Experimental Study of Mobility in the Soccer Field with Application to Real-Time Athlete Monitoring

Vijay Sivaraman[†], Sarthak Grover[†], Alexander Kurusingal[†], Ashay Dhamdhere[†], Alison Burdett[‡]

[†]School of EE&T, University of New South Wales, Sydney, NSW 2052, Australia

Emails: {vijay@unsw.edu.au, shahifaqeer@gmail.com, alex@ee.unsw.edu.au, ashay@unsw.edu.au}

[‡]Toumaz Technology Limited, Abingdon, Oxfordshire, UK. Email: {alison.burdett@toumaz.com}

Abstract— Live monitoring of athletes during sporting events can help maximise performance while preventing injury, and enable new applications such as referee-assist and enhanced television broadcast services. A major challenge is the extraction of athlete physiological data in real-time, since the radio range of body-worn sensor devices is limited, necessitating multi-hop routing mechanisms. However, little is known about the highly dynamic operating conditions on a soccer field under which communication protocols need to operate.

In this work we conduct field experiments in which we outfit first-division soccer players with sensor devices and record their inter-connectivity during a real game. Our first contribution profiles the key properties of the dynamic wireless topologies arising in the soccer field, and highlights the consequences for routing mechanisms. We show that the topology is in general sparse, with short encounters and power-law distributed inter-encounters. Importantly, the co-ordinated movement of players in the field gives rise to significant correlations amongst links, an aspect that can potentially be exploited by routing. Our second contribution develops a model for generating synthetic topologies that mirror connectivity in a real soccer game, and can be used for simulation studies of routing mechanisms. Its novelty lies in explicitly modelling the underlying auto-correlation and cross-correlation properties of the links, from which derived measures such as inter-encounter times and neighbourhood distributions follow. Our study is an important first step towards understanding and modelling dynamic topologies associated with sports monitoring, and paves the way for the design of real-time routing algorithms for such environments.

I. INTRODUCTION

Advances in sensing and communications technologies are enabling new low-cost and lightweight devices that allow measurement and remote monitoring of an individual's vital physiological signs such as ECG, temperature and oxygen saturation levels. Such technology, though designed primarily for the healthcare industry, is being adapted to the massively popular and growing field of sports science, specifically for the purpose of athlete monitoring.

Biomedical technology has long been used by professional coaches and trainers in striving to push their athletes' bodies to the edge of its capabilities. However, much of this examination of the body has been performed under laboratory conditions where results attained in the artifical environment may not parallel those observed in competition [1]. Devices are now starting to emerge in the market that are making the leap from monitoring athletes in training (e.g. SPI Elite [2] platform from GPSports) to monitoring them during competition (e.g. e-AR [3] and VxSport [4]). We are partnering with Toumaz Technologies in the UK who are manufacturing a platform called Sensium[TM][5] that integrates low power wireless technology with miniaturised sensors and lightweight flexible batteries [6]. This platform, weighing just a few grams, will allow non-intrusive collection and real-time wireless transmission of athlete physiological data during competition.

We seek to apply the above wearable platforms to monitoring athletes in field sports, specifically soccer. Soccer is a hugely popular sport throughout the world, and attracts large financial investment, particularly in Europe. Several soccer clubs in the UK have expressed great interest in monitoring their athletes on the field, predominantly to reduce the risk of injury and improve player substitution decisions. Soccer organisers have also expressed some interest in using real-time position and impact information for referee-assist services, and television channels are eager to augment live broadcasts with player parameters (e.g. heart-rate during clutch events, speed and acceleration, impact levels during collisions, etc.) so as to heighten the level of engagement for audiences.

While hardware platforms for athlete monitoring are maturing rapidly, there is much research needed in developing communication protocols that can operate under the unique conditions arising in the soccer field: (a) Rapid acceleration and impact are part of the sport, and this restricts the monitoring device to be small, lightweight, unobtrusive and non-protruding so that the players' degree of freedom is not limited. This is in contrast to devices tried in sports such as rowing [7] or cross country skiing [8] that have form-factor akin to a mobile phone. Consequently, monitoring devices for soccer can be expected to have extremely limited battery power and restricted radio range, placing severe energy and reach constraints on the communication protocols. (b) The playing area in soccer is very large at over $4000m^2$. Given the limited transmission range of body-worn devices, coupled with attenuation effects arising from attachment to the human body (profiled in the next section), real-time extraction of player data would require multi-hop routing. One-hop communication from the device to base-station, such as proposed for ice-hockey in [9], or the protocols proposed in [10] for monitoring team-sports such as basketball and volleyball having a small playing area, would not suffice for soccer. (c) Soccer players move very rapidly in the field, and this makes the topology highly dynamic at short time-scales (seconds). Designing routing mechanisms that can deliver data to base-stations within stringent time and energy constraints over multiple hops in this time-varying environment promises to be challenging.

In order to design appropriate communication protocols (specifically for multi-hop routing), it is essential to have a good understanding of the dynamics of the wireless topology that arises in the soccer field. To the best of our knowledge no such data or characterisation from real soccer games is available in the literature today. We therefore undertake experimental work in which we outfit first-division soccer club players with sensor devices and collect data on their connectivity (with each other and with base-stations around the field) over multiple games. Using the collected data we make two important contributions: (1) We provide a **stochastic characterisation** of key aspects of the topology, such as the number of wireless neighbours of a player (indicating the number of alternate routes that may be available) and distributions of the encounter and inter-encounter times between players (indicating the length of time for which routes may persist or vanish). Additionally, we show that several links can exhibit correlations with each other, i.e. the presence of link between one pair of players can affect the probability (positively or negatively) of link between another pair. Our study is the first to quantify these metrics for a soccer game, providing valuable input to the design of routing algorithms for emerging athlete monitoring applications. (2) We develop a novel **mathematical model** for generating dynamic topologies that stochastically match empirical traces. Our model explicitly considers the underlying auto-correlation and cross-correlation structure of links, from which derived metrics such as inter-contact times and neighbourhood distribution follow. Unlike existing (individual and group) mobility models in the literature that rely on certain movement patterns, our model is the first to directly generate connectivity topology for arbitrarily specified link (auto- and cross-) correlations. Our study sets the stage for design of multi-hop routing algorithms that can in the near future enable real-time physiological monitoring of athletes during live competition games.

The rest of this paper is organised as follows: Section II describes the experimental set-up to monitor soccer players during a game. Statistics of the observed data are presented in Section III, while Section IV highlights the correlations between links. A model for generating dynamic topologies with correlated links is presented in Section V, and a summary and directions for future work are outlined in Section VI.

II. EXPERIMENTAL SETUP

With the objective of gaining an understanding of wireless connectivity in a soccer field, we outfitted all players of the University of New South Wales Football Club (UNSWFC) first-division men's team with wireless monitoring devices. We collected data over two trial games held on 28 October 2008 and 5 February 2009. In this paper we use the more comprehensive data we were able to collect in the second game, since by then we were able to improve on aspects such as device attachment to the body and base-station placement.

A. Mounting the Device on the Body

The body-worn devices we used were the MicaZ motes [11] from Crossbow technologies. These are off-the-shelf devices
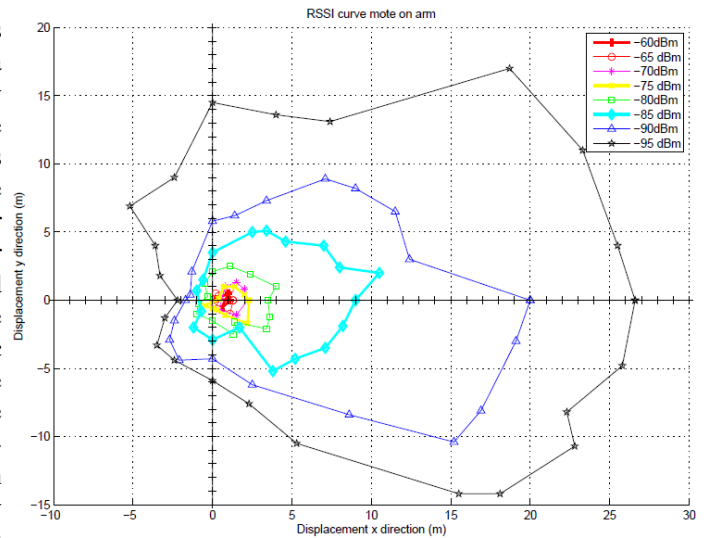

Fig. 1.   RSSI contour for arm worn monitoring device

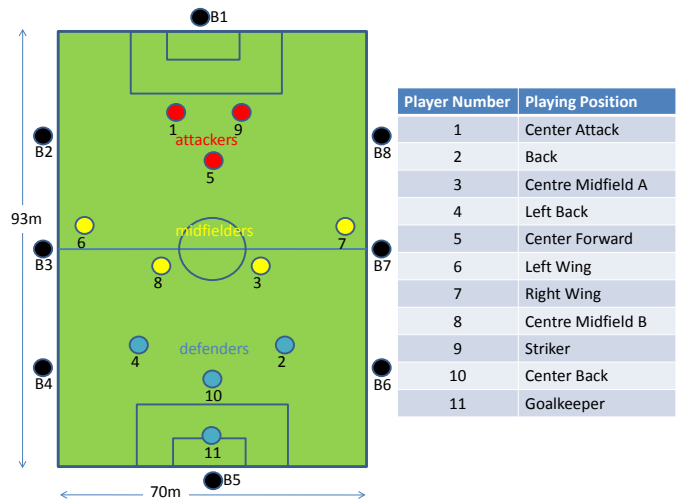| Player Number	Playing Position
1	Center Attack
2	Back
3	Centre Midfield A
4	Left Back
5	Center Forward
6	Left Wing
7	Right Wing
8	Centre Midfield B
9	Striker
10	Center Back
11	Goalkeeper

Fig. 2.   Player positions on the soccer field

operating in the 2.4 GHz band that are readily available today. We intend to replace these with emerging platforms custom-built for body-area-networking as they become available. Though the MicaZ motes were not designed for body-worn applications, they have been used before for body health monitoring, such as in Harvard's Code Blue project [12], and in our own prior work [13] in profiling the body channel for patients with chronic illnesses.

Finding a good way to mount sensor nodes on athletes turns out to be quite a challenge. Taking into consideration aspects such as attenuation of the wireless signal by the body, ease and stability of attachment, and possibility of damage to the device itself, we decided to go with an arm mounted attachment using an arm-band. We conducted experiments to profile the propagation of wireless signals around the athlete in an open field (see our paper [14] for a detailed study), and show in Fig. 1 a contour map of the received signal strength (RSSI) at various distances and directions from the body-mounted mote device, which transmits data every second at the highest available power level of 1mW. In the plot the athlete wearing the device is located at $(0,0)$ and is facing north, while the device is attached to the subject's right arm. Not surprisingly,
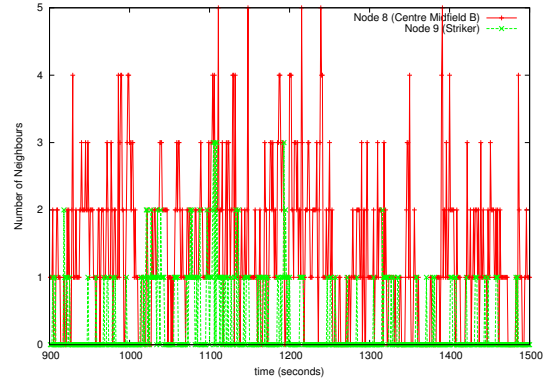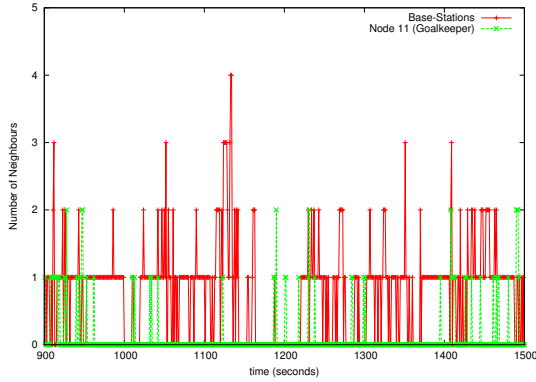
(a) Base-stations (Aggregated) and Node 11 (Goalkeeper) (b) Node 8 (Centre Midfielder B) and Node 9 (Striker)

Fig. 3. Number of neighbours versus time for (a) base-stations and goalkeeper, and (b) centre midfielder B and striker

the reach is much larger to the right (at around 27m the signal fades to below -95dBm) than to the left (signals barely reach beyond 3m), and there is a 10-15m reach to the front and back of the player. We also tried other mounting positions (e.g. back), but found such mountings to either cast a larger "shadow" region of poor signal, or create discomfort for the athlete due to clothing impediments or increased chance of injury/damage during a fall. We therefore proceeded with an arm-mounted position for all our subsequent studies.

B. Game Layout and Data Collection

The game was played on a full size field with dimensions 93m x 70m. Each of the 11 players wore a monitoring device on their arm, and 8 base-stations were positioned (at a height of about 1m from the ground) along the sidelines of the playing area. Fig. 2 shows the nominal playing positions and associated node identification numbers. Unfortunately the devices worn by players 2 (back) and 4 (left back) were damaged during play and we could not obtain data from them, as was base-station B3 which got hit by the ball.

We implemented software on each of the body-worn devices such that it broadcasts, once every second, at the highest available power level of 1mW, a packet containing its unique identifier and a sequence number, during the entire measurement period. All devices (body-worn as well as base-stations) that successfully receive this packet record this event in their on-board memory. As the game proceeds, each node (and base-station) will be cataloguing which other nodes it could hear at each time instant. To prevent collisions in-the-air, each second is divided into 11 slots each of approximate duration 90ms, and each of the 11 body-worn devices is given a unique such slot for transmission every second. Just prior to commencement of the game, the master base-station sends a clock synchronisation message to all nodes, upon receipt of which each node starts recording connectivity data in on-board memory. Data collection stops after 25 minutes, and at the end of the game data from each node is extracted by the master base-station for off-line analysis.

III. PROFILING PLAYER CONNECTIVITY

The data collected above tells us how the wireless connectivity between players evolves from second-to-second during

the game. In this section we analyse this data and highlight several aspects that are pertinent to the design of routing protocols for real-time extraction of player physiological data. Though we recognise that each soccer game is different and data acquired from repeat trials would undoubtedly yield a different composition of results, our aim is to highlight key common characteristics and trends associated with player connectivity arising in a real first-division soccer game. We have developed a Java GUI animation that displays the connectivity data collected for this game, and can be viewed at our web-site [15]. The data was collected for 25 minutes; in this paper we will use only the data from the 10 minute interval from 900s to 1500s, since during that period there were no substitutions and no play stoppages. We strongly encourage the reader to observe the Java GUI animation, specifically for the interval 900s-1500s, to get a better feel for the data obtained from the experiment.
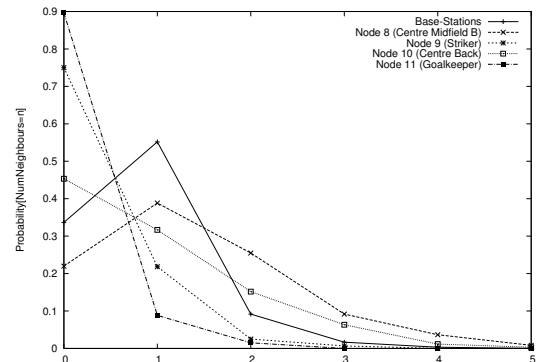

Fig. 4. Neighbour distribution for base-stations and nodes 8, 9, 10, 11

A. Number of Neighbours

In Fig. 3 we show the time-evolution of the number of neighbours for selected nodes. Specifically, for each selected node, we show at each instant of time, the number of other nodes whose transmissions are successfully received. To give a flavour of the diversity we pick players from forward (node9: striker), middle (node 8:centre midfield B), and backward (node 11: goalkeeper) playing positions, as well as the (aggregation of all 8) base-stations. Important observations that emerge from these plots are:

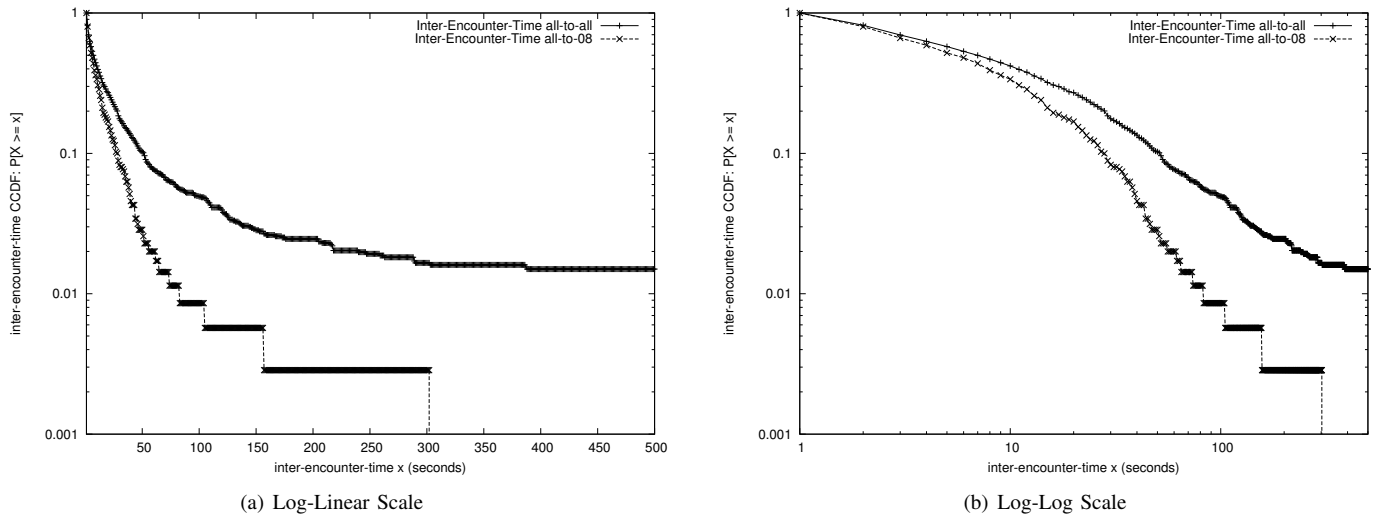| (a) Log-Linear Scale | (b) Log-Log Scale |

Fig. 5. Distribution of inter-encounter time on (a) log-linear scale and (b) log-log scale

- The number of neighbours changes very rapidly, literally from second-to-second, which is not surprising given that soccer players move very fast. Routing therefore has to contend with a highly dynamic topology wherein routes may not persist for long.
- The average number of neighbours is generally low, indicating that the topology is in general sparse (this can also be confirmed visually in our Java GUI). This is because the range of the body-worn device is small (due to battery current-draw limitations and body attenuation effects) compared to the playing area of the soccer field. Fig. 4 plots the probability distribution of the number of neighbours for several nodes, and shows that it is often 0 or 1, indicative of a very sparse topology.
- Connectivity varies with playing position: for example, Fig. 3(b) shows that the midfielder better connectivity (due to prime location in the centre of the field) than the striker or goalkeeper (Fig. 3(a)), both of whom are more likely to be at the extremes of the field. This is also confirmed in Fig. 4 which shows the midfielder is more likely than not to have at least one link at any time, whereas the extreme positions (striker, back, and goalkeeper) have no connectivity very often. Routing can exploit this information to bias its choice of next-hop towards nodes that are more richly-connected on average.
- Connectivity to the base-stations (put together, as shown in Fig. 3(a)) is in general quite poor, which implies that multi-hop routing will be essential if real-time delivery of data from players to the base-stations is required.

B. Encounter Duration and Inter-Encounter Time

Another metric that is known to have an important bearing on the route-selection algorithm in mobile ad-hoc networks is the inter-encounter time [16] (also known as inter-meeting or inter-contact time) between nodes. Most prior studies have relied on exponentially distributed inter-encounter times for tractable anaysis of routing performance; however, recent studies such as [17] have shown that non-exponential behaviour can lead to unbounded routing delays. To see which model best fits the soccer field environment, in Fig. 5 we show the distribution of the inter-encounter time amongst all pairs of nodes, as well as between all transmitters and a specific receiver 8 (the centre midfield B, chosen for its rich connectivity), as obtained from our experimental data.

Fig. 5(a) shows the Complementary Cumulative Distribution Function (CCDF) of the inter-encounter time on log-linear scale. The non-linear nature of the curve, particularly at time scales ranging from 1 to 200 seconds, indicates that the inter-encounter delays do not follow an exponential distribution at such time-scales. In Fig. 5(b) we therefore depict the inter-encounter time on log-log scale, and notice that in the range of 10-100 seconds the inter-encounter delay curve (over all pairs of nodes) is roughly linear, indicative of power-law behaviour in that range. The power-law exponent in this region is estimated at around $\alpha \approx 1.6$. Though [17] estimates analytically that $\alpha < 2$ leads to unbounded routing delays, it does so by extrapolating the inter-encounter delay tail to infinity as a power-law. Our experimental data shows that the curve flattens out (on log-linear scale) beyond around 200 seconds, and in this region inter-encounters are better modelled as exponential. This combination of power-law and exponential behaviour is consistent with reported mixtures [18] seen in inter-meeting times for regular human activity, and result in bounded routing delays unlike the pessimistic estimates in [17].

Another aspect in which the soccer field environment departs from typical mobile ad-hoc networks is the length of time for which two nodes are in continuous contact with each other for exchanging routing messages and data packets. In Fig. 6 we plot the CCDF of the duration for which encounters last (on log-linear scale) over all pairs of nodes and with node 8 (centre midfielder B) as receiver. It is seen that encounters are in general very short (over 90% of encounters lasts no more than 4 seconds), and their duration falls exponentially (the curves

are near-linear on log-linear scale). A routing algorithm cannot therefore assume sufficient contact time with a neighbour in order to be able to forward all its stored messages, or sufficient time for bidirectional communication with the neighbour.
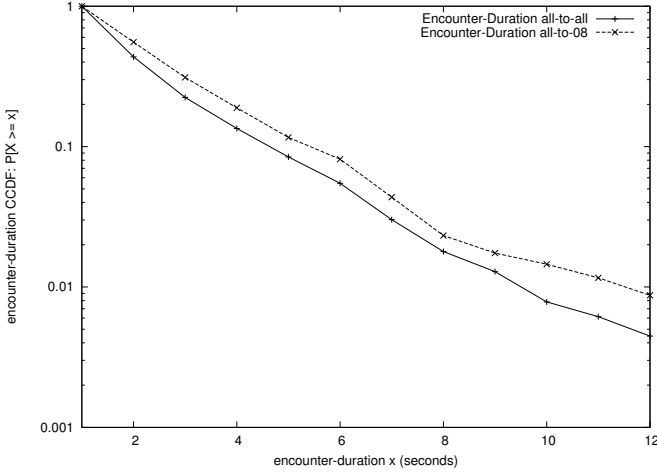

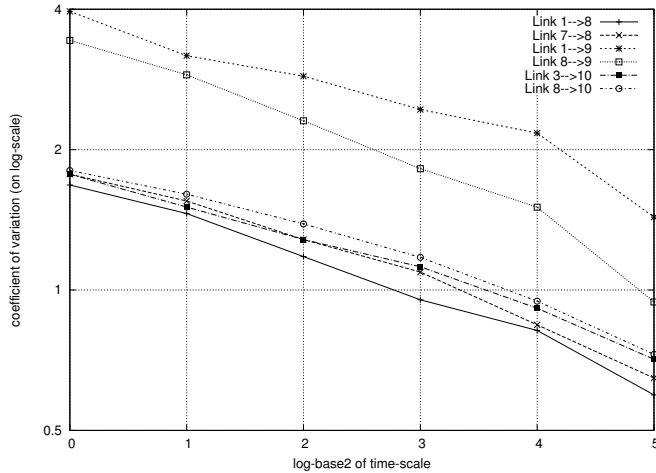Fig. 6. Distribution of encounter duration (log-linear scale)

Fig. 7. Log-log plot of link coefficient-of-variation

To characterise the encounter and inter-encounter distributions and their auto-correlations in a succinct way (which we later employ in our model in section V), we borrow a technique used for the analysis of long-range dependent (LRD) traffic. Considering a link between a pair of nodes, at a given time step, we use a 1 to depict presence of the link and 0 its absence. For this link, we therefore have from our experimental data a time-sequence of 0s and 1s. We consider this sequence in blocks of 2^s samples, for given s, and for this resulting sequence we compute the mean, variance, and coefficient-of-variation $\beta(s)$ (in effect these metrics are computed at time-scale 2^s). Log-log plots of $\beta(s)$ versus s are routinely used in the literature to depict self-similarity and to estimate the corresponding Hurst parameter $H \in [0.5, 1)$. In Fig. 7 we show such a plot for several links (we picked two links each from centre, forward, and backward playing positions), and

observe that the curves can be approximated as straight-lines with slope $-(1 - H)$, yielding a Hurst parameter $H \approx 0.75$. This single-parameter captures in a succinct way the link auto-correlations, and will be used in the connectivity model we develop in section V.

IV. LINK CORRELATIONS

Unlike many mobile ad-hoc networks in which we can reasonably assume that users move independently, in a soccer game we would expect player movements to have significant correlations. For example, when the team is attacking the opponent's goal, several players in the forward and midfield positions can be expected to move towards the opponent's goal simultaneously, and conversely when the home goal is being attacked the defenders and midfielders will likely fall back towards the home goal to protect it. This leads to correlations amongst links, an aspect which can have a significant impact on the performance of routing algorithms.

Correlations are computed as follows: if x_t is a binary variable that is 1 or 0 depending on whether link x is present or absent at time step t, then the cross-correlation at time lag k between two links x and y is given by [19, Sec 12.1.2]:

$$\rho_{xy}(k) = \frac{\frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(y_{t+k} - \bar{y})}{\sigma_x \sigma_y}, \quad k = 0, \pm 1, \pm 2, \ldots \tag{1}$$

where n is the number of sample points, $\bar{x}$ is the estimated mean and σ_x the estimated standard deviation of x.

In Fig. 8(a) we show the correlation between node 3's (centre midfield A) and node 10's (centre back) links to node 8 (the centre midfield B) for lags in the range $[-20, 20]$ seconds. Two things are noteworthy from this plot: (a) the correlations are positive, meaning that when node 3 is close to node 8, node 10 is also likely to be close to node 8; this suggests nodes 3 and 10 move in a co-ordinated way quite often, and (b) the correlations are high (> 0.2) for lag close to 0, and decay rapidly as the lag moves away from 0. This is not surprising, because the fast nature of the game implies that the locations of the players can vary significantly from one minute to the next, making them nearly independent. In Fig. 8(b) we show the correlation between node 1's (centre attack) and node 10's (centre back) with node 8 (centre midfield B). This time we notice that the correlations are predominantly negative (< -0.2 for lags close to 0), which is understandable: when the team is attacking, the midfielder is more likely to be close to the striker and far from the defender, while the converse is true when the team is defending their own goal. Again we notice that the anti-correlations decay with time due to the rapid movement of players in the game.

Having seen specific examples of correlated and anti-correlated links, let us now examine how pervasive the correlations are amongst all links. We have data from 9 transmitters (as stated earlier the devices on two players got damaged and the data was lost) and 10 receivers (9 players and a "virtual" base-station which aggregates the data from all base-stations), giving us a total of 81 possible uni-directional links. During the
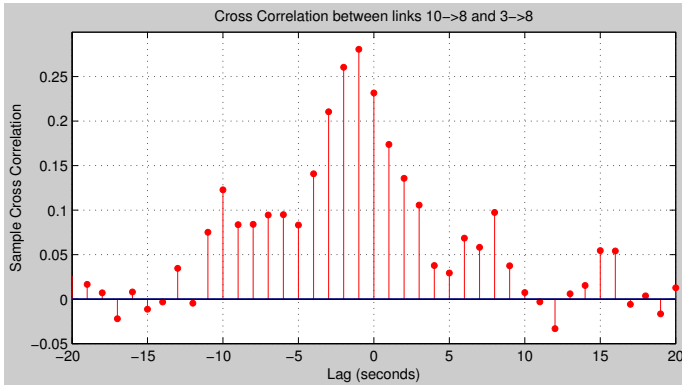
(a) correlation between links $3 \to 8$ and $10 \to 8$
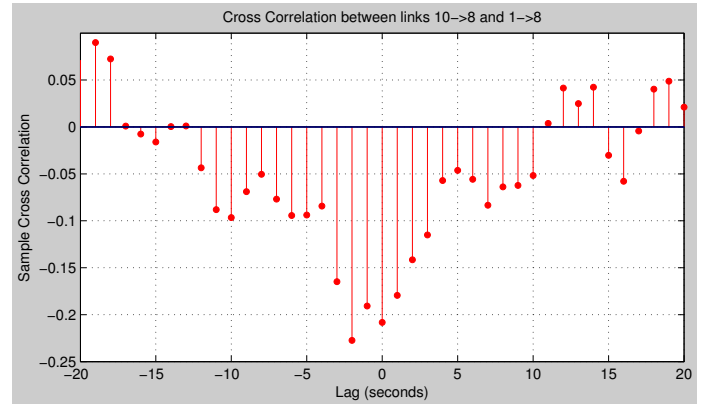
(b) correlations between links $1 \to 8$ and $10 \to 8$

Fig. 8. Correlation of (a) link $3 \to 8$ with $10 \to 8$ and (b) link $1 \to 8$ with $10 \to 8$, as a function of lag (in seconds)
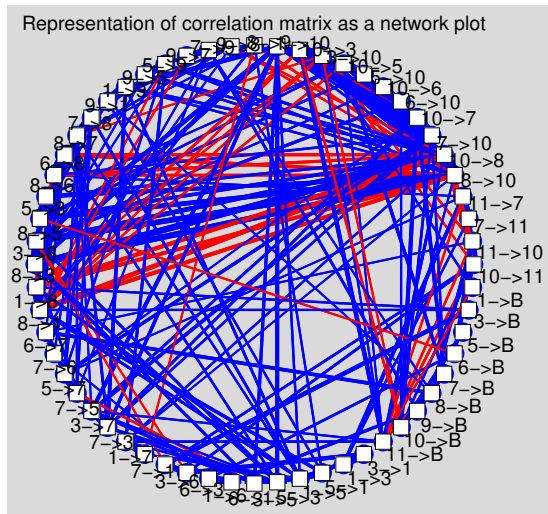

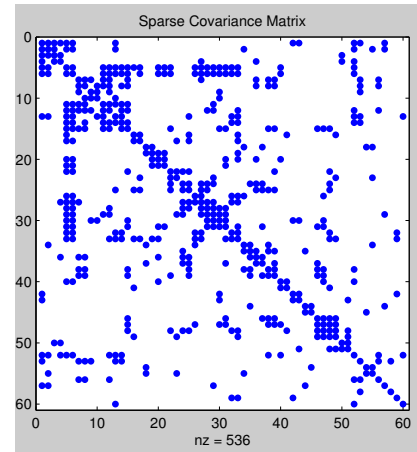Fig. 9. Correlations depicted via inter-connections

Fig. 10. Correlations shown as a matrix

10 minute measurement period, 60 of these links had sufficient (at least 10) packet receptions to be statistically significant, and so we compute pair-wise cross-correlations for these 60 links. To eliminate random chance of correlated values, we also estimate the P-value [20] (used for statistical hypothesis testing) for each pair, and only retain those that are statistically significant (i.e. have $P \leq 0.05$). To help the reader visualise the correlations, we place the 60 links as nodes on a circle in Fig. 9 and draw a line between two nodes if they have significant correlation: blue lines depict positive correlation while red lines depict negative correlation, and the higher the correlation (or anti-correlation), the thicker the line. Also, links have been ordered on the circle so that the two directions of the link are adjacent to each other (so that correlations between the two directions of a link do not clutter the plot). We see that often links to a given player (say player 8, the centre midfielder B) can have strong correlations with each other, but some links with distinct end-points can also be correlated (for example, link $1 \to 10$ is correlated with links $8 \to 3$, $5 \to 3$ and $7 \to 3$, indicating that when an attacker 1 is close to a defender 10, many players are likely to be huddled closer to the ball, and hence more likely to be connected to a midfield player such as

3). To give an idea of how widespread the correlations are, in Fig. 10 we show the elements of the 60x60 correlation matrix that are significant. The matrix is seen to be relatively sparse, which suggests that determining the entire matrix (needed as input to our model) may not be as onerous as one may think.

V. A CONNECTIVITY MODEL FOR SOCCER PLAYERS

We now develop a model that can generate synthetic dynamic topologies with similar stochastic properties to those observed empirically. Such a model would be useful in generating long traces to simulate the performance of different routing strategies for soccer player monitoring, and also allow key parameters such as link auto- and cross-correlations (which may depend on a team's playing style) to be varied to study their impact on routing performance.

A. Prior Work on Mobility and Connectivity Modelling

There are two broad approaches to modelling the topology arising in an ad-hoc network: the first models the **mobility** of individual nodes in terms of their physical location, and deduces connectivity between two nodes at any time instant based on their proximity to each other. Mobility models abound in the literature (see [21], [22] for a survey), and can consider nodes independently (e.g. Random Waypoint model, Levy Walk model [23], etc.) or as groups (e.g. Reference Point

Group Mobility model and Pursue Mobility model). While it may be possible to fit the movement of soccer players to one of the mobility models in the literature (indeed one such attempt is made in [24] using a learning automata fed by player position information obtained via optical tracking from multiple cameras), such an approach has limitations: (a) it requires empirical data on player location at each time instant, which is not easy to obtain, (b) it often makes assumptions about how nodes behave, which may be difficult to validate, and (c) it typically assumes a circular transmission range, which ignores body-orientation effects in the athlete monitoring context.

The approach we choose to take in this paper is to model directly the evolution of **connectivity** (aka topology), rather than inferring it from a mobility model. There is little prior work based on this approach: one example is [25] that proposes a statistical encounter-based model in the context of delay tolerant networks (DTNs). However, their model assumes links to be independent, which is inadequate for capturing correlations that we have shown to exist in team sports such as soccer. The model we present next overcomes this important limitation.

B. Model Requirements

We seek a model that takes the following inputs: (a) Number of players, base-stations, and links, (b) Mean and variance for each link (the link is binary in each time-step: 0 if down and 1 if up), (c) Auto-correlation of the links (to keep the model simple we assume that all links have similar auto-correlations), specified via the Hurst parameter (section III-B) or auto-regressive coefficients (discussed in the next subsection), and (d) Cross-correlation between each pair of links, specified as a covariance matrix.

The model should output for each successive time-step the connectivity topology, i.e., the set of links that are up at that time-step. If an empirical trace is available from which the input parameters were derived, then the generated topology should statistically match the empirical trace in the following metrics: (a) for each link, the on/off (0/1) distribution, (b) the distribution of the number of active links in the network, (c) the distributions of encounter durations and inter-encounter times, and (d) the correlations between every pair of links.

C. The Model

We use W to denote the total number of links (player-to-player as well as player-to-base). The covariance matrix (which is an input to the model) is denoted by $\mathbf{C}$, and is of dimension $W \times W$. Element C_{ij} denotes the covariance between the i-th and j-th links, and is related to the correlation defined in Eq. (1) by $C_{ij} = \sigma_x \sigma_y \rho_{ij}(0)$ (note that to reduce complexity our model directly incorporates correlation at lag 0 only; correlations at other lags will follow from cross-correlations at lag 0 combined with the auto-correlations of the links). Further, each diagonal entry C_{ii} corresponds to the variance of the binary variable associated with the i-th link.
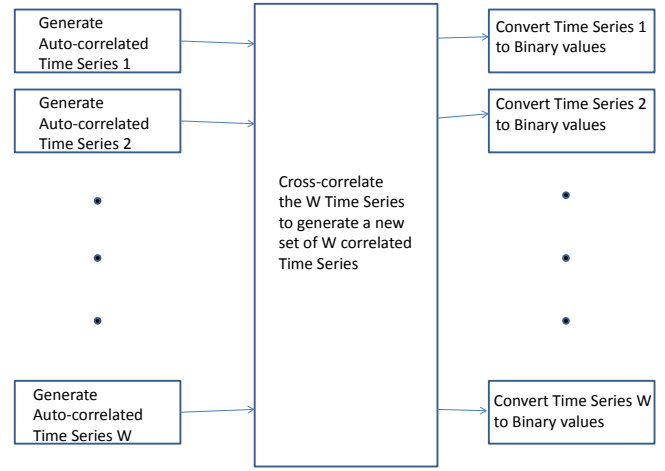

Fig. 11. Flow diagram of our model for generating time-varying topology

A valid covariance matrix is required to be symmetric (i.e. $C_{ij} = C_{ji}$) and positive definite (aka $\mathbf{C} > \mathbf{0}$).

The general flow of our model is shown pictorially in Fig. 11, and broadly consists of three steps: step 1 generates independent random variables, one per link, with the appropriate auto-correlation, step 2 mixes them to create the correct cross-correlations, and step 3 converts them from continuous to discrete (binary) values so they correspond to links being up or down in each time-step. These steps are elaborated next.

1) Generating Auto-Correlated Time Series: The first step is to generate W independent time-series of link variables with desired auto-correlation. Different methods could be used for generating the time-series based on how the auto-correlation is specfied. We try two methods based on analysis of the field data we collected. The **first** method uses the long-range dependent characterisation of link connectivity we presented in Fig. 7. In this approach the link auto-correlation can be specified very succinctly by a single number: the Hurst parameter H, which for our experimental data was $H \approx 0.75$ across all links. We generate long traces of normalised fractional Gaussian noise (fGn) (with zero mean and unit variance) for this H, using the filtering method developed in [26]. The W fGn time-series thus generated each have the requisite auto-correlation properties; subsequent steps will cross-correlated them, and shift/scale them to have the appropriate link-specific mean and variance.

The **second** method we use to generate the auto-correlated time-series assumes a linear stationary auto-regressive (AR) model [19] of appropriate order. An order p AR process derives the sample $x(t)$ at time-step t as:

$$x(t) = \sum_{k=1}^{p} a_k x(t-k) + w(t) \tag{2}$$

The auto-correlation in the above process stems from the fact that the sample at time-step t is a weighted sum of the previous p samples, with an additional random noise component that has zero mean and constant variance. Based on the auto-

correlation properties of links at different lags, we estimated that an AR process of order $p = 20$ matched our experimental data well. We then used the Yule-Walker method (`aryule` in Matlab) to estimate the AR coefficients, which were then applied as a filter to sequences of random white Gaussian noise to yield the desired auto-correlated time-series.

2) Cross-Correlating the Time Series: Having generated W sequences of independent variables with appropriate auto-correlations, this step introduces the cross-correlations as per the specified covariance matrix $\mathbf{C}$. The general idea is to take appropriate linear combinations of the W independent random variables to generate a new set of W random variables that have the desired cross-correlations. To this end we first determine the Cholesky decomposition of the covariance matrix, i.e. find the lower-triangular matrix $\mathbf{L}$ such that $\mathbf{C} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$ where $\mathbf{L}^{\mathbf{T}}$ denotes the transpose of $\mathbf{L}$. The symmetric positive definite nature of $\mathbf{C}$ ensures that such decomposition exists and can be computed relatively easily (using `chol` in Matlab). However, for computation stability it is desirable to have $\mathbf{L}$ as sparse as possible. To this end we tried several methods to permute the rows and columns of $\mathbf{C}$ to make it more diagonally dominant, and chose the symmetric approximate minimum degree permutation (`symamd` in Matlab) to obtain the most sparse Cholesky decomposition $\mathbf{L}$.

Given a vector of W uncorrelated random variables $\mathbf{x(t)} = (x_1(t) \ \dots \ x_W(t))$, we generate a vector of W correlated random variables $\mathbf{y(t)} = (y_1(t) \ \dots \ y_W(t))$ with covariance as per matrix $\mathbf{C}$ using:

$$\mathbf{y(t)^T} = \mathbf{Lx(t)^T} \tag{3}$$

where $\mathbf{L}$ is the Cholesky decomposition of $\mathbf{C}$.

3) Converting to Binary Variables: Random variable $y_i(t)$ above corresponds to link-i at time-step t, and already has requisite correlation with $y_i(t')$ (i.e. auto-correlation) as well as with $y_j(t)$ (i.e. cross-correlation with other links). In this step we convert continuous-valued $y_i(t)$ to corresponding binary values $z_i(t)$ by comparing with threshold T_i, i.e. $z_i(t) = 1$ if $y_i(t) > T_i$, and 0 otherwise. The threshold T_i is chosen so that

$$P[y_i(t) > T_i] = P[z_i(t) = 1] \tag{4}$$

The right side is the mean value $E[z_i(t)]$ of the link, which is available as input to the model. Random variable $y_i(t)$ is a linear combination of Gaussian variables $x_j(t)$, with weights known from the Cholesky decomposition matrix and the fGn/AR parameters, and therefore $y_i(t)$ is also Gaussian with known variance. Using tabulated values of the CDF of the normal distribution, the threshold T_i in Eq. (4) can easily be computed, and this threshold is then used for converting the model output to binary.

D. Validating the Model

To validate the model we compared its synthetic output with the empirical trace obtained in the field. Parameters estimated from the empirical trace, such as mean and variance of each link, their auto-correlations, and the covariance matrix, were

fed as input to the model. The trace output by the model (i.e. binary time-series for each link) was subjected to the Kolmogorov-Smirnov (K-S) "goodness of fit" test and found to match the distribution seen in experiment for all links. Moreover, the statistical metrics directly controlled by the model, such as link mean, variance, auto-correlations, and cross-correlations, were found to match well.
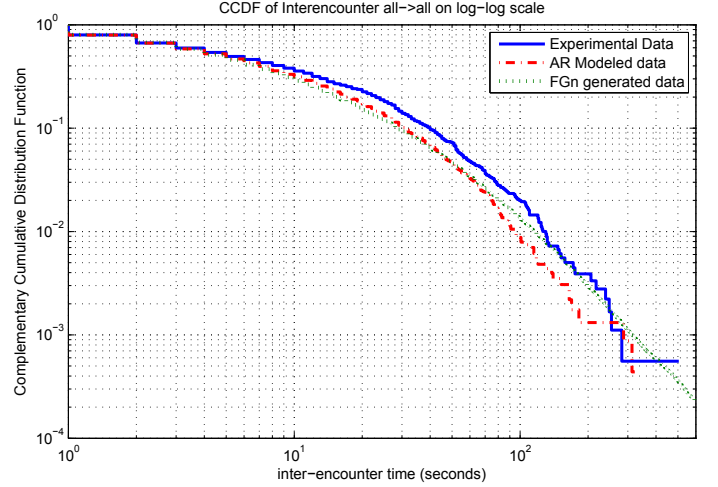

Fig. 12. CCDF of inter-encounter times from Experiment and Model
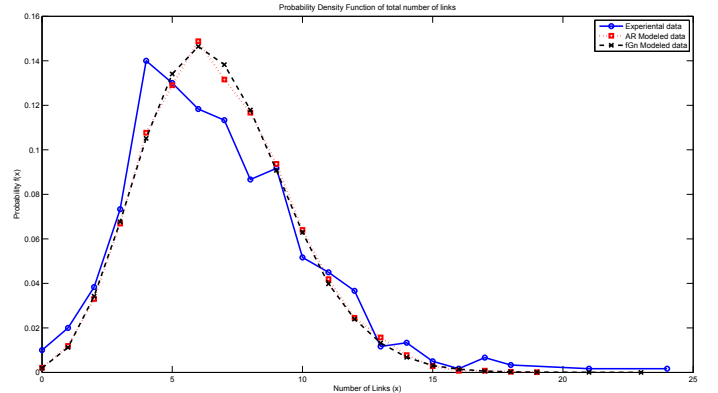

Fig. 13. PDF of total links from Experiment and Model

We show that metrics that are not directly controlled by the model corroborate well with experiment. One such important metric is the inter-encounter time, which is a by-product of link auto-correlation. In Fig. 12 we show that the CCDF of the inter-encounter time seen in model trace data matches very well with experiment, confirming that our model has captured auto-correlations correctly. Another important metric is the number of links in the topology at any time-instant, which in turn is influenced by the cross-correlations amongst links. The PDF of the number of links shown in Fig. 13 again shows that our model matches well with experiment, affirming that the cross-correlations are also captured correctly by our model. Several other metrics (such as node degrees) seen in our model output were also found to match well with empirical data, and are omitted here due to lack of space.

E. Using the Model

Our model is fairly general: it takes as input the mean and variance of individual links, their auto- and cross-correlations, and outputs an arbitrary-length time-series of dynamic topologies with desired stochastic properties in terms of number of links, inter-encounter times, neighbour distributions, etc. Our model does not make any assumptions specific to the operating environment, and as such can be applied to model dynamic topologies arising in any mobile ad-hoc or delay tolerant network studies.

Deducing the input parameters to the model, in particular the cross-correlations between links, requires access to sufficient experimental data. Even then, estimating the parameters can be tricky: for example, the same soccer team plays each game differently depending on their strategy and their opponent. Nevertheless, we think reasonable approximations can often be made: for example, we can expect that links between players in similar positions (e.g. defending) are more highly correlated with each other than with a link between players in different positions (e.g. defender and forward), or that a midfielder's link to a left-wing player will generally be negatively correlated with his link to a right-winger. We believe that capturing even a few key correlations (in a sparse covariance matrix) can give us much more realistic dynamic topologies for routing studies as compared to using overly simplistic models that ignore correlation effects.

VI. Conclusions

In this paper we have studied the wireless connectivity that arises amongst soccer players in the field. Understanding such connectivity is important for designing algorithms for multi-hop routing of athlete physiological data in real-time to base-stations positioned around the field. We outfitted a first-division soccer club team with wireless devices and collected empirical data on their connectivity during multiple games, something that to the best of our knowledge has not been undertaken before. Our first contribution profiles the connectivity data to highlight several aspects relevant to the design of appropriate routing mechanisms, such as that the topology is sparse and highly dynamic, and that inter-encounter times have a power-law distribution. Importantly, we identified that links in the soccer field can have significant correlations, due to the inter-dependent way in which players move, and this can impact routing. For our second contribution we therefore develop a novel model for synthetic generation of dynamic topologies arising in the soccer field. Our model is quite general, and considers both auto-correlations within and cross-correlations between links. It is shown to generate traces that match empirical data in key stochastic metrics, and forms the basis of routing studies for environments such as the soccer field that exhibit correlated mobility.

Our current work in [27] develops multi-hop routing schemes appropriate for soccer player monitoring. We will also undertake further experimental work replacing the Crossbow motes with wearable body-area networking devices as they become available.

References

[1] S. Armstrong, "Wireless Connectivity for Health and Sports Monitoring: A Review," *British Journal of Sports Medicine*, vol. 41, no. 5, pp. 285–289, 2007.

[2] GP-Sports, "SPI Elite." [Online]. Available: http://gpsports.com

[3] Sensixa-Ltd., "e-AR." [Online]. Available: http://www.sensixa.com

[4] VxSport, "VxLog Model 100 and 220." [Online]. Available: http://www.vxsport.com

[5] Toumaz-Technologies-Ltd., "Digital Plaster using Sensium^{TM}," online. [Online]. Available: http://www.toumaz.com

[6] Power-Paper, "Power Patch Platform." [Online]. Available: http://www.powerpaper.com

[7] R. Smith and C. Loschner, "Biomechanics Feedback for Rowing," *Journal of Sports Sciences*, vol. 20, no. 10, pp. 783–791, Feb 2002.

[8] J. Hallberg, S. Svensson, A. Ostmark, *et al.*, "Enriched media-experience of Sport Events," in *Proceedings of the 6th Workshop on Mobile Computing Systems and Applications*, December 2004.

[9] T. Konberg, C. Ohult, and J. Delsing, "Measuring breathing and heart rate data with distribution over wireless IP networks," in *Instrumentation and Measurement Technology Conference*, 2003.

[10] J. Afonso, H. Silva, P. Oliveira, J. Correia, and L. Rocha, "Design and Implementation of a Real-Time Wireless Sensor Network," in *SENSORCOMM*, October 2007, pp. 496–501.

[11] *MPR/MIB Users Manual*, Crossbow Technology, Inc, June 2007, available at http://www.xbow.com.

[12] V. Schnayder *et al.*, "Sensor Networks for Medical Care," Division of Engineering and Applied Science, Harvard University, Tech. Rep. TR-08-05, 2005.

[13] S. Xiao, A.Dhamdhere, V. Sivaraman, and A.Burdett, "Transmission Power Control in Body Area Sensor Networks for Healthcare Monitoring," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 1, pp. 37–48, Jan 2009.

[14] A. Kurusingal, A. Dhamdhere, and V. Sivaraman, "Modeling Signal Strength of Body-Worn Devices," in *IEEE LCN*, Denver, CO, USA, Oct 2010.

[15] V. Sivaraman, "Athlete Monitoring Project." [Online]. Available: http://www.ee.unsw.edu.au/~vijay/athlete

[16] H. Cai and D. Eun, "Toward Stochastic Anatomy of Inter-Meeting Time Distribution under General Mobility Models," in *ACM MobiHoc*, Hong Kong, 2008.

[17] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms," in *IEEE Infocom*, Barcelona, Spain, Apr 2006.

[18] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnovic, "Power Law and Exponential Decay of Inter-Contact Times Between Mobile Devices," in *ACM MobiCom*, Montreal, Canada, Sep 2007.

[19] G. Box, G. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed. Wiley, 2008.

[20] G. Dallal, "Historical background to the origins of p-values and the choice of 0.05 as the cut-off for significance," 2007. [Online]. Available: http://www.jerrydallal.com/LHSP/p05.htm

[21] F. Bai and A. Helmy, *Wireless Adhoc and Sensor Networks*. Kluwer Academic Publishers, 2004, ch. A Survey of Mobility Models in Wireless Adhoc Networks, pp. 1–30.

[22] T. Camp, J. Boleng, and V. Davies, "A Survey of Mobility Models for Ad Hoc Network Research," *Wireless Communication and Mobile Computing (WCMC): Special Issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, vol. 2, no. 5, pp. 483–502, 2002.

[23] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "On the Levy-Walk Nature of Human Mobility," in *IEEE INFOCOM*, Phoenix, AZ, USA, Apr 2008.

[24] A. H. Jamalian, A. R. Sefidpour, M. T. Manzuri-Shalmani, and R. Iraji, "SME: Learning Automata-Based Algorithm for Estimating the Mobility Model of Soccer Players," in *IEEE Intl. Conf. Cognitive Informatics (ICCI)*, Lake Tahoe, CA, Aug 2007, pp. 462–469.

[25] F. Tan, Y. Borghol, and S. Ardon, "EMO: A Statistical Encounter-Based Mobility Model for Simulating Delay Tolerant Networks," in *IEEE WoWMoM*, Newport Beach, CA, USA, Jun 2008.

[26] D. Ostry, "Synthesis of Accurate Fractional Gaussian Noise by Filtering," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1609–1623, Apr 2006.

[27] A. Dhamdhere, H. Chen, A. Kurusingal, V. Sivaraman, and A. Burdett, "Experiments with Wireless Sensor Networks for Real-Time Athlete Monitoring," in *IEEE SensApp*, Denver, CO, USA, Oct 2010.