Double Regression: Efficient Spatially Correlated Path loss Model for Wireless Network Simulation

Seon Yeong Han
Dept. of Computer Science,
Korea Advanced Institute of Science
and Technology (KAIST)
Email: shan8017@kaist.ac.kr

Nael B. Abu-Ghazaleh
Computer Science Department,
Binghamton University
Email: nael@cs.binghamton.edu

Dongman Lee
Dept. of Computer Science
Korea Advanced Institute of Science
and Technology (KAIST)
Email: dlee@cs.kaist.ac.kr

Abstract—The accuracy of wireless network packet simulation critically depends on the quality of the wireless channel models. These models directly affect the fundamental network characteristics, such as link quality, transmission range, and capture effect, as well as their dynamic variation in time and space. Path loss is the stationary component of the channel model affected by the shadowing in the environment. Existing path loss models are inaccurate, require very high measurement or computational overhead, and/or often cannot be made to represent a given environment. The paper contributes a flexible path loss model that uses a novel approach for spatially coherent interpolation from available nearby channels to allow accurate and efficient modeling of path loss. We show that the proposed model, called Double Regression (DR), generates a correlated space, allowing both the sender and the receiver to move without abrupt change in path loss. Combining DR with a traditional temporal fading model, such as Rayleigh fading, provides an accurate and efficient channel model that we integrate with the NS-2 simulator. We use measurements to validate the accuracy of the model for a number of scenarios. We also show that there is substantial impact on simulation behavior (e.g., up to 600% difference in throughput for simple scenarios) when path loss is modeled accurately.

I. INTRODUCTION

Simulation is widely used for performance evaluation in wireless and mobile network research due to its flexibility, controllability and observability compared to testbeds or emulation studies. However, the validity of simulation studies has been criticized due primarily to the poor accuracy of the wireless channel models [1], [2], [3], [4], [5], [6], [7].

In response to this criticism, more accurate and realistic temporally fading channel models, e.g., Rayleigh-Ricean fading [8] and Nakagami fading [9], have been implemented in network simulators. However, these models account for the fast fading component of the channel model which exhibits temporal correlation but little spatial correlation [10]. On the other hand, the stable component of a signal is determined by shadows from large objects in an environment; this component, i.e., path loss component, is still commonly modeled assuming idealized distributions. However, empirically, path loss has been show to be spatially correlated as RF shadows from large objects tend to affect nearby channels similarly. The path loss component determines the mean signal strength and therefore significantly affects link quality [11].

A. Path Loss Models are Important

Compared to the efforts in modeling temporal fading, less attention has been paid to obtaining an accurate mean signal strength for network packet simulation. Existing simulators typically use simple disc-shape path-loss models, such as the free space model or the two ray ground model, in which path-loss increases as a function of distance. However, in practice, the path loss depends on the shadowing in the environment and how it affects the channel between a sender and a receiver. Disc-shape path-loss models do not account for shadowing and are therefore not realistic. Statistical approaches to account for shadowing, such as the log-normal shadowing model, are also not effective because they do not account for the spatial correlation and the temporal stability of the path loss –often an independent path loss value is generated with every packet transmission even if no mobility is present.

The use of imprecise models can dramatically affect the simulation leading to inaccurate results and false conclusions. With respect to path loss, it impacts the perceived signal power when a transmission occurs, which impacts both the act of transmission in CSMA protocols, as well as the success of transmission at the receiver with or without the presence of interfering transmissions. Moreover, spatial and temporal correlation can influence the loss pattern. These effects at the channel level can be compounded as they interact with the upper protocol layers; for example, the packet loss pattern affects protocols such as TCP which adapt their sending rate with every lost packet. There is a great need for accurate models of path loss to enable accurate simulation of wireless networks.

B. Site Specific Path Loss Estimation Approaches

There are two major approaches to accurate path loss estimation: (1) site surveys use extensive measurement to map the signal power from every location to every other location in a site [12]; and (2) Ray tracing: instead of measurement, signal propagation in a site is tracked in detail using geographical and material information specific to the site; the direct, reflected, refracted, diffracted, and scattered rays between a sender and a receiver are summed to estimate the average received signal strength for a specific location [13], [14], [15]. Both these approaches are accurate, but require significant measurement or computational overhead, which makes them unsuitable for use in a packet simulator. Moreover, they cannot easily be extended to model different environments.

C. Proposed Approach: Double Regression for spatially coherent path loss modeling

In this paper, we propose a path-loss model for wireless network simulation capable of spatially coherent estimation of

path loss in the presence of both sender and receiver mobility. The model exploits the spatial correlation of path loss: since path loss depends on the shadowing from the environment, if two links are close enough that they share the same environment, their path loss components are correlated. The spatial correlation of path-loss has been empirically observed in both outdoor and indoor scenarios [16], [17], [18].

To obtain a plausible path-loss value for a new link, we use a novel regression algorithm to estimate the channel from nearby links. Known links are stored in an R-tree (a data structure for spatial indexing [19]) indexed both by sender and receiver location. When generating a new link, we search the R-tree for known nearby links. We then interpolate twice from these nearby links, once to account for the difference in receiver location and once for the sender; the algorithm is called *Double Regression* (DR). If there are insufficient reference data (known links that are sufficiently close) for regression, we generate the path loss value from a log-normal distribution with mean and variance obtained by interpolation so that the spatial correlation between existing neighbors and the random variables is kept. Finally, the newly generated link is stored in the R-tree to serve as a known link. As a result, unlike existing models, the path loss remains spatially coherent (correlated to nearby channels) and temporally consistent (path loss values are preserved over time, mirroring the long-term nature of shadowing which determines the path loss).

We envision the model to be used in a way where a set of known channels are built over time initially starting from stochastic values, and then using interpolation as more channels are known. However, it is also possible to seed the model with measurement or ray-tracing channels creating simulations that are site specific. When seeded with known channel values, the model can "fill-the-gap" with realistic values between the known measured ones, allowing faster site surveys or ray-tracing analysis where only a few channels are measured or ray-traced.

We integrate the path-loss with a temporal fading channel model (Rayleigh fading) to provide a complete channel model, and show that different path loss estimates cause substantially different link quality estimates. We show that even in a simple two-hop scenario, the use of spatially coherent models can result in up to 600% difference in estimated throughput over existing models when there is only 10dBm difference in path loss. These experiments emphasize the fact that accurate modeling of path loss is critical to plausible channel models.

The remainder of the paper is organized as follows. Section II overview some background and related work. In Section III presents some background regarding regression. In Section IV, we present the proposed double regression algorithm for spatially correlated path-loss estimation. We present our experimental evaluation of the channel model and the assessment of its impact on the accuracy of simulation studies in Section V. Finally, Section VI presents some concluding remarks.

II. BACKGROUND AND RELATED WORK

Wireless propagation is a complex phenomena: multiple versions of the transmitted signal that take different paths through the environment, suffering deflection, refraction, and dispersion, combine to make the received signal. Thus, existing channel models attempt to model the overall received signal as the sum of two components: (1) path loss models: which determines the long term average received signal power. This component depends on the RF shadows from large static features of the environment (e.g., buildings); and (2) Small-scale multi-path fading models: these models attempt to capture the time-varying nature of the signal due to the multi-path effect. This component is modeled using a zero mean multi-path model, such as the Rayleigh fading model or the Nakagami-m fading model.

A. Typical Simulator Models

To implement a complete channel model for a packet simulation, the two models (path loss and small scale fading) are needed, summing up to give the instantaneous received signal power. This section overviews two commonly used combinations of path loss models and small-scale multi-path models: ideal path loss model with Rayleigh fading model (IPL-Rayleigh) and log-normal shadowing model with Rayleigh fading model (LNSM-Rayleigh).

The ideal path loss model in the IPL-Rayleigh can be the free space model, the two-ray ground model, or any other isotropic distance dependent path loss model. IPL-Rayleigh is widely used because of its simplicity [8], [9]. Due to the small-scale multi-path fading channel model, it has a realistic channel in terms of temporally correlated fading. However, a parameter of the fading channel model, the mean power, is obtained from the ideal path loss models; thus, the mean power and therefore the average link quality of a link do not reflect a real site with characteristic shadowing effects.

The LNSM-Rayleigh is more realistic because the path loss is obtained by the log-normal shadowing model, which is widely accepted as a site-specific stochastic model [20]. However, the LNSM model assumes independent and identically distributed path loss for points at the same distance from a sender; it does not consider the spatial correlation between nearby points [16], [18]. Gudmundson showed that the high space correlation between two nearby receivers, but showed that this correlation exponentially decreases with distance between them [16]. Jalden et al. measured signal strength along a route of 90 meters in an indoor scenario, and found out path loss at close links are highly correlated [18]. A precise RT also supports the spatial correlation. Wang, Tameh and Nix discovered the same exponentially decreasing spatial correlation by using RT [21]. Thus, this model is not appropriate for the case where spatial correlation exists: we show later that spatial correlation has significant impact on simulation performance.

B. Site Surveys and Ray Tracing

One approach to providing highly accurate path loss exponent is to use a site survey. A site survey is a process where signal power measurements are performed throughout an area which will be used as the basis for simulation. However, this process is extremely time consuming and expensive; often it is only feasible to perform measurements in limited number of locations, and the quality of measurement depends on equipment and skill [12]. Due to this reasons, several methods to reduce the number of required measurement in planning of

access points have been proposed [22], [23]. Stepanov et al use a commercial site survey tool (WinPROP) and present the significance of using accurate propagation model in MANET study [24].

In the second approach, called Ray Tracing (RT), detailed geographical information about a site is used. In particular, the direct, reflected, refracted, diffracted, and scattered rays between a sender and a receiver are summed to estimate the average received signal strength for a specific location [13], [14], [15]. RT is highly accurate if the precise geographic information (GIS) (e.g., materials, facets, and edges) about the environment, as well as the sender and receiver is available. However, when the typical number of facets and edges are used, the computational complexity easily gets into the trillions of rays [25]. This computational cost will be prohibitive for packet simulation especially in the presence of mobility (since the path loss would have to be continuously re-estimated).

Dricot and Doncker attempted to implement the ray-tracing model in a slowly moving environment in a network simulator [26]. Because of the change of the environment, they re-calculated the ray-tracing result every 0.05sec, which was assumed to be coherent time in the 2.4GHz channel. Even though they used the same result without re-calculation for 0.05 sec, they noticed that the simulation runtime is increased 100 fold. There is a rich set of studies for increasing the efficiency of the ray-tracing, however, the computational complexity remains very high [27]. For example, one solution for this problem is to store the precomputed ray-tracing result for a whole city, if the location of the sender is fixed [27] (e.g., base station). However, this is not applicable to ad-hoc networks where both senders and receivers move. Due to computational inefficiency and the expensive cost of detailed GIS data, RT is not used as a path-loss model for packet simulators.

C. Efficient Spatially Coherent Models

Our goal is to develop efficient spatially coherent models that can also be used to model specific sites. Related to our work, Xu et al [28] use regression to estimate link quality to add spatial coherence to temporal correlation. They do not consider separate regression from the source and destination side. Moreover, they attempt to estimate the full link quality, rather than just the stationary spatially coherent component representing path loss. In particular, a sliding window of recently perceived link qualities are maintained and as time passes, the quality of the link can drift by arbitrary amounts.

The Sum-of-Sinusoids (SOS) model was initially developed by Rice [29] and advanced and extended by others [30], [31], [21]. For two arbitrary links, the autocorrelation is determined by the distance between the senders and the receivers as each link's path-loss is a Gaussian random variable. Autocorrelation is achieved by selecting the frequency set defined in the SOS formula carefully. Like DR, SOS can generate a spatially correlated log-normal shadowing model and is computationally efficient. However, unlike DR, SOS is completely stochastic and cannot be used to create site specific models.

III. BACKGROUND: REGRESSION

The proposed approach uses regression to estimate the path loss of a link given nearby links whose path loss is known (e.g., from measurement or those previously generated by the simulation). Regression is a process of least square fitting used to estimate a dependent value (also called a response variable) from one or more independent variables (also called explanatory variables) [32]. The regression equation is a function of independent variables. The regression coefficients in the regression equation are estimated by a set of observations of the dependent variables with the corresponding independent variables.

A general regression equation is expressed by

$$p = \beta_0 + \sum_{i=1}^{k} \beta_i \Phi_i \qquad (1)$$

where β_i represents the regression coefficient, and Φ_i is a function of the independent variable. The most simple form of regression equation is a linear regression that has the following equation:

$$p = \beta_0 + \beta_1 x + \beta_2 y \qquad (2)$$

In this case, $\Phi_1 = x$ and $\Phi_2 = y$. A factorial regression equation involves the mixture of two independent variables as the following:

$$p = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy \qquad (3)$$

In this case, $\Phi_1 = x$, $\Phi_2 = y$, and $\Phi_3 = xy$. Given a set of m observations $\mathbf{O} = \{(x_j, y_j, p_j) | j = 1, 2, .., m\}$, the regression coefficient can be obtained by

$$\beta = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{p} \qquad (4)$$

where $\mathbf{p} = (p_1, p_2, .., p_m)^T$ and Φ is a $m \times k$ matrix of $[\Phi_i(x_j, y_j)], i = 0, 1, .., k$ and $j = 1, 2, .., m$.

IV. DOUBLE REGRESSION: SPATIALLY CO-HERENT ESTIMATION OF PATH LOSS

Our goal in this paper is to develop a plausible model for the long term stable component of a wireless channel. Plausibility for this path loss component entails spatial correlation such that nearby channels have correlated path loss. If there are no neighboring links (defined by emperically observed correlation distance), the estimated path loss can be independent.

This section explains how our proposed model generates a plausible estimated of path-loss in a computationally efficient way. This paper follows the notation used in log-normal shadowing description by Rappaport [10]: i.e., $X_\sigma(t_x, t_y, r_x, r_y)$ is a normally distributed random variable $N(0, \sigma^2)$ that models the path-loss component of a link $t(x, y) \rightarrow r(x, y)$. The path loss component represents the long term average signal power on a channel (to eliminate the zero mean multi-path fading component).

A. Path Loss Estimation using Double Regression

Intuitively, DR applies regression from nearby channels to create an estimate of path loss that is correlated to these channels. Note that the estimate includes an independent component whose weight depends on the number and distance from the nearby channels used in the estimate to produce correlation consistent that observed empirically.

The primary difficulty in using regression is that the notion of distance in regression is a scalar value reflecting Euclidean

distance between the reference and projected points. In contrast, channels are line segments in 2-D or 3-D space with no clear definition for distance between them. Our approach solves this problem by applying regression in two steps. First, we apply regression based on the distance from the destination of the nearby links to the destination of the estimated link producing an intermediate estimate. In the second step, we apply regression based on the sender distances to the intermediate estimate to find the final path loss. The remainder of this subsection formalizes this process.

Consider a path loss offset $X(x, y, u, v)$ of a link $(x, y) \rightarrow (u, v)$ where (x,y) and (u,v) represents the location of a sender and a receiver in a 2-D space respectively. Let the location coordinates x, y, u, v be the independent variables of the regression and the X be the dependent variable. When a set of neighboring links are given (from previous estimates, through measurement or pre-calculated ray tracing), the $X(x, y, u, v)$ for a new link needs to be estimated. To estimate a pathloss for a new link from existing channel data, only channels that are in the sufficiently close distance from the new link should be considered. To distinguish those reference data, we define a correlation distance threshold, D_n. If (x, y) is static, and m observations for $(u_1, v_1), .., (u_m, v_m)$ are given within D_n, a regression algorithm helps to estimate an offset for a new link $(x, y) \rightarrow (u', v')$ based on the assumption that the relationship between the independent variables (i.e., locations) and their dependent variables (i.e., the path-loss) among the observations continues to hold between the new independent variable (u', v') and its dependent variable $X(x, y, u', v')$. The assumption relies on the dependent variables having spatial correlation. We now describe the DR algorithm.

B. Algorithm

First, observations are stored in an R-tree [19] so that a link can be efficiently indexed based on its location. A leaf node of the R-tree contains the location of the sender, the receiver and the observed pathloss. Since a path-loss is symmetric, a link with opposite direction and the same path-loss value is also stored in the R-tree. When a new estimation is requested, reference links within a correlation distance threshold (D_n), up to a limited number (L) are obtained from the R-tree. If the number of reference links is not limited, the computation time can increase as the number of measurements increases. The distance between two links $(x, y) \rightarrow (u, v)$ and $(x', y') \rightarrow (u', v')$ is calculated as follows.

$$
\begin{aligned}
dist &= \sqrt{dt^2 + dr^2} \\
dt &= \sqrt{(x - x')^2 + (y - y')^2} \\
dr &= \sqrt{(u - u')^2 + (v - v')^2}
\end{aligned}
$$

Let us consider a set of reference links **RL** and a link τ that has a sender A and a receiver B. We denote each coordinate of a node by $id(e)$, where $e = x, y, u,$ or v. Then, $\tau = A \rightarrow B = (x, y) \rightarrow (u, v)$, and $X_\tau = X(A, B) = X(x, y, u, v)$. The goal is to estimate the path loss offset X_τ denoted by $\hat{X}_\tau$ by using the reference links **RL**.

The proposed estimation algorithm is as follows.

1) Look up the senders for each r_l in the **RL**, and make a reference sender set $\mathbf{S}_\tau$.

2) Make an observation set for each sender $s_{\tau i}$ as $\mathbf{O}_{s_{\tau i}} = \{(r_{s_\tau ij}(u), r_{s_\tau ij}(v), X(s_{\tau i}, r_{s_\tau ij}))|j = 1, 2, .., J\}$, where J is the number of receivers for $s_{\tau i}$.
3) Apply Eq 4 to the observation set $\mathbf{O}_{s_{\tau i}}$ to get the regression coefficient βs.
4) Apply Eq 2 or 3 to the location of the target receiver (u, v) to get the estimated value $\hat{X}_s(s_{\tau i}, B)$.
5) Repeat the estimation over all $s_{\tau i} \in \mathbf{S}_\tau$.
6) Make an observation set $\mathbf{O}_{s_\tau} = \{(s_{\tau i}(x), s_{\tau i}(y), \hat{X}(s_{\tau i}, B))|i = 1, 2, ,, .I\}$, where I is the size of $\mathbf{S}_\tau$.
7) Apply Eq 4 to the observation set $\mathbf{O}$ to get the regression coefficient β.
8) Apply Eq 2 or 3 to the location of the target sender A to get the estimated value $\hat{X}(A, B)$.

The above algorithm applies regression twice, one for receivers and one for senders. This method enables estimating any link, including those whose sender and receiver are different from end nodes of reference links.

Note that β in Equation 4 is obtained only when the $\Phi^T \Phi$ is invertible. For the case when Equation 4 is not applicable, i.e., if there are insufficient neighbors through the process, we use virtual neighbors in locations which are far enough to have no correlation with existing measurements. Then, we apply interpolation with the virtual neighbors (which contribute no weight due to their distance) and existing neighbors to estimate the path loss. Since a virtual neighbor does not need to be correlated with any neighbor, the path loss for an imaginary neighbor can have normal distribution $N(0, \sigma^2)$.

C. Discussion

DR applies regression when nearby known channels are available. These channels could be obtained from measurement data to produce a site specific simulation. In such scenarios, the accuracy of the DR estimate relative to the site from which the measurement data is obtained depends on the presence of nearby measured (or ray-traced) links; absent such links, there is no spatial correlation from measured links and the estimated path loss is independent in the initial phase of the simulation where the R-tree is sparsely populated.

DR always generates "plausible" path loss models even when the provided measurement or ray-trace data is sparse or even non-existant; the generated path loss values will be consistent with path loss distribution in real environments and are thus representative of a hypothetical real environment. However, if the measurement (or ray-trace) data is sufficiently dense, DR will generate site-specific path loss representative of the actual environment from which the measurement data is obtained. The measurement or ray-tracing data simply causes the generated path loss values to correlate to this data making the simulated area similar to the measured area. .

V. PERFORMANCE EVALUATION

The model is validated against a measurement experiment in Section V-A. We show that DR can also be applied without measurement results, and compare it with the Sum-of-Sinusoids (SOS) model in Section V-B. Spatial coherence
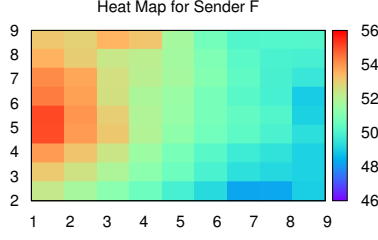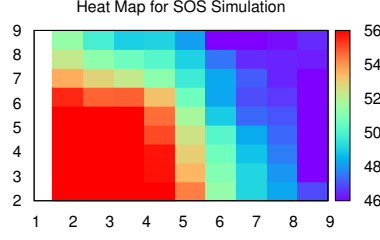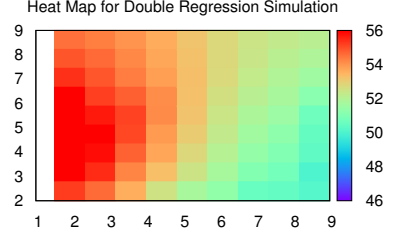
Fig. 1: Measurement

Fig. 2: SOS simulation

Fig. 3: DR simulation

is evaluated in V-C and V-D shows that DR can reduce site-survey cost. Finally, we show the impact of plausible path loss models on network performance and stability in Section V-E.

A. Measurement-based Validation of DR

In this study, we measure the path loss in a lecture hall that has dimensions of 10m by 10m by 3m. We carry out measurement experiments from six sender locations. The sender broadcast packets every 10msec with transmission power of 16dBm, 802.11g in 2.471GHz frequency, and 6Mbps data rate. To eliminate the multi-path effect (the temporal fading component of the signal) and orientation effect [33], [34], we averaged the signal power in a small area of 1.8m rotating the receiver antenna in all directions.

To extract the offset $X(t_i, r_i)$ from the measured $S(t_i, r_i)$, d_0, $P(d_0)$, β and σ need to be determined. First, P is roughly obtained by $S - 95$ [35]. P_0 is obtained from the measurement by setting d_0 to be 1. σ is determined depending on β. So, β is chosen to make the zero mean and the smallest σ. In our testbed, the values $P_0 = -33dB, \beta = 2.8, \sigma = 8$ result from this approach.

Figure 1 depicts the signal strength map (heat map) when the sender is located at point $(1, 5)$ on the figure. The (x, y) coordinate represents the location of a receiver and the color represents the RSSI value. We compared the simulated path loss with the measurement of the corresponding physical location as well as path loss values generated by the SOS model (we set the SOS parameters as recommended [21]).

For the DR experiment, we use DR to interpolate each link from a preset number of neighboring measurements. Figure 2and 3 show the corresponding heat maps obtained from SOS and DR respectively. DR clearly generates the more similar heat map to the measurement obtained one. The reason is that our model estimates signal strength at each location from a few reference links that are obtained through measurement, while the SOS method estimates them with only overall statistics and is not able to incorporate measurement values.

For a more quantitative evaluation, we calculated the root mean square (RMS) of the estimation error defined as $x_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - m_i)^2}$, where n is the number of measured links, x_i and m_i are the simulated value and the measured value for the ith link, respectively. If the RMS for all 174 measured links are investigated, SOS RMS value is

SOS		0.522			
DR	L	2	3	4	5
	x_{RMS}	0.239	0.243	0.254	0.35
	| RL|	12	18	24	30
	x_{RMS}	0.258	0.24	0.25	0.25

TABLE I: Comparison of RMS of estimation errors

0.522, while DR achieves 0.24 RMS with only two reference links for each estimate.

Table I shows the RMS of estimation errors for various number of links used in the DR. In all cases, the RMS values achieved by DR are smaller than that of SOS method. If $L = 5$, i.e., 5 neighbors of a target receiver are involved in the estimation, it will include some far nodes that actually are uncorrelated with the target. Thus, the reference links for a target are chosen based on distance in practice. The number of reference links does not affect the estimation error significantly.

B. DR without Measurement: Correlated Gaussian Process

DR can also generate a geographically-independent space when no measurement data is available to seed the model. The geographically-independent space is particularly useful when the size of network increases, and physical mapping to a specific site is not required.

First, a distance threshold (D_n) can be set to a decorrelation distance [21] because points at this distance away from each other experience weak correlation in path loss. In this case, the R-tree is originally empty. We can generate some initial reference links (e.g., on a grid D_n meters apart) to seed the R-tree, or see it on demand whenever a new link is needed and no nearby reference links exist. When no nearby links exist for a new channel, a Gaussian random variable from $N(0, \sigma)$ is used to assign the path loss at that point. The new estimate is added to the R-tree such that future nearby channels can use it in generating their estimate.

Figure 4 depicts a realization of the virtual space by factorial regression, where the sender is fixed at the middle of the topology and the (rx, ry) coordinate represents the location of a receiver. The number of reference neighbors is set to 3, and the distance threshold D_n is set to 6 meters that is typical in an urban area [21]. We first generated channels at grid points D_n
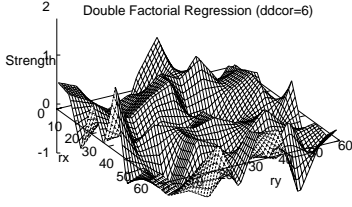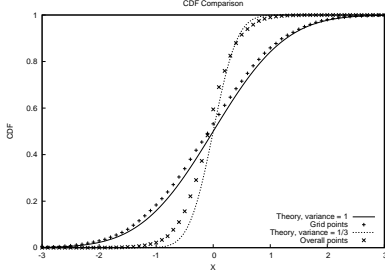
Fig. 4: Double factorial regression model

Fig. 5: CDF of double factorial regression

Fig. 9: Impact of the number of Initial Measured Links

apart; those points are generated independently because of the lack of nearby reference links. The spatial correlation among the receivers, i.e., the similarity among the neighbor receivers, is plotted. We obtained similar results when a receiver is fixed and the (x, y) coordinate represents the location of a sender. As a result, a small location change for either senders or receivers does not cause abrupt change of path loss, and the new link has plausible path loss consistent with empirical results that show spatial correlation of path loss. While the spatial variation changes smoothly in most areas, it changes abruptly at the grid points which are generated randomly (i.e., with no correlation).

Figure 5 compares the CDF of overall space to that at the grid points. The distribution of grid points shows 3 times larger variance than the distribution of overall space since they were generated independently. Averaging n random variables for the target receiver and sender results in a reduced variance by $\sqrt{n} \times \sqrt{n}$ by the central limit theorem. Since we used an average of 3 reference links, the reduction in variance relative to the grid points follows the expected distribution.

Finally, Figure 6 shows the joint autocorrelation with respect to dt and dr. The DR generated path loss shows a good agreement with theory [21]

C. Spatial Coherence Analysis

In this section, we analyze DR with respect to the spatial correlation among the generated path loss estimates. Primarily, when no nearby links are available for regression, DR uses an uncorrelated Gaussian estimate for path loss. The density of the observations is varied by changing the correlation distance threshold (D_n), which defines the distance beyond which links are no longer considered neighbors. When D_n is large, simulating dense measurements, more neighbors will be included in the DR estimate. When it is small, simulating
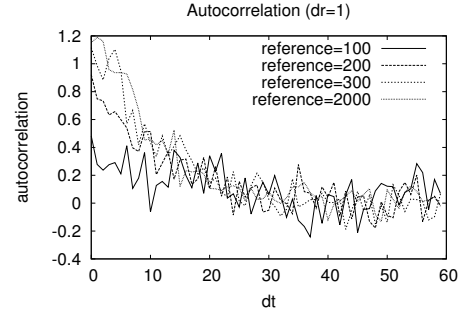
sparse measurements, many points will have no measured neighbors and would have to be generated using the Gaussian distribution, increasing randomness.

Figure 7 and 8 show autocorrelation when the receiver distance (dr) and transmitter distance (dt) between the reference and estimated link increases. When the distance between links is small (e.g., 1), the autocorrelation should be high (closer to 1). A lower autocorrelation indicates that the estimated path loss is not correlated in space. The topology size is 400x400, and 2000 reference observations are generated from $N(0, \sigma^2)$. When $D_n = 30$ is used, explicit exponential autocorrelation is shown, which means if there are sufficient neighbor links, DR generates spatially correlated path loss. However, due to the random variables involved in the regression process when distance from reference links is large or insufficient nearby reference links exist, the autocorrelation does not look smooth. Each point represents an average of 720 simulation runs.

We vary the number of reference links in the same size of topology. If fewer reference is used, more links are generated independently because they do not have sufficient neighbors for interpolation. Figure 9 shows that the autocorrelation is not seriously affected when the number of initial channels is reduced to 300.

Figure 10 shows the autocorrelation when between points that are 1 and 5 meters apart as the number of links in the simulation increases. At first, there is no reference link (x=0). As a new link appears and an estimation for the link is stored in the R-tree, there will be more data that can be referred by a new link. When the number of links is small, the autocorrelation is low. Figure 10 shows that when the number of links reaches 2000 (x=2000), the autocorrelation converges to match the empirical values. Each point represents the average of 30 simulation runs.

D. Using DR to Interpolate Measurement Data

Site Surveys require a large number of measurements. Similarly Ray Tracing requires many computationally intensive computations. In this section we evaluate the success of DR to interpolate when only a few number of measurements is available. If it is successful, DR can significantly lower the cost of site surveys and Ray Tracing.

We use a measurement data set from the CRAWDAD repository measuring path loss in part of the Colorado University campus [36]. The data contains RSS measurements collected using the CU Wide Area Radio Testbed (CU-WART).
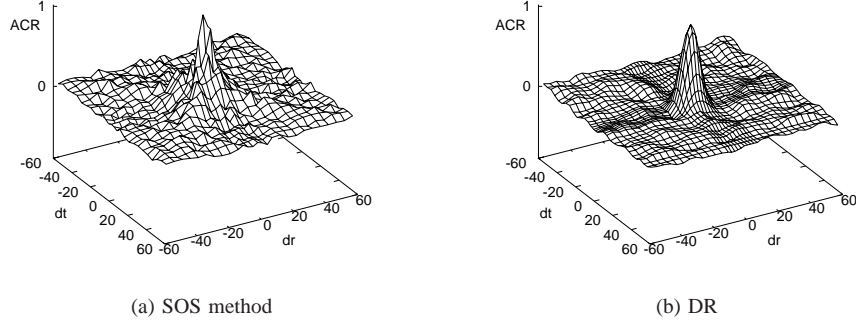
(a) SOS method

(b) DR

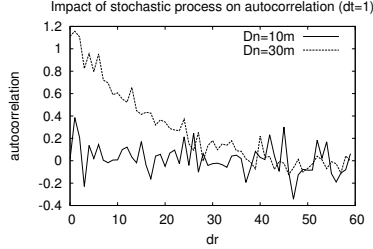Fig. 6: Joint Autocorrelation for SOS and DR

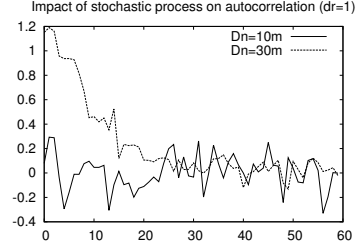Fig. 7: Autocorrelation with receiver distance

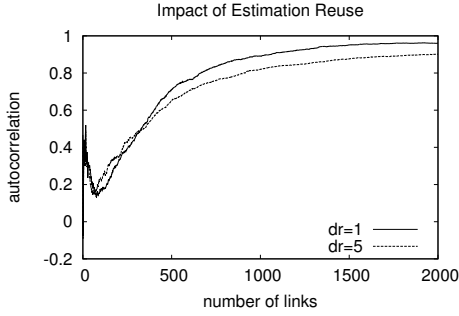
Fig. 8: Autocorrelation with transmitter distance

Fig. 10: Path loss stability as the number of generated links increases

Figure 11a shows a heatmap of the measured RSS. The total number of used links is about 900. The unmeasured areas are shown as white. To show our proposed method can regenerate a RSS map with small number of measurement data, we reproduce a heatmap of RSSI with 1/4 sample measurement that are randomly selected. Figure 11b that the proposed method can regenerate realistic path loss.

E. Impact on Network Stability

The path loss models the long time average of the signal power; it has substantial impact on the simulation results, affecting the signal and interference components of SINR and how they change as the nodes move. This impact percolates up the protocol stack to the routing and application layers; the use of inaccurate path loss models can lead to substantial inaccuracy in the simulation. To demonstrate this effect, in

this section, we evaluate the impact of path loss models on the behavior of two representative simulation studies.

One of the most notable impacts of the spatial correlation is on network stability. A representative reactive routing protocol for a mobile ad-hoc network, Dynamic Source Routing (DSR), detects a route failure after consequent link failures, then re-issues DSR route discovery packets. As mobile nodes move, the path loss and the link qualities between nodes keep changing, which causes route failure and rediscovery process.

We investigated how many times the re-discovery process occurs in the i.i.d. log-normal shadowing path loss model, which is commonly used in simulators such as NS-2, and the DR path loss model in a general mobile ad-hoc network scenario. The scenario includes 20 nodes in a 30mx30m topology in which each node moves at maximum speed of 1m/sec. To make sure that the network includes connection and disconnection through multi-hops, the path loss exponent is set to 5. Since the nodes move, a Rayleigh fading channel is used in addition to the path loss model to account for multi-path fading.

Fig. 12 shows the number of route discovery DSR packets normalized by the number of data packets. When the DR path loss model is used, the link quality of a pair of mobile nodes will change slowly, because the path loss is highly correlated with the distance. Thus, a small change of location probably not causes a route failure. However, i.i.d shadowing path loss model does not guarantee the consistent link quality as nodes move even slowly, because the path loss for different location is determined independently. As a result, the i.i.d shadowing path loss model overestimates route failures and re-discovery processes as the standard deviation increases, while the DR model shows very stable behavior. A less stable routing layer
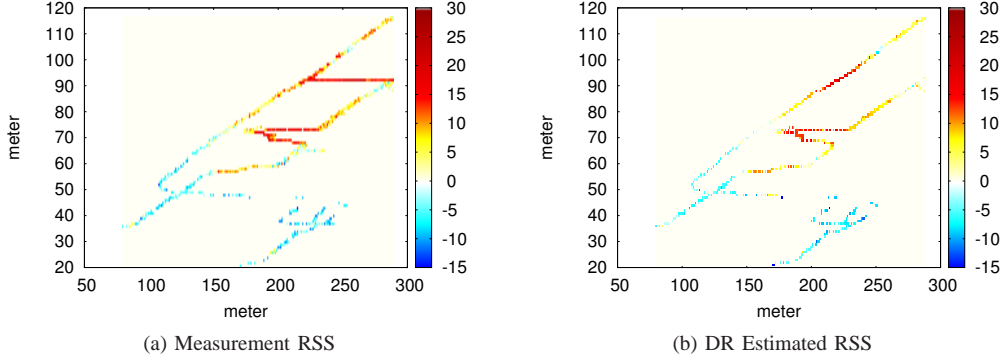
(a) Measurement RSS

(b) DR Estimated RSS

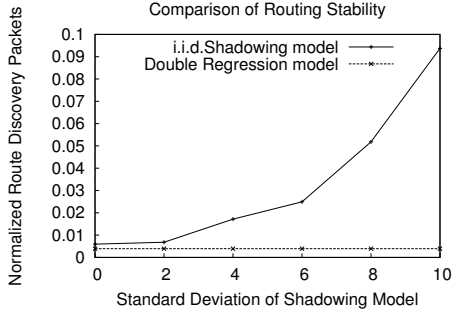Fig. 11: Estimated RSS using DR for the CU Wireless Trace

Fig. 12: Impact of spatial correlation on routing stability
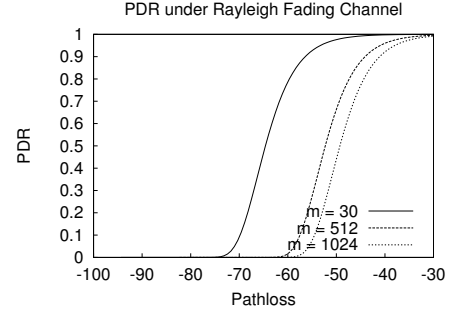

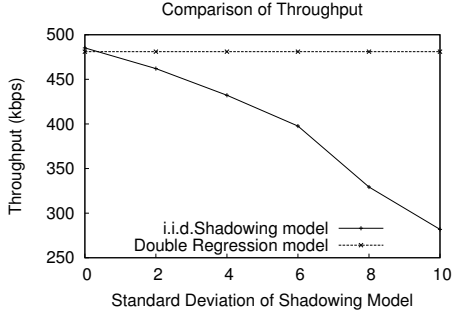Fig. 14: PDR: Rayleigh fading and different packet sizes

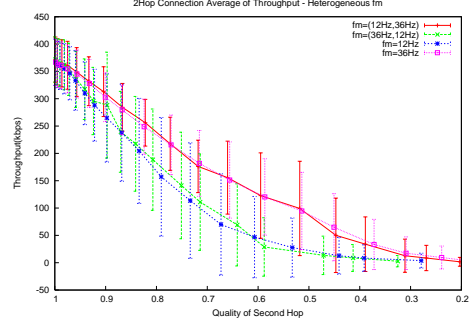Fig. 13: Impact on Throughput

Fig. 15: Impact of PDR on throughput

will reduce the throughput, because the time and channel capacity are waisted by the routing overhead. Our investigation of the impact on the throughput showed that using i.i.d shadowing channel model will underestimate the throughput by up to 42%. This result highlights the significance of using correct channel model in a simulation-based research.

F. Impact on Throughput

This section shows the impact of the realistic path-loss model on throughput when we feed it to a temporally fading model (Rayleigh fading). The symbol loss rate in the Rayleigh fading channel is: $1 - e^{-\rho^2}$, where $\rho = \frac{R_{threshold}}{R_{RMS}}$. Figure 14 depicts the delivery ratio of packet size of 30, 512, and 1024 bytes. X-axis represents attenuated signal strength due to path loss, $R_{RMS} = \sqrt{10^{(A(d)-30)/10}}$, and the threshold

$R_{threshold} = \sqrt{10^{(-90dBm-30)/10}}$. Since the signal strength under the Rayleigh fading channel fluctuates over time, the packet delivery ratio increases gradually around the threshold as the attenuated signal strength due to path loss increases. Note that small difference in path-loss results in big difference in PDR. For example, if the packet size is 1024, if the PDRs when the attenuated signal = -50dBm and -40dBm are 0.4 and 0.8 respectively.

The impact of realistic path loss on throughput can be shown by the impact of different PDR on throughput. We consider a two-hop connection A-B-C. We keep the PDR of A-B at a relatively high (0.83) and investigate the throughput as the PDR of B-C varies. Figure 15 depicts the impact of the various PDR on the throughput when various fading rapidity (f_m) of the Rayleigh fading channel is used for each hop. If the

PDR of the second hop are 0.4 and 0.8, the throughput are 25kbps and 150kbps when second hop's $fm = 12Hz$, i.e., the throughput difference is up to 600%. This result supports the significance of estimating accurate path-loss in a multi-hop ad-hoc connection.

VI. Conclusion

Accurate and efficient channel models are a critical need for improving the fidelity of wireless network simulation. In this paper, we considered the problem of path loss channel models; this is the component of the channel model that estimates the long term average signal power between a sender and a receiver. Path loss is affected by the RF shadows in the environment. Existing stochastic models are inaccurate because they do not consider spatial correlation. In addition, they are difficult to configure to parameters representative of realistic environments such as corridors or building shadows. On the other hand, accurate models can be built using either detailed site-surveys (requiring a prohibitive number of measurement experiments) or ray tracing (which is computationally very difficult and requiring precise GIS). Thus, there is a need for accurate and efficient path loss models.

In response to this need, we proposed a novel solution for a geographically accurate, efficient, and flexible path loss model. The key idea of the proposed model is to estimate a plausible value for new channels using regression from nearby known channels. The algorithm uses a highly efficient double regression procedure: in the first step, we use regression to estimate a channel by moving one of the communication ends, while the second step moves the other. The known channels can be those obtained from measurement experiments, ray-tracing, or previous models generated by the simulation. By using regression, we preserve spatial correlation as measured in previous studies, leading to plausible channel models.

References

[1] D. Kotz, C. Newport, and C. Elliott, "The mistaken axioms of wireless-network research," Dept. of Computer Science, Dartmouth College, Tech. Rep. TR2003-467, July 2003.

[2] L. Perrone and Y. Yuan, "Modeling and simulation best practices for wireless ad hoc networks," in *Simulation Conference, 2003. Proceedings of the 2003 Winter*, vol. 1, Dec. 2003, pp. 685–693.

[3] J. Liu, Y. Yuan, D. M. Nicol, R. S. Gray, C. C. Newport, D. Kotz, and L. F. Perrone, "Simulation validation using direct execution of wireless ad-hoc routing protocols," in *Proc. PADS*, 2004.

[4] D. Couto, "A high-throughput path metric for multi-hop wireless routing," in *MobiCom*, 2004.

[5] J. Bicket, D. Aguayo, S. Biswas, and R. Morris, "Architecture and evaluation of an unplanned 802.11b mesh network," in *MobiCom '05: Proceedings of the 11th annual international conference on Mobile computing and networking*, 2005.

[6] A.Adya, P.Bahl, J.Padhye, A.Wolman, and L.Zhou, "A multi-radio unification protocol for ieee 802.11 wireless networks," Microsoft Research, Tech. Rep., Jul. 2003.

[7] R.Draves, "Comparison of routing metrics for static multi-hop wireless networks," in *SIGCOMM*, 2004.

[8] R. Punnoose, P. Nikitin, and D. Dtancil, "Efficient simulation of ricean fading within a packet simulator," in *Vehicular Technology Conference – Fall*, Sep. 2000.

[9] ISI, "ns: Change history," http://www.isi.edu/nsnam/ns/CHANGES.html.

[10] T. S. Rappaport, *Wireless Communications : Principles and Practice*. Prentice Hall, 2002.

[11] W. Jakes Jr., *Microwave Mobile communications.* Wiley, 1974.

[12] S. ZVANOVEC, P. PECHAC, and M. KLEPAL, "Wireless lan networks design:site survey or propagation modeling?" in *WIRELESS LAN NETWORKS DESIGN*, 2003.

[13] D. I. Laurenson, "Indoor radio channel propagation modelling by ray tracing techniques," Ph.D. dissertation, The University of Edinburgh, 1994.

[14] M. Nidd, S. Mann, and J. Black, "Using ray tracing for site-specific indoor radio signal strengthanalysis," in *Vehicular Technology Conference, 1997 IEEE 47th*, vol. 2, May 1997, pp. 795 – 799.

[15] G. E. Athanasiadou and A. R. Nix, "A novel 3-d indoor ray-tracing propagation model:the path generator and evaluation of narrow-band and wide-band predictions," in *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, vol. 49, july 2000.

[16] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," in *Electron. Lett.*, vol. 27, Nov. 1991, pp. 2145–2146.

[17] N. Patwari, Y. Wang, and R. O'Dea, "The importance of the multipointto -multipoint indoor radio channel in ad hoc networks," 2002. [Online]. Available: citeseer.ist.psu.edu/patwari02importance.html

[18] N. Jalden, P. Zetterberg, B. Ottersten, A. Hong, and R. Thoma, "Correlation properties of large scale fading based on indoor measurements," in *Wireless Communications and Networking Conference, 2007 WCNC 2007 IEEE*, March 2007, pp. 1894–1899.

[19] N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger, "The r*-tree: an efficient and robust access method for points and rectangles," in *INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA*. ACM, 1990, pp. 322–331.

[20] J.-M. Dricot and P. D. Doncker, "Integrated ad-hoc and cellular networking in indoor or faded environments," in *IWCMC '09: Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing.* New York, NY, USA: ACM, 2009, pp. 333–337.

[21] A. N. Zhenyu Wang, Eustace K. Tameh, "Simulating correlated shadowing in mobile multihop relay/ad-hoc networks," 2006. [Online]. Available: http://ieee802.org/16

[22] J. Robinson, R. Swaminathan, and E. W. Knightly, "Assessment of urban-scale wireless networks with a small number of measurements," in *Proceedings of the 14th ACM international conference on Mobile computing and networking*, ser. MobiCom '08. New York, NY, USA: ACM, 2008, pp. 187–198. [Online]. Available: http://doi.acm.org/10.1145/1409944.1409967

[23] A. Konak, "Estimating path loss in wireless local area networks using ordinary kriging," in *Winter Simulation Conference (WSC), Proceedings of the 2010*, Dec. 2010, pp. 2888 – 2896.

[24] I. Stepanov, D. Herrscher, and K. Rothermel, "On the impact of radio propagation models on manet simulation results," *University Stuttgart, Fakultat 5, Germany, Computer Science Archive*, 2005.

[25] M. Catedra, J. Perez, F. Saez de Adana, and O. Gutierrez, "Efficient ray-tracing techniques for three-dimensional analyses ofpropagation in mobile communications: application to picocell andmicrocell scenarios," in *IEEE Antennas and Propagation Magazine*, Apr 1998, pp. 15–28.

[26] J. Dricot and P. Doncker, "High-accuracy physical layer model for wireless network simulations in ns," in *International Workshop on Wireless Ad-hoc Networks*, 2004. [Online]. Available: citeseer.ist.psu.edu/dricot04highaccuracy.html

[27] V. Sridhara and S. Bohacek, "Realistic propagation simulation of urban mesh networks," *Computer Networks*, vol. 51, no. 12, 2007.

[28] Y. Xu and W.-C. Lee, "Exploring spatial correlation for link quality estimation in wireless sensor networks," in *PERCOM '06: Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications.* Washington, DC, USA: IEEE Computer Society, 2006, pp. 200–211.

[29] S.O.Rice, "Mathematical analysis of random noise," *Bell System Technology Journal*, vol. 23,24, pp. 282–332,46–156, 1944,1945.

[30] M. Patzold, U. Killat, F. Laue, and Y. Li, "On the statistical properties of deterministic simulation modelsfor mobile fading channels," *Vehicular Technology, IEEE Transactions on*, vol. 47, pp. 254 – 269, Feb. 1998.

[31] X. Cai and G. B. Giannakis, "A two-dimensional channel simulation model for shadowing processes," *IEEE Trans. on Vehicular Technology*, November 2003.

[32] I. StatSoft, "General regression models (grm)," Website, 1984, http://www.statsoft.com/textbook/stgrm.html.

[33] A. Gehring, M. Steinbauer, I. Gaspard, and M. Grigat, "Empirical channel stationarity in urban environments," in *EPMCC*, 2001.

[34] G. Gaertner and V. Cahill, "Understanding link quality in 802.11 mobile ad hoc networks," *IEEE Internet Computing*, vol. 8, no. 1, 2004.

[35] Wildpackets, "Converting signal strength percentage to dbm values," 2002. [Online]. Available: http://www.wildpackets.com/elements/whitepapers/Converting_Signal_Strength.pdf

[36] C. Phillips and E. W. Anderson, "CRAWDAD trace $cu/cu_w art/2010/path_l oss(v.2011 - 10 - 24)$," Downloaded from $http : //crawdad.cs.dartmouth.edu/cu/cu_w art/2010/path_l oss$, Oct. 2011.