compared the direction angle change of human mobility to the known distribution functions (see Figure 7). By the visual analysis, we can identify that the Lognormal distribution function has the better fit to real data. This can be verified by Table II, where the smaller value of the MSE and K-S tests was of the Lognormal distribution function. Furthermore, we verify that the Gamma, Lognormal and Weibull cumulative distribution functions adjusted with real data containing values upper to 180, but these values were not shown in Figure 7.

The pause time is another measure of interest to be observed in mobility, since it affects the performance of MANET protocols, as seen in [1], [13]. Thus, in Figure 8 we present the epdf of pause time of obtained real data. Moreover, we can observe that the mean value of pause time was approximately 3.6 seconds and most of the values were up to 10 seconds. However, there were a few pauses with great length.

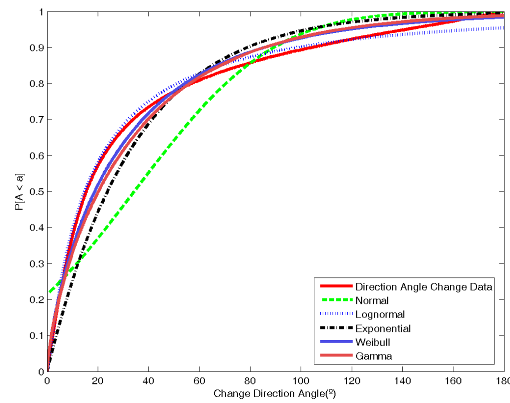

Figure 7. Empirical cumulative distribution function (ecdf) of direction angle change component and a comparison with Exponential, Gamma, Lognormal, Normal and Weibull cumulative distribution functions.

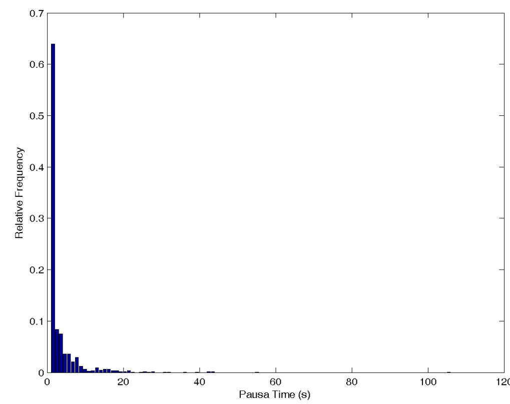

Figure 8. Empirical probability distribution function (epdf) of pause time measure.

In Figure 9, we present the ecdf of pause time of real data and some distribution functions. In addition, we can observe through the MSE, that the Lognormal distribution has the better fit to real data. Similar observation were also found in [1], [2], [3], [13], [14]. However, by the K-S test the ecdf of pause time follows an Exponential distribution. Thus, this measure should be better investigated.

With the purpose of investigating the occurrence of temporal dependencies in the movement components, we present the autocorrelation of each movement component in Figure 10. In this figure, we can observe that the velocity has a high correlation, mainly, in the initial lags values (1-10). Thus, we can say that velocity has a high temporal dependency. It happens also with the angle component, since the initial lags values have an autocorrelation between 0.5 and 0.3. However, it does not happen with the acceleration, where the initial lags values have negative autocorrelation and the remaining lags values are around zero. Therefore, we can say that the acceleration has no temporal dependence.

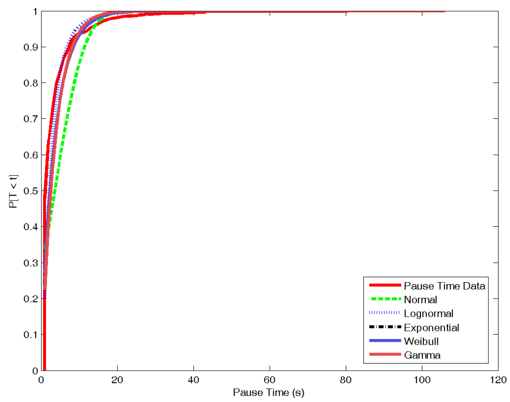

Figure 9. Empirical cumulative distribution function (ecdf) of pause time measure and a comparison with Exponential, Gamma, Lognormal, Normal and Weibull cumulative distribution functions.

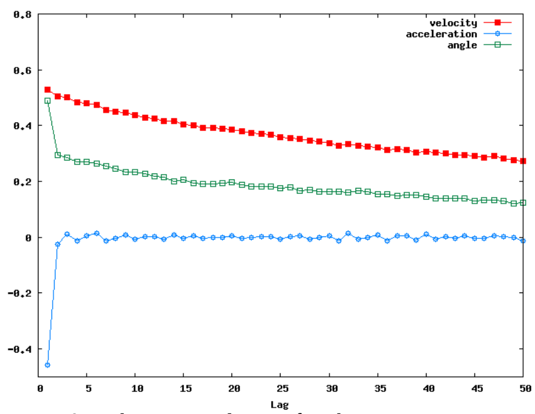

Figure 10. The autocorrelation of each movement component.

From the achieved results, we can see that the human movement in the scenario investigated presented a smooth variation in the velocity and direction angle change, which allows us to say that this movement is smooth and without abrupt changes, unlike the behavior of random mobility models as the RWP model.

V. Conclusions

In this paper, we investigate the human mobility based on a detailed and careful analysis of pedestrian movement components into a specific environment. Thus, several measures of interest were used to obtain more detailed information over this mobility.

A case study was realized in which individual human mobility was captured in a park. From the proposed methodology movement components of real data were obtained. The empirical distribution of these components were computed and compared to the Exponential, Gamma, Lognormal, Normal and Weibull distribution functions. Besides, we identify that the behavior of velocity and acceleration components follow a Normal distribution, the direction angle change component and the pause time measure are better represented to Lognormal distribution function. Thus, these information should be considered in the representation of these scenarios by mobility models.

As future works, we want to evaluate the accuracy of several mobility models with the real data obtained in this work, and to evaluate the impact of this human mobility in the performance of MANETs.

References

[1] M. Kim, D. Kotz, and S. Kim, "Extracting a Mobility Model from Real User Traces," in *Proc. of the IEEE INFOCOM'06*, Barcelona, Spain, April 2006, pp. 1–13.

[2] V. Lenders, J. Wagner, and M. May, "Analyzing the Impact of Mobility in Ad Hoc Networks," in *Proc. of the ACM REALMAN'06*, Florence, Italy, 2006, pp. 39–46.

[3] A. Chaintreau, P. Hui, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007.

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

[5] "CRAWDAD - Community Resource for Archiving Wireless Data At Dartmouth," URL http://crawdad.cs.dartmouth.edu/, last access in September, 2008.

[6] "HAGGLE - A European Union funded project in Situated and Autonomic Communications," URL http://www.haggleproject.org/, last access in September, 2008.

[7] T. Camp, J. Boleng, and V. Davies, "A Survey of Mobility Models for Ad Hoc Network Research," *Wireless Communications and Mobile Computing (WCMC)*, vol. 2, no. 5, pp. 483–502, 2002.

[8] F. Bai, N. Sadagopan, and A. Helmy, "The IMPORTANT framework for analyzing the Impact of Mobility on Performance Of RouTing protocols for Ad hoc NeTworks," *Elsevier Ad Hoc Net.*, vol. 1, pp. 383–403, 2003.

[9] J.-Y. Le Boudec and M. Vojnovic, "Perfect Simulation and Stationarity of a Class of Mobility Models," in *Proc. of the INFOCOM'05*, Miami, USA, 2005, pp. 72–79.

[10] M. L. J. Yoon and B. Noble, "Random Waypoint Considered Harmful," in *IEEE INFOCOM'03*, San Francisco, USA, apr 2003, pp. 1312–1321.

[11] C. Bettstetter, "Mobility Modeling in Wireless Networks: Categorization, Smooth Movement, and Border Effects," *ACM Mobile Computing and Communications Review*, vol. 5, no. 3, pp. 55–66, 2001.

[12] C. A. V. Campos and L. F. M. de Moraes, "A Markovian Model Representation of Individual Mobility Scenarios in Ad Hoc Networks and Its Evaluation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, 2007, 14 pages.

[13] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "Human Mobility Patterns and Their Impact on Delay Tolerant Networks," in *Proc. of the ACM HotNets VI*, Atlanta, Georgia, USA, 2007.

[14] ——, "On the Levy-walk Nature Human Mobility," in *Proc. of the IEEE Infocom'08*, Phoenix, Arizona, USA, 2008, pp. 1597–1605.

[15] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri, "Towards Realistic Mobility Models for Mobile Ad hoc Networks," in *Proc. of the ACM MobiCom'03*, San Diego, USA, 2003, pp. 217–229.

[16] D. Lelescu, U. C. Kozat, R. Jain, and M. Balakrishnan, "Model T++:: an Empirical Joint Space-time Registration Model," in *Proc. of the ACM MobiHoc'06*, Florence, Italy, 2006, pp. 61–72.

[17] J. Yoon, B. D. Noble, M. Liu, and M. Kim, "Building Realistic Mobility Models from Coarse-Grained Traces," in *Proc. of the ACM MobiSys'06*, Uppsala, Sweden, 2006, pp. 177–190.

[18] M. Boc, A. Fladenmuller, and M. D. de Amorim, "Otiy: Locators Tracking Nodes," in *Proc. of the ACM CoNext'07*, New York, NY, USA, December, 2007, pp. 1–12.

[19] J. Yoon, M. Liu, and B. Noble, "A General Framework to Construct Stationary Mobility Models for the Simulation of Mobile Netwoks," *IEEE Trans. Mob. Comput.*, vol. 5, no. 7, pp. 860–871, jul 2006.

[20] S. Kurkowski, T. Camp, and M. Colagrosso, "MANET Simulation Studies: the Incredibles," *SIGMOBILE Mobile Computing and Communications Review*, vol. 9, no. 4, pp. 50–61, 2005.

[21] "Trimble GPS Pathfinder Office software," URL http://www.trimble.com/pathfinderoffice.shtml, last access in September, 2008.

[22] R. Jain, *The Art of Computer Systems Performance Analysis*. New York, NY, USA: John Wiley and Sons, 1991.