

Maximum Likelihood Topology Maps for Wireless Sensor Networks Using an Automated Robot

Ashanie Gunathillake

School of Electrical Engineering
and Telecommunication,

University of New South Wales,
Sydney, Australia

NICTA, Eveleigh, NSW 2015, Australia

ashanie@student.unsw.edu.au

Andrey V. Savkin

School of Electrical Engineering
and Telecommunication,

University of New South Wales,
Sydney, Australia

a.savkin@unsw.edu.au

Anura P. Jayasumana

Department of Electrical and
Computer Engineering,

Colorado State University,
Fort Collins, CO 80523 USA

anura.jayasumana@colostate.edu

Abstract— Topology maps represent the layout arrangement of nodes while maintaining the connectivity. As it is extracted using connectivity information only, it does not accurately represent the physical layout such as physical voids, shape, and relative distances among physical positions of sensor nodes. A novel concept Maximum Likelihood-Topology Maps for Wireless Sensor Networks is presented. As it is based on a packet reception probability function, which is sensitive to the distance, it represents the physical layout more accurately. In this paper, we use a binary matrix recorded by a mobile robot representing the reception of packets from sensor nodes by the mobile robot at different locations along the robots trajectory. Maximum likelihood topology coordinates are then extracted from the binary matrix by using a packet receiving probability function. Also, the robot trajectory is automated to avoid the obstacles and cover the entire network within least possible amount of time. The result shows that our algorithm generates topology maps for various network shapes under different environmental conditions accurately, and that it outperforms the existing algorithms by representing the physical layout of the network more accurately

Keywords—Localization, Packet Receiving Probability, Signal Propagation, Topological Map, Wireless Sensor Network

I. INTRODUCTION

Many Wireless Sensor Network (WSN) protocols require the location coordinates or map of the sensor nodes, as the data collected by the sensors are useful when considered in the context of the location from which it was collected. The location information can be obtained by the incorporation of hardware components such as Global Position Systems (GPS). However, cost prevents the use of expensive hardware devices in large-scale networks and also they do not work in all environments.

Therefore, one of the major challenges in WSN is to determine the physical coordinates of the sensor nodes while minimizing the hardware cost. To address this issue, numerous localization algorithms have been proposed in the literature. These can be divided into two categories, namely range-based and range-free localization algorithms. Range-based algorithms use special hardware components to measure range-based parameters such as received signal strength (RSS) [1], time of arrival (TOA) [2], angle of arrival (AOA) [3]. These algorithms are affected by noise, fading of the signals and interference [4] and as a result, their accuracy decreases in environments with

obstacles. Range-free algorithms rely on the information about the connectivity of the sensors and known location of some of the sensors (anchor nodes) instead of special hardware. However, the accuracy of these algorithms is highly dependent on the number of anchor nodes and their distribution [5] [6].

A topology preserving map is an attractive alternative to the physical map of the network. Topology map is a representative of arrangement of node that preserves the connectivity. Topology map are obtained by Singular Value Decomposition (SVD) of Virtual Coordinate System (VCS) [7], in which sensor node is identified by a vector that contains the distance in hops to a set of anchor nodes. Even though topology maps are based only on the connectivity, they have been demonstrated to preserve the general shapes of voids and boundaries. They are non-linearly distorted versions of physical maps.

The objective of this research is to come up with a topology preserving map that is closer to the actual physical map, but without requiring the expense associated with localization based on actual physical distance measurements. In the process, we generalize the concept of topology map, by using measurements other than direct hop distances to obtain the map. To this end, we propose a Maximum Likelihood Topology Map (ML-TM) of a sensor network. We consider the problem of generating a map of the network by using a mobile robot. To avoid the disadvantages described above of geographic localization, It does not attempt to measure the actual physical distances. Unlike the case with topology preserving maps in [7], it does not rely solely on the connectivity. Here we use a binary matrix to calculate the maximum likelihood topology coordinates that reduce the dependency of the output on range-based parameters. The robot moves in the area where the network is deployed, and gathers a binary matrix based on the packets received from the nodes from different locations. Then, the topology coordinates are calculated by the binary matrix and a packet receiving probability function, which is sensitive to the distance. Received Signal Strength Indicator (RSSI) based algorithms extract the distances from received power, which encounters significant errors due to RF communication effects. The proposed scheme does not require such error prone and unreliable distance measurements. On the other hand, range-free algorithms use a

hop count matrix, where one hop may correspond to widely different physical distances depending on the node location and its RF environment, e.g. a node may move some distance without affecting the hop distances. Thus, to overcome those problem, we use a binary matrix recorded by a mobile robot, instead of using range-based parameter values, and a probability function sensitive to the distance to convert the binary matrix to a distance matrix. By using a distance sensitive conversion, we try to eliminate the drawbacks in range-free algorithms. Moreover, the trajectory of the mobile robot is automated. The robot traverses the area in which the network is deployed, while listening to transmissions from different locations. During the robot movement, it records a binary matrix in some locations of its trajectory based on packet reception from sensor nodes. Finally, using the liner-decay packet receiving model proposed in Section III, the maximum likelihood topology coordinates of each sensor nodes are calculated. From the results it can be seen that the method is being able to provide a topological map of nodes that is significantly closer to the actual physical layout than existing hop-based topology preserving maps. The proposed topology map preserve the dimensions and shapes of features such as physical voids and network boundaries. Also, it outperforms the RSSI geographical localization and hop based topology maps. The outline of this paper is as follows. Section II reviews the background of our work. Section III discusses the details of our proposed algorithm and Section IV presents the results. Finally, Section V presents the conclusions and future work.

II. BACKGROUND

Prior work in the area of network mapping can be grouped into two categories, namely physical localization and topology mapping schemes for networks.

A. Localization

The most relevant are relative localization techniques, which calculate the actual positions of nodes using anchor nodes and range-based parameters. In range-based techniques special hardware devices are used to measure the range-based parameters such as signal strength, time of arrival and/or angle of arrival. In RSSI based localization algorithms, theoretical or empirical models are used to estimate the distance from the signal strength of receiving packets [8]. The theoretical models rely on an RF signal transmission loss model to calculate the distance between two nodes [9] [10]. The empirical models use a two steps process to obtain the location. First, they create an offline RSSI database, using anchor nodes. Secondly, it determines the coordinates of non-anchor nodes by matching the received signal strength to a record of the database [1] [4].

The localization methods based on time difference of arrival (TDOA) estimates the coordinates of an unknown node by anchor nodes' coordinates and the time difference of arrival from those anchors to the node. To be able to calculate the coordinates, it needs at least four anchors [11]. Furthermore, in time of arrival (TOA) based localization, the anchor nodes broadcast a signal and the sensor nodes that receive these broadcasts, use

the time difference of arrivals. This requires the sensor timers to be synchronized. To overcome this requirement, researchers have proposed the use of Round-Trip TOA (RTOA) [12] and Two Way TOA (TW-TOA) [13]. Again, TOA localization does not address issues associated with obstacles and RF signal transmission affects, and achieving accurate synchronization is difficult.

In range-free localization algorithms, the locations of nodes are obtained without using any special hardware. They rely on the connectivity information of nodes. First, it gets the distance in hops and then maps the hop distance to geometric distance [6] [5] using anchor node locations. Therefore, the accuracy of these algorithms highly depends on the number of anchor nodes and their deployment. The key issue of range-free algorithms is the distance estimation, i.e. the mapping of the hop distance to geometric distance.

B. Mapping Schemes for Networks

Topological mapping techniques are fundamentally different to localization techniques because the mapping algorithms are concerned with the relative arrangement of the nodes. They are not concerned with the exact location of the nodes. When a topology map is isomorphic to the physical layout, it is a good substitute for the actual physical maps for many purposes. In this case, the mapping schemes expect the relative distances to be accurate, but not the physical distances.

In [14], several unsupervised learning algorithms have been proposed that use eigenvalue decomposition for obtaining a lower dimensional embedding of the data. It provides a unified framework for extending Multi-Dimensional Scaling (MDS) [15], Isomap [16], Local Linear Embedding (LLE), and Laplacian Eigenmaps (LE) [14]. MDS [15] is a commonly used statistical technique in information visualization for exploring similarities or dissimilarities in higher-dimensional data from the complete distance matrix (similarity matrix), which is defined as the matrix of all the pairwise distances between points/nodes. Isomaps [16] is an extension of MDS to geodesic distance-based topology map generation. Again, the geodesic distances are actual distances between nodes, which require expensive error prone distance estimators such as RSSI or time of arrival (TOA). Moreover, LLE and LE both use an iterative approach to preserve the neighborhood distances, the realization of which is infeasible in energy-limited WSNs [7].

Topology Preserving Maps (TPM) [7] are obtained by starting with a Virtual Coordinate System (VCS) in which each node is characterized by the hop distances to a set of anchor nodes. A Singular Value Decomposition (SVD) is then used to recover the layout maps from the VCS. In Virtual Coordinate System (VCS), the layout information such as physical voids, shape etc. are absent. To overcome that, Singular Value Decomposition (SVD) of VCs is used in this method. Furthermore [7] shows that transformation for topological map from virtual coordinates can be generated using a subset of nodes. However, when the number of nodes increase, the time required to generate the virtual coordinate matrix also increases. And also, the result, highly depends on the anchor distribution.

III. MAXIMUM LIKELIHOOD TOPOLOGY MAPS (ML-TM)

This section describes Maximum Likelihood Topology Map (ML-TM), a novel method to create topology maps of networks that more accurately characterizes the physical features than the traditional TMs. Unlike these schemes which start with a hop-distance based coordinate system, ML-TM relies on the set of locations from which packets can be received from a given node to estimate its position. Consider a mobile robot that traverses a sensor field. At a given time, it is able to receive packets from nodes in its vicinity. The probability of receiving a packet is sensitive to the distance, and we exploit this property. The robot keeps track of which nodes it received packets from at different times, and then finds the maximum likelihood position of each node.

The following subsections describe the probability function that we used and how the coordinates are calculated.

A. The Packet Receiving Probability Function

This function describes the probability of receiving packets from a sensor when robot is at a particular distance. Let, $S(d)$ be the probability value when robot is at distance d from the sensor. Then, $S(d)$ satisfies the following constraints:

$$\begin{aligned} 0 &\leq S(d) \leq 1 \quad \forall d \\ S(d_1) &\leq S(d_2) \quad \forall d_1 \geq d_2 \\ S(d) &= 0 \quad \forall d > R \end{aligned} \quad (1)$$

where R is some given distance.

The results presented in Section IV are based on the following specific $S(d)$.

$$\begin{aligned} S(d) &:= p_0 \quad \forall d \leq r \\ S(d) &:= 0 \quad \forall d \geq R \\ S(d) &:= \frac{p_0(R-d)}{(R-r)} \quad \forall r < d < R \end{aligned} \quad (2)$$

where $0 < p_0 \leq 1$, $0 < r < R \leq R_c$ are some given constants. R_c is the communication range of a sensor node. It is obvious that the function (2) satisfies all the conditions (1).

Note that the receiving probability function we use here is an intermediate model between RSSI [17] and VC [7], both of which satisfies the condition (1) as in Figure 1. VC maps a range of distance values from 0 to R to a unit one hop, which is the major cause for the map being different from a physical map. On the other hand, RSSI is based on an exact assumed relationship between the signal strength and the distance, but such a model is not accurate due to fading, interference and noise, and hard to estimate the parameters for different environmental situations. Hence, in theory, the proposed approach can accommodate both models, but in practice RSSI model is time varying, environment dependent, etc. Therefore, here we consider an intermediate level between RSSI model and VC model to obtain our topological map.

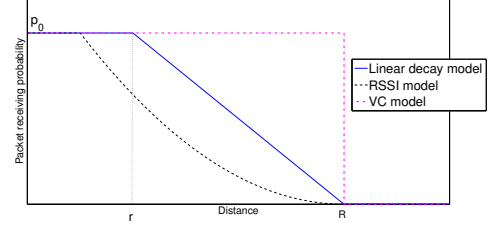


Fig. 1: Packet receiving probability function for different models

B. Calculating Topological Coordinates

This section describes the information gathering by mobile robot and prediction of the ML topology coordinates.

Let $(x_R(t), y_R(t))$ be the coordinates of the mobile robot at time t . The network consists of n stationary sensors labeled $i = 1, 2, \dots, n$, whose locations are not known. The robot is assumed to be able to receive packets at $L_1, L_2, \dots, L_N$ locations on the robot trajectory at discrete time instances, $t_1, t_2, \dots, t_N$. Robot stops in these time slot t_i at the location L_i and listen to its local neighborhood nodes and then moves on to the next position in the trajectory. The robot can here from a node from multiple locations. Hence we introduce a binary matrix M of order $n \times N$ by the following rule:

$$\begin{aligned} M(i, j) &= 1, \text{ if the robot receives a packet from the} \\ &\quad \text{sensor } i \text{ at the time } t_j; \\ M(i, j) &= 0, \text{ if the robot does not get a packet from the} \\ &\quad \text{sensor } i \text{ at the time } t_j. \end{aligned}$$

Such a binary matrix M is called the packet receiving matrix. Table I refers to the M matrix of the network shown in Figure 2.

Then, based on the robots trajectory $(x_R(t), y_R(t))$, the receiving probability function $S(d)$ and the packet receiving matrix M , we introduce the function $P(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$. $P(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$ is the probability to obtain the packet receiving matrix M for the robots trajectory $(x_R(t), y_R(t))$ when the sensors $1, 2, \dots, n$ are located at the points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, respectively. The probability values are obtained using the function $S(d)$ where d is the distance between the robot and the sensor at the time of sending the packet.

Definition: A number set $(x_1^{opt}, y_1^{opt}, x_2^{opt}, y_2^{opt}, \dots, x_n^{opt}, y_n^{opt})$ is said to be the set of most likely locations of the sensors $1, 2, \dots, n$ based on the robots trajectory $(x_R(t), y_R(t))$, the receiving probability function $S(d)$ and the receiving matrix M if,

$$\begin{aligned} P(x_1^{opt}, y_1^{opt}, x_2^{opt}, y_2^{opt}, \dots, x_n^{opt}, y_n^{opt}) &\geq \\ P(x_1, y_1, x_2, y_2, \dots, x_n, y_n) \end{aligned} \quad (3)$$

for any $(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$.

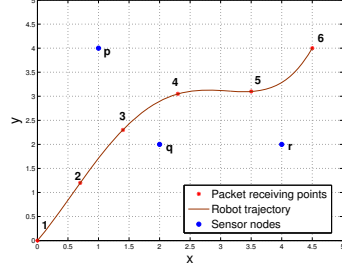


Fig. 2: Sensor network with three nodes

TABLE I: The M matrix

	t_1	t_2	t_3	t_4	t_5	t_6
p	0	0	1	0	0	0
q	0	1	1	1	0	0
r	0	0	0	0	1	0

TABLE II: $P_i(x_i, y_i)$ calculation for the network in Figure 2

Grid Vertex	Point 1		Point 2		Point 3		Point 4		Point 5		Point 6		$P_i(x_i, y_i)$
	d_{ij}	$Z(d_{ij})$	d_{ij}	$Z(d_{ij})$	d_{ij}	$Z(d_{ij})$	d_{ij}	$Z(d_{ij})$	d_{ij}	$Z(d_{ij})$	d_{ij}	$Z(d_{ij})$	
(0,0)	0	0	1.44	0.7064	2.64	1	3.78	1	4.68	1	6.02	1	0
(1,1)	1.41	0.69	0.36	0.14	1.36	0.3368	2.42	1	3.2	1	4.6	1	0.0325
(1,3)	3.16	1	1.81	0.90	0.76	0.6518	1.3	0.6316	2.50	1	3.64	1	0.3708
(4,2)	4.4	1	3.2	1	2.72	0	1.97	0.985	1.21	0.5833	2.06	1	0

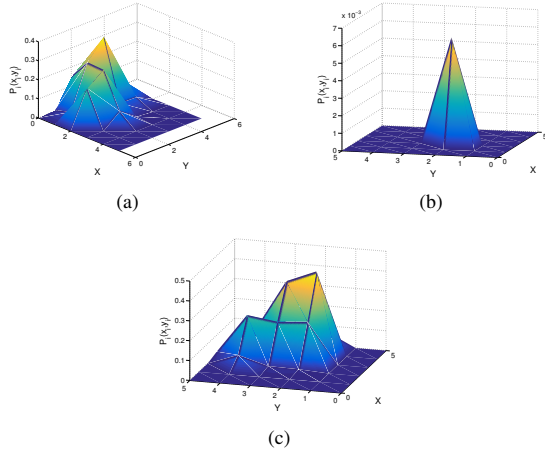


Fig. 3: (a) $P_i(x_i, y_i)$ distribution of node p, (b) $P_i(x_i, y_i)$ distribution of node q and (c) $P_i(x_i, y_i)$ distribution of node r

In other words, the most likely location of the sensors are the points on the plane that maximize the probability of producing the measured packet receiving matrix M for the robot's trajectory.

Our goal is to find optimal solution for the location of the sensor in topological map. The algorithm is as follows:

$$P_i(x_i, y_i) = Z(d_{i1})Z(d_{i2})...Z(d_{iN}) \quad (4)$$

Furthermore, it is obvious that $P_i(x_i, y_i) = 0$ for any node i for all (x_i, y_i) that are outside the R - neighbourhood of the robots trajectory $(x_R(t), y_R(t))$. Hence, we can take a small step grid in the R - neighbourhood of the robots trajectory

Algorithm 1: Calculating ML Topology Coordinates

Result: Topology coordinates of sensor nodes

```

1  $N \leftarrow \text{total\_number\_of\_nodes};$ 
2  $T \leftarrow \text{Initializing\_time};$ 
3 while  $t_i < T$  do
4   sensors transmit signal;
5   robot update the M matrix;
6 end
7 divide R-neighborhood of robot trajectory into grids;
8 for each node  $i \in N$  do
9   calculate  $P_i(x_i, y_i)$  for vertices in the grid;
10  find maximum  $P_i(x_i, y_i)$  ;
11   $\text{coordinate}_i \leftarrow \text{vertex coordinate of maximum}$ 
     $P_i(x_i, y_i);$ 
12 end
```

$(x_R(t), y_R(t))$, and for any vertex of the grid and for any $i = 1, 2, \dots, n$, calculate the value of $P_i(x_i, y_i)$ based on the formulas (4),(6). Then we take the grid vertex (x_i^{opt}, y_i^{opt}) delivering the maximum value of $P_i(x_i, y_i)$ among all vertices of the grid. This is an approximation of the optimal location of the sensor node i .

Let vectors $m_1, m_2, \dots, m_n$ be the rows of the matrix M. Hence the vector m_i is the packet receiving vector of the sensor node i describing receiving/not receiving by the robot signals from the sensor node i . Also, let $m_i(1), m_i(2), \dots, m_i(N)$ denote the elements of the vector m_i . Furthermore, for all $i = 1, 2, \dots, n$, introduce the function $P_i(x_i, y_i)$ which is the probability of obtaining the packet receiving vector m_i for the robots trajectory $(x_R(t), y_R(t))$ under the assumptions that the sensor node i is located at the point (x_i, y_i) and

the probability for the robot to receive a packet from the sensor node i is described by the function $S(d)$ where d is the distance between the robot and the sensor node i at the time of sending the packet. Since, the location of sensor node i does not depend on the location of sensor node j , we can obtain $P(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$ as,

$$P(x_1, y_1, x_2, y_2, \dots, x_n, y_n) = P_1(x_1, y_1)P_2(x_2, y_2)\dots P_n(x_n, y_n) \quad (5)$$

Therefore, to find an optimal location of the sensors delivering the maximum value of $P(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$, we just need to find independently locations (x_i, y_i) delivering maximum values of $P_i(x_i, y_i)$. Moreover, to find $P_i(x_i, y_i)$, we defined a function $Z(d_{ij})$ as,

$$Z(d_{ij}) = \begin{cases} S(d_{ij}) & \text{if } m_i(j) = 1 \\ 1 - S(d_{ij}) & \text{if } m_i(j) = 0 \end{cases}$$

where

$$d_{ij} := \sqrt{(x_i - x_R(t_j))^2 + (y_i - y_R(t_j))^2} \quad (6)$$

In a binary matrix, there are one of two events that is captured i.e. either one or zero. If it is one, there is a $S(d_{ij})$ probability of locating sensor node i in a distance of d_{ij} from the robot location. On the other hand, if it is zero, there is a $1 - S(d_{ij})$ probability of locating sensor node i in a distance of d_{ij} from the robot location. This is similar to the coin-tossing example. Thus we can calculate $P_i(x_i, y_i)$ as in equation(4).

Let us consider an example to illustrates the calculation of maximum likelihood sensor coordinates using the probability function and the binary matrix. Figure 2 shows a sensor network with three location unknown nodes (i.e. p, q, r) and a random trajectory of a mobile robot. The points 1 to 6 represent the locations that robot receives packet from nodes at time $t_1 < t_2 < t_3 < t_4 < t_5 < t_6$ respectively. Table I shows the binary matrix recorded by the robot at each time instance. Then the R neighborhood of the robot trajectory is divided into a grid. For the simplicity, we consider the grid size to be 1. Consider the row vector of M matrix corresponds to node p. Table II illustrates the calculations of $P_i(x_i, y_i)$ for some grid vertices. First calculate the distances from grid vertex to all the robot locations 1 to 6. Then using the probability function (with $p_0=1$, $r=0.2$ and $R=2$) find the $S(d_{ij})$ value for all distances and calculate the $Z(d_{ij})$ as in equation 6. Afterwards, $P_i(x_i, y_i)$ can be calculated as in equation 4. Figure 3 shows the $P_i(x_i, y_i)$ distribution over all the grid vertices of three sensor nodes. Finally, choose the grid vertex delivering the maximum value of $P_i(x_i, y_i)$ and that is the optimal location of the sensor node. Here the optimal locations of the three sensor nodes are, $p \equiv (1, 3), q \equiv (2, 2), r \equiv (4, 2)$.

C. Trajectory of the Mobile Robot

This section describes the selection of the mobile robot trajectory. Our objective is to cover the entire network by a mobile robot within a least possible time. Therefore we need to decide whether the robot follows a pattern or it randomly

walk around the network. But, in some cases random walk takes longer time to cover the entire network. Hence we decided to go for a trajectory following a pattern such as square curve, triangle curve or a sine curve as shown in Figure 6. A proper trajectory selection will be doing in our future works.

IV. RESULT

The performance of the proposed ML-TM algorithm is evaluated in this section. MATLAB simulation software was used for the computations. The model we used to simulate the received signal is described below.

A. Receive Signal Strength Model

The received signal strength can be modeled as having two components, path loss and shadowing [18]. Therefore, the commonly used propagation model of RF signals is given in equation (7).

$$P_{rx,i}(t) = P_{tx,j} - 10\epsilon \log d_{ij}(t) + v_i(t) \quad (7)$$

where, the received signal strength at node i at time t is $P_{rx,i}(t)$, the transmitted signal strength of the signal at node j is $P_{tx,j}$, the path-loss exponent is ϵ , the distance between node i and node j at time t is $d_{ij}(t)$ and the logarithm of shadowing component on node i at time t is $v_i(t)$. The shadowing values are selected from a normal distribution with zero mean and standard deviation parameter described as in Table III.

However, this model is not suitable for a network with some obstacles. In [19], they proposed a MultiWall-Multifloor Model for RF communication. In this model the variation of the absorption against the thickness of the medium which signal traverse, is not considered. Therefore we updated the equation (7) by using the Lambert-Bouguer law. Let $L_{ob,i}(t)$ is the loss due to signal absorption from obstacles exist in the line of sight of node i and j at time t , then the RF signal propagation model is as in equation (8).

$$P_{rx,i}(t) = P_{tx,j} - 10\epsilon \log d_{ij}(t) - L_{ob,i}(t) + v_i(t) \quad (8)$$

The absorption coefficient and the thickness of the obstacle medium, which signal traverses are α and $d_o(t)$ respectively. Then $L_{ob,i}(t)$ can be calculated as,

$$L_{ob,i}(t) = \sum_{k=1}^n 10\alpha d_o(t) \log(e) \quad (9)$$

where, e is the exponent and n is the number of obstacles exist in between node i and node j .

B. Sensitivity of the Algorithm

The sensitivity of the output against the algorithm parameters is considered in this section. The results of the proposed algorithm depends on the receiving probability function. Therefore, we examine the probability curve for different communication ranges. The p_0 value, which the highest packet receiving probability in the network, is set as 0.95. First, the probability curve is examined in the same environment (free space environment) by changing the communication range from 5m to 25m. The result is shown in Figure 4. It can be seen that the r value increases by small amounts when R_c increases and $R \approx R_c$. Secondly,

r and R values are checked against the different environments by keeping the transmitting power in a constant value. The result is shown in Figure 5 and, it can be seen that r and R values decrease when the environment is full with obstacles. Therefore, r and R values must be selected by considering the communication range of sensor nodes and the environment that sensor network is deployed in. The three different network setups are described below [20] [21].

Case 1: Suburban area with $\epsilon=2.8$.

Case 2: Heavy tree density area with $\epsilon=4.6$.

Case 3: Light tree density area with $\epsilon=3.6$.

For the output evaluation we use an error parameter E_{total} define as in equation (10).

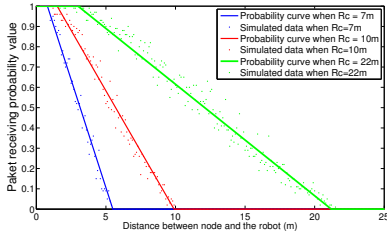


Fig. 4: Packet receiving probability curve for different communication ranges

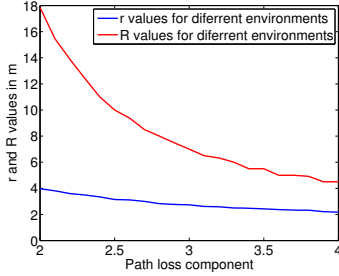


Fig. 5: r and R values for different environment conditions

$$E_{total} = \frac{\sum_{i=1}^n |e_i|}{\sum_{i=1}^n |N_{p,i}|} \times 100\% \quad (10)$$

$$e_i = N_{p,i} - \{N_{p,i} \cap N_{t,i}\}$$

where, $N_{p,i}$ and $N_{t,i}$ are the one hop neighbor set of node i in physical and topological map respectively. This error indicator represents the node connectivity by considering the correlation of physical and topological map distances. Therefore, the distortion rate is considered and find out the connectivity of the nodes.

The network used for the simulation is shown in Figure 8(a) and Table III presents the simulation parameters. Also, Table IV presents the best values, i.e. E_{total} is minimum, for the parameters r and R in the receiving probability function and it justifies the above mention conclusion.

TABLE III: SIMULATION PARAMETERS

Parameter	Value
Transmitted power P_{tx}	-50dB
Sensitivity P_{rx}	-80dB
Suburban area	$\epsilon = 2.7$, $\sigma = 9.6$, $p_0 = 0.95$, $R = R_c = 9.5m$, $r = 2m$
Light tree density area	$\epsilon = 3.6$, $\sigma = 8.2$, $p_0 = 0.95$, $R = R_c = 6m$, $r = 1.2m$

TABLE IV: BEST p_0 AND R VALUES FOR THE THREE CASES

	Case 1	Case 2	Case 3
r	2m	1.2m	1m
R	9.5m	6m	5m

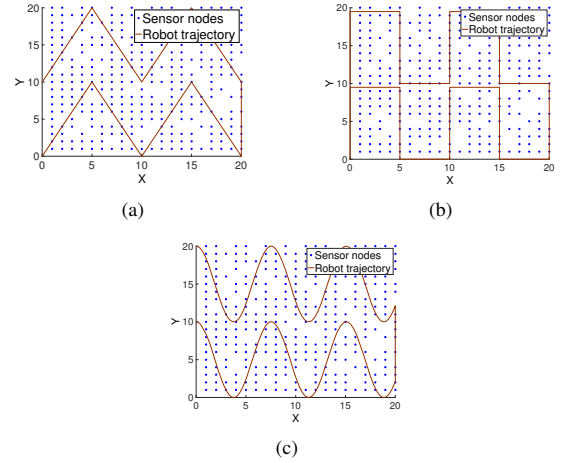


Fig. 6: (a) $P_i(x_i, y_i)$ Triangular, (b) $P_i(x_i, y_i)$ Square and (c) $P_i(x_i, y_i)$ Sin curve robot trajectory patterns

C. Output of the algorithm against robot trajectory pattern

In this section we examine the output of the algorithm by changing the robot trajectory pattern. Here we consider the mobile robot moves with a constant speed along a triangular curve, a square curve and a sine curve as shown in Figure 6. As in the figure, four cycles of the patterns are used to cover the entire network (we kept the number of cycles as the constant for the comparison).

For the comparison, two parameters are considered. First one is the accuracy of the output and the second is the traveling length of the robot to cover the entire network area. The reason we considered the travel length is, when robot travels in a long path, it required more time to cover the entire network.

The results are shown in Table V. From the result it can be seen that the accuracy of the three patterns are same, but in triangular curve pattern length is less than other two. Hence robot required less time to move along the triangular curve

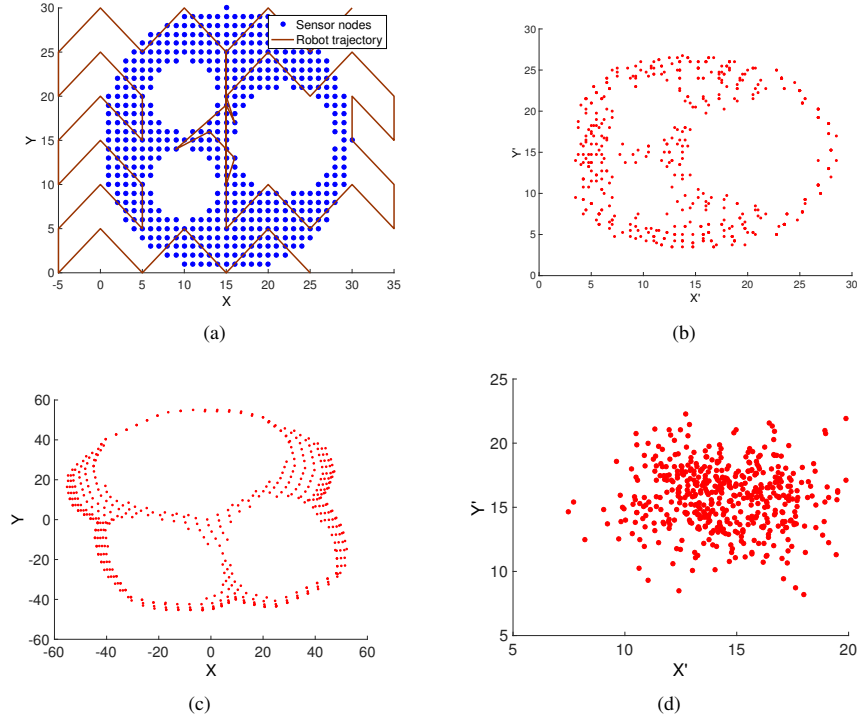


Fig. 7: (a) Circular-shaped network with 496 nodes, (b) Proposed topological map, (c) SVD based TPM and (d) RSSI based map

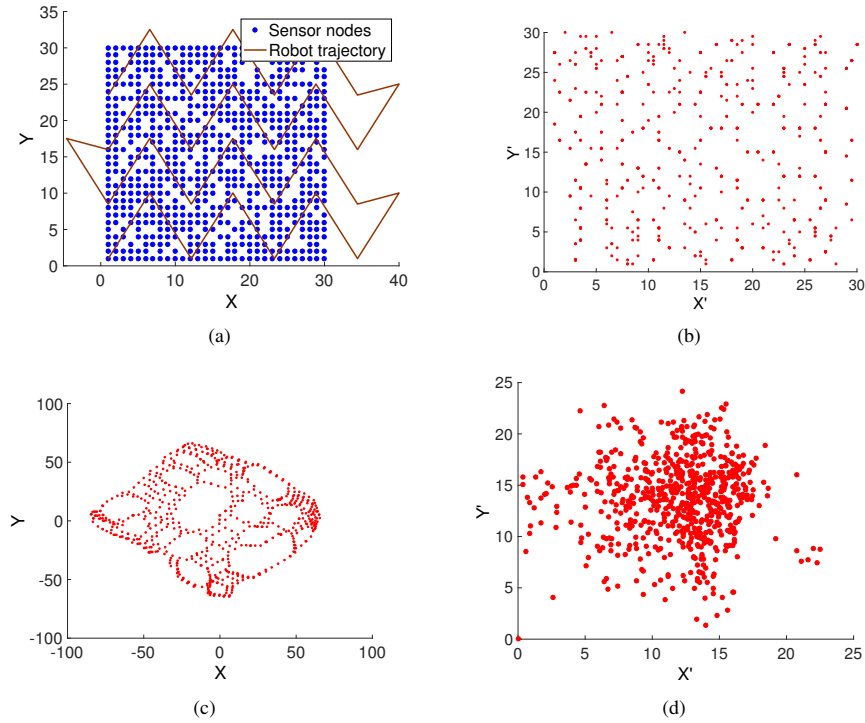


Fig. 8: (a) Sparse grid network with 700 nodes, (b) Proposed topological map, (c) SVD based TPM and (d) RSSI based map

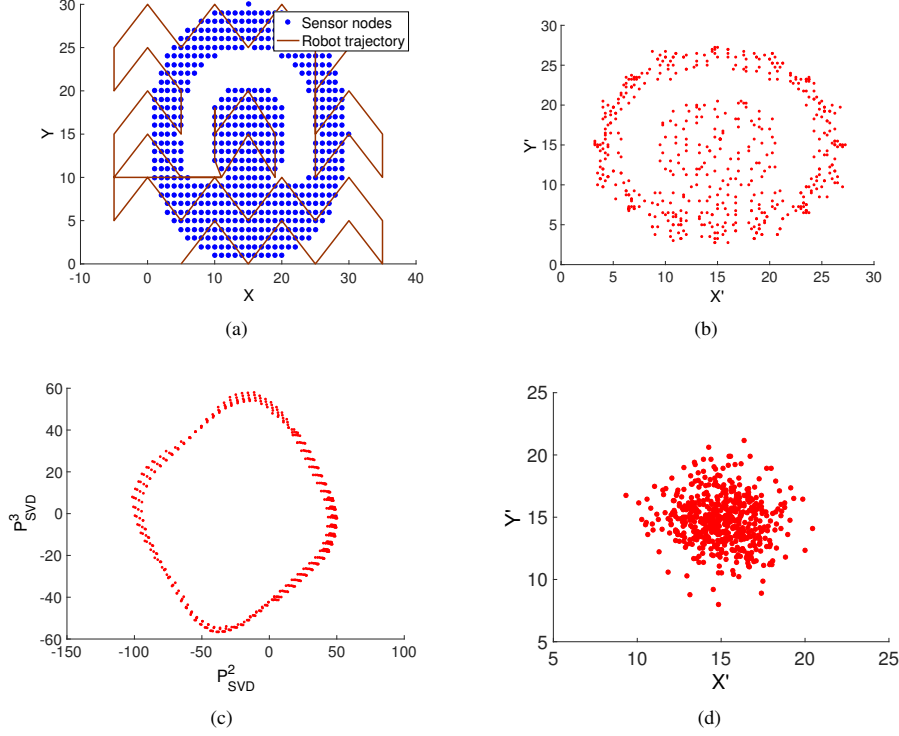


Fig. 9: (a) Concave void network with 554 nodes, (b) Proposed topological map, (c) SVD based TPM and (d) RSSI based map

TABLE V: OUTPUT OF THE ALGORITHM AGAINST ROBOT TRAJECTORY pattern

	E_{total}	Trajectory Length
Triangular	7%	99.4 m
Square	7%	130 m
Sine	7%	124.5 m

trajectory pattern. Therefore, triangular pattern is used in the future performance comparison.

D. Performance Comparison

The performance of the proposed ML-TM is compared with the SVD based TPM [7] and RSSI location method based on the Triangle Centroid Localization [9]. SVD based TPM was selected, because it is the most recent and more relevant work done to our proposed algorithm. Also, to compare our result with range-based algorithm, RSSI based localization was selected. By this comparison, we can see how our method eliminates the error occurs due to RF communication effects (i.e. noise, fading etc.) and anchor selection. The physical maps we selected for the comparison is shown in Figure 7 -9. Here we have different shapes of networks with obstacles and without obstacles. The Figure 7(a) is a 496 sensor nodes circular-shaped network in a suburban area with three physical obstacles (concrete barriers). The Figure 8(a) is a sparse grid

TABLE VI: E_{total} FOR TOPOLOGY MAPS IN FIGURE 7 -9

Figure	E_{total}		
	Proposed	SVD based	RSSI based
	ML-TM	TPM	Map
7(a)	5%	10%	40%
8(a)	7%	14%	33%
9(a)	6%	19%	50%

deployed in a light tree density area with 700 sensor nodes. The Figure 9(a) is a 554 sensor nodes network with a concave void (concrete barriers) in a suburban area. Table III presents the simulation parameters.

Figure 7 - 9, clearly demonstrate the effectiveness of the proposed algorithm. Without any prior knowledge of geographical information, the generated topological maps have captured physical voids and boundaries of the actual physical network. Moreover, it can be seen that the RSSI method mapping is less accurate when the environment is noisy and full of obstacles. Similarly, in SVD based TPM, the orientation of the network is completely lost. It can be seen more clearly in Figure 9.

Moreover, to check the isomorphic nature of the actual physical map and the topological maps, E_{total} is calculated and presented in Table VI. From the result, it can be seen that the number of nodes located in incorrect places is less than or equal to 7%. However in RSSI and SVD based TPM, the maximum

error is 50% and 19% respectively. Thus, the proposed ML-TM error is a minor value comparative to the other two methods. Also, irrespective of the presence or absence of obstacles, the proposed ML-TM provide a more accurate map. Furthermore, it does not depend on the anchor selection or deployment.

V. CONCLUSION

We presented a novel ML-TM algorithm to generate topology map for WSNs without using any special hardware device embedded to sensor nodes. ML-TM is a more accurate map to represent physical layout with voids/obstacles compared to other alternative topology maps. It uses a automated mobile robot that moves on the network to extract the information from sensor nodes, and map it to a different coordinate system by using a signal receiving probability function sensitive to the distance. This function is an intermediate level between of RSSI curves and VC hop-distance approximation. Therefore we have been able to release the effects on modeling RSSI curves for different environmental situations. The result shows that the error percentage is less than 7% error in ML-TM and it outperforms the RSSI geographical localization and hop based topology maps. Also, this method can be used in various network types under different environmental conditions to represent the physical map. Therefore, it can be used as an alternative to geographical map in the automation of sensor network protocol.

Moreover, ML-TM can be used in 3D-WSNs without making any changes to the algorithm. The use of ML-TM in various applications such as routing, target tracking, boundary detection is under investigation.

ACKNOWLEDGMENT

This research is supported in part by Data61|CSIRO Australia.

REFERENCES

- [1] W. Chengdong, C. Shifeng, Z. Yunzhou, C. Long, and W. Hao, "A rssi-based probabilistic distribution localization algorithm for wireless sensor network," in *Information Technology and Artificial Intelligence Conference (ITAIC), 2011 6th IEEE Joint International*, vol. 1, pp. 333–337, Aug 2011.
- [2] Y. Wang and K. Ho, "Tdoa source localization in the presence of synchronization clock bias and sensor position errors," *Signal Processing, IEEE Transactions on*, vol. 61, pp. 4532–4544, Sept 2013.
- [3] S. Kotwal, S. Verma, S. Suryansh, and A. Sharma, "Region based collaborative angle of arrival localization for wireless sensor networks with maximum range information," in *Computational Intelligence and Communication Networks (CICN), 2010 International Conference on*, pp. 301–307, Nov 2010.
- [4] X. Wang, S. Yuan, R. Laura, and W. Langb, "Dynamic localization based on spatial reasoning with rssi in wireless sensor networks for transport logistics," in *Sensors and Actuators*, pp. 421–428, 2011.
- [5] H. Tian, C. Huang, B. M. Blum, J. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking, MobiCom '03*, pp. 81–95, 2003.
- [6] D. Niculescu and B. Nath, "Dv based positioning in ad hoc networks," in *Kluwer journal of Telecommunication Systems*, vol. 22, p. 267280, Jan. 2003.
- [7] D. Dhanapala and A. Jayasumana, "Topology preserving maps ;extracting layout maps of wireless sensor networks from virtual coordinates," *Networking, IEEE/ACM Transactions on*, vol. 22, pp. 784–797, June 2014.
- [8] H. Zhang, J. Zhang, and H. Wu, "An adaptive localization algorithm based on rssi in wireless sensor networks," in *Cloud Computing and Intelligent Systems (CCIS), 2012 IEEE 2nd International Conference on*, vol. 03, pp. 1133–1136, Oct 2012.
- [9] B. Mukhopadhyay, S. Sarangi, and S. Kar, "Novel rssi evaluation models for accurate indoor localization with sensor networks," in *Communications (NCC), 2014 Twentieth National Conference on*, pp. 1–6, Feb 2014.
- [10] R. Al Alawi, "RSSI based location estimation in wireless sensors networks," in *2011 17th IEEE International Conference on Networks (ICON)*, pp. 118–122, Dec 2011.
- [11] C. Savarese, J. Rabaey, and K. Langendoen, "Robust positioning algorithms for distributed ad-hoc wireless sensor networks," *USENIX Annual Technical Conference*, pp. 317–327, 2002.
- [12] H. Wymeersch, J. Lien, and M. Win, "Cooperative localization in wireless networks," *Proceedings of the IEEE*, vol. 97, pp. 427–450, Feb 2009.
- [13] M. Gholami, S. Gezici, and E. Strom, "Improved position estimation using hybrid tw-toa and tdoa in cooperative networks," *Signal Processing, IEEE Transactions on*, vol. 60, pp. 3770–3785, July 2012.
- [14] Y. Bengio, J.-F. Paiement, and P. Vincent, "Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering," in *In Advances in Neural Information Processing Systems*, pp. 177–184, MIT Press, 2003.
- [15] Y. Shang, W. Rumi, Y. Zhang, and M. Fromherz, "Localization from connectivity in sensor networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 15, pp. 961–974, Nov 2004.
- [16] J. Tenenbaum, V. deSilva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, p. 23192323, Dec. 2000.
- [17] D. Turner, S. Savage, and A. C. Snoeren, "On the empirical performance of self-calibrating wifi location systems," in *Proceedings of the 2011 IEEE 36th Conference on Local Computer Networks, LCN '11*, (Washington, DC, USA), pp. 76–84, IEEE Computer Society, 2011.
- [18] P. Pathirana, N. Bulusu, A. Savkin, and S. Jha, "Node localization using mobile robots in delay-tolerant sensor networks," *Mobile Computing, IEEE Transactions on*, vol. 4, pp. 285–296, May 2005.
- [19] M. Lott and I. Forkel, "A multi-wall-and-floor model for indoor radio propagation," in *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*, vol. 1, pp. 464–468 vol.1, 2001.
- [20] S. Seidel, T. Rappaport, S. Jain, M. Lord, and R. Singh, "Path loss, scattering and multipath delay statistics in four european cities for digital cellular and microcellular radiotelephone," *Vehicular Technology, IEEE Transactions on*, vol. 40, pp. 721–730, Nov 1991.
- [21] V. Erceg, L. Greenstein, S. Tjandra, S. Parkoff, A. Gupta, B. Kulic, A. Julius, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *Selected Areas in Communications, IEEE Journal on*, vol. 17, pp. 1205–1211, Jul 1999.