On the Importance of Information Quality for Wireless Sensor Networks

Toks Adebutu and Dr. Lionel Sacks

University College London

Abstract: The primary objective of wireless sensor networks is to gather as much information as possible about whichever environment they are deployed in. Considerable efforts have been made by researchers to limit redundancy whilst maintaining a suitable level of detail in the information that is obtained. What is lacking however is an investigation into the impact of poor quality information on overall network performance. This work proposes that in order for a wireless sensor network to function properly, there is an optimum range of disparate sources of information each node should have and a variety of good quality information sources are necessary for satisfactory network performance.

1. Introduction.

At the onset of the sensor network research effort, the Defense Advanced Research Projects Agency (DARPA) defines a sensor network as "… a deployment of massive numbers of small, inexpensive, self-powered devices that can sense, compute, and communicate with other devices for the purpose of gathering local information to make global decisions about a physical environment[1]."

Given the above definition, a wireless sensor network is only as good as the information it produces. In order to obtain a high quality network, efforts need to be made to ensure the quality of the local information that is gathered. Considerable efforts have been made in the areas of spatial and temporal data aggregation, to optimise the available power and network bandwidth, whilst obtaining most if not all relevant environmental information. Information assurance[1] has also been significantly developed, but little effort has been made to ensure the quality of information that is being aggregated or assured.

2. Discussion

Many wireless sensor networking applications utilise nearest neighbour information to may some calculation or inference. Nodes within the network have variable levels of connectivity and this connectivity has an impact on the quality of inferences that any given node makes. Analysing the example of the iterative averaging algorithm [2] **-** a position estimating algorithm that utilises the average position estimate of each node's nearest neighbours to compute the position estimate for that node. The quality of this estimate is affected by two main factors.

The number of neighbours – disparity in the sources of information has a direct impact on the value of inferred data to the network. At first glance, this can be translated to mean the greater the number of neighbours a given node has, the better the quality of its inference. In practice, there is an optimum range of neighbours that results in the computation of the 'best' inference. This behaviour can be inferred from Figure 1 and 2. Figure 1 shows the percentage of position determining nodes located within 15m of their absolute position. At low ranges and beacon densities, the majority of nodes are in small isolated clusters. As the range increases, the connectivity between cluster increases, until above some threshold, the number of 'unique' clusters within the network decreases, to a point where most nodes in the network can see a large portion of the other nodes in the network. As the number of unique clusters reduces, resulting in large numbers of nodes having the same information which is invariably of bad quality. Figure shows the low, medium and high connectivity situations.

---

[1] "Information operations that protect and defend information and information systems ensuring their availability, integrity, authentication, confidentiality, and non repudiation. This includes providing for restoration of information systems by incorporating protection, detection, and reaction capabilities"

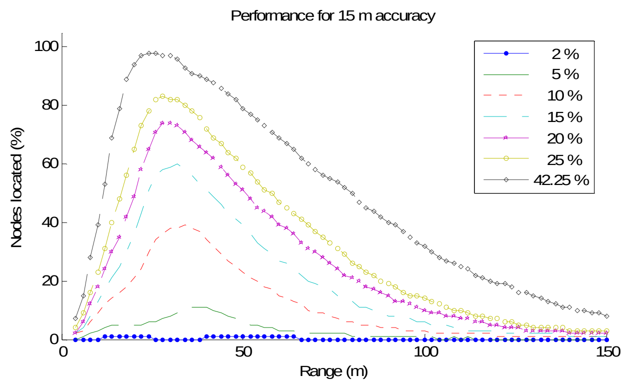

Figure 1 Algorithm performance, based on the percentage of position determining nodes with error in positioning error less than 15m for a network of 400 nodes with 4, 6, 8, 9, 10, 12 to 100 in increments of 4, then 121, 144, 156, 169 beacons in a 400x400 grid, with range in metres on the x-axis and percentage of nodes located on the y-axis.

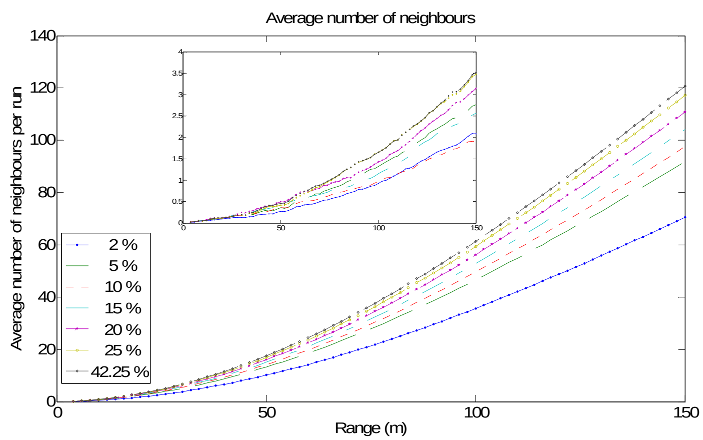

Figure 2 The average number of neighbours of all nodes for separate runs of the above simulation, with the inset showing the standard deviation in the number of neighbours between runs.

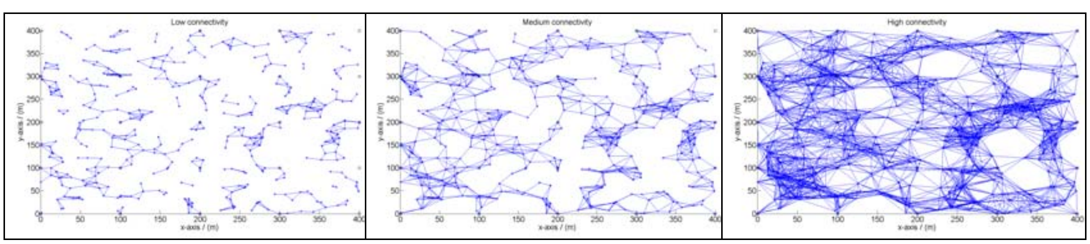

Figure 3 Network connectivity as node communication range is increased. From left to right, we start with a highly fragmented network. Moving towards a network in which each node is aware of every other node.

Proximity to good sources of information – as shown in [2], the availability of good sources of information in the proximity of a position determining node translates to that node having the ability to make good inferences as to its situation within the network. In the simple 2 dimensional localisation case, a node needs to have at least three good sources of information in close proximity, thereby providing it with a range of information in both dimensions. The requirement is for at least three, because two sources will effectively place it on a line between both quality information sources, which would lead to large errors in at least one3 if not both axes, depending on the absolute distance to these information sources. This proximity requirement should not be viewed simply in terms of a single hop. Quality information sources within 2 to 3 hops also serve.

Looking at Figure 4, we see that there is a relationship between the overall concentration of position aware nodes and the percentage of nodes that are able to estimate their position with satisfactory accuracy. As the concentration of quality information sources increases, we see that the quality of position estimates approaches saturation.

Figure 5 shows 2 trends, one at low (<5%) and the other at higher beacon densities. At low beacon densities, the best performance is obtained at low frequencies. This is because at these low ranges (<10m) in a low beacon concentration network, the dominant contributing factor to node localisation is proximity to position aware nodes, as there are still a large number of isolated node clusters, resulting in the poor performance, although best performance is attained at low ranges. As the beacon density increases, this effect is reduced and the availability of disparate sources of information becomes the dominant factor.

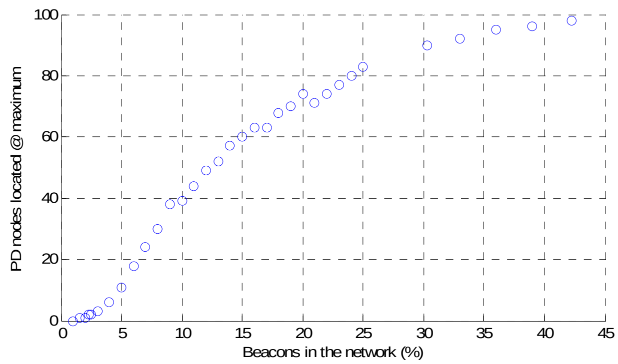

Figure 4 The maximum percentage of nodes located for different beacon concentrations.

3. Applications.

This results presented above, have implications in a number of areas, both internal and external to wireless sensor networking. Simplistically, the ideas presented here can be used to measure and maintain optimum network connectivity, making it also useful in the area of power management, as highly connected nodes are likely to have their transmit power reduced so they can only see a more manageable set of nodes. This will significantly increase the node lifetime as communication is the single largest power drain on wireless sensor nodes.

This adaptive power architecture is also highly desirable in mesh networks, where having a large number of nodes within range will resulted in a limited transmission capabilities as collision happen more frequently, resulting in inefficient usage of the available bandwidth.

A number of algorithms will also benefit from this scheme. The example of localization algorithms has been dealt with in detail above. Additionally, clustering and data aggregation algorithms will benefit, because anomalous data can result in the formation of unrepresentative node clusters and result in incorrect environmental representation as well as reducing the efficiency of data aggregation and compression algorithms.

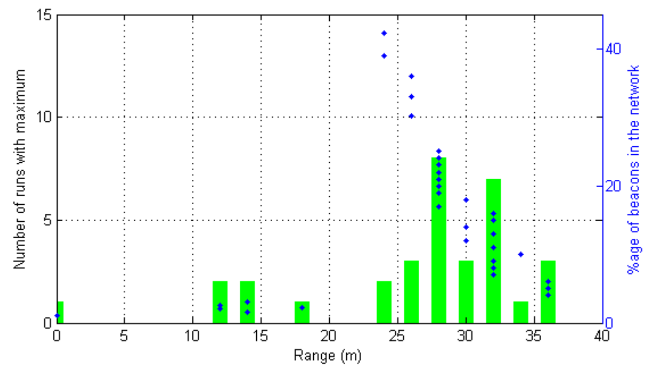

Figure 5 Histogram of frequency range at which the maximum number of nodes are located on the same axis as the frequency at which the maximum number of nodes are located for various beacon concentrations.

4. Future Work.

Each point in the result shown above is for individual runs of the simulation. Work is currently in progress to adaptively change the communication range, so that each node's connectivity can be tweaked, so that it satisfies both the disparate information source criteria and the connectivity to good quality source criteria. In addition, a quality metric is being introduced, so that any given node has this metric for itself and its neighbours. This can help nodes discriminate between poor and good quality information nodes. As we see from Figure 2, the higher the number of good quality nodes the better the overall network performance, so nodes with poor quality information will favour good quality information nodes, thereby improving their own quality. This develops a self-reinforcing cycle, whereby the overall quality of information within the network as a whole is improved, resulting in better performing wireless sensor networks, up to a saturation point.

5. Conclusions.

The quality of information that individual nodes within sensor networks use to make computations is of paramount importance. The overall effectiveness of the network is dependent on the quality of information that is produced and the use of a few erroneous reading can disrupt this. Consequently, wireless sensor network nodes need to have a validity metric in order to alert each other to possible errors.

References.

[1] Stephan Olariu, Qingwen Xu. Information Assurance in Wireless Sensor Networks. In *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium* (Denver, Colorado, 4th-8th April, 2005). vol. 13, no. 13, pp. 236a.

[2] Toks Adebutu and Lionel Sacks. Resource Constrained Position Estimation for Wireless Sensor Networks. In *Proceedings of the 2004 London Communications Symposium* (London, UK, September 13th-14th, 2004). LCS02004. Communications Engineering Doctorate Centre, London, United Kingdom, pp. 101 –104.

[3] Deb, B., Bhatnagar, S., and Nath, B. 2003. Information assurance in sensor networks. In *Proceedings of the 2nd ACM international Conference on Wireless Sensor Networks and Applications* (San Diego, CA, USA, September 19th, 2003). WSNA '03. ACM Press, New York, NY, pp. 160-168.