Data-driven and GIS-based Coverage Estimation in a Heterogeneous Propagation Environment

Nihesh Rathod[α,β], Renu Subramanian[α], and Rajesh Sundaresan[α,β]
[α]Department of Electrical Communication Engineering,
[β]Robert Bosch Centre for Cyber-Physical Systems,
Indian Institute of Science, Bangalore, India. 560012

***Abstract*—We provide a data-driven coverage estimation technique that employs machine-learning based regression ideas for exploiting commonality of antenna-gain and other parameters across measurements made in multiple propagation environments. We then show how readily available geographic information system (GIS) data could be exploited for quick classification of geographic areas into various propagation environments, and how this could enable quick and automated estimation of coverage for faster and more efficient deployment of Internet of Things.**

***Index Terms*—coverage, GIS, heterogeneous propagation environment, Internet of Things (IoT) deployment.**

I. Introduction

The traditional network deployment strategy, based on deploy-first and survey-later, is neither scalable nor efficient in terms of manpower and material resource usage. This strategy is bound to fail given the projected scale of IoT expansion. Automated network deployment is therefore a necessity. Instead of relying on surveying a particular deployment region afresh for getting an estimation of signal coverage, prior knowledge of the terrain, based on Geographic Information System (GIS) data and measurements in similar environments, can lead to quick estimation of coverage at an early stage of design. An ability to estimate the received signal strength indication (RSSI) without actual deployment saves valuable resources and leads to both rapid and more efficient network deployment. This paper is an attempt to demonstrate that quick estimation of coverage is indeed possible based on readily available GIS data and extensive measurement samples in example propagation environments.

Often, one estimates channel parameters such as pathloss exponents, antenna gain parameters for a specific transmitter-receiver antenna pair, frequency-dependent decay parameters, etc., based on extensive RSSI measurements. In typical Contiki protocol implementations, these are made available to the upper layers only on correctly received packets. The estimation of RSSI is therefore limited to the extent that it depends only on correctly received packets. On the other hand, many network designers all too often design their networks based only on packet error rate measurements without ever relying on RSSI measurements. Also, it is often the case that the same type of transmitters and receivers are used across measurements over multiple propagation environments. One could then exploit the knowledge that the antenna gain-related parameters are the same across measurements even though the propagation environments across measurements may have been different.

Our first goal in this paper is to demonstrate that a scheme inspired by logistic regression, yet different from it, in combination with a regression-based estimation algorithm that minimises the error between measured RSSI on correctly received packets and predicted RSSI, provides a much better estimate of the received signal strength than the one that relies only on measured RSSI on the correctly received packets. The logistic-regression like scheme is attuned to the communication theoretic model of transmission over a Rayleigh fading channel, yet exploits the knowledge that the antenna-gain parameters are common across measurements in different propagation environments.

Our second goal in this paper is to demonstrate that quick estimation of coverage is possible based on readily available GIS data and measurement data in example environments. To demonstrate this we have built a tool that processes open-source GIS data to classify a large heterogeneous candidate deployment area into various regions with different propagation characteristics. The estimated parameters from the first part of the paper are then applied to each of the smaller component regions. These are then put together in a natural way to estimate the RSSI in a heterogeneous propagation environment, resulting in a heat map that enables easy visualisation of coverage and coverage holes. See Figure 1 for the building blocks of our tool. We use the Indian Institute of Science (IISc) campus as a test bed to explain our ideas, algorithms and describe the outcomes.

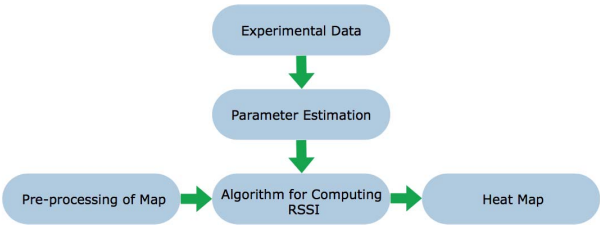

Fig. 1: Building blocks of the coverage prediction tool.

A part of this work was presented by one of the authors as an invited talk at the 9th International Conference on COMmunication Systems & NETworkS, Bangalore, India, January 2017.

We would like to thank Rashmi Ballamajalu, Network Engineering Lab[α] for helping us in conducting the experiments. This work was supported by the Robert Bosch Centre for Cyber-Physical Systems, Indian Institute of Science. N. Rathod was supported by a Cisco Research Scholarship.

978-1-5386-4727-1/18/$31.00 ©2018 IEEE

The rest of the paper is organised as follows. In Section II, we explain our data driven approach for joint parameter estimation and show its superiority over two simpler estimation schemes, one based only on RSSI-from-correctly-received-packets and another based only on packet error rate. In Section III, we describe our tool for quick estimation of coverage. We show how to extract useful terrain information from a GIS database and tessellate the area into different propagation environments. We then describe the RSSI computing algorithm with examples, and demonstrate the outcome in the form of a heat map on an example deployment.

II. The data driven approach

Our data-driven approach uses a regression methodology commonly used in machine learning (ML). We assume that each gathered data point has associated with it several *factors* that affect an *outcome* which is the quality of reception. The factors considered in this paper are transmitter power, transmitter height, receiver height, carrier frequency, and propagation environment. As for the last factor, in an earlier work [1], the authors had classified the IISc campus into five different radio propagation environments or regions – open area, building, road, moderately wooded, heavily wooded. Associated with each data point is also an outcome – whether a transmitted packet was received correctly and, if yes, the value of the received signal strength indication (RSSI). We will study several regression-based approaches to estimate how these factors affect the outcome, and will compare their estimation capabilities. Often only the packet error rate or only the RSSI of correctly received packets are used. The first two approaches use only RSSI of correctly received packets or only packet error rate. The third approach and another variant uses both, and have significant improvement in performance.

A. The regression methodologies

The model for received energy, when the transmitter and receiver are placed in a specified homogeneous propagation environment, is well-established, see [2, p. 83]:

$$P_{Rx} = C \cdot P_{Tx} \cdot h_{Tx}^2 \cdot h_{Rx}^{\gamma} \cdot d^{-\eta_r} \cdot f^{-\kappa_r}. \tag{1}$$

Here P_{Rx} is the received power, C is a constant that depends on transmit and receive antenna gain factors, P_{Tx} is the transmitted power, h_{Tx} and h_{Rx} are transmitter and receiver heights, respectively, γ is an exponent that specifies how received power improves with receiver antenna height, d is the distance between transmitter and receiver, η_r is the region-dependent path-loss exponent, typically between 2 and 6, and f is the carrier frequency of operation. Finally κ_r is a region-dependent parameter (between 2 and 3) that tells how fast the received energy decays with increasing frequency in region r.

Observe that some of these parameters are common across regions for a given transmitter-receiver pair, e.g., C and γ, while others are region-specific, e.g., η_r and κ_r. The regression approach estimates all these parameters in a unified way.

Assuming N_0 is the thermal noise power, the signal-to-noise ratio (SNR) is, see [3, p. 173]:

$$\mathsf{SNR} = \frac{P_{Rx}}{N_0} = \frac{C \cdot P_{Tx} \cdot h_{Tx}^2 \cdot h_{Rx}^{\gamma} \cdot d^{-\eta_r} \cdot f^{-\kappa_r}}{N_0}. \tag{2}$$

Atop this, we assume that uncoded transmitted symbols suffer from Rayleigh fading. (Extension to Ricean fading or other fading models are straightforward and merely increases the number of parameters of the model. We restrict attention to Rayleigh fading to highlight our approach in the simplest of settings.) We may then view SNR as the average signal-to-noise ratio across fading instances.

We now explain our approaches all of which work with the same data. This data may have been collected over different regions, and may not have been time-stamped.

In the first approach, only RSSI measurements and that too of only correctly received packets are taken into account. This is often the case in common implementations of the Zigbee protocol, e.g., Contiki. Suppose that there are M_r correctly received packets in region r, where $r = 1, \ldots, 5$. Let $RSSI(n)$ be the measured received power for the nth correctly received packet. Denote by $P_{Rx}(n)$ the true received power when the transmit parameters are $P_{Tx}(n)$, $h_{Tx}(n)$, when the receiver height, receiver distance, and frequency of operation are $h_{Rx}(n)$, $d(n)$, and $f(n)$, respectively, and the region of operation is $r(n)$. Let us collectively denote all these factors by $z(n)$. Using these factors, we obtain $P_{Rx}(n)$ from the formula (1). We then attempt to solve the regression problem:

$$\min \sum_{r=1}^{5} \sum_{n=1}^{M_r} \Big| RSSI(n) - P_{Rx}(n) \Big|^{\zeta} \tag{3}$$

where the minimisation is over parameters $C > 0$, $\gamma \in [1, 2]$, $\eta_r \in [2, 6]$, $\kappa_r \in [2, 3]$, $r = 1, \ldots, 5$. Let us reiterate that this involves a joint optimisation across all collected data. $\zeta = 1$ yields absolute error loss between predicted and measured RSSI while $\zeta = 2$ yields squared error loss. (The approach extends to other loss functions such as $|\log RSSI(n) - \log P_{Rx}(n)|$.

In the second approach, we take inspiration from the machine learning technique of logistic regression [4, Ch. 4.4], although we emphasise that our technique is different from the classical logistic regression, to exploit information on the number of incorrectly received packets.

Under a Rayleigh fading assumption, we know that the probability of error of a BPSK transmission with average signal-to-noise ratio SNR is, see [5, eqn. (3.19)],

$$\Pr\{Error\} = \frac{1}{2}\left(1 - \sqrt{\frac{\mathsf{SNR}}{1+\mathsf{SNR}}}\right) \sim \frac{1}{4\mathsf{SNR}}, \tag{4}$$

where the approximation holds when the SNR is high. Let us denote the signal-to-noise ratio by $\mathsf{SNR}(z)$ when the factor vector is z, so that

$$p(z) = \frac{1}{2}\left(1 - \sqrt{\frac{\mathsf{SNR}(z)}{1+\mathsf{SNR}(z)}}\right).$$

Let $y(n)$ be the indicator function whose value is 1 when nth packet is in error and 0 otherwise, when the corresponding transmission factor is $z(n)$. Assuming independent transmissions, which is a good assumption when there is sufficient time separation between transmissions or if the data is without time-stamps, $\{y(n)\}_{n \geq 1}$ is an independent sequence of Bernoulli

random variables with parameters $\{p(z(n))\}_{n\geq 1}$, and hence the likelihood of the observed sequence of packet errors corresponding to the sequence of factors is:

$$\Pr\{y(1),\ldots,y(N)\} = \prod_{n=1}^{N} p(z(n))^{y(n)}(1-p(z(n)))^{1-y(n)}$$

where N is the total number of transmitted packets, taking both correctly and incorrectly received packets into account. The negative log likelihood is then:

$$-\ln\Pr\{y(1),\ldots,y(N)\} = \sum_{n=1}^{N}\Big(-y(n)\ln(p(z(n))-(1-y(n))\ln(1-p(z(n)))\Big). \quad (5)$$

Under the second approach, the goal is to maximise likelihood (or minimise the negative log likelihood) over parameters $C > 0$, $\gamma \in [1,2]$, $\eta_r \in [2,6]$, $\kappa_r \in [2,3]$, $r = 1,\ldots,5$.

Our third approach is to use a combined objective of maximising likelihood and minimising the RSSI estimation error jointly as follows:

$$\min \sum_{n=1}^{N}\Bigg(-y(n)\ln(p(z(n))-(1-y(n))\ln(1-p(z(n))) + (1-y(n))\cdot|RSSI(n)-P_{Rx}(n)|^{\zeta}\Bigg), \quad (6)$$

where yet again the minimisation is over parameters $C > 0$, $\gamma \in [1,2]$, $\eta_r \in [2,6]$, $\kappa_r \in [2,3]$, $r = 1,\ldots,5$. One could further optimise the relative weightages assigned to these two objectives, but this is out of the scope of this work. As we will see, there is already significant improvement with equal weightage. The optimisations were done in Python.

We remark, in passing, that if all $z(n)$ were the same, then the maximum likelihood estimate is that weighting of the factors which brings $p(z(n))$ closest to $\frac{1}{N}\sum_{n=1}^{N} y(n)$ is relative entropy divergence; i.e. $\min D\left(\frac{1}{N}\sum_{n=1}^{N} y(n) \,||\, p(z(n))\right)$, where $D(p \,||\, q)$ is the binary relative entropy. In the general case, $z(n)$ varies from sample point to sample point and our approach accounts for this variation in an appropriate way.

B. Experimental Data

We conducted extensive field experiments, to capture the environmental effects on signal propagation and decay, in the five aforementioned regions or propagation environments. In each region, the transmitters and receivers were placed at different distances as indicated in Figure 2. The transmitters and receivers were kept at heights 1 m, 2 m, and 3 m. Receivers placed at all the 3 heights listened to the channel simultaneously at any given time. A total of 1200 packets were transmitted from a given transmitter height. Only one transmitter was allowed to transmit in any collection period to avoid packet collisions. The same procedure was repeated for each transmitter height. There are thus nine combinations of transmitter and receiver heights for a given distance. Given that there are 22 distance-region pairs (see Figure 2), the number of data points $N = 22 \times 9 \times 1200 = 237,600$.

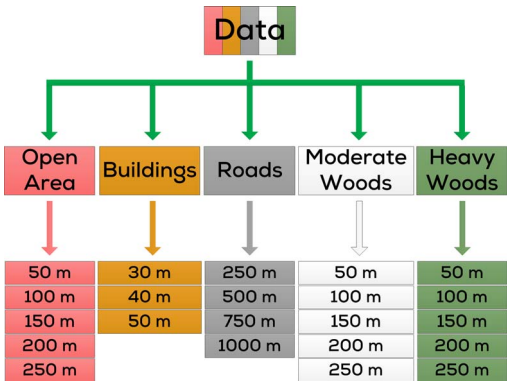

Fig. 2: Regions of experimentation and distances between transmitters and receivers. The colour-coding corresponds to the colours in Figure 3.

C. Cross-validation

Ten-fold cross-validation is presented in Table I below. Data was divided in ten random subsets for each environment, nine of which were used for estimating the parameters and the tenth was used for validation. The results of this procedure are listed in Table I. The values in "Error I" column are observed when using logistic regression without RSSI term. Columns "Error II" and "Error IV" has error values corresponding to regression on just RSSI term 3 with absolute error loss ($\zeta = 1$) and squared error loss ($\zeta = 2$), respectively. Columns "Error III" and "Error V" show the errors values under combined optimisation in (6) with absolute error loss ($\zeta = 1$) and squared error loss ($\zeta = 2$), respectively. Clearly, the regressions that exploit information on both lost packets and RSSI are far superior to RSSI-alone or packet-error-rate-alone regressions, except in 2 out of the 22 cases. The large estimation error of 12.3 dB and 10.3 dB for $\zeta = 1$ and $\zeta = 2$, respectively, in moderately wooded area at 250 m is due to a high packet error rate encountered there. This might have been due to some shadowing although we noticed no visible blockage at the location.

D. Remarks

The above results indicate that the estimation can improve significantly by employing all available data The ML methodology inspired by logistic regression enables good use of packet error rate information. The loss functions used are mere examples, and one could equally well explore other loss functions. While we have demonstrated that equal weights for RSSI estimation loss and negative log-likihood loss already provides an improvement, one could consider further optimising the weights assigned to these objectives. Finally, as more data arrives, the above technique is easily amenable to incremental updates – one can make an incremental move from the current set of best parameters à la stochastic approximation.

III. RSSI Computation in a Heterogeneous Region

We now embark on the second part of the paper which develops an RSSI estimation tool for heterogeneous regions. We have already discussed how our data-driven approach

TABLE I: Performance comparison. O = open area, B = buildings, R = roads, M = moderate woods, H = heavy woods.

Reg.	Dist. (m)	Meas. (dBm)	Error I (dB)	Error II (dB)	**Error III (dB)**	Error IV (dB)	**Error V (dB)**
O	50	-59.4	-47.4	4.7	**2.4**	5	**1.6**
O	100	-63.1	-53.3	4.5	**-1.1**	5.1	**-2.2**
O	150	-69.4	-52.5	3.7	**1.1**	3.6	**0.3**
O	200	-72.9	-53	3.1	**1.3**	3	**0.2**
O	250	-75.7	-53.4	4.7	**2.1**	4.8	**1**
B	30	-87.6	-7.6	4.3	**1.2**	4.6	**0.3**
B	40	-92.8	-4.7	5.3	**2.4**	5.1	**2.5**
B	50	-92.9	-8.8	4.7	**-2.3**	4.9	**-2.3**
R	250	-79.4	-15.6	-6.1	**-4.2**	-5.1	**-2**
R	500	-91.2	-7.6	3.3	**1.0**	3.9	**2.3**
R	750	-97.8	-5.1	2.2	**2.0**	3	**2.9**
R	1000	-99.0	-3.7	3.0	**2.8**	3.3	**2.8**
M	50	-65.7	-27.2	4.8	**-2.3**	4.8	**-2**
M	100	-69.2	-26.7	3.8	**-4.1**	4.5	**-4.7**
M	150	-82.9	-19.6	4.5	**1.5**	4.4	**0.1**
M	200	-88.2	-7.5	3.6	**2.5**	2.4	**0.6**
M	250	-98.3	-2.4	12.9	**12.3**	10.4	**10.3**
H	50	-86.1	-2.4	4.9	**2.5**	5.5	**3.1**
H	100	-90.2	-3.8	4.7	**-0.7**	4.7	**-0.2**
H	150	-93.4	-3.4	4.9	**-1.4**	5	**-0.6**
H	200	-95.1	-5.1	5.0	**-2.7**	5	**-0.2**
H	250	-99.0	0.7	2.3	**-0.4**	1.7	**-1.5**

estimates parameters for transmission and reception within a region. We now see how to take it to a heterogeneous propagation environment in an automated fashion to enable fast deployment.

Overview: See Figure 1 for an overview of the building blocks of the tool. The tool takes as input from the user the frequency of operation, transmission power, and receiver noise power (N_0) for which coverage needs to be predicted. A spatial resolution can also be set based on the carrier frequency. The user can also click on a map to select one or more transmitter locations. Two additional inputs to the tool are (1) the geographic information system (GIS) data which provides high level information about the area of operation and the user-selected transmitter location(s), and (2) experimental data from measurements. The GIS data is handled by a map pre-processor that classifies the area of operation into appropriate propagation environments or regions and generates a colour-coded image; we will discuss this soon. The data-driven approach takes the experimental data and converts it to useful antenna-related, frequency-related, and pathloss-related parameters; we already discussed this in the previous section. These are then sent to the central RSSI computing engine. Finally, a heat map is generated and superimposed on the colour-coded map after RSSI calculations. In the subsections that follow, we provide details on each of the blocks.

A. Pre-processing Of Map

In an earlier work, the IISc campus (Figure 3) was broadly classified into five different regions namely Open area, Buildings, Roads, Moderate woods (sparsely thick trees) and Heavy woods (denser thin trees); see [1]. For predicting signal strength in such a diverse area, one needs to find the exact boundaries of these regions. One also needs good measurements in each of these areas. We have already discussed the extensive measurements carried out by us and how they have been used to generate good propagation parameters.

Figure 3 shows the "Base Image", the extracted "Colour Image" and the generated "Greyscale Image". A base map of the area of interest can be obtained from different sources like Google Maps, Bing Maps, Open Street Map (OSM), etc. We have used data from OSM since it is free and widely supported. Quantum GIS, a free, open-source GIS application which supports creation, editing and visualisation of geospatial data, was used to extract the five regions from the base map. In Figure 3, steps 1-4 indicate extraction of open areas, buildings, roads and heavy woods, respectively. Remaining areas are taken as moderately wooded. Unique colours are assigned for identifying these regions. The "Colour Image" is then obtained by combining all these layers. The white parts in the image represent the moderately wooded regions. Each colour in this image has three components (Red, Green, and Blue). For ease of computation, this colour image is then converted to a "Greyscale Image", the last image in Figure 3. Each of the five regions are represented by different shades of grey.

Even though the colour image of the sectionalised map is displayed to the user, it is converted to greyscale image before performing RSSI calculations. A re-sizing of the original image is then done. It decreases the number of pixels in the image, based on the frequency of operation, to speed up RSSI computation. Maps are descriptive figures with possibly different x and y scales. Actual distance between two points may differ from the distance between their respective pixels on the map. So, scaling from pixel distance to actual distance must be done for accurate RSSI calculations.

RSSI variations are predominantly due to changes in the environment through which the signal propagates. The pre-processing on the base image helps our algorithm capture the different environments accurately thereby facilitating an easy computation of RSSI at the receiver, as we will soon describe.

B. Algorithm for RSSI computation in a heterogeneous region

As already, indicated, the user can select multiple transmitter locations just by clicking on the map. Then predicted received power P_{Rx} is computed independently at each pixel for each transmitter. Some computation savings will be highlighted.

From (1), the received power P_{Rx} is

$$P_{Rx} = C \cdot P_{Tx} \cdot h_{Tx}^2 \cdot h_{Rx}^{\gamma} \cdot d^{-\eta_r} \cdot f^{-\kappa_r} \tag{7}$$

Using the data driven approach described in Section II, C, γ, η_r and κ_r for the five different regions in the test-bed IISc campus have been inferred.

Let the transmitter Tx be located in a certain pixel, say pixel 1. We assume that Tx is located at the centre of this pixel. Consider identifying the received power at a candidate pixel of interest. We assume that the receiver is at the centre of this pixel. Let d be the distance between the transmitter and this candidate receiver location.

To compute the predicted received power at this pixel, we simply use the formula (7), with C and γ as per measurement, but with η_{eff} and κ_{eff} as given in (8) and (9), respectively, given below. The two images in Figure 4 illustrate the idea.

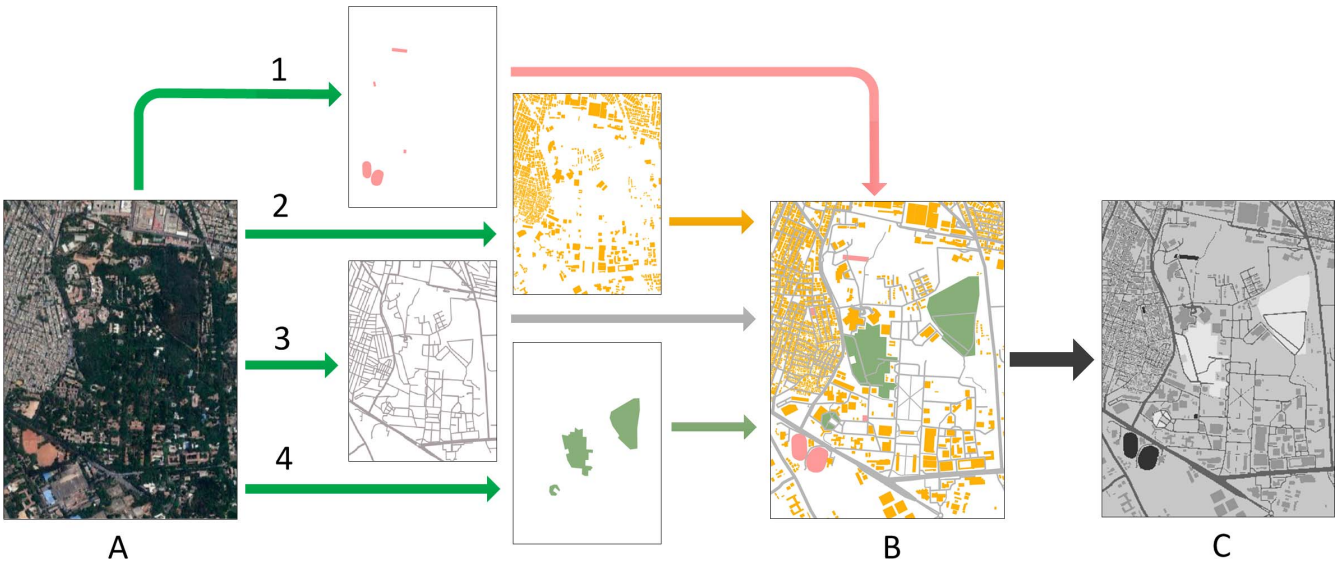

Fig. 3: Different stages of map processing where A = Base Image, B = Colour Image and C = Greyscale Image

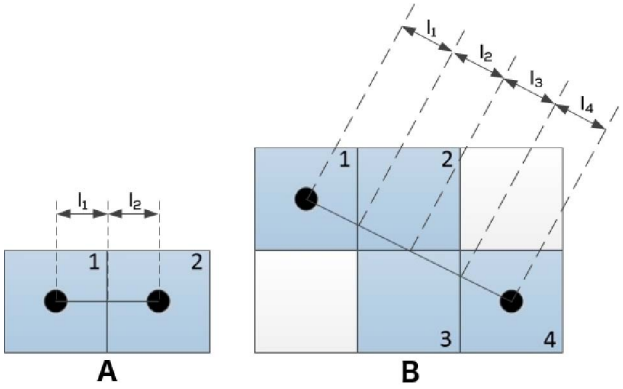

Fig. 4: Ray tracing for (A) adjacent pixels and (B) non-adjacent pixels

Draw a line between the transmitter and the receiver and identify all the pixels through which the line passes. Suppose there are i such pixels. Identify the lengths of the line segments in each pixel. Let these be $l_k, k = 1, \ldots, i$. Associate with each pixel a region $r_k,\ k = 1, \ldots, i$. In Figure 4(A), we have $i = 2$, and in Figure 4(B), we have $i = 4$.

We take κ_{eff} to be the weighted average of the individual κ_{r_k}'s, weighted by the lengths of the line-segments:

$$\kappa_{\text{eff}} := \frac{\sum_{k=1}^{i} l_k \kappa_{r_k}}{\sum_{k=1}^{i} l_k}. \tag{8}$$

To compute the effective pathloss parameter in Figure 4(A) with $i = 2$ segments, we take, with $d = l_1 + l_2$, [2, p. 85]

$$d^{-\eta_{\text{eff}}} = l_1^{-\eta_{r_1}} \cdot \left(\frac{l_1 + l_2}{l_1}\right)^{\eta_{r_2}},$$

The intuition comes from Huygen's principle that there is an imaginary transmitter at the pixel interface that radiates into the second region at exactly that power which it receives from the first region. This can be extended. To compute the effective pathloss parameter in Figure 4(B) with $i = 4$ segments, we deduce, this time with $d = l_1 + l_2 + l_3 + l_4$,

$$d^{-\eta_{\text{eff}}} = {l_1}^{-\eta_{r_1}} \cdot \left(\frac{l_1 + l_2}{l_1}\right)^{-\eta_{r_2}} \cdot \left(\frac{l_1 + l_2 + l_3}{l_1 + l_2}\right)^{-\eta_{r_3}} \cdot \left(\frac{l_1 + l_2 + l_3 + l_4}{l_1 + l_2 + l_3}\right)^{-\eta_{r_4}}.$$

Generalising, for a line segment that passes through i pixels, we take η_{eff} to be as calculated from

$$d^{-\eta_{\text{eff}}} = {l_1}^{-\eta_{r_1}} \cdot \prod_{j=2}^{i} \left(\frac{\sum_{k=1}^{j} l_k}{\sum_{k=1}^{j-1} l_k}\right)^{-\eta_{r_j}}. \tag{9}$$

For each transmitter, we compute the RSSI at each pixel as above. We then take the maximum value across transmitters, and associate the pixel to the corresponding transmitter. A receiver in this pixel will receive the strongest signal from that associated transmitter. The pixel is then coloured according to the predicted received power. The actual algorithm proceeds by expanding around the transmitter in concentric ℓ_∞-circles, and then stops when the RSSI is below a threshold, say -100 dBm, in an attempt to minimise the computational load. The following algorithm summarises the steps. It assumes the existence of a subroutine $RSSI((y, x), (y', x'))$ which returns the RSSI (as obtained using the method above) when the transmitter is at pixel (y, x) and the receiver is at pixel (y', x').

C. Heat Map

Figure 5 shows the coverage for five transmitter locations (marked 1-5 in Figure 5(A)) inside the IISc campus. Transmitters 1-5 are kept, respectively, in an open area, on a road junction, inside a building, in a moderately wooded area and in a heavily wooded area. Receiver sensitivity was set to -100 dBm and locations with RSSI less than -100 dBm from all five transmitters were taken to be out of coverage. Black pixels in Figure 5(B) are areas whose RSSI is larger than

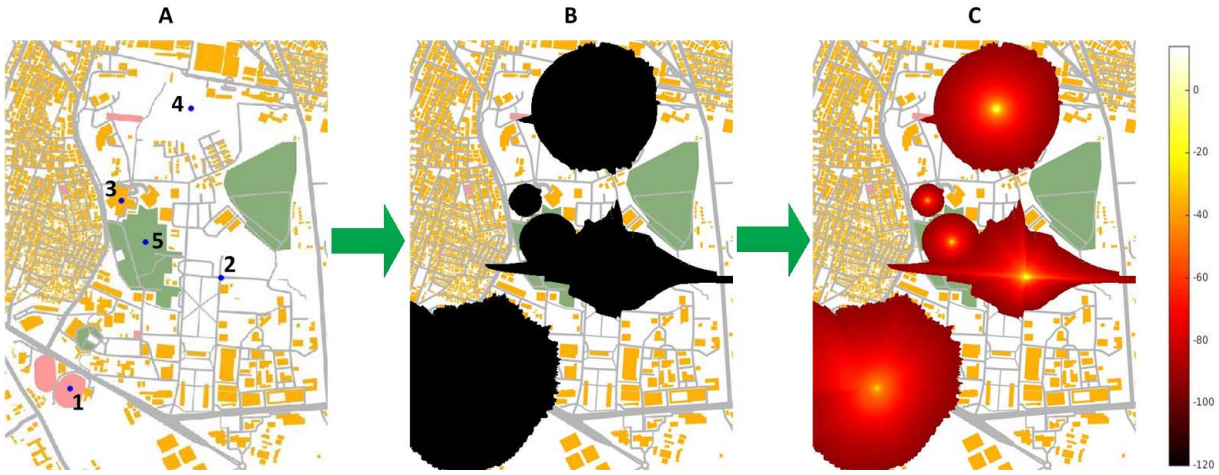

Fig. 5: Coverage for transmitters in different environments for transmitter and receiver height of 1 m where A = Tx Locations, B = pixels with RSSI $<$ -100 dBm and C = Heat Map

Algorithm for RSSI computation

Data: A set of N transmitters
$T = \{(y_1, x_1), (y_2, x_2), \ldots, (y_N, x_N)\}$ where each cartesian pair $(y_i, x_i), 1 \leq i \leq N$ represents a transmitter location in the Euclidean plane.
Result: $MatrixRSSI_{Y \times X \times N}$ where
Y = number of pixels in y-direction
X = number of pixels in x-direction

Initialization;
for $i \leftarrow 1$ **to** N **do**
  Set $S = \{y_i, x_i\}$;
  for $j \leftarrow 1$ **to** $\max\{X, Y\}$ **do**
    Create a set R_1 *of all the pixels such that* $x \in [x_{i-j}, x_{i+j}]$ *and* $y \in [y_{i-j}, y_{i+j}]$;
    $R_2 = R_1 - S$;
    for *each pixel* $p = (p_y, p_x) \in R_2$ **do**
      $MatrixRSSI(u, v, i) \leftarrow RSSI((y_i, x_i), (p_y, p_x))$;
    end
    $S = R_1$;
    if $\forall p = (p_y, p_x) \in R_2, RSSI((y_i, x_i), (p_y, p_x)) \leq$ *-100 dBm* **then**
      break;
    end
  end
end

-100 dBm (a negative image). A heat map is generated for only these locations as shown in Figure 5(C). As the transmitters are placed in different environments, their coverage areas as well as their coverage patterns are different. Being in an open area, transmitter 1 coverage area is more than those of the other transmitters. Transmitter 2 shows longer range along road-lines, but the RSSI degrades faster in the other directions. As transmitter 3 is inside a building, its coverage area is the least. Transmitter 4 has higher coverage area as compared to transmitter 5 because the area around transmitter 5 is more thickly wooded than the area around transmitter 4. All these observations are qualitatively appealing. In addition, it is evident from the coverage pattern that heterogeneity affects signal propagation and in turn the coverage area, and our GIS-enabled data-driven RSSI estimation tool has captured this heterogeneity in a quantitative fashion.

IV. Conclusion

We demonstrated that coverage estimation can improve significantly by employing all available information. We proposed an ML method inspired by logistic regression to make the joint use of packet error rate information as well as RSSI measurements on correctly received packets. As indicated earlier, our method can be easily adapted to other loss functions, different weights for packet error rate and RSSI-estimation-error loss functions, and also incremental updating. We also showed how this unified estimation procedure can be effectively used, along with readily available open-source GIS data and automated classification of regions into various propagation environments, in order to estimate coverage in a heterogeneous propagation environment. The heat map enables easy visualisation of coverage as well as coverage holes, thereby leading to better, more efficient, and faster deployment of IoT services.

References

[1] N. Rathod, P. Jain, R. Subramanian, S. Yawalkar, M. Sunkenapally, B. Amrutur, and R. Sundaresan, "Performance analysis of wireless devices for a campus-wide IoT network," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2015 13th International Symposium on*, IEEE, 2015.

[2] M. Yacoub, *Foundations of Mobile Radio Engineering*. Taylor & Francis, 1993.

[3] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, Springer New York, 2013.

[5] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.