On the Applicability of Mobility Metrics for User Movement Pattern Recognition in MANETs

Elmano Ramalho Cavalcanti
Systems and Computing Department
Federal University of Campina Grande
Campina Grande, Paraíba - Brazil
elmano@copin.ufcg.edu.br

Marco Aurélio Spohn
Federal University of Fronteira Sul
Chapecó, Santa Catarina - Brazil
marco.spohn@uffs.edu.br

ABSTRACT

In this paper we propose a set of mobility metrics, which are employed in the generation of supervised classification learning methods through the decision tree algorithm, with the goal to recognize user movement patterns in mobile ad hoc networks. Hundreds of scenarios produced by several well-known mobility models were employed for training and testing the supervised algorithms. The most suitable classification model showed an accuracy of 99.20 % and Kappa index of 0.991, which indicates a high level of agreement between the classification model and real classification.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Wireless communication*
; C.4 [**Performance of Systems**]: [Measurement techniques]

General Terms

Theory

Keywords

Mobility model, mobility metric, pattern recognition, supervised classification, learning algorithm

1. INTRODUCTION

The study of human mobility applied to the mobile communications industry is among the leading topics of research nowadays. Several theoretical and experimental studies have demonstrated that user mobility is an essential factor in the performance of protocols from different network layers [1,7]. User's movement patterns can be described by mobility models. In recent years, several mobility models have been proposed for modeling individual and group movement in mobile ad hoc networks (MANETs).

MobiWac'13, November 3–8, 2013, Barcelona, Spain.

http://dx.doi.org/10.1145/2508222.2508228.

In order to compare the various possible mobility patterns, a range of mobility metrics have been designed recently. They aim at measuring user motion characteristics, such as the level on movement randomness, the degree of spatial movement dependence among nodes, and the existence of geographic restriction on the scenario (e.g., streets) [3].

In this context, a prominent application of mobility metrics is to assist in the construction of inference models that are able to classify any movement trace into some well-known mobility model (i.e., by choosing the mobility model more capable of reproducing the characteristics of motion described in the trace file). Such classification models can be employed, for example, in determining transportation modes of mobile users (e.g., stationary, walking, riding, or driving), which can be used by adaptive protocols and context-aware applications [15, 19].

After detailing the concepts of mobility models and metrics (Section 3), we propose a rich set of new mobility metrics (Section 4) that are employed together on devising some supervised classification methods for user mobility traces (Section 5.1). The comparison of the classification models, as well as the results and discussion of the selected models are described in Section 5.2. The conclusions of the study are outlined in Section 6.

2. RELATED WORK

Mousavi et al. [14] used the metrics degree of spatial dependence (DSD), degree of temporal dependence (DTD), and the relative speed in the development of a classifier model based on the k-nearest neighbor algorithm. Moreover, the mobility models Random Direction, Random Walk, Gauss-Markov, RPGM, and Manhattan were employed for training and testing the model. Nevertheless, the use of DSD and DTD metrics has been discouraged due to their critical limitations as shown in one of our recent studies [6].

Mun et al. [15] proposed a classification model for determining the type of user movement based on spatial-temporal characteristics of GSM data (number of unique cell IDs and residence time in a cell footprint) and *WiFi* beacons (signal strength variance and duration of dominant access point in view). The collected data were used to generate a decision tree able to classify user movement onto three states: stopped, walking, and driving.

A step forward was taken by Reddy et al. [19] who have developed a system able to classify the transportation mode of a user who carries a mobile phone with built-in GPS receiver. There are five identifiable transportation modes: stationary, walking, running, biking, or in motorized transport. The

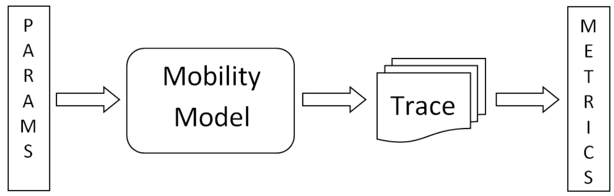

Figure 1: Mobility model seen as an input/output process.

authors also used a decision tree as the classification model, which achieved an accuracy level of 93.6 %. The former and the last works differ from this paper since instead of classifying user movements (i.e., traces) into a specific movement state or type of transportation mode, our model takes into account mobility metrics in order to classify traces into a well-known mobility model (i.e., by choosing the mobility model more capable of producing similar traces).

3. BACKGROUND

3.1 Mobility Models

A mobility model can be seen as an I/O process (Figure 1). The input, $\langle M_i \rangle$, consists on mobility model's parameters such as network lifetime, the length and width of the geographic area, network size (i.e., number of nodes), and the average node speed. The output, $\langle \phi \rangle$, consists on mobility trace files each containing details about the movements of all nodes during the network lifetime. From this file, one can compute mobility metrics.

The first generation of mobility models for MANETs were introduced in the late 90's. The main representative for that generation is the Random Waypoint (RWP), which is still broadly used in simulation-based works. In the Random Waypoint a node randomly chooses a destination point and a constant speed at which it moves until it reaches that destination. After that, the node may stay still for some time (in case a pause time is defined) before starting a new movement.

Among other first generation models, there are the Random Walk, Random Direction, RPGM [9], Gauss-Markov [13], and Manhattan [3]. Random Walks is based on a random choice of direction and speed, with no pause time. In Random Direction, nodes walk until they reach the simulation boundary, where they stop for some time, and then select a new target walking direction. Using stochastic process to model node speed, Gauss-Markov overcomes a limitation (i.e., abrupt speed changes) found in the previous random models. Lastly, in the Manhattan model nodes follow specific paths (e.g., streets) distributed in a grid-based scenario. However, all the aforementioned models are considered synthetic (i.e., based only on mathematical modeling).

The second mobility model generation started around year 2005, when real user movement traces were used for building and validating them. Among such models, there are the mobility model based on communities (CMM) [17], SLAW (Self-similar Least Action Walks) [12], and Smooth [16]. CMM is based on the theory of social networks, taking into account how people come together and move according to their social relations, which is estimated from what the authors call the *social attractiveness*.

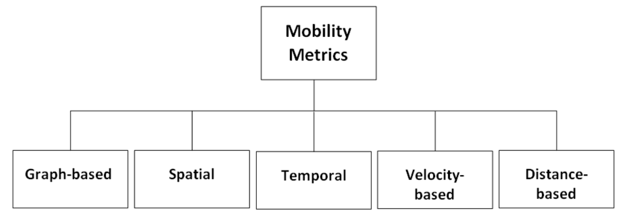

Figure 2: A new classification of mobility metrics for MANETs.

SLAW is a complex model that leverages on several statistical features found in the evaluation of real human walks, such as pause time power-law distributions, inter-contact time, and trip length, as well as restriction on node's mobility within confined areas, and fractal waypoints. Smooth was proposed as a simple alternative to generate realistic traces similar to SLAW, but using simpler input parameters. A good survey of mobility models is covered by Roy [20].

3.2 Mobility Metrics

After several mobility models had been proposed, there was a need for better analyzing and comparing them. For this reason, mobility metrics were introduced aiming at evaluating any mobility model, and capturing features of user movement patterns.

The metrics may be classified according to the type of information employed in their computation. Many of them are derived from graph theory, such as the vertex degree corresponding to the number of a node's neighbors in a MANET. Other metrics, which are based on link/path measurements between pairs of nodes, also belong to this group, which is called graph-based mobility metrics. Another group consists on velocity-based metrics, which employs measures related to the velocity vector (e.g., magnitude, angle), such as the relative speed [3]. If the metric takes into account the distance between nodes as the key computation factor, it is classified as distance-based metric, such as the degree of node proximity [5]. Likewise, if time is the prime factor, the metric is considered a time-based metric (e.g., average link lifetime). Lastly, the metrics that address jointly the node location and network area are labeled as spatial metrics. Figure 2 depicts this classification.

4. NEW MOBILITY METRICS

Before defining the metrics we first present the basic terminology used hereafter. N denotes the number of nodes in the wireless mobile network, T is the lifetime of the network, in units of time, and R is the radio communication range. The main movement variables of a node are depicted in Figure 3. Next, we present the new mobility metrics grouped according to the proposed classification.

4.1 Velocity-based metrics

4.1.1 *Speed-Angle Rate (SAR)*

Consider that from time step $t = 0$ to $t = T$ a node (i.e., mobile user) will present a set of values of magnitude (speed) and angle of its velocity. Let $\langle V^i \rangle = \{v_1^i, v_2^i, \cdots, v_{p-1}^i, v_p^i\}$ be the sequence of the p-th speeds of node i during a period of time, where $v_k^i \neq v_{k+1}^i$, $k, p \in \mathbb{N}$, $0 < k \leq p-1$, and $p \geq 1$. Let also $\langle A^i \rangle = \{a_1^i, a_2^i, \cdots, a_{q-1}^i, a_q^i\}$ denote the sequence

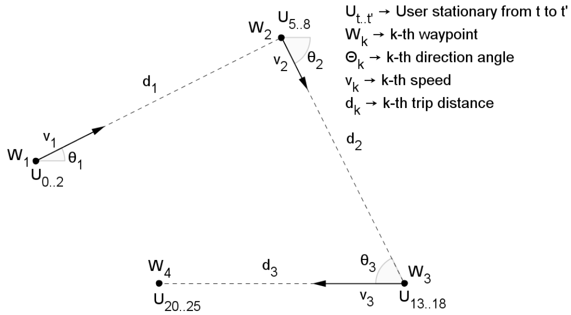

Figure 3: Illustration of a node movement.

of q direction angles that node i had during the same period of time, such that $a^i_k \neq a^i_{k+1}$, $k, q \in \mathbb{N}$, $0 < k \leq q-1$, and $q \geq 1$.

Given that the minimum time that a speed or angle change occurs is of one time-stamp, then the maximum amount of changes of speed/angle is T. Thus, the cardinality of the sets $\langle V^i \rangle$ and $\langle A^i \rangle$ is always less than or equal to T (i.e., $\overline{\langle V^i \rangle}$, $\overline{\langle A^i \rangle} \leq T$). Since $\overline{\langle V^i \rangle} = p$ and $\overline{\langle A^i \rangle} = q$, the rate $\frac{p}{q}$ refers to the amount of times the node speed changes for each angle change. As both p and q varies from 1 to T, it follows that $\frac{1}{T} \leq \frac{p}{q} \leq T$. We call $\frac{p}{q}$ as the Speed-Angle Rate (SAR).

In scenarios where the user's speed varies more frequently than the angle, $SAR > 1$, which is the default case (e.g., in vehicular mobility the vehicle speed changes several times along a straight line segment). On the other hand, if the node speed remains constant even in presence of direction changes (i.e., curves), then $SAR < 1$. Although unlikely in realistic scenarios, the third and last case, $SAR = 1$, is valid in many synthetic mobility models (e.g. Random Waypoint), where a new speed and angle are selected for each new movement, which remain constant till the end of the trip.

4.1.2 Speed and Angle Coefficient of Variation (SVC and AVC)

There are several units for speed measure, angle, and time (Table 4.1.2). With regards to speed, if the mobile node is a pedestrian, m/s is often more appropriate, whereas Km/h or mph are used for ground vehicles. Moreover, the angle of a vector can be expressed in degrees, radians, or grads, while the time can be measured in seconds (typically in MANETs) or in minutes (more adequate for Delay Tolerant Networks, DTNs). Due to this diversity, velocity-based mobility metrics should be independent of unit (dimensionless), in order to avoid misleading conclusions.

Mobile node's variables	Usual measurement units
speed	$\frac{m}{s}$, $\frac{Km}{h}$, Mph
direction angle	degrees, radians, grads
pause time	seconds, minutes

One of the measures used to characterize the variability of a variable that can be represented by different units of measure is the coefficient of variation (CV), which is defined as the ratio of the standard deviation to the mean. CV is a normalized measure of dispersion and is free of scales (i.e., dimensionless). Since the magnitude and angle of the velocity are ratio variables, the CV can be used without restrictions.

Let μ_v denotes the average between all nodes' speed during T, and σ_v be the standard deviation of these values. The speed coefficient of variation (SCV) is given by σ_v/μ_v.

Analogously, the angle coefficient of variation (ACV) is given by σ_a/μ_a.

4.2 Distance-based Metrics

4.2.1 Average Trip Length (ATL)

The displacement between two consecutive waypoints of a node is known as a trip (also known as a flight). Let $\langle W^i \rangle = \{w^i_1, w^i_2, \cdots, w^i_{n-1}, w^i_n\}$ be the waypoints of node i during a period of time (as illustrated in Figure 3). The displacement of the trip from w_k to w_{k+1} is given as follows:

$$ATL(w^i_k, w^i_{k+1}) = Dist((x^i_{w_k}, y^i_{w_k}), (x^i_{w_{k+1}}, y^i_{w_{k+1}})) \quad (1)$$

where $x^i_{w_k}$ and $y^i_{w_k}$ are the x and y-coordinate of node i at the k-th waypoint, and $Dist(w^i_k, w^i_{k+1})$ the Euclidean distance between two consecutive waypoints.

4.2.2 Average Path Length (APL)

We define *pathlength* as the total distance accomplished by a node between two consecutive waypoints. In some mobility models, as in the Random Waypoint, the displacement between two waypoints will be equal to the distance between them (i.e., $ATL = APL$). This occurs because nodes always keep the same speed and direction angle until reach the next waypoint. Obviously, independent of the movement behavior, $ATL \leq APL$.

Let $t_{w^i_k}$ and $t_{w^i_{k+1}}$ be the time steps when node i reaches the k-th and (k+1)-th waypoints. The total distance traveled by i between w^i_k and w^i_{k+1} is as follows:

$$APL(w^i_k, w^i_{k+1}) = \sum_{j=t_{w^i_k}}^{t_{w^i_{k+1}}} Dist((x_j, y_j), (x_{j+1}, y_{j+1})) \quad (2)$$

4.3 Temporal Metrics

4.3.1 Pause time Variation of Coefficient (PVC)

This metric is computed analogously to the speed coefficient of variation. If μ_p is the average node pause time between all nodes during the network lifetime, and σ_v the standard deviation, we have that $CVP = \sigma_p/\mu_p$.

4.3.2 Relative Link Duration (RLD)

Link duration (LD), also known as link lifetime and contact time, is a well-known metric that refers to the total amount of time where there is communication between pairs of nodes; i.e., when nodes are distant from each other up to R meters. However, LD is an absolute value, which value depends on the network lifetime (T). A good mobility metric should not rely on a single input parameter. For example, one should expect LD in a group-based mobility model (e.g., RPGM) to be greater than LD for an individual random mobility (e.g., RWP). However, if we compare a scenario of 1000 seconds of a RWP mobility trace file with a scenario of 100 s of RPGM trace file, LD is likely to be greater in RWP compared to RPGM. To avoid this drawback, we propose

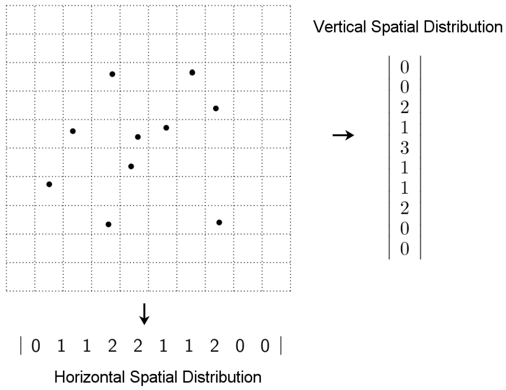

Figure 4: Node Spatial Distribution in a square NxN matrix (N= number of nodes).

the relative link duration (RLD), which is equal to the ratio between LD and the network lifetime T.

4.4 Spatial Metrics

4.4.1 Degree of Network Spatial Distribution (DNSD)

Let $D = [c_{i,j}]$, $1 \le i, j \le N$, denotes a square matrix $N \times N$, where N is the number of nodes in the network. Let us assume that $c_{i,j}$ is a rectangular region (cell) in the scenario where there may be a number between 0 to N nodes. Matrix D represents the node spatial distribution in the network. Formally,

$$D = \begin{vmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{vmatrix}$$

Figure 4 shows the node distribution at a certain time. The additional row and column matrices in the figure express the horizontal and vertical node distribution, respectively. These matrices are called Horizontal Distribution Matrix (HDM) and Vertical Distribution Matrix (VDM), where $[h_{1,j}]$, $1 \le j \le N$ and $h_{1,j} = \sum_{i=1}^{N} c_{i,j}$ and MDV $=$ $[h_{i,1}]$, $1 \le i \le N$ and $h_{i,1} = \sum_{j=1}^{N} c_{i,j}$.

Taking into account this representation, the lower and upper bounds of DNSD will occur in cases of minor and major uniformity, respectively. The DNSD is upper bound if all elements of HDM and VDM are equal to 1, meaning that in each row and column should exist exactly one node. On the other hand, the lower bound is reached when all nodes are in the same cell. An example of lower bound configuration is given below.

$$D = \begin{vmatrix} N & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{vmatrix}$$

Thereby, let $L = [c_{1,j}]$, $c_{1,j} = 1$, $1 \le j \le N$ be a row matrix wherein all values equal to 1. Analogously, let $C = [c_{i,1}]$, $c_{i,1} = 1$, $1 \le i \le N$ be a column matrix wherein all values are also equal to 1. Thus, the vertical distribution deviation is given as the difference between the matrices VDM and C (in absolute terms):

$$\begin{vmatrix} N \\ 0 \\ \vdots \\ 0 \end{vmatrix} - \begin{vmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{vmatrix} = \begin{vmatrix} (N-1) \\ 1 \\ \vdots \\ 1 \end{vmatrix}$$

The sum of the vertical deviation is equals to $(N-1) + (N-1).1 = 2(N-1)$. Likewise, the sum of horizontal deviation is also $2(N-1)$. As a result, the total deviation of node distribution is equal to $4(N-1)$. Thus, we define the degree of spatial distribution at time t as follows:

$$DNSD(t) = 1 - \log(DEV(t) + 1)/\log(MAX + 1) \quad (3)$$

where $MAX = \log(4(N-1))$ is the maximum theoretical deviation of node distribution, and $DEV(t)$ is the sum of all elements of HDM and VDM matrices at time step t. The DNSD value will be the average between $DNSD(t)$, $0 < t \le T$ (e.g., Figure 4 $DNSD(t) = 1 - \log(14/36) \approx 0,59$). The use of a logarithmic scale is suitable due to the wide range of possible values[1].

4.4.2 Degree of Spatial Accessibility (DSA)

Considering the same modeling employed for computing DNSD, the degree of spatial accessibility is given as the proportion of visited cells by the total number of cells. Note that a cell $c(i, j)$ is said to be visited if at least one node was placed in the cell at some moment.

In geographic restricted mobility models, there are regions on the map where a node can never be. Consequently, the DSA will be lower in those models than in random models (e.g., Random Waypoint), where a node may be anywhere. Thus, the benefits of this metric are twofold: a) to distinguish between geographic restricted and geographic unrestricted mobility models, and for somehow quantifying the user movement freedom level for a given scenario.

Let $x(i, j)$ be an indicator random variable that informs whether a cell was visited by at least one user, which means that $x(i, t) = 0$ if $c(i, j) = 0$ or $x(i, t) = 1$ if $c(i, j) > 0$. Thus, we define the degree of spatial accessibility of a network at time t as follows:

$$DSA(t) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} x(i, j)}{N^2} \quad (4)$$

4.5 Graph-based metrics

4.5.1 Relative Node Degree (RND)

The node degree (ND) represents the number of neighbors of a node, having a direct impact on the analysis of mobile networks. If $G(i, t)$ denotes the node degree of node i at time t, N the network size, and T the network lifetime, ND is defined as follows:

$$\overline{ND} = \frac{1}{N \cdot T} \sum_{i=1}^{N} \sum_{t=1}^{T} G(i, t) \quad (5)$$

However, since ND is an absolute value, it seems not appropriate for mobility classification purposes. For instance, a scenario provided by a random mobility model may show

[1] It can be proven that the number of non-negative integer n x n matrices with sum of elements equal to n is given by the formula $\binom{n^2+n-1}{n}$ http://oeis.org/A054688.

greater ND than a group-based model, depending on the network size (N) and area (A). For this reason, we propose the definition of a relative and normalized node degree metric, based not only on the number of neighbors, but also on N and A. Let ρ be the node density of the network, defined as N/A. Considering an homogeneous random node distribution in a wireless network, the expected number of neighbors of a node is $\rho\pi R^2 - 1$ [4]. Hence, the relative node degree ($\overline{RND}$) is defined as follows:

$$\overline{RND} = \frac{ND - (\rho\pi R^2 - 1)}{(N-1) - (\rho\pi R^2 - 1)} \tag{6}$$

$$= \frac{ND - (\rho\pi R^2 - 1)}{N - \rho\pi R^2} \tag{7}$$

where (N-1) is the maximum node degree. RND is a normalized metric within interval $[0, 1]$, with lower bound when RND is equal to the expected node degree in a completely homogeneous network node distribution, and upper bound when $RND = N - 1$; i.e., when all nodes are always neighbors to each other.

4.5.2 Neighborhood Probability (NP)

NP represents the probability that two randomly selected nodes of the network are neighbors at an arbitrary instant of time. In a low node density network, NP should present lower values than in scenarios with a higher node density. Taking $E(i, j, t)$ as a boolean variable equal to 1 when there is a communication link between nodes i, j, and 0 otherwise, the NP between them is given by the equation 8, whereas the NP of the network is computed according to equation 9.

$$PCN(i, j) = \frac{1}{T}\sum_{t=1}^{T} E(i, j, t) \tag{8}$$

$$PCN = 100 \times \frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N} PCN(i, j) \tag{9}$$

4.5.3 Degree of Link Changes (DLC)

The number of link changes (LC) for a pair of nodes i and j is the number of times the link between them transitions from down to up [3]. Let $E(i, j, t)$ be a boolean variable equal to 1 when there is a communication link between nodes i, j and 0 otherwise. Additionally, let $E(i, j)$ denotes a boolean value that is equal to 1 if there was at least one communication link between i and j during the network lifetime, and 0 otherwise. Thus, LC is define as follows:

$$LC(i, j) = \frac{1}{P}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sum_{t=1}^{T} C(i, j, t) \tag{10}$$

where P is the number of node pairs i, j such that E(i,j) = 1, and C(i,j,t) = 1 if E(i,j,$t-1$) = 0 and E(i,j,t) = 1 or E(i,j,$t-1$) = 1 and E(i,j,t) = 0.

However, as the ND drawback aforementioned, LC share a similar limitation as it is an absolute value. We propose a relative and normalized link change metric, what we call the degree of link changes (DLC):

$$DLC = \frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N} LC(i, j). \tag{11}$$

5. CLASSIFICATION MODEL OF MOBILITY PATTERNS

5.1 Materials and Methods

Ten mobility models were used for constructing the classification model, briefly described in Table 1Mobility trace files for Smooth and CMM were computed by running the code made publicly available by their authors, whereas for the other models we used the BonnMotion tool [2] (version 2.0). For building the classification model, the RapidMiner tool was employed [18], an open source software for knowledge discovery, machine learning, and data mining.

The configuration of the mobility models's variables are described in Table 2. In our approach, each variable under consideration is assigned a list of possible (common) values, resulting in thousands of combinations of scenarios for each model. For each mobility model, we have chosen 50 of these combinations at random for generating the mobility trace files (a similar methodology procedure was done by Kurkowski [10]). As a result, 500 random generated trace files were taken into account in this study.

The classification method employed was based on the supervised learning algorithm of decision tree (DT), which is considered one of the most widespread and consolidated supervised classification algorithms [11].

For each trace file we compute all metrics outlined in Section 4, adding a label indicating the class (i.e., mobility model's name). The tabular data were stored in a sheet, wherein each line contains the values of the metrics and the label.

Besides the proposed metrics, we added three more metrics from our recent works [5,6]: (a) the Improved Degree of Spatial Dependence (IDSD), a spatial metric able to capture both movement and pause correlation among mobile nodes; (b) the Improved Degree of Temporal Dependence (IDTD), a metric that captures to which extent the current node speed depends on its past moving pattern; and the Degree of Node Proximity (DNP), another spatial metric.

After this pre-processing step, we started the DT algorithm, which consists of two phases: training and testing. During the training phase, the classifier model is built from the input data. After that, an unlabeled data is used in the testing phase, which should be classified by the model. The stratified ten-fold cross-validation approach was employed for validating the model [11]. In this technique, the data is randomly and uniformly spread into 10 parts (stratified sampling). Each part is used as a holdout set and the other nine parts are used to train the model, totalizing ten combinations for testing. For each one, the error rate is calculated on the holdout set, and thus the learning procedure is executed 10 times using different training sets. Finally, the 10 error estimations are averaged to yield an overall error estimate.

5.2 Results and Discussion

We have developed several classification models, each based on a specific class of mobility metrics. The objective was to determine the level of quality in pattern recognition of user movement when using solely metrics of a certain category. For evaluating the models we used the performance metrics Accuracy and Kappa index $\hat{k}$. The $\hat{k}$ statistics indicates the level of agreement between the classification model and real classification (i.e., the one known in advance). As outlined

Table 1: Characteristics of the selected mobility models.

Characteristic	RWP	RPGM	GM[1]	Manhattan	*Column*	RD[2]	RW[3]	CMM	SLAW	Smooth
Entity movement	X		X	X		X	X		X	X
Collective movement		X			X			X		
Temporal dependence			X					X	X	X
Geographic restricted				X				X	X	X
1st generation (synthetic)	X	X	X	X	X	X	X			
2nd generation								X	X	X

[1] Gauss-Markov. [2] *Random Direction.* [3] *Random Walk.*

Table 2: Mobility Models Configuration

Parameter	Values	Parameter	Values
Simulation time	900 s	Number of nodes	25, 50, 75, 100, 125, 150, 175, 200
Scenario length	1, 1.5, 2 km	Scenario width	1, 1.5, 2 km
Min. speed	1, 2, 3 m/s	Max. speed	5, 10, 20, 30 m/s
Group length[1]	5, 10, 20, 25	Number of rows[2]	5, 10, 20
Transmission range	200 m	Number of columns[2]	5, 10, 20
Turn probability[4]	0.25, 0.5, 0.75	Speed std. deviation[4]	2, 4, 6 m/s
Alpha distance	2, 3, 4	Memory parameter[7]	0.2, 0.4, 0.5, 0.6, 0.8
Min. pause time	1, 5, 10	Max. pause time[8]	5, 10, 50, 100, 200, 500 s
Clustering range[8]	20, 50, 100, 200	Number of waypoints[8]	20, 40, 60, 80, 100

[1] For RPGM, *Column* and CMM. [2] For Manhattan and CMM. [3] For SLAW and Smooth. [4] For Manhattan. [5] For CMM. [6] For RPGM and *Column*. [7] For Gauss-Markov. [8]For SLAW and Smooth.

in Table 3, both accuracy and $\hat{k}$ are strongly affected when more mobility metrics are used in the supervised algorithm. The best classification model was achieved when all metrics were taken into account, reaching an accuracy of 99.20% and 0.991 Kappa. We next describe the model's DT algorithm.

The DT algorithm (Figure 5) starts classifying movement traces in which the average trip length (ATL) is zero as derived from the Gauss-Markov model. ATL=0 only when nodes move continuously, never stopping, which is exactly what happens for Gauss-Markov (GM). The speed coefficient of variation (SCV) is close to zero only when its standard deviation is also close to zero, indicating a nearly constant speed. Among the evaluated models, only SLAW's node speed is almost always constant (as provided in the authors' code).

The speed angle rate (SAR) is greater for both Smooth and Manhattan models, because in these models nodes usually change the speed several times during the trip between two waypoints, while the direction keeps unchanged (in Smooth) or it has only a few changes (in Manhattan). However, the angle distribution is uniform in Smooth, whereas it is variable in Manhattan, depending on factors such as turn probability and number and layout of the roads in the simulation area. Thus, the angle coefficient of variation (ACV) is greater in Smooth (> 0.53) than in Manhattan.

The improved degree of spatial dependence (IDSD) metric, which measures the level of movement correlation among nodes, was the key factor for detecting the RPGM, a group-based model, while the IDTD distinguishes CMM trace files from the models where node speed values are independent of past values (especially the random models).

The column model showed lower values of the degree of spatial accessibility (DSA) due to its group movement constraints. The degree of node proximity (DNP) was capable of distinguishing Random Direction (RD) traces from Random Walk (RW) and Random Waypoint (RWP) ones. In RD the number of users is greater at the edges than in the central area, resulting in lower values of DNP compared to RW and RWP.

In the next step of the algorithm, DSA and degree of node spatial distribution (DNSD) were used for recognizing RW and RWP traces among others, with few cases of misclassification. Finally, the degree of link changes (DLC) was lower for RWP than RW. As RWP is known for showing a non-uniform distribution, with a higher node density in the central area, as a result the number of link changes is smaller than in Random Walk.

The classification model performance metrics are show in Table 4. A class recall is defined as the ratio between the number of correctly documents (i.e., mobility trace files) classified and all documents belonging to the class (i.e., mobility model). Precision is the ratio between the number of correctly classified documents and all documents considered by the model as belonging to that class [8].

6. CONCLUSIONS

In this paper we proposed a new and comprehensive set of mobility metrics, which were employed for building a user mobility pattern recognition model through a decision tree supervised learning algorithm. Hundreds of movement trace files from ten well-known mobility models were randomly generated and taken into account by the learning algorithm. The Decision Tree supervised classification model showed an accuracy of 99.2% and a Kappa index of 0.991. The DT algorithm can be used as some sort of oracle for classifying an unknown mobility trace into a know mobility model. The

Table 3: Comparison of classification models based on different sets of mobility metrics.

Classification Model	Performance Metrics *Accuracy*	*Kappa*
A) Employing velocity-based metrics only (i.e., AVC, SVC, SAR, IDSD and IDTD)	80.00%	0.7778
B) Employing distance-based metrics only (i.e., APT, ATL and DNP)	61.40%	0.5711
C) Employing spatial metrics only (i.e., DNSD and DSA)	55.40%	0.5044
D) Employing temporal metrics only (i.e., RLD and PVC)	55.60%	0.5067
E) Employing graph-based metrics only (i.e., RND and DLC)	50.00%	0.4444
F) Employing all mobility metrics	99.20%	0.9911

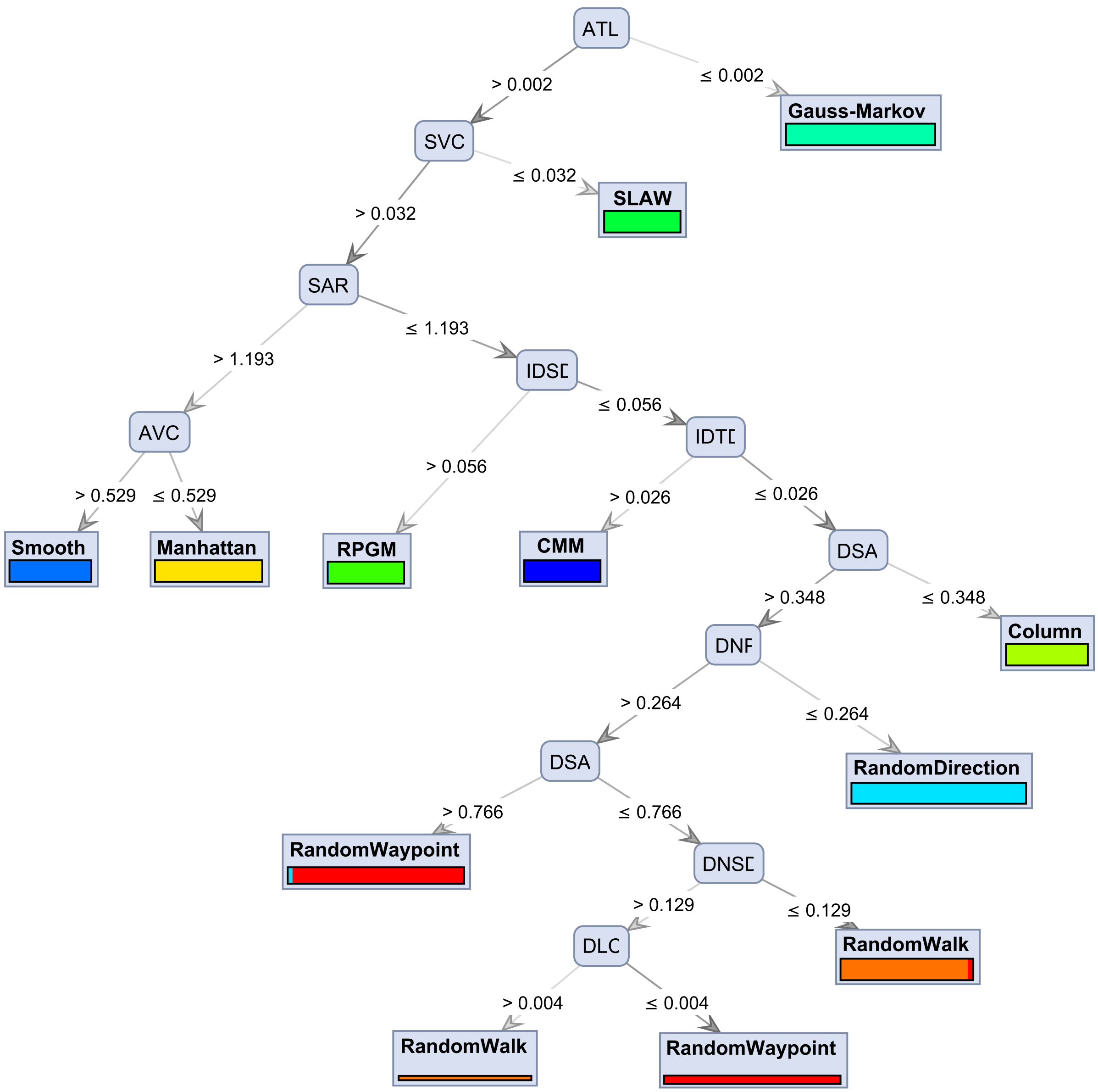

Figure 5: Decision tree for classifying mobility traces.

Table 4: **Multiclass Classification Performance (Confusion Matrix).**

		TRUE CLASSIFICATION										
		CMM	Smooth	RD	GM	SLAW	RPGM	Column	Manhattan	RW	RWP	*Precision*
P	CMM	500	0	0	0	0	0	0	0	0	0	100%
R	Smooth	0	500	0	0	0	0	0	0	0	0	100%
E	RD	0	0	480	0	0	0	0	0	0	0	100%
D	GM	0	0	0	500	0	0	0	0	0	0	100%
I	SLAW	0	0	0	0	500	0	0	0	0	0	100%
C	RPGM	0	0	10	0	0	500	0	0	0	0	98.04%
T	Column	0	0	0	0	0	0	500	0	0	0	100%
I	Manhattan	0	0	0	0	0	0	0	500	0	0	100%
O	RW	0	0	0	0	0	0	0	0	500	20	96.15%
N	RWP	0	0	10	0	0	0	0	0	0	480	97.96%
	class recall	100%	100%	96%	100%	100%	100%	100%	100%	100%	96%	

proposed mobility metrics which have not been employed by the algorithm (i.e., APL, PVC, RLD and RND) will be subject for future research. Also, we intend to consider yet other mobility models as well as additional classification algorithms.

7. REFERENCES

[1] W. Alasmary and W. Zhuang. Mobility impact in ieee 802.11p infrastructureless vehicular networks. *Ad Hoc Networks*, 10(2):222 – 230, March 2012.

[2] N. Aschenbruck, R. Ernst, E. Gerhards-Padilla, and M. Schwamborn. Bonnmotion: a mobility scenario generation and analysis tool. In *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques*, SIMUTools '10, pages 51:1–51:10, Brussels, Belgium, June 2010. ICST.

[3] F. Bai, N. Sadagopan, and A. Helmy. IMPORTANT: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks. In *Proc. 22nd INFOCOM*, pages 825–835, San Franciso, CA, USA, 2003. IEEE.

[4] C. Bettstetter. On the minimum node degree and connectivity of a wireless multihop network. In *Proc. of the 3rd ACM International Symposium on Mobile ad hoc Networking & Computing*, MobiHoc '02, pages 80–91, New York, NY, USA, 2002. ACM.

[5] E. R. Cavalcanti and M. A. Spohn. Degree of node proximity: a spatial mobility metric for manets. In *Proceedings of the 9th ACM international symposium on Mobility management and wireless access*, MobiWac '11, pages 61–68, New York, NY, USA, 2011. ACM.

[6] E. R. Cavalcanti and M. A. Spohn. On improving temporal and spatial mobility metrics for wireless ad hoc networks. *Inf. Sci*, 188:182–197, 2012.

[7] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on the design of opportunistic forwarding algorithms. IEEE Transactions on Mobile Computing, 6(2):606–620, 2007.

[8] R. Feldman and J. Sanger. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2007.

[9] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang. A group mobility model for ad hoc wireless networks. In *Proc. 2nd ACM MSWiM*, pages 53–60, Seattle - WA, USA, 1999.

[10] S. Kurkowski. *Credible Mobile Ad Hoc Network Simulation-Based Studies*. PhD thesis, Colorado School of Mines, 2006.

[11] D. T. Larose. *Discovering Knowledge in Data - An Introduction to Data Mining*. Wiley, 2004.

[12] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong. SLAW: A new mobility model for human walks. In *Proc. 28th INFOCOM*, pages 855–863, Rio de Janeiro, Brazil, 2009. IEEE.

[13] B. Liang and Z. Haas. Predictive distance-based mobility management for pcs networks. In *Proc. 18th Int. Conf. on Computer Communications*, INFOCOM '99, pages 1377–1384. IEEE, April 1999.

[14] S. Mousavi, H. Rabiee, M. Moshref, and A. Dabirmoghaddam. Mobility pattern recognition in mobile ad-hoc networks. In *Proceedings of the 4th Mobility Conference: International Conference on Mobile Technology, Application & Systems*, pages 302–309. ACM, 2007.

[15] M. Y. Mun, D. Estrin, J. Burke, and M. Hansen. Parsimonious mobility classification using gsm and wifi traces. In *HotEmNets*. ACM, 2008.

[16] A. Munjal, T. Camp, and W. Navidi. SMOOTH: a simple way to model human mobility. In A. Helmy, B. Landfeldt, and L. Bononi, editors, *MSWiM*, pages 351–360. ACM, 2011.

[17] M. Musolesi and C. Mascolo. Designing mobility models based on social network theory. *Mobile Computing and Communications Review*, 11(3):59–70, 2007.

[18] Rapid-I. Rapidminer - open-source data mining with the java software rapidminer. http://rapid-i.com/, 2013.

[19] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using mobile phones to determine transportation modes. *ACM TOSN*, 6(2), Feb. 2010.

[20] R. R. Roy. *Handbook of Mobile Ad Hoc Networks for Mobility Models*. Springer, 2011.