Data-Gathering Wireless Sensor Networks: Organization and Capacity

Introduction

We define the many-to-one throughput capacity as the *per source data throughput*, when all or many of the sources are transmitting to a single fixed receiver or sink.

The throughput capacity of a wireless network was first studied by Gupta and Kumar, key results of which include that the achievable per node throughput is $\theta(W/\sqrt{n \log n})$, where W is the transmission capacity and n is the total number of nodes in the network. A *sink* / *destination* is located at the center of the network / circle. Each node is not only a source of data, but also a relay for some other sources to reach the sink

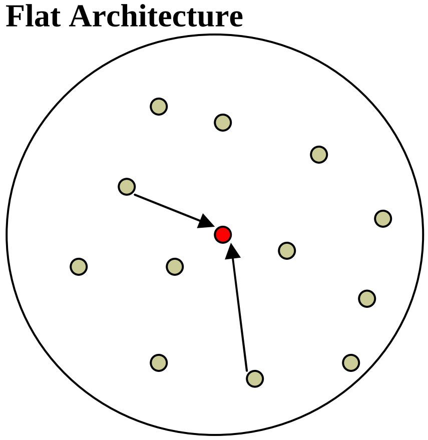
Network Model

Since many-to-one communication causes the sink to become a point of traffic concentration, the throughput achievable per source node in this case is reduced. Here we consider the case where every source gets an equal (on average) amount of original data (not including relayed data) across to the sink.

Random network a network where the nodes are randomly placed following a uniform distribution and we have no direct control over the exact location of the nodes.
Arbitrary network a network where we can determine the exact locations of the nodes Note that an arbitrary network is thus a particular instance of the random network with a very low probability of occurring.

The approach is to derive the throughput limit that is achievable with high probability as the number of sensors goes to infinity under the random deployment.

Flat Architecture



Nodes communicate with the sink via possibly multi-hop routes by using peer nodes as relays. With fixed transmission range, nodes closer to the sink will serve as relay for a larger number of sources. We assume that all sources use the same frequency to transmit data, thus sharing time. However, the results apply as long as there is a single shared resource, e.g., time, frequency, and so on.

The sources share the resource (time) by transmitting following a schedule that consists of time slots. The same analysis and same results could be obtained if we considered different resources, such as frequency or codes.

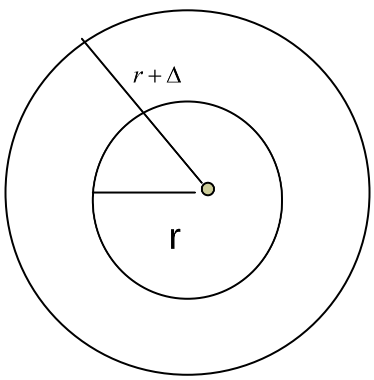
We will simply assume that a node has enough of its own packets buffered so no time slot is wasted

The field has an area of 1 which can then be scaled with the size of the area. Nodes share a common wireless channel using omni-directional antennas. Nodes use a fixed transmission power and achieve a fixed transmission range.

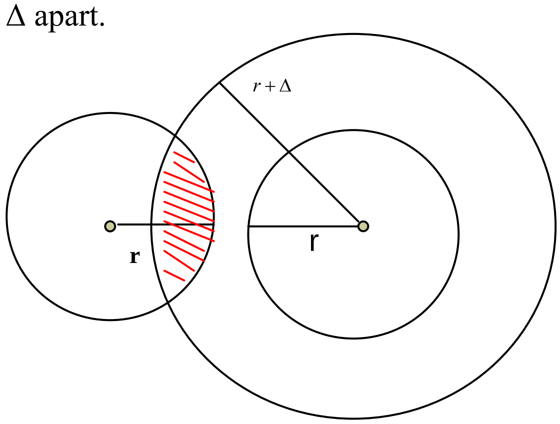
Let X_k and X_l be two sources with distance d_{kl} between them. The transmission from X_k to X_l will be successful if and only if

$$d_{kl} \leq r \qquad d_{km} > r + \Delta \qquad \Delta \geq 0$$

for any source X_m that is simultaneously transmitting.



r is the transmission range.
A node will interfere with another node that is receiving if they are within distance $r + \Delta$ of each other.
A node may interfere with another node that is transmitting if it is within distance $2r + \Delta$ of that node. If the intended receiver is located within the overlapping area the transmission will fail because of interference, hence the two nodes need to be at least $2r+\Delta$ apart.



This implies that no nodes can receive more than one transmission at a time. We assume that no node can transmit and receive at the same time.

The capacity will be derived as a function of the transmission range, assuming the transmission range can provide connectivity. This is because while in the one-to-one case it has been proven that reducing the range of transmission to increase spatial reuse increases the capacity of the network

Upper Bound

The maximum per node throughput in a wireless network featuring many-to-one communication outlined by the network model is upper bounded by W/n

$$n\lambda \leq W \quad \longrightarrow \quad \lambda \leq \frac{W}{n}.$$

This means that each source can only use up to 1/n of the resources under this model.
This result has following consequences

$\lambda = W/n$ can be achieved when every source can directly reach the sink.

$\lambda = W/n$ is not achievable if not every source can directly reach the destination and $\Delta > r$.

$\lambda = W/n$ may be achieved in an arbitrary network when not every source can directly reach the destination and $\Delta < r$.

here the bound is achieved by carefully positioning nodes in the network. For a randomly deployed network this bound cannot be achieved with high probability.

If a network has randomly deployed sources and the transmission range r is such that not all sources can directly reach the sink, then with high probability the throughput upper bound $\lambda = W/n$ is not achievable.

The number of simultaneous transmissions, denoted by *t*, that can be accommodated is

$$t < \frac{1}{A_{r+\Delta}} \leq \frac{1}{\pi r^2}.$$

A randomly deployed network using multi-hop transmission for many-to-one communication can achieve throughput

$$\lambda \geq \frac{\hat{W}}{n} \frac{\pi r^2 - \sqrt{\epsilon}}{4\pi r^2 + 4\pi r\Delta + \pi\Delta^2 + \sqrt{\epsilon}}$$

with high probability, when no knowledge of the traffic load is assumed ε positive but arbitrarily small.

A randomly deployed network using multi-hop transmission for many-to-one communication can achieve

$$\lambda \geq \frac{W}{\sum_{h=1}^{h=\lceil 2+\frac{\Delta}{r}\rceil} l_h^+ n_h^+}$$

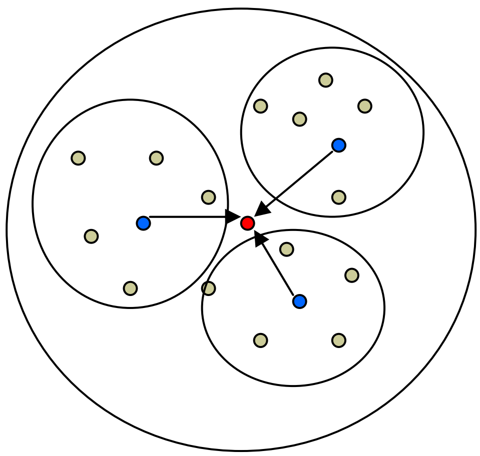
with high probability, when knowledge of the traffic load is assumed. l^+_h *and* n^+_h *are the upper bounds on the number of virtual sources per actual source and the number of actual sources respectively, that are h hops away from the sink with high probability.*

A randomly deployed network using multi-hop transmission for many-to-one communication can achieve a throughput arbitrarily close to

$$\frac{W}{n(2-\pi r^2)}$$

when knowledge of the traffic load is assumed and $\Delta = 0$.

Hierarchical networks



The second architecture is hierarchical where clusters are formed so that sources within a cluster send their data (via a single hop or multi-hop depending on the size of the cluster) to a designated node known as the *cluster head*. We assume that the cluster heads serve as simple relays and no data aggregation is performed. The communication between nodes and cluster heads and communication between cluster heads and the sink are on separate frequency channels so that the two layers do not interfere. The cluster heads are extra nodes introduced.

W refers to the transmission capacity of the channel in a flat network. In a hierarchical network *W* will refer to the transmission capacity of the channel used within clusters. *W'* will refer to the transmission capacity of the channel used from the heads to the sink. *H* denotes the number of clusters (heads) introduced. Each cluster head will create a cluster containing the sources closest to it. Within each cluster the communication is either via a single hop or via multi-hop, while the communication from cluster heads to the sink is assumed to be done via a single hop on a different channel. Thus cluster heads are assumed to have much higher transmission power than source nodes. We assume that cluster heads cannot transmit and receive simultaneously.

We assume there is at least a distance of $2(2r + \Delta)$ between any two cluster heads. We also assume that each cluster covers an area of same size.

We refer to the throughput achieved within a cluster (as opposed to that obtained in the entire network) as *λ'*. When cluster heads have the same transmission capacity *W* as the sources then *W/ n* remains to be the upper bound.

In order to achieve *W/n* our assumptions on the formation of clusters implies

$$H \leq \frac{1}{\pi\,(2r+\Delta)^2}.$$

which means that the range of transmission *r* must satisfy

$$\frac{20r^4+36\Delta r^3+25\Delta^2 r^2+8\Delta^3 r+\Delta^4}{r^2-\sqrt{\epsilon}\left(4r^2+4r\Delta+\Delta^2-\frac{1}{\pi}\right)} \leq \frac{1}{\pi}.$$

We note that as the density of the network increases, the *r* needed for connectivity decreases.

If single-hop communications are also used within each cluster then, *λ' =W/nH*. In this case we would need

$$\frac{W}{n/H}n(1-\tfrac{1}{H}) \geq W$$

which means *H* >= 2.

In a network using clustering, where cluster heads have the same transmission capacity W as the sources, there exists an appropriate number of clusters H and an appropriate range of transmission r that would allow the network to achieve λ' = W/n with high probability as n goes to infinity.

If the transmission capacity of the cluster heads is *W'*, assuming *W0 > W* then

In a network using clustering, where cluster heads have transmission capacity W' , there exists an appropriate number of clusters H and an appropriate range of transmission r, as n goes to infinity, that allows the network to achieve λ' = W0 /n with high probability. W'/ n is also the upper bound on throughput in this scenario.

The results showed higher throughput can be achieved by using clustering. The cost is the extra nodes functioning as cluster heads. They require a bigger transmission range/rate and a greater energy reserve and a second channel so that their transmissions do not interfere with the transmissions within the cluster. The idea is that while previously the only way to achieve *λ' = W n* was with direct transmission, where *all n* nodes need to be able to reach the sink in a single hop, that result can be achieved with only a handful of "enhanced" nodes. Moreover the number of these enhanced nodes does not depend on *n*.

Energy Consumption

If to a network with *n* nodes we add nodes acting as relays, the minimum amount of energy consumed in the network does not change for a given transmission range. However, by introducing extra nodes, a smaller range of transmission is sufficient to ensure connectivity.

The results show that a flat network consumes less energy if the area of the network is large, and a hierarchical network consumes less energy if the area is small

Based on the results from [1] and the results from the previous section, small networks would benefit from the use of clusters, which reduces the energy consumption and increases capacity. However, in large networks a trade-off exists. If the capacity of the flat network is enough for the application, then one should design the network to use multi-hop transmission in order to save energy. If higher capacity is needed then a hierarchical architecture should be used at the expense of energy consumption.

At smaller network size *as r* increases, the energy consumption decreases. At a bigger scale, as *r* increases the energy consumption increases because the dominant part of the energy consumption becomes related to the square of the distance, meaning that we are better off with many small hops than a few large ones. This observation is important because it is generally accepted that smaller hops are better than large ones when it comes to energy consumption, so the answer depends on the scale of the network.

Practical Implications

It has been shown that in the case of many-to-one communication higher per sensor throughput is achieved by having a traffic load-aware MAC. We need a MAC scheme that will allocate resources proportional to the amount of communication each sensor has to perform.

We have assumed that all clusters are of roughly the same size. However, if the deployment of cluster heads is random, then there is no obvious way to ensure that the actual outcome of the deployment will satisfy our assumption. One possible solution is to add redundancy and deploy more cluster heads than needed, but only use a subset of them based on some selection algorithm. This allows us to create a more even distribution of selected cluster heads.

1] E. J. Duarte-Melo, M. Liu, Energy efficiency of many-to-one communications in wireless networks, in: IEEE Midwest Symposium on Circuits and Systems (MWSCAS), Tulsa, Oklahoma, 2002.