A Feasibility Check for Geographical Cluster Based Routing under Inaccurate Node Localization in Wireless Sensor Networks*

Hannes Frey[1] and Ranjith Pillay[2]

1 **University of Paderborn, Germany,** `hannes.frey@uni-paderborn.de`
2 **Amrita Vishwa Vidyapeetham, Kollam, India,** `ranjith@arl.amrita.edu`

Abstract

Localized geographic single path routing along a wireless network graph requires exact location information about the network nodes to assure message delivery guarantees. Node localization in practice however is not exact. Errors ranging from several centimeters up to several meters are usual. How to perform localized routing in practice when such errors are prevalent?

In this work we look at a promising routing variant which does not completely overcome this practical problem but which mitigates it. The concept does away with trying to find node positions as precise as possible but allows inaccuracies from the very beginning. It partitions the plane by a regular mesh of hexagons. The only information which is of interest is in which cell of that partitioning a node is located in. Using this node embedding, a virtual geographic overlay graph can then be constructed.

To find the node positions we apply three variants of multidimensional scaling, two of them being a node localization approach which has been well studied in the context of sensor networks and one which we apply here for the first time in that context. Using the location information we get from these localization approaches we embed the nodes into the clusters their location falls into. We define two graph metrics to assess the quality of the overlay graph obtained by the embedding. Applying these two metrics in a simulation study, we show that cluster based routing is an eligible approach to support localized geographic routing when location errors are prevalent.

1998 ACM Subject Classification C.2 [Computer-Communication Networks]: Network Protocols – *Protocol verification*

Keywords and phrases Geographical clustering, face routing, planar graph routing, localization, multi dimensional scaling, simulation study

Digital Object Identifier 10.4230/OASIcs.KiVS.2011.145

1 Introduction

We consider geographic routing along geographical clusters which are resulting from a regular hexagon tiling of the plane. With that approach a node does not need to know its exact location. It is sufficient that each node locates itself in the geographical cluster it is actually located in.

As we briefly describe in Section 2, in conjunction with routing, geographical clustering has been considered to improve robustness of localized routing methods. In these routing

* This work was funded in part by the Erasmus Mundus External Cooperation Window (EMECW) of the European Commission.

17th GI/ITG Conference on Communication in Distributed Systems (KiVS'11).
Editors: Norbert Luttenberger, Hagen Peters; pp. 145–156
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

approaches, it is assumed that nodes already know their exact physical location (e.g. by means of GPS). This, however, is not a realistic assumption.

In this work we consider geographical clustering under the assumption that nodes can perform erroneous range measurements with their immediate neighboring nodes. In Section 3 we apply multidimensional scaling, one of the well known localization techniques applied in sensor networks, to estimate node locations based on range measurements. We apply three different multidimensional scaling variants, a non-metric, a classical, and a probabilistic one. The latter variant has not been used in sensor networks so far. Using the position estimates we get from multidimensional scaling, we then embed the nodes in hexagonal geographical clusters. To reduce the number of nodes placed in a wrong cluster we propose a simple heuristic to align the hexagonal grid based on anchor node positions.

The proposed approach is a heuristic and the quality depends on the number of anchor nodes and the extent of the range measurement errors. In Section 4 we define two quality metrics, the connectivity of an overlay graph one can construct over the geographical clusters and the consistency of the view of each node on the cluster structure. Both metrics are of interest when geographical cluster based localization is applied in combination with geographic routing.

With these metrics we performed a simulation study on the quality of geographical clustering under range measurement errors. In Section 5 we first derive a model to express range measurement errors which is used in our simulations. The results of the simulations are then summarized in Section 6. In Section 7 we finally conclude the findings and draw some open directions which could be followed in the sequel.

2 Related Work

2.1 Multidimensional Scaling

Multidimensional Scaling (MDS) [2, 15, 12] is a method used to depict the spatial structure, normally in 2 or 3 dimensions of distance-like data between points in space. MDS aims to find a geometric representation of the data, such that the distances between data points fit as close as possible to the given dissimilarity (or similarity) information.

In the domain of sensor networks MDS has been used to determine node positions. In this work we consider the following well known centralized MDS localization method[1]. Nodes which are neighbors in the wireless network determine their mutual distances and communicate these distances to one central node. This node collects these distance values in a so called dissimilarity matrix and then completes this matrix by the shortest paths between all possible node pairs. The matrix then contains an estimate of the distance between all possible node pairs. The dissimilarity matrix is used as MDS input to determine node coordinates. After that the central node communicates the coordinates back to the network nodes.

Depending on the MDS method applied two kind of node localization can be considered: range free and range based. In range based localization, inter nodal distance measurements are determined through different ranging techniques, viz. RSSI, time of arrival (TOA), or time difference of arrival (TDOA). The goal is to estimate the true Euclidean distance between the neighboring nodes. Whereas, in range free localization Euclidean distances are not considered. The only information available is whether two nodes are connected or not.

[1] Distributed methods of multidimensional scaling based localization can be found in the literature as well but are not subject of study of this work.

Range-free localization based on MDS can be found in [12]. Starting with the given network connectivity information, the shortest path in hop counts between each possible pair of nodes is determined. So called non-metric multidimensional scaling is then applied which allows a dissimilarity matrix just consisting of the hop counts. The resulting coordinates are shifted, rotated and/or reflected about an axis so as to get the anchor nodes (nodes with known location information) as much as in their right locations. The result is a placement of nodes which compares quite well with the nodes' true physical locations.

Range-based localization based on MDS can be found in [11]. The starting point is the Euclidean distance measurements between neighboring pairs of nodes through some range measurement technique. The missing distances are completed by the path lengths of the Euclidean shortest paths between all node pairs. The dissimilarity matrix is then fed into so called classical metric MDS which expects a dissimilarity matrix of Euclidean interpoint distances.

A third MDS variant which we consider in this work is probabilistic MDS. It hitherto has not been explored for wireless sensor localization. Unlike the other two MDS methods, probabilistic MDS [8] can take as input a variance value associated with each distance measurement along with the distance (dissimilarity) matrix. This makes probabilistic MDS more feasible to use in cases where the standard deviation of a ranging measurement technique is known beforehand. It was shown in [8] that probabilistic MDS performs better over classical MDS in node positioning when variances are associated with distance measurements.

2.2 Geographical Cluster Based Routing

Geographical cluster based routing was originally introduced in [5]. It is a combination of routing and topology control based on a virtual geographic overlay graph. To support correct routing, the overlay graph may not contain intersecting links. The original publication described a way to remove intersecting links which was then further refined in a follow up publication [3]. Moreover, the basic idea of geographical cluster based routing was one year later rediscovered in [14]. The publication describes a slightly different way to remove intersecting links. In [13], a follow up of that publication, the same algorithm was simulated with another routing metric. Besides these two works in that year one more algorithm was introduced in [9] which is an alternative to remove intersecting overlay links.

While these works focus on removal of intersecting links, one plain alternative approach is preventing intersecting links in the first place such that no intersections have to be removed later on. This is the variant considered in [4] which forms also the object of investigation of this work. The details of that approach are as follows: As assumed in all geographical cluster based routing variants, nodes have to know their physical location on a two dimensional plane. The plane is partitioned by a regular mesh of hexagons. Each hexagon defines a geographical cluster. Using its location information each node can assign itself to the cluster it is located in. We denote such assignment as $v_1 \in S$. Moreover, $C(v)$ denotes the cluster, a node v is assigned to.

Using the cluster assignment, the geometric overlay graph considered here is then constructed as follows. The centers of the clusters which contain at least one physical network node form the nodes of the overlay graph. The links between the clusters are defined by the physical links. Two clusters C and D are connected if they share a hexagon boundary and if there exists at least one pair of nodes $v \in C$ and $w \in D$ which are connected in the physical network graph. We term these links *short links*, since these overlay graph links are the ones with the smallest Euclidean length.

It is easy to see that the resulting overly graph is always planar. However, simplicity of the scheme comes at a cost. In sparse networks the overlay graph might be disconnected. However, if the network is dense enough – for instance in sensing covered networks with the double range property defined in [16] – then connectivity of the network can always be assured [4].

Routing along the planar overlay graph is done according to a combined greedy face routing strategy. Whenever possible a message is forwarded to a cluster which is closer to the destination cluster compared to the cluster the message is currently residing in. If no such cluster is found, right hand or left hand rule is applied to traverse a sequence of clusters until a cluster again closer to the destination cluster is found where greedy routing can be resumed. Under right hand or left hand rule a message is forwarded to the overlay graph link which in clockwise or counterclockwise is lying next to the last overlay link visited by the message.

3 Subject of Study

As it will be detailed in Section 5 we assume that nodes which are neighbors in the wireless network determine their mutual distances using RSSI based ranging. In the way already explained in Section 2 we then apply multidimensional scaling (MDS) on one central node to determine the node locations. We consider the three mentioned MDS variants: the non-metric, the metric (also termed classical), and the probabilistic one. The node locations we get out of these MDS variants are then used to determine the clusters the network nodes are located in and each node in the network is then informed by the central node about the cluster it is located in.

This basically explains our subject of study. However, two aspects need further explanation: the way the central node determines exactly the cluster a node is located in, and the way the central node computes the additional variance input required by the probabilistic MDS variant.

To determine the cluster a node is located in, we consider three anchor nodes which already know their positions. In our evaluation we select from the random node set the three nodes which are farthest from each other as anchor nodes. We then align the hexagonal grid with a heuristic to get all the anchor nodes as close to the cell centers. Starting with an initial alignment of the grid, for each anchor node, the hexagon it is located in and the six neighboring hexagons are determined. Then, for each anchor a and for each hexagon h of the seven hexagons determined for a, the grid is shifted such that a is the center of h. Then the grid is rotated around a to get the other anchor nodes closest to the center of one of the seven hexagons determined for them. Among all iterations the grid placement which minimizes the placement error is chosen, while the placement error is the sum of the squared errors of the distance of the anchor nodes from their cell centers.

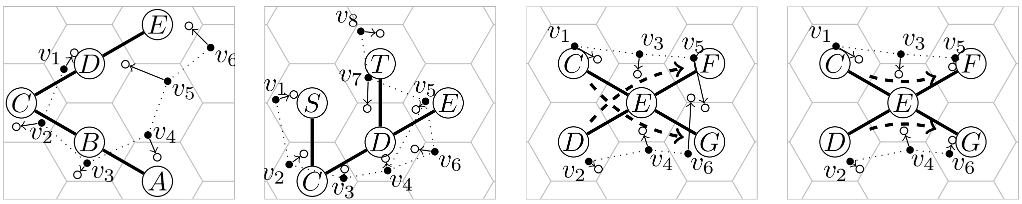
The variance input for probabilistic MDS is determined with the following heuristic. The actual Euclidean distances between anchor nodes is known a priori. From the disparity matrix fed into MDS we also know the estimated erroneous distances between the anchor nodes. For the three anchor node pairs, the statistical variance of the erroneous distances from the actual true Euclidean distances is then computed.

4 Evaluating Cluster Based Localization

If all network nodes are assigned to the right cluster they are physically located in, then the scheme described in [6] supports guaranteed message delivery from one cluster C to another cluster D; of course, provided that there exists at least one path in the overlay graph which connects C and D and provided that all nodes exchange adjacent clusters they see with all neighbor nodes which are in the same cluster. If we consider, however, that nodes may be wrongly assigned to clusters they are not physically located in, then any localized routing scheme following the links connecting neighboring clusters may fail although the underlying physical network is connected. We observe two phenomena which might occur. One is that the overlay graph might be disconnected and the other one is that the nodes in the same cluster might have an inconsistent view on the adjacent clusters. Both the phenomena and the evaluation metrics we use to account for them are detailed in the next two subsections.

4.1 Number of Partitions

Consider the example depicted in Figure 1a. The true node locations are depicted with solid dots. These are pointing to the determined node locations which deviate from true locations due to localization errors. In the example, the network nodes are connected by a network link – the dotted lines – if the Euclidean distance between the nodes' true locations are less or equal than the cluster diameters. This results in the depicted physical connected network over $v_1, \ldots, v_6$ and the depicted overlay graph over the clusters $A, \ldots, E$. Obviously, the overlay graph is connected. However, sending a message from A to E for instance, following the links will not be successful. A message can get from A to B via the physical link v_4v_3, from B to C via v_3v_2, from C to D via v_2v_1. However, there is no physical link which supports sending a message directly from cluster D to E.



(a) Network partitions. **(b)** Inconsistent views. **(c)** Critical case. **(d)** Benign case.

Figure 1 Potential failure sources in localized, cluster-based routing.

The example shows that just checking connectivity of the overlay graph defined by the cluster centers and the short links one can draw between the cluster centers is not sufficient to see if localized routing along this structure is always successful. To account for such phenomena in our empirical study, given an overlay graph we count the *minimal number of connected cluster partitions* we can construct over the set of clusters. Thereby, a given set of clusters $\mathcal{S}$ is termed *connected* if for any two clusters $C, D \in \mathcal{S}$ there exists a path $v_1v_2 \ldots v_k$ in the physical network such that $C(v_1) = C$, $C(v_k) = D$, and $C(v_i)$ and $C(v_{i+1})$ are either the same or have a hexagon edge in common.

In the network depicted in Figure 1a communication between any pair of clusters in $\{A, B, C, D\}$ is possible based on the path $v_1v_2v_3v_4$. However, no communication is possible between a cluster in $\{A, B, C, D\}$ and cluster E. Thus, in this example we count two connected partitions, one partition is $\{A, B, C, D\}$ and the other one is $\{E\}$.

4.2 Number of Inconsistent Clusters

Consider now the example depicted in Figure 1b. There is a direct connection from cluster S to C via the physical link v_1v_2, from cluster C to D via v_3v_4, from D to E via v_4v_5, from

E back to D via v_5v_7, and then from D to T via v_7v_8. Thus, the clusters S, C, D, E, and T form a single partition. However, the success of planar graph routing based on that overlay graph depends on the actual selection of the physical nodes located in those clusters.

Suppose starting at node v_1 a message is to be sent from cluster S to T. Since S has no adjacent cluster closer to T, face routing has to be employed. This leads to routing along the clusters C, D, and E using the physical network path $v_2v_3v_4v_5$, for instance. Once the message arrives at v_5 in cluster E, the next hop decided at v_5 may either be v_4, v_6, or v_7. If v_5 decides on v_4 or v_6 the message will traverse the clusters D and C back to S and will eventually be dropped at S since nodes v_4 and v_6 know about adjacent clusters C and E but not about T. If in contrast v_5 decides on v_7 the message will travel as well back from cluster E to cluster D, however, at cluster D the message arrives at a node which has a direct connection into cluster T. Thus, the message will eventually reach its destination cluster.

The example shows that in case of location errors it is not sufficient to exchange information about adjacent clusters within a cluster. In the example the node sets $\{v_4, v_5\}$ and $\{v_7\}$ are forming two partitions of the physical nodes in cluster D which can not communicate without resorting to relaying nodes outside of cluster D.

If the overlay graph is connected in the sense defined in the previous subsection and if none of the involved clusters are hosting partitions of the physical nodes then face routing along the overlay links will always find a path to the destination cluster. This follows immediately from the correctness proof in [7]. Thus, testing for clusters with physical node partitions is a reasonable metric for measuring if routing errors might occur with cluster based localization under location errors.

The routing failure depicted in Figure 1b originates from the fact that the inconsistent view of the nodes in cluster D leads to intersecting overlay graph links. In the example, the link DE exists twice. It exists due to the connection of v_5 with v_7 and due to the connection of v_5 with v_4, v_6. Besides intersection due to such edge doubling, intersections may also occur in the way depicted in Figure 1c. In this example the physical network path $v_1v_3v_5$ produces the overlay path CEG, while the path $v_2v_4v_6$ results in the overlay path DEF. The physical network nodes v_3 and v_4 located in cluster E are forming two partitions in that cluster. A message traveling along the edge CE according to the left or right hand rule will always be forwarded along the edge EG, while a message traveling along DE will always be forwarded along the edge EF. Thus, the overlay paths CEG and DEF are basically forming two intersecting links CG and DF. It is well known that face routing in case of such intersecting links may fail. We term this as an inconsistent view and more so, an inconsistent view with duplicate overlay graph links as exemplified in Figure 1b a *critical inconsistent view*.

Besides the critical ones we can also observe what we denote *benign inconsistent views* which will not lead to a routing failure. In the example in Figure 1d the physical nodes v_3 and v_4 in cluster E are forming two partitions in that cluster as well. The difference however with the case in Figure 1c is that the links $\{EC, EF\}$ seen by v_3 and the links $\{ED, EG\}$ seen by v_4 can be interpreted as face boundary links of two faces which are touching each other in one single point. A message traveling along the overlay edge DE according to the left or right hand rule will always be sent to the overlay edge EG next. It will thus never travel inside the upper area of the face having CE and EF on its boundary. The same applies to a message traveling along the overlay edge CE. It will never travel inside the lower area of the face having DE and EG on its boundary.

5 Model Assumptions

5.1 Modeling RSSI Measurements

We consider ranging based on RSSI[2] measurements. Neighboring nodes exchange a sequence of data packets and compute an average over the RSSI values they get for each data packet. We assume that the number of exchanged data packets is sufficient such that fading and interference will average out. In that case the measured average RSSI between two node pairs at different locations but same transmitter receiver separation may still vary due to shadowing and multipath propagation caused by surrounding environmental clutter. We use the well established log-normal shadowing [10] to model this effect. Expressed in dB the signal attenuation $PL(d)[\mathrm{dB}]$ encountered when sender and receiver are separated by a distance d is then given by $PL(d)[\mathrm{dB}] = \overline{PL}(d_0)[\mathrm{dB}] + 10n \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma$, where n is the path loss exponent, $\overline{PL}(d_0)[\mathrm{dB}]$ is the average path loss in free space for reference distance d_0, and X_σ is a zero-mean Gaussian distributed random variable with standard deviation σ.

Given a transmitter output power $P_t[\mathrm{dBm}]$ and signal attenuation due to log-normal shadowing over a transmitter receiver separation d, the average over the RSSI values at the receiving node can then be expressed in dBm as $RSSI(d)[\mathrm{dBm}] = P_t[\mathrm{dBm}] - PL(d)[\mathrm{dB}]$.

5.2 Modeling Ranging Errors

Once a node has determined the RSSI average for a given transmitter, the value has to be mapped to a distance value. We consider the technique of inverting the average large-scale path loss formula. Given path loss exponent n, reference distance d_0, path loss at reference distance $\overline{PL}(d_0)[\mathrm{dB}]$, and transmitter receiver separation d, the theoretical average large-scale path loss $\overline{PL}(d)[\mathrm{dB}]$ is $\overline{PL}(d)[\mathrm{dB}] = \overline{PL}(d_0)[\mathrm{dB}] + 10n \log_{10}\left(\frac{d}{d_0}\right)$. Thus, given transmitter output power $P_t[\mathrm{dBm}]$ the average RSSI value $\overline{RSSI}(d)[\mathrm{dBm}]$ not taking shadowing into account amounts $\overline{RSSI}(d)[\mathrm{dBm}] = P_t[\mathrm{dBm}] - \overline{PL}(d)[\mathrm{dB}]$. By inverting this function we can estimate the transmitter receiver separation $d(x)$ from measured RSSI averages x by $d(x) = d_0 \cdot 10^{\frac{P_t[\mathrm{dBm}] - \overline{PL}(d_0)[\mathrm{dB}] - x}{10n}}$. Substituting x with the expression $RSSI(d)[\mathrm{dB}]$ we use to model measured RSSI averages under shadowing then eliminates $P_t[\mathrm{dBm}]$ and $\overline{PL}(d_0)[\mathrm{dB}]$, yielding the erroneous distance estimate $\tilde{d}$ for a given transmitter receiver separation d as $\tilde{d} = d_0 \cdot 10^{\frac{10n \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma}{10n}} = d \cdot 10^{\frac{X_\sigma}{10n}}$.

6 Evaluation

We present a simulation study on the quality of the MDS based geographical clustering methods as described in Section 3. We consider how localization errors at node level affects correctness of geographical clustering. Moreover, using the metrics derived in Section 4, we express the feasibility of the resulting geographical clustering for geographic routing. The simulations were done with MATLAB using the built in function for non-metric and classical MDS. In the following we term those functions NMDS and CMDS, respectively. In addition, PROSCAL [8] was used to perform probabilistic MDS.

Every depicted measurement point is the average of 100 independent simulation runs. In each simulation run the network nodes are placed uniformly random distributed on a

[2] The received signal strength indicator (RSSI) expresses the power received at the receiver. Many wireless transceivers provide an RSSI value for each received packet and usually a mapping from RSSI values to radio frequency level in dBm can be found in the data sheets.

square field. Then for each node pair the path loss between two nodes is randomly selected according to the log normal model. Connectivity of the wireless network is then tested by checking if all nodes can communicate directly or via intermediate hops. Two nodes are defined as connected if the average RSSI for transmission between the two nodes is on or above the nominal receiver sensitivity. Non-connected networks are discarded.

6.1 Simulation Parameters

For choosing evaluation parameters, we consider TmoteSky sensor nodes which are equipped with the Chipcon CC2420 transceiver as a reference architecture and we consider the empirical average RSSI value estimates for CC2420 transceivers listed in [1]. The latter reports (among others) results of an indoor measurement with one CC240 transmitter and one CC2420 receiver. From standard formulas (see [10]) a reference distance $d_0 = 1$m and path loss at reference of $\overline{PL}(d_0)$[dB] ≈ 40[dB] can be determined. Applying the least squared error method[3] described in Example 4.9 of [10] on the empirical mean RSSI values listed in table 2 of [1] we determine a path loss exponent $n \approx 2.7$ and a standard deviation $\sigma \approx 3.3$ for the log normal shadowing model.

We conduct simulations with 50 nodes. The simulations are adjusted to path loss exponent 2.7 and vary the log normal standard deviation around the estimated value 3.3. The cluster diameter is adjusted to 100 which is close to the average maximum communication range of the nodes given by the CC2420 nominal receiver sensitivity of -94[dBm].

To adjust the node density we keep the number of nodes fixed and vary the size of the rectangular simulation field. Given k nodes, average maximum communication range d, and a square field length l, under uniform node distribution the expected neighbors δ in a node's average maximum communication range can (despite boundary effects) roughly be expressed as $\delta = (k-1) \cdot \frac{\pi d^2}{l^2}$. Thus, for a desired network density δ we estimate the square field length by solving for l.

6.2 Results

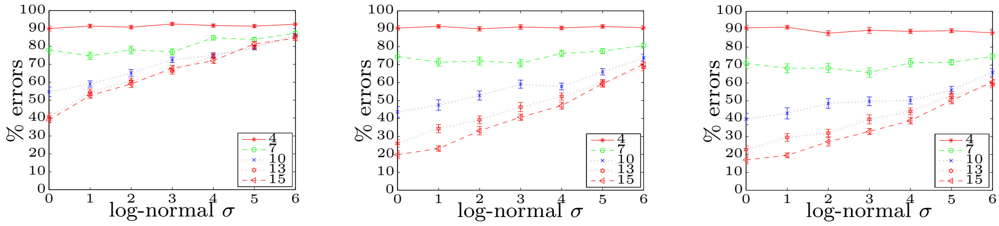
6.2.1 Assignments to Wrong Clusters

The first question we look at is how far the physical node position errors of the MDS variants propagate into the geographical clustering structure. We consider the Geographical clustering of nodes using the physical node positions determined by MDS and then embedding the node into the cluster structure as described in Section 3. We consider the clusters the nodes are placed according to that method and the actual clusters with respect to their true physical positions. For the latter we use the same alignment and orientation of the hexagonal grid computed by the method from Section 3.

Figures 2a, 2b, and 2c show depending on σ of the log normal path loss, the percentage of nodes which are placed into a cluster they are physically not located in. Of course, a higher σ means a higher variance in distance measurement errors which then negatively affects the MDS precision and consequently clustering as well.

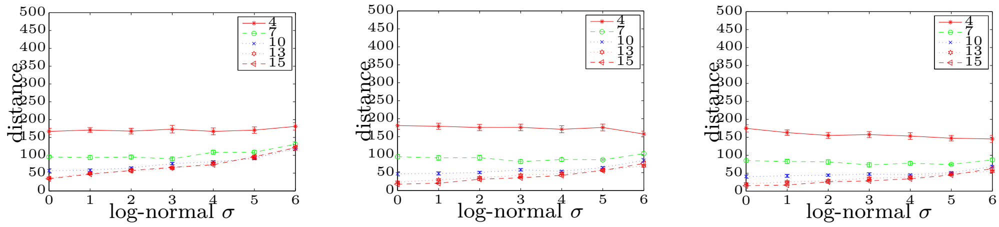
What can be seen as well is how improved quality of physical location information positively affects the assignment of nodes into the right cluster: clustering based on probabilistic

[3] The method finds the path loss exponent n which minimizes the sum over the squared distances between path loss due to log normal shadowing and the empirical values. Having found n, σ is determined by the statistical variance of the empirical and the mean values given by the log normal model.



(a) non-metric **(b)** metric **(c)** probabilistic

Figure 2 Percentage of node placement errors over log-normal σ for different node densities.



(a) non-metric **(b)** metric **(c)** probabilistic

Figure 3 Average distance from original cell over log-normal σ for different node densities.

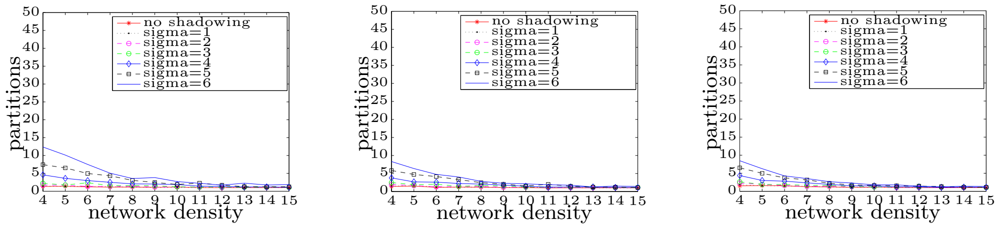
MDS performs slightly better than the one based on classical MDS, and classical MDS performs visibly better than the non-classical one.

Finally, for all methods, increasing the network density reduces the wrong assignment of nodes to clusters. This can be explained by the fact that completing the distance matrix by computing shortest paths among all node pairs has larger distance estimate errors when the network is sparse. Of course, the shortest path length may deviate from the Euclidean distance between the nodes. In the case where the network density tends to infinity, the shortest path becomes the true Euclidean distance between the nodes, and MDS would produce error free node locations, and no node placement errors would occur. In our measurement we went to a network density up to 15 which of course is not an infinite network density, so still results in node placement errors.

6.2.2 Closeness to the Correct Cluster

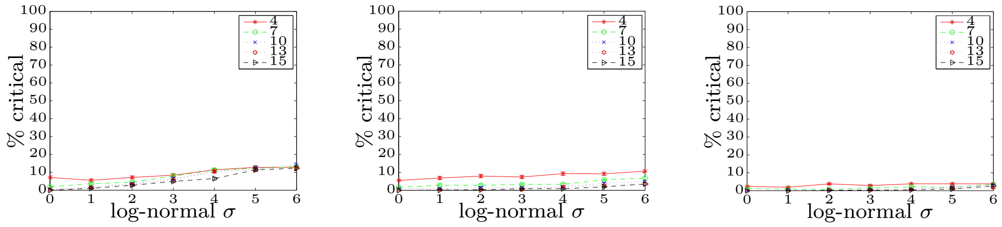
Figures 2a, 2b, and 2c show that particularly in sparse networks almost all nodes are located in a wrong cluster. So the question arises if the considered MDS based geographical clustering makes sense at all if used to reflect the actual geographical positions of nodes. To answer this question in Figures 3a, 3b, and 3c we consider the average Euclidean distance of the center of cluster the node is assigned by the considered MDS based geographical clustering methods and the center of the cluster the node is actually physically located in. Here again we use the alignment and rotation of the hexagon grid resulting from the MDS based geographical clustering methods to decide the cluster a node is physically located in.

We observe that though many assignments to wrong clusters may exists, on an average the deviation from the actual clusters are just local fluctuations. Of course, we consider a cluster diameter of 100. Thus, the Euclidean distance between the centers of two neighboring hexagons is given by $\frac{\sqrt{3}}{2} \approx 87$. The numbers depicted on the y-axis of the Figures 3a, 3b, and 3c show that on an average error distances range within a few hops from the true cluster



(a) non-metric **(b)** metric **(c)** probabilistic

Figure 4 Network partitions over network density for different log-normal σ values.



(a) non-metric **(b)** metric **(c)** probabilistic

Figure 5 Percentage of critical inconsistent clusters over log-normal σ for different node densities.

in low density networks. In high density networks the value is clearly below 87 which means that the combined MDS and clustering methods often hit the right cluster.

The intuition behind the observation in Figures 3a, 3b, and 3c is that the applied MDS strategies do not find a perfect node embedding which reflects the true physical node locations but of course, for all variants it is highly unlikely that a node suddenly happens to be located in another area of the network than its neighboring nodes.

6.2.3 Support of Localized Routing

When considering the Figures 2a, 2b, and 2c, also the question arises if MDS based geographical clustering makes sense at all to be used in combination with geographical cluster based routing. To asses the quality of the overlay graph we apply the metrics which we defined in Section 4. Figures 4a to 5c show though there are local fluctuations in the right cluster assignment, the resulting structure is in fact surprisingly a robust one for geographical cluster based routing.

The average number of network partitions depicted in Figures 4a, 4b, and 4c is about 1 for all MDS variants when σ of the log normal shadowing is low. Only in low network densities performance visibly degrades for all schemes when σ increases. At this, the frequency of partitions is the highest for non-metric MDS, is better for classical MDS and, is the best for the probabilistic one. This is again in accordance with the fact that probabilistic MDS results in node positions closest to the physical node positions, which in terms of precision is then followed by classical MDS and then by non-metric MDS which considers only hop count but not the Euclidean distance.

The same trend can be observed in Figures 5a, 5b, and 5c for the percentage of clusters where nodes have an inconsistent critical view. Within the considered parameter range the fraction of clusters with such inconsistencies is around 10% for non-metric and classical MDS in the worst case. Probabilistic MDS achieves even better results in those worst case parameter settings with a fraction of clusters with inconsistencies never exceeding 5%.

7 Conclusion

Geographical cluster based routing is an interesting research track to make the conceptual idea of localized routing robust to the stochastic vagaries of real wireless networks. In this work we were interested in the effects of location errors which occur when information about node positions is determined by the wireless network infrastructure itself. In conclusion: The percentage of nodes assigned to wrong clusters can be high under large localization errors. However, in our study we observed that even for very large localization errors, the number of network partitions and inconsistent clusters is low. Thus, though message delivery is not guaranteed, geographical cluster based routing is feasible under inaccurate node localization. Here, among the three variants observed, geographical clustering with probabilistic MDS turned out to be the best variant.

In this work we investigated a centralized localization scheme. Though it seems curious at the first look to apply a centralized scheme first which already learns all nodes and then resort to a localized routing method later, this actually makes sense. The wireless links may change over time while the node positions are not (in a static wireless sensor network). Once nodes are assigned to geographic clusters, localized routing further requires only local maintenance to adapt to network dynamics such as changing links, and node failures.

We derived a model for location errors based on a well established path loss model and we derived the model parameters based on previous work which cited testbed results. This gives us high confidence that our conclusions actually hold for real sensor network deployments too. However, to get to the final conclusion that geographical cluster based routing in conjunction with certain localization scheme makes sense, it will be of high interest to complement our results with testbed experiments.

In this work we defined graph metrics to asses the quality of the geographical cluster based overlay graph with respect to localized routing. We selected this approach to have results independent of a particular routing algorithm. In fact, the results of the network partitions and critical inconsistent views depicted in our plots are worst case statements. In case of more than one overlay graph partition, two randomly chosen communication partners may happen to be in the same partition and routing is still successful. And, though there are clusters with critical views in the overlay graph, for two randomly chosen communication end points the path followed by a concrete face traversal algorithm may happen to visit only clusters with no or with benign inconsistent views. Even when a cluster with a critical inconsistent view is visited this does not necessarily mean that routing fails; there is only a chance that it may fail. An interesting future research direction is to complement our findings with measurements where actual localized geographic routing variants are applied over the constructed overlay graphs. Compared to the already promising low fraction of network partitions and critical inconsistent views observed in this work, we expect even better outcomes with respect to the success rate of those routing protocols.

Finally, it is important to note that if we allow several non connected node partitions in one cluster (as it may happen under localization errors as described before) the geographical cluster based routing approach as described, does not assure that a node in a certain cluster can be reached. In the case that there are more than one node partitions in one cluster, a message may eventually reach its destination cluster but happen to arrive in any of the partitions. If we are just interested in reaching a particular cluster, this is not an issue. If we are interested in reaching a particular node and when the node is not in the partition where the message arrived, a recovery has to take place to eventually reach the right partition in that cluster. The latter problem opens an interesting track of future research.

References

1 K. Benkic, M. Malajner, P. Planinsic, and Z. Cucej. Using RSSI value for distance estimation in wireless sensor networks based on ZigBee. In *Proceedings of the 15th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 303–306, 2008.

2 Ingwer Borg and Patrick J.F. Groenen. *Modern Multidimensional Scaling – Theory and Applications*. Springer Series in Statistics, 2005.

3 Hannes Frey. Geographical cluster based routing with guaranteed delivery. In *2nd IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS 2005)*, Washington, DC, USA, November 7–10 2005.

4 Hannes Frey and Daniel Görgen. Geographical cluster based routing in sensing-covered networks. In *Proceedings of the 2nd International Workshop on Wireless Ad Hoc Networking (WWAN 2005)*, pages 885–891, Columbus, Ohio, USA, June 6–9 2005. IEEE Computer Society.

5 Hannes Frey and Daniel Görgen. Planar graph routing on geographical clusters. *Ad Hoc Networks, Special issue on Data Communication and Topology Control in Ad Hoc Networks*, 3(5):560–574, September 2005.

6 Hannes Frey and Daniel Görgen. Geographical cluster based routing in sensing-covered networks. *IEEE Transactions on Parallel and Distributed Systems: Special issue on Localized Communication and Topology Protocols for ad hoc Networks*, 17(4), September 2006.

7 Hannes Frey and Ivan Stojmenović. On delivery guarantees of face and combined greedy face routing in ad hoc and sensor networks. In *Proceedings of the ACM Annual International Conference on Mobile Computing and Networking (Mobicom)*, Los Angeles, USA, September 2006.

8 David B. MacKay and Joseph L. Zinnes. A probabilistic model for the multidimensional scaling of proximity and preference data. *Marketing Science, Special Issue on Consumer Choice*, 5(4), 1986.

9 Sumesh J. Philip, Joy Ghosh, Hung Q. Ngo, and Chunming Qiao. Routing on overlay graphs in mobile ad hoc networks. In *Proceedings of the IEEE Global Communications Conference, Exhibition & Industry Forum (GLOBECOM'06)*, 2006.

10 Theodore S. Rappaport. *Wireless Communications: Principles & Practice*. Prentice Hall, 2002.

11 Yi Shang and Wheeler Ruml. Improved MDS-based localization. In *Proceedings of the 23th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'04)*, number 4, pages 2640–2651, 2004.

12 Yi Shang, Wheeler Ruml, Ying Zhang, and Markus P. J. Fromherz. Localization from connectivity in sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 15(11):961–974, nov 2004.

13 Hector Tejeda, Edgar Chavez, Juan A. Sanchez, and Pedro M. Ruiz. Energy-efficient face routing on the virtual spanner. In *Proceedings of the 5th International Conference on AD-HOC Networks & Wireless (ADHOC-NOW'06)*, pages 101–113, 2006.

14 Hector Tejeda, Edgar Chavez, Juan A. Sanchez, and Pedro M. Ruiz. A virtual spanner for efficient face routing in multihop wireless networks. In *Proceedings of the IFIP TC6 11th International Conference on Personal Wireless Communications*, pages 459–470, 2006.

15 W.S. Torgerson. Multidimensional scaling: I. theory and method. *Psychometrica*, 17:401–419, 1952.

16 Guoliang Xing, Chenyang Lu, Robert Pless, and Qingfeng Huang. On greedy geographic routing algorithms in sensing-covered networks. In *Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing (MobiHoc '04)*, pages 31–42, Roppongi Hills, Tokyo, Japan, May 24–26 2004.