Data-Gathering Wireless Sensor Networks: Organization and Capacity

Enrique J. Duarte-Melo, Mingyan Liu
Electrical Engineering and Computer Science Department
University of Michigan, Ann Arbor, MI 48109-2122
{ejd,mingyan}@eecs.umich.edu.

Abstract

In this paper we study the transport capacity of a data-gathering wireless sensor network under different communication organizations. In particular, we consider using a flat as well as a hierarchical/clustering architecture to realize many-to-one communication. The capacity of the network under this many-to-one data gathering scenario is reduced compared to random one-to-one communication due to the unavoidable creation of a point of traffic concentration at the data collector/receiver. We introduce the overall bound of $\lambda = \frac{W}{n}$ and show under what conditions it can be achieved and under what conditions it cannot. When those conditions are not met, we constructively show how $\lambda \approx \frac{W}{4n}$ or $\lambda \approx \frac{W}{2n}$ is achieved with high probability as the number of sensors goes to infinity. We also show how the introduction of clustering can improve the throughput to the point where $\lambda = \frac{W}{n}$ is achievable. We also discuss the trade-offs between achieving capacity and energy consumption, how transport capacity might be affected by considering in-network processing and the implications this study has on the design of practical protocols for large-scale data-gathering wireless sensor networks.

Keywords

Wireless sensor networks, capacity, energy consumption, many-to-one communications, clustering, in-network processing

I. Introduction

This paper studies the transport capacity of many-to-one communication in a data-gathering wireless sensor network. The rapid advances in micro-electromagnetic systems (MEMS) and wireless technologies have enabled the integration of sensing, actuation, processing and wireless communication capabilities into tiny sensor devices. These sensors can then be deployed in large amounts to self-organize into networks that serve a wide range of purposes. The main objective of this study is to understand the fundamental scalability of the large-scale wireless sensor networks used for monitoring and data gathering. Such understanding is essential in the deployment of such networks and the development of efficient protocols for these networks.

The reason for considering many-to-one type of communication among other possible forms of communication is because many-to-one and many-to-few are communication modes that commonly take place in a data-gathering wireless sensor network. Consequently they are the system-level abstractions that capture and represent the nature of communication in a wide range of sensor network applications, where the ultimate destination of data is a single control center (subsequently referred to as the *sink* or *receiver*) or a few connected control centers that have the resource to process the data and the authority to issue/take actions. Such applications include various data-gathering, monitoring and surveillance sensor networks, such as field imaging or monitoring where snapshots of the field are reconstructed from the sensing data over a period of time. At the same time, many statistical signal processing algorithms and studies that utilize sensor networks for detection and tracking are based on a hierarchical structure that uses *clusters* [1], [2]. Communications within each cluster are again of the many-to-one type, i.e., data flows from each sensor to the cluster head where they can be processed, compressed, aggregated and relayed. More broadly, clustering is arguably one of the most frequently proposed and used method to organize communications in a large scale network [3], [4]. Thus we can easily envision many future large sensor networks to have many-to-one communication overall at the network level, as well as within clusters in local areas. Hence fully understanding the scalability properties and implications of many-to-one communication is of great importance in the design and configuration of networks employing such communications.

Within the context of many-to-one communication, possible organizations of the network include the flat and hierarchical organizations. In a flat organization all nodes/sensors act as peers in transmitting and relaying data for one

This study is funded by NSF ITR grant number ANI-0112801.

another. In a hierarchical network, layers of clusters are formed. Nodes send their data to the cluster heads who then relay the data to either the higher layer cluster head or the sink (such as in [3]). In this paper we will examine and compare both organizations in terms of transport capacity and energy consumption.

We define the transport capacity as the *maximum achievable per source data throughput*, when all or many of the sources are transmitting to a single fixed receiver or sink. This is an important measure in determining the scalability of a large-scale wireless sensor network. The throughput capacity of a wireless network was first studied by Gupta and Kumar in [5], key results of which include that the per node throughput achievable is on the order of $\frac{W}{\sqrt{n \log n}}$, where W is the transmission capacity and n is the total number of nodes in the network. Study in [6] further pointed out that the actual throughput achievable is significantly affected by traffic patterns and the MAC used. [7] showed that the capacity obtained by [5] can be improved if mobility is used to reduce the number of hops needed to reach the destination. Another related paper is [8], which showed that when delay and complexity are ignored and infinitely long and complex codes are used the capacity obtained by [5] can be increased via relay sensors.

The result obtained by [5] is based on the assumption that communications are one-to-one, and that sources and destinations are randomly chosen. It does not apply to scenarios where there are communication hot spots in the network. 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 affected. In this paper we are only interested in the case where every source gets an equal (on average) amount across to the sink. This is because otherwise throughput can be maximized by having only the sensors closest to the destination transmit. Equal share of throughput from every sensor is desired for applications like imaging where each sensor represents a certain region of the whole field and data from each part are equally important. However, when there is significant amount of in-network processing (e.g., data compression) then this is no long true since the amount of processed data may vary from source to source. We will discuss this more in section V-B.

Creating a tessellation of the network has been used in [5] to create a lower bound to guarantee with high probability that the network will not be segmented. However there is no guarantee that the range of transmission obtained in this manner will provide the highest capacity. Furthermore, 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, it is not immediately clear if this remains true in the many-to-one case (in fact it will be shown that it does not). For these reasons, the capacity studied in this work will be derived as a function of the range of transmission, assuming the transmission range can provide connectivity. Finding a range of transmission that guarantees connectivity and maximizes capacity under many-to-one communication is a research problem on its own and is out of the scope of this paper. An expression of what range guarantees connectivity as n goes to infinity is given in [9]. Recently Xue and Kumar also derived bounds on the number of neighbors needed to guarantee connectivity [10].

The rest of the paper is organized as follows. Section II presents our network model. Section III gives our main results on capacity analysis in a flat network along with discussion. IV presents results on capacity of a hierarchical network. Section V discusses issues related to trade-offs between capacity and energy consumption, in-network processing, and design implications of our results on efficient protocol development. Section VI concludes the paper.

II. Network Model

We consider a network deployed in a field of circular shape. There are n nodes/sources (we will use *nodes*, *sources* and *sensors* interchangeably in subseqent discussions) in the network. One *sink/destination* is situated at the center of the network. Each source is not only a "source" of data, but can also be a relay for some other sources to reach the sink.

Throughout the paper we will refer to a network where the nodes are randomly placed following a uniform distribution as a *randomly deployed network* or a *random network*. In such a network we have no direct control over the exact location of the nodes. We will refer to a network where we can determine the exact locations of the nodes as an *arbitrary network*. Note that an arbitrary network is thus a particular instance of the random network with a very low probability of occurring. Considering this, two possible ways of deriving capacity limit arise. One is to consider the best possible deployment and determine its capacity. Although this will be a true upper bound, it is not a very useful or insightful one since that particular deployment outcome is likely to be of a very small probability as a result of the random deployment strategy. The other, which is the approach we take in this paper, is to derive the capacity limit that is applicable with high probability as the number of sensors goes to infinity.

We consider two network organizational architectures. The first one is a flat architecture where nodes communicate with the sink via possibly multi-hop routes by using peer nodes as relays. Intuitively, with fixed transmission range, nodes closer to the sink will serve as relay for a larger number of sources. As mentioned before, we will assume that all sources use the same frequency to transmit data, thus sharing time. However our results apply so long as there is a single shared resource, e.g., time, frequency, and so on. We also assume that a sensor cannot receive and transmit at the same time, and that the sink can only hear from one sensor at a time.

The second architecture is hierarchical where clusters are formed. Under this architecture, 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*. The cluster head can potentially perform data aggregation and processing and then forward data to the sink. In this study, we will assume that the cluster heads serve as simple relays and no data aggregation is performed. We will also assume that 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. In general clusters can be formed by selecting a few nodes in the network to serve as cluster heads [3], or by adding specific cluster head nodes [11] into the network. For simplicity of presentation, we will assume that cluster heads are extra nodes introduced while the total of n source nodes remain constant. This is solely for clarity of discussion and does not affect our conclusion.

Throughout the paper we will assume that the sources share the resource (time) by transmitting following a schedule that consists of time slots. Note however that the same analysis and same results could be obtained if we considered a different resource, such as frequency or codes. This schedule determines what subset of node can transmit simultaneously during which time slot, thus reflecting space reuse. This is somewhat reminiscent of certain dynamic TDMA schemes that generate a local time schedule. Note that how this schedule is generated is left unspecified in this paper. We will examine schedules that can guarantee an equal average throughput from all sources in the network. Our results will be built on the existence of such a schedule that may be obtained either in a centralized or distributed way. We will not be concerned with causality in packet relaying as we are only interested in the average transmission rate of sources. We will simply assume that a node has enough of its own packets buffered so that when it is time for it to transmit and it happens not to have a route-through packet, they will transmit one from the buffer.

We assume the field has an area of 1. This is done to simplify the resulting expressions, which can then be easily scaled with the size of the area. The field is assumed to have a circular shape. This is done to simplify the handling of boundary effect as we will discuss in the next section. Nodes share a common wireless channel using an omni-directional antenna. We assume nodes use a fixed transmission power and achieve a fixed transmission range. We adopt the following commonly used interference model. Let X_i and X_j be two sources with distance $d_{i,j}$ between them. Then the transmission from X_i to X_j will be successful if and only if

$$d_{i,j} \leq r \text{ and } d_{k,j} > r + \Delta, \quad \Delta \geq 0 \tag{1}$$

for any source X_k that is simultaneously transmitting. Subsequently r will be known as the transmission range and Δ as the interference range.

The sink situated deterministically at the center of this field is the ultimate receiver of all data generated by sources in the network. The effect of positioning the sink close to the edge of the network is discussed in the next section. Our network scenario is depicted in Fig. 1.

Throughout the paper W will refer 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 destination.

III. Capacity in a Flat Network

In this section we present capacity results for the flat network along with respective proofs. We begin by outlining a trivial upper bound on throughput capacity. We then show when this bound can be achieved. For conditions under which this is not achievable, we construct a few lower bounds on achievable capacity. We end this section by a discussion on the effect of placing the sink on or close to the boundary of the network.

A. A Trivial Upper Bound

Theorem 1: The maximum per node throughput in a wireless network featuring many-to-one communication outlined by the network model is upper bounded by $\frac{W}{n}$.

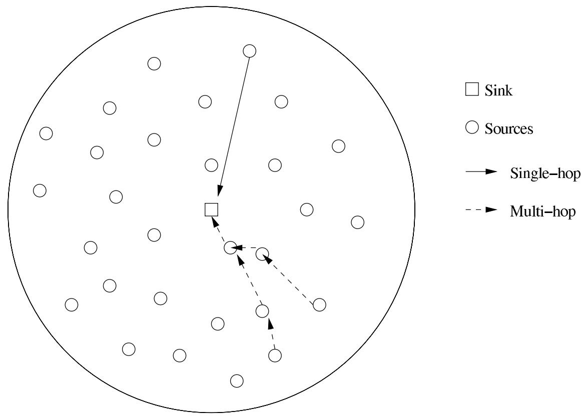

Fig. 1. Many-to-one Network Scenario

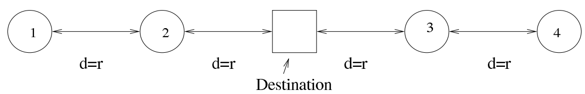

Fig. 2. $\lambda = \frac{W}{n}$ is achievable

proof: The maximum throughput is achieved when the sink is 100% busy, shared by n sources. Since W is the capacity of the shared channel, the maximum throughput can be achieved by any single source, λ, must be such that

$$n\lambda \leq W \quad \longrightarrow \quad \lambda \leq \frac{W}{n}. \tag{2}$$

Thus $\frac{W}{n}$ bits/second per source is an upper bound for the maximum throughput that can be achieved on average by each source in the network. ■

The above result means that each source can only use up to one n^{th} of the resources under our network model.

Corollary 1: $\lambda = \frac{W}{n}$ can be achieved when every source can directly reach the sink.

proof: When all sources can directly (via a single hop) reach the sink, one can simply use TDMA or round robin to schedule source transmission. n slots will be needed, one for each source. Thus each source gets an equal share of the channel resulting in $\lambda = \frac{W}{n}$. ■

Corollary 2: $\lambda = \frac{W}{n}$ is not achievable if not every source can directly reach the destination and $\Delta > r$.

proof: Consider a source node which is two hops away from the destination. Let d be the distance from this source to the destination. Then $r < d \leq 2r$ under the interference and transmission model. The interference range is then $r + \Delta$. Since $\Delta > r$, we have $d < 2r < r + \Delta$. Thus the sink is within the interference range of this source node. This means that during this node's transmission the sink has to be idle. From Theorem 1 we know that the bound $\frac{W}{n}$ is only achieved when the sink has zero idle time. Therefore $\lambda = \frac{W}{n}$ can not be achieved in this case. ■

This result points to the fact that when $\Delta > r$ there is no schedule that will allow the sink to be busy all the time. This in turn prevents the network from achieving the upper bound on capacity.

Corollary 3: $\lambda = \frac{W}{n}$ may be achieved in an arbitrary network when not every source can directly reach the destination and $\Delta < r$.

proof: We prove this corollary by construction. Consider a network with four sources and one sink on a straight line as shown in Fig. 2. The distance between any two adjacent sources (sink) is precisely r. We show below that there exists a schedule of length 4 that allows the sink to receive one packet from each of the 4 sources, therefore achieving $\frac{W}{n}$ per source.

Using the labels shown in Fig. 2, since $\Delta < r$, the first slot can be shared by Sources 1 and 3. Thus the sink receives one packet from Source 3 and Source 2 receives one packet from Source 1. The second slot is shared by Sources 2 and 4, with the sink receiving a packet from Source 2 and Source 3 receiving a packet from Source 4. In the third slot Source 3 relays the packet from Source 4 to the sink. In the fourth slot Source 2 relays the packet from Source 1 to the sink. The same schedule then repeats. ■

Unfortunately here the bound is achieved by carefully positioning nodes in the network. For a randomly deployed networks this capacity can not be achieved with high probability. The following subsection examines this issue further.

B. An Upper Bound for Random Multiple-Hop Networks

In this subsection we will show that when the sink cannot directly receive from every source in the network, and assuming that the channel allocation does not take into account difference in traffic load, then $\lambda = \frac{W}{n}$ is not achievable with high probability regardless of the value of Δ. We will determine the upper bound on capacity by finding out the maximum number of simultaneous transmissions.

Denote by A_r the area of a circle of radius r, i.e., $A_r = \pi r^2$. Let random variable V_r denote the number of nodes within an area of size A_r and assume a total area of 1. We then have the following lemma.

Lemma 1: In a randomly deployed network with n nodes,

$$\text{Prob}\left(nA_r - \sqrt{\alpha_n n} \leq V_r \leq nA_r + \sqrt{\alpha_n n}\right) \rightarrow 1 \quad \text{as} \;\; n \rightarrow \infty, \tag{3}$$

where the sequence $\{\alpha_n\}$ is such that $\lim_{n\rightarrow\infty} \frac{\alpha_n}{n} = \epsilon$, ϵ positive but arbitrarily small.

proof: The mean of V_r is nA_r and the variance σ^2 is $nA_r(1 - A_r)$, using Chebychev's inequality we have

$$\text{Prob}\left(nA_r - \sqrt{\alpha_n n} \leq V_r \leq nA_r + \sqrt{\alpha_n n}\right) \geq 1 - \frac{\sigma^2}{\alpha_n n} = 1 - \frac{A_r(1 - A_r)}{\alpha_n}. \tag{4}$$

The second term on the RHS of Equation (4) goes to zero since $\alpha_n \rightarrow \infty$ as $n \rightarrow \infty$. Thus the proof is complete. ■

This lemma shows that the number of nodes in a fixed area is bounded within $\sqrt{\alpha_n n}$ of the mean where α_n goes to infinity as $n \rightarrow \infty$ but $\lim_{n\rightarrow\infty} \frac{\alpha_n}{n}$ is arbitrarily small.

Theorem 2: 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 capacity $\lambda = \frac{W}{n}$ is not achievable.

proof: The proof is based on the maximum number of simultaneous transmissions achievable within the network. For every transmitter-receiver pair, there is an interference area around the receiver, within which no nodes can transmit in order for the receiver to receive successfully. In particular this area is a circle of radius $r + \Delta$ centered around the receiver. Denoted by $A_{r+\Delta}$, this area satisfies

$$A_{r+\Delta} \geq \pi r^2, \tag{5}$$

where equality holds when $\Delta = 0$. Since every receiving sensor needs the same amount of space, the number of simultaneous transmissions, denoted by t, that can be accommodated is

$$t < \frac{1}{A_{r+\Delta}} \leq \frac{1}{\pi r^2}. \tag{6}$$

Note the first inequality is strict because circles cannot create a perfect tessellation in a two dimensional area. Regardless of how the circles are arranged, there will always be some uncovered area. Denote by μ_m the minimum uncovered area per transmitter-receiver pair as a result of a node arrangement that minimizes the total uncovered area. It can be shown that a very good approximation of μ_m is $r^2\left(2\sqrt{3} - \pi\right)$ (details can be found in the Appendix). Then

$$t \leq \frac{1}{\pi r^2 + \mu_m}, \tag{7}$$

where equality holds when $\frac{1}{\pi r^2 + \mu_m}$ is an integer. The length of a schedule, denoted by s, that ensures all sources have a chance to transmit has to satisfy the following

$$s \geq n(\pi r^2 + \mu_m). \tag{8}$$

Again equality holds when (7) holds and $n(\pi r^2 + \mu_m)$ is an integer. Denote by l the number of sources that use a node that's one hop away from the sink as their relay, including the relaying node itself. In order to maximize throughput each of the node that's one hop away from the sink has to get an equal share of the total traffic load. Using Lemma 1 l can be bounded with high probability as follows:

$$\frac{1}{\pi r^2 + \sqrt{\frac{\alpha_n}{n}}} = \frac{n}{n\pi r^2 + \sqrt{\alpha_n n}} \leq l \leq \frac{n}{n\pi r^2 - \sqrt{\alpha_n n}} = \frac{1}{\pi r^2 - \sqrt{\frac{\alpha_n}{n}}} \tag{9}$$

$$\text{as } n \to \infty, \quad \frac{1}{\pi r^2 + \sqrt{\epsilon}} \leq l \leq \frac{1}{\pi r^2 - \sqrt{\epsilon}}. \tag{10}$$

A node that is one hop away from the sink will need to carry traffic from l source nodes. So as $n \to \infty$, with high probability

$$\begin{aligned} l \cdot \lambda &= \frac{W}{s} \\ \lambda &\leq \frac{W}{\frac{1}{\pi r^2 + \sqrt{\epsilon}} n(\pi r^2 + \mu_m)} \\ &\leq \frac{W(\pi r^2 + \sqrt{\epsilon})}{n(\pi r^2 + \mu_m)}. \end{aligned} \tag{11}$$

Note that since μ_m is positive and fixed for fixed r, and $\sqrt{\epsilon}$ can be made arbitrarily close to zero, as $n \to \infty$

$$\lambda \leq \frac{W(\pi r^2 + \sqrt{\epsilon})}{n(\pi r^2 + \mu_m)} < \frac{W}{n} \tag{12}$$

Thus $\lambda = \frac{W}{n}$ is not achievable. ■

(12) indicates that λ is approximately (as $n \to \infty$, ϵ close to zero) bounded by $\frac{W\pi r^2}{n(\pi r^2 + \mu_m)} = \frac{W}{1.014n}$. This is not a significant improvement on the trivial upper bound. Nevertheless it is an important and interesting result, as it shows that the trivial upper bound is not achievable with high probability using multi-hopping. It is not immediately clear whether or not this slightly tighter bound is achievable with high probability for a random topology. This is subject to further study. Also, as mentioned before, we have assumed that the channel allocation does not take into account difference in traffic load. In the next subsection we will first construct achievable capacity using multi-hopping following the same assumption, then we will assume that channel allocation takes into account the difference in traffic load and see that the achievable capacity increases.

C. Achievable Capacity

The following theorems construct capacities that can be achieved with high probability in a randomly deployed network that follows the conditions outlined in Section 2. Again our results are as a functions of r and we assume the transmission range r is large enough to guarantee connectivity. In constructing these bounds we will assume that the routing and relaying scheme is such that each of the one hop away nodes carries an equal share of the overall traffic. This is feasible given our network model outlined in Section II.

Theorem 3: A randomly deployed network using multi-hop transmission for many-to-one communication can achieve throughput $\lambda \geq \frac{W}{n} \frac{\pi r^2 - \sqrt{\epsilon}}{4\pi r^2 + 4\pi r\Delta + \pi\Delta^2 + \sqrt{\epsilon}}$ with high probability, when no knowledge of the traffic load is assumed and ϵ is as given in Lemma 1.

proof: Consider a source that is at least $2r + \Delta$ away from the closest border of the network. The area of interference is thus a circle of radius $r' = 2r + \Delta$ centered at this source. Using Lemma 1, with high probability the number of interfering neighbors including the source, k_1, is

$$nA_{r'} - \sqrt{\alpha_n n} \leq k_1 \leq nA_{r'} + \sqrt{\alpha_n n}. \tag{13}$$

Consider the entire network represented as a connected graph G(V, E), *with edges connecting nodes that are within each other's interference range*. Then the highest degree of this graph is $k_1 - 1$, since k_1 is the number of nodes within any interference area. Using the known result from graph theory, see for example [12], [13], that the chromaticity of such a graph is upper bounded by the highest degree plus one, $k_1 - 1 + 1 = k_1$ in this case. Therefore there exists a schedule

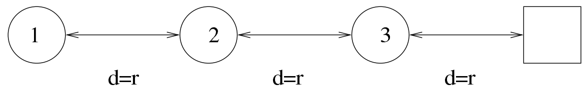

Fig. 3. Chain Network

of length at most $s \leq k_1$ that allows all sources to transmit. The sources one hop away from the sink will have to carry the traffic of the entire network. The number of these one hop sources, k_2, is bounded with high probability as follows (Lemma 1): $nA_r - \sqrt{\alpha_n n} \leq k_2 \leq nA_r + \sqrt{\alpha_n n}$. Therefore we have:

$$\begin{aligned} \frac{n}{nA_r - \sqrt{\alpha_n n}}\lambda &\geq \frac{n}{k_2}\lambda = \frac{W}{s} \geq \frac{W}{k_1} \geq \frac{W}{nA_{r'} + \sqrt{\alpha_n n}} \\ \frac{1}{A_r - \sqrt{\alpha_n / n}}\lambda &\geq \frac{W}{nA_{r'} + \sqrt{\alpha_n n}} \\ \text{as } n \to \infty, \quad \lambda &\geq \frac{W}{n} \cdot \frac{A_r - \sqrt{\epsilon}}{A_{r'} + \sqrt{\epsilon}} \\ &= \frac{W}{n} \cdot \frac{\pi r^2 - \sqrt{\epsilon}}{4\pi r^2 + 4\pi r\Delta + \pi\Delta^2 + \sqrt{\epsilon}} \\ (\text{since } \sqrt{\epsilon} \text{ arbitrarily close to } 0) &\approx \frac{W}{4n\left(1 + \Delta\left(\frac{1}{r} + \frac{\Delta}{4r^2}\right)\right)}. \end{aligned} \tag{14}$$

■

In this proof we considered a source that is at least $2r + \Delta$ away from the closest border of the network because such a source is fully surrounded by interfering neighbors on all sides. Therefore a source with this characteristic has the largest number of interfering neighbors in the network. Note that the bound increases as r increases. This is because as r increases the effect of a fixed Δ diminishes, i.e., Δ becomes a smaller percentage of r. Conversely, an increase in Δ decreases the bound as a larger number of interfering neighbors will affect a given source. When $\Delta = 0$ we can approximately achieve $\lambda \approx \frac{W}{4n}$, which seems to be independent of r. This is because implicitly we required r to be such that there is connectivity in the network to construct this proof. Given this requirement is met, then the achievable throughput capacity is independent of the transmission range.

The above bound is constructed by finding a schedule of length $s \leq k_1$ and assuming each source gets an equal share of the bandwidth, represented by W/s. As we have discussed, the nodes closer to the sink carry more traffic. Equally sharing the bandwidth necessarily means that nodes further away from the sink will waste some of the assigned slots. Intuitively allocating more share to the nodes that carry more traffic should result in higher throughput capacity. The next theorem examines the existence of such a schedule and derives a new constructive lower bound based on this schedule.

Before proceeding to Theorem 4, it helps to introduce a new concept *virtual sources*. This is best illustrated with an example. Consider a simple network consisting of three sources and a sink, shown in Fig. 3. The distance between all adjacent nodes is r. Note that regardless of the value of Δ, when one source transmits, it interferes with all other sources in this network. Therefore only one source can transmit at a time. The number of interfering neighbors for any of the sources is two, which is the highest degree of the graph that represents the interference relationship in this network. Thus a schedule of length 3 allows all sources to transmit once during the schedule. The load on the source closest to the sink, Source 3, is 3λ, since it carries the traffic of all three sources. The achievable throughput is then calculated as $3\lambda = \frac{W}{3}$, thus $\lambda = \frac{W}{9}$.

On the other hand, it is easy to see with this example that λ could achieve $\frac{W}{6}$. The way the schedule was calculated previously assigned the same share of the resources to all the sources. Since we used the source with the highest need of resource (the one carrying the most traffic) to calculate the amount of resource needed, every other source is wasting resource. In our example we are giving every source the possibility of making three transmissions. Source 3 does indeed need all three transmissions, but Source 2 only needs two and Source 1 only needs one, hence a total of six transmissions.

Now consider a similar network only this time we have three sources that can reach the sink, shown in Fig. 4. We create a schedule where each one of the sources gets to transmit once and once only. However this time not all sources

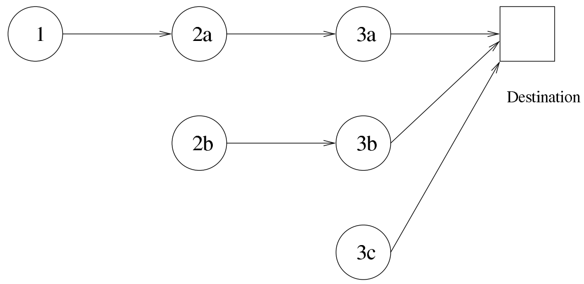

Fig. 4. Virtual Sources

generate data. Using labels shown in Fig. 4, Source 1 generates a packet and transmits it to Source $2a$. Source $2a$ relays the packet to Source $3a$, who then relays it to the sink. Then Source $2b$ generates a packet and transmits it to Source $3b$, who relays it to the sink. Finally source $3c$ generates and transmits a packet to the destination. We can view each column of sources in this network as an equivalent of a single source in the previous example, i.e., $2a$ and $2b$ combined is equivalent to 2 in Fig. 3, $3a, 3b$ and $3c$ combined is equivalent to 3, in terms of interference and traffic load. We will define Sources $2a-2b$ and $3a-3c$ as *virtual sources* in that they each represents one actual source in the network but they are co-located in one physical source. Adopting this concept, in this network the highest number of interfering neighbors is 5 (with a total of 6 virtual sources all in one interference area) and therefore there exists a schedule of length 6 that enables every virtual source to transmit once. Since the traffic load is the same for all virtual sources, the resources will be shared equally and no source will be wasting its share. In this case we get $\lambda = \frac{W}{6}$. Note that this is the best λ that could be obtained in our previous example in Fig. 3.

This concept allows us to define a "traffic load-aware" schedule in the following way.

1. For each source node, create one virtual source for every source node whose traffic goes through this node, including itself;
2. Counting all the virtual sources we can determine the number of interfering neighbors (virtual sources) k. The new maximum degree of the interference graph is then $k-1$;
3. A schedule of length $s \leq k$ exists which is equally shared among virtual sources;
4. The achievable throughput per node is simply the share obtained by any virtual source in the network, i.e., $\lambda = \frac{W}{s} \geq \frac{W}{k}$.

Here again we are not specifying how to find such a schedule but rather we establish the existence of such a schedule. The concept of virtual sources is used to prove the following theorem.

Theorem 4: 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^+ is the upper bound on the number of virtual sources per actual source that is h hops away from the sink with high probability. n_h^+ is the upper bound on the number of actual sources that are h hops away from the sink with high probability.

proof: Let us consider a source that is arbitrarily close to the sink. We now derive the number of virtual sources differentiated by the number of hops to the sink. By the geometry of the circles of radius $r, 2r, \cdots, ir$ that are centered around the sink we get the entries in Table I with high probability as $n \rightarrow \infty$ (Using Lemma 1). Then the number of interfering neighbors, k, in terms of virtual nodes with high probability satisfies

$$k \leq \sum_{h=1}^{h=\lceil 2+\frac{\Delta}{r}\rceil} l_h^+ n_h^+. \tag{15}$$

The inequality is due to the fact that the right hand side is summed to the next integer greater or equal to $2+\Delta/r$, and the fact that this is the number of interfering neighbors calculated by considering a node that is arbitrarily close to the sink. Note such a choice leads to the maximum number of interfering neighbors (virtual sources). Thus a schedule

TABLE I
TRAFFIC LOAD

h	n_h	l_h
1	$[nA_r - \sqrt{\alpha_n n}, nA_r + \sqrt{\alpha_n n}]$	$[\frac{1}{A_r+\sqrt{\epsilon}}, \frac{1}{A_r-\sqrt{\epsilon}}]$
2	$[3nA_r - 2\sqrt{\alpha_n n}, 3nA_r + 2\sqrt{\alpha_n n}]$	$[\frac{1-A_r-\sqrt{\epsilon}}{3A_r+2\sqrt{\epsilon}}, \frac{1-A_r+\sqrt{\epsilon}}{3A_r-2\sqrt{\epsilon}}]$
...	...	...
i	$[(2iA_r - A_r) - i\sqrt{\alpha_n n}, (2iA_r - A_r) + i\sqrt{\alpha_n n}]$	$[\frac{1-(i-1)^2A_r-\sqrt{\epsilon}}{(2iA_r-A_r)+i\sqrt{\epsilon}}, \frac{1-(i-1)^2A_r+\sqrt{\epsilon}}{(2iA_r-A_r)-i\sqrt{\epsilon}}]$

of length $s \leq k$ exists, and the per node throughput with high probability is

$$\begin{aligned} \lambda &= \frac{W}{s} \geq \frac{W}{k} \\ &\geq \frac{W}{\sum_{h=1}^{h=\lceil 2+\frac{\Delta}{r} \rceil} l_h^+ n_h^+}. \end{aligned} \quad (16)$$

■

The above result is not a direct or explicit function of r and Δ, since the expression of the sum depends on the value of Δ. The following corollary is probably more interesting by assuming $\Delta = 0$.

Corollary 4: 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 $\Delta = 0$.

proof: This is a direct result of Theorem 4 when $\Delta = 0$.

$$\begin{aligned} \lambda &\geq \frac{W}{l_1^+ n_1^+ + l_2^+ n_2^+} \\ \text{as } n \to \infty, \quad &\geq \frac{W}{n\left((\pi r^2 + \sqrt{\epsilon})\left(\frac{1}{\pi r^2+\sqrt{\epsilon}}\right) + (3\pi r^2 + 2\sqrt{\epsilon})\left(\frac{1-\pi r^2+\sqrt{\epsilon}}{3\pi r^2-2\sqrt{\epsilon}}\right)\right)} \\ \text{(since } \sqrt{\epsilon} \text{ arbitrarily close to } 0,) \quad &\approx \frac{W}{(2-\pi r^2)\, n}. \end{aligned} \quad (17)$$

■

Note that when traffic load is taken into account in scheduling, the achievable capacity of the network is almost doubled, compared to the throughput $\lambda \approx \frac{W}{4n}$ obtained previously. Also note that even with $\Delta = 0$, r still has an effect on λ. As r increases so does λ. This is because regardless of Δ, as r increases, the number of sources one hop away from the destination increases and thus each source carries a smaller load.

D. Discussion

There are some circumstances that might affect our results. Note that the previous subsections have intentionally avoided the boundary effect by positioning the sink close at the center and limiting r to a value that allows the node close to the sink to have its interference range within the network area. If the sink is close to the edge, the area close to the sink (one or two hops away) will be smaller as they can only reach the destination from a limited range of directions.

Let A_1 be the area that contains the sources that are one hop away and A_2 be the area that contains the sources that are two hops away. Let l_1 be the number of virtual sources per actual source node that is one hop away and l_2 is the number of virtual sources per actual source node that is two hops away. For the purpose of this discussion let $\sqrt{\epsilon} \approx 0$, which means that $l_h^- \approx l_h \approx l_h^+$ as $n \to \infty$. Then the number of interfering neighbors, assuming $\Delta = 0$, for a source arbitrarily close to the destination is $k = l_1 n A_1 + l_2 n A_2$, where $l_1 = \frac{n}{nA_1} = \frac{1}{A_1}$ and $l_2 = \frac{l_1 - 1}{A_2/A_1} = \frac{1-A_1}{A_2}$. Thus $k = n(2 - A_1)$. As A_1 decreases the number of interfering neighbors increases and so does the length of the schedule. In turn the achievable capacity decreases. Thus placing the sink near the border of the network reduces the achievable throughput capacity by reducing the area close to the sink. Furthermore in this case it is not clear whether the source node with the highest number of interfering neighbors is the one arbitrarily close to the destination. Therefore the achievable capacity might be reduced even further. On the other hand, since these results are lower bounds on throughput capacity, one may still achieve higher throughput under certain conditions.

Another issue worth discussing is decentralized scheduling, which would be highly desirable. So far we only assumed that certain schedules exist but the generation of these schedules is highly non-trivial. However, distributed methods that compute schedules with similar performance are not impossible. The traffic-load awareness discussed above naturally exists in many MAC schemes simply because nodes with lower traffic load will compete for the channel less frequently and therefore nodes with higher traffic load will get more share. Transmission rate control schemes, such as proposed in [14], combined with MAC seems promising and would be an interesting future study.

IV. Capacity in a Hierarchical Network

In this section we outline the results for the hierarchical network along with respective proofs. As mentioned in the network model, we will consider a hierarchical network by introducing extra nodes as cluster heads. By doing so we obtain more clarity in the resulting expressions. H denotes the number of clusters (heads) introduced. Each cluster head will create a cluster containing 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 will consider the following placement of the cluster heads. Note in practice we may or may not be able to control the location of these heads. However considering an ideal placement helps us construct the achievable capacity of the network. We assume cluster heads are placed with high probability in such a way that as $n \to \infty$ the number of nodes in each cluster approaches $\frac{n}{H}$. In order to avoid boundary problems, there is at least a distance of $2(2r + \Delta)$ between any two cluster heads. By doing so the clusters essentially form a voronoi tessellation [15] of the field, where every cluster (or voronoi cell) contains a circle of radius $2r + \Delta$. Consequently sources located near the boundary between two clusters will not have a higher number of interfering neighbors (in terms of virtual sources), due to low traffic load, than the ones closer to the cluster heads. Thus in applying bounds obtained earlier to each cluster we do not have to be concerned with the boundary effect, and previous results are directly applicable. These assumptions also imply that each voronoi cell effectively has the same area. Using Lemma 1 again we have with high probability each cluster size is within $\sqrt{\alpha_n n}$ of $\frac{n}{H}$, where α_n is such that $\lim_{n\to\infty} \frac{\alpha_n}{n} = \epsilon$.

A. Main Results

We will refer to the capacity achieved within a cluster (as opposed to the capacity obtained in the entire network) as $\lambda^{'}$. Note that since a cluster head needs to split its time between transmission and reception, $\lambda^{'}$ is the per node throughput achieved during the portion of time that the cluster head is receiving. The bounds on $\lambda^{'}$ are immediately available from our results in the previous section by considering the cluster head as the sink and a total of $\frac{n}{H}$ sources in the network. Note that in general the achievable throughput $\lambda^{'}$ is a function of H. Intuitively, from previous results we expect each cluster to achieve a higher throughput due to the reduced number of sources in a cluster.

The question of interest is whether there exists an appropriate number of clusters H that would allow the network to achieve $\lambda = \frac{W}{n}$ with high probability using clustering, when cluster heads have the same transmission capacity W as the sources. That $\frac{W}{n}$ remains to be the upper bound is obvious considering the fact that the sink cannot receive at rate more than W, and that there are n sources in the network.

In order to achieve the maximum capacity $\lambda = \frac{W}{n}$, the sink has to be busy all the time, which implies that at any given moment one of the cluster heads must be transmitting. Since each cluster has the same size, every cluster head would need to transmit the same amount of data and requires the same amount of time. If this limit is achieved, it follows that each head transmits $\frac{1}{H}$ percentage of the time, leaving $1 - \frac{1}{H}$ as the percentage of time devoted for receiving from sources within the cluster. In order to achieve $\frac{W}{n}$, total throughput achieved within clusters must be at least W:

$$\lambda^{'} n \left(1 - \frac{1}{H}\right) \geq W. \tag{18}$$

Using Theorem 3, we have

$$\lambda^{'} \geq \frac{W}{\frac{n}{H} + \sqrt{\alpha_n n}} \cdot \frac{\pi r^2 - \sqrt{\epsilon}}{4\pi r^2 + 4\pi r\Delta + \pi\Delta^2 + \sqrt{\epsilon}}. \tag{19}$$

Thus

$$\lambda^{'} n \left(1 - \frac{1}{H}\right) \geq \frac{W}{\frac{n}{H} + \sqrt{\alpha_n n}} \cdot \frac{\pi r^2 - \sqrt{\epsilon}}{4\pi r^2 + 4\pi r\Delta + \pi\Delta^2 + \sqrt{\epsilon}} n \left(1 - \frac{1}{H}\right). \tag{20}$$

Therefore if the following holds, (18) will hold:

$$\text{as } n \to \infty, \quad \frac{W}{n\left(\frac{1}{H} + \sqrt{\epsilon}\right)} \cdot \frac{\pi r^2 - \sqrt{\epsilon}}{4\pi r^2 + 4\pi r\Delta + \pi\Delta^2 + \sqrt{\epsilon}} n \left(1 - \frac{1}{H}\right) \geq W. \tag{21}$$

After some algebra it can be shown that we need the following to satisfy (21) and (18):

$$H \geq \frac{5\pi r^2 + 4\pi r\Delta + \pi\Delta^2}{\pi r^2 - \sqrt{\epsilon} + (4\pi r^2 + 4\pi r\Delta + \pi\Delta^2)\sqrt{\epsilon}}. \tag{22}$$

Since $r > 0$, $\Delta \geq 0$ and $\sqrt{\epsilon}$ is arbitrarily close to 0, H is bounded from below in order to achieve capacity. At the same time our assumptions on the formation of clusters implies

$$H \leq \frac{1}{\pi (2r + \Delta)^2}. \tag{23}$$

Thus $\lambda = \frac{W}{n}$ can be achieved if

$$\frac{5\pi r^2 + 4\pi r\Delta + \pi\Delta^2}{\pi r^2 - \sqrt{\epsilon} + (4\pi r^2 + 4\pi r\Delta + \pi\Delta^2)\sqrt{\epsilon}} \leq \frac{1}{\pi (2r + \Delta)^2}, \tag{24}$$

which means that the range of transmission r must satisfy

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

In the case of $\Delta = 0$ and letting $\sqrt{\epsilon} \approx 0$, we need $r < \sqrt{\frac{1}{20\pi}}$. Therefore there is a range of transmission that allows us to achieve $\lambda = \frac{W}{n}$ as $n \longrightarrow \infty$. Note that as the density of the network increases, the r needed for connectivity decreases. In fact as n goes to infinity the authors of [9] show that $r(n) = \sqrt{\frac{log(n) + \gamma_n}{\pi n}}$ ensures connectivity. Since we are dealing with a fixed area size, increasing n increases our density, therefore when $n \to \infty$ it is always possible to use the number of heads H needed to achieve $\lambda = \frac{W}{n}$.

Note (18) holds when the amount cluster heads receive from sources equals the amount they transmit to the sink. Strict inequality is also feasible but that would imply that sources send more to the cluster head than they can delivery to the sink, which would eventually lead to overflow.

If $\lambda^{'}$ is greater than the lower bound used above then $\lambda = \frac{W}{n}$ can be achieved with even less heads. For instance if we use Corollary 4 , $\lambda^{'} \geq \frac{W}{\frac{n}{H}(2 - \pi r^2)}$, which leads to

$$\begin{aligned} 1 - \frac{1}{H} &\geq \frac{\frac{W}{n}}{\frac{W}{\frac{n}{H}(2 - \pi r^2)}} \\ H &\geq 3 - \pi r^2. \end{aligned} \tag{26}$$

If single-hop communications are also used within each cluster, $\lambda^{'} = Wn/H$, then we would need $\frac{W}{n/H} n(1 - \frac{1}{H}) \geq W$, which means $H \geq 2$.

Note that in both cases the minimum requirement on H is independent of n. The reason for this is that both the capacity of the multi-hop and the capacity for the single hops are of order W/n, which in this case relates to the communication on the first layer within clusters and the second layer between clusters. The analysis above allows us to state the following theorem:

Theorem 5: 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 allows the network to achieve $\lambda = \frac{W}{n}$ with high probability as $n \to \infty$.

Now we consider the case where the transmission rate of cluster heads $W' > W$. We want to know if there exists an appropriate number of heads H that allows the wireless network to achieve $\lambda = \frac{W'}{n}$, which would also be the upper bound on capacity in this case. $\frac{W'}{n}$ is the upper bound because the sink cannot receive at rate more than W', and that there are n sources in the network. The rest of the analysis is very similar to the previous one. We therefore skip the reasoning and state that in order to achieve the capacity we need

$$\lambda' n \left(1 - \frac{1}{H}\right) \geq W'. \tag{27}$$

Using Theorem 3 we can show if the following holds then (27) will hold:

$$\frac{W}{n\left(\frac{1}{H} + \sqrt{\epsilon}\right)} \cdot \frac{\pi r^2 - \sqrt{\epsilon}}{4\pi r^2 + 4\pi r\Delta + \pi\Delta^2 + \sqrt{\epsilon}} n \left(1 - \frac{1}{H}\right) \geq W' \text{ as } n \to \infty, \tag{28}$$

which means we need

$$H \geq \frac{5\pi r^2 + 4\pi r\Delta + \pi\Delta^2}{\pi r^2 - \sqrt{\epsilon} + (4\pi r^2 + 4\pi r\Delta + \pi\Delta^2)\sqrt{\epsilon}} \cdot \frac{W'}{W}. \tag{29}$$

Following the same reasoning as before, H is lower bounded and and has to satisfy $H \leq \frac{1}{\pi(2r+\Delta)^2}$. For the same reason as before, as $n \to \infty$ there always exists an r that will enable us to use the H needed to achieve $\lambda = \frac{W'}{n}$. Once again if λ' is greater than the lower bound used above then $\lambda = \frac{W'}{n}$ can be achieved with even less heads. Using Corollary 4, $\lambda' \geq \frac{W}{\frac{n}{H}(2 - \pi r^2)}$, we need

$$\begin{aligned} 1 - \frac{1}{H} &\geq \frac{\frac{W'}{n}}{\frac{W}{\frac{n}{H}(2-\pi r^2)}} \\ H &\geq \left(3 - \pi r^2\right)\frac{W'}{W}. \end{aligned} \tag{30}$$

In this case, H remains independent of n for the same reason discussed before. However, it is dependent on W'. This is because in order to achieve higher throughput (due to $W' \geq W$), we need smaller clusters. The above analysis allows us to state the following theorem.

Theorem 6: 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 $n \to \infty$, that allows the network to achieve $\lambda = \frac{W'}{n}$ with high probability. $\frac{W'}{n}$ is also the upper bound on throughput capacity in this scenario.

B. Discussion

The results in this section are important as they show higher throughput can be achieved by using clustering. However this capacity comes at a cost, which is the extra nodes functioning as cluster heads. These extra nodes will require a bigger transmission range/rate and a greater energy reserve to handle the transmissions required.

We showed the minimum requirement on the number of clusters needed to achieve the capacity. The feasibility of this obviously depends on the size of the network and the range of transmission r. Throughout the paper we have assumed that r is sufficiently large to ensure connectivity without quantifying it. Here we use the same assumption. It is reasonable to expect that as long as the network is large enough, this minimum number of clusters should be able to be accommodated.

Instead of introducing new nodes, one could have used some of the sources as cluster heads assuming they use a different channel and higher transmission power for communicating with the sink. Minor changes would occur in our equations following this model. However, the end conclusion would be the same. The difference is that H would then represent the number of sources to be substituted instead of the number of nodes to be introduced.

The result of using more than the required number of heads is that each head will increase its idle time. That can be an advantage in a more practical scenario where that "idle" time can be used for synchronization or the exchange of control messages. So while the result presented in the previous section is theoretically valid, in a practical scenario we will need to increase the number of heads.

It is important to note that we could build a hierarchical network where both the communication within clusters and between cluster heads and the sink is performed using multiple hops, i.e., multi-hop on both layers not interfering with each other. In that case H is approximately $\sqrt{n}$. There are two reasons for this. One is related to energy consumption. The energy consumption is directly related to the number of transmissions needed to transmit the message. Note that it is impossible to change H from $\sqrt{n}$ to reduce the number of transmissions in one layer without increasing the number of transmission in the other layer. Thus the number that minimizes the energy consumption is approximately $\sqrt{n}$. The other reason is related to capacity. Following our results, it is possible to derive a lower bound on the capacity of this network. If the function that is obtained as lower bound is plotted with H in the X axis and λ in the Y axis, a knee will appear at approximately $H = \sqrt{n}$. This knee would be a good operating point. Note that this scenario is also bounded above by $\lambda = \frac{W}{n}$.

V. Discussion

A. Energy Consumption

In this subsection we attempt to reveal certain trade-offs between energy consumption and achievable capacity. The previous section derived asymptotic bounds on the transport capacity as the number of sensors in the network grows infinitely. These limiting results do not directly apply to a deployment with fixed, finite number of sensors. More specifically, we were able to bound the number of sensors in a fixed area A to be within $\sqrt{\alpha_n n}$ of mean nA with high probability as $n \longrightarrow \infty$ (Lemma 1). Such a result does not exist when n is fixed. However, if we imagine a perfectly *typical* deployment that happens to have precisely nA sensors in an area of A, then all the previous results would apply to a network with fixed n by simply replacing the interval (of half width $\sqrt{\alpha_n n}$) around the mean nA by the mean itself. This is obviously an ideal imaginary scenario since for any random deployment of n sensors, the probability of having precisely nA sensors in an area of A indeed diminishes as n becomes large. Nevertheless, this is the network scenario we are going to assume in this subsection for the following reasons. Firstly, such a perfectly typical network can be viewed as the average of a large number of random deployments. Secondly and more importantly, this allows us to compare our capacity results with the energy consumption results of [16] and discuss the trade-offs under a finite setting. Consequently, the results presented in this subsection are averages.

We briefly restate the assumptions we made on [16]. In that paper, we considered the energy consumed under ideal conditions, by assuming that when a node is neither transmitting nor receiving it would be asleep and does not consume any energy. Also, the energy model used was such that the energy consumed in transmitting b bits was $E_t(r) = (e_t + e_d r^{\alpha})b$, where e_t and e_d are specifications of the transceiver used by the nodes, and r is the transmission range. Note that we did not consider power control, therefore for a given scenario r was fixed. α depends on the characteristics of the channel, with typical values of 2 and 4. The channel considered was time invariant, thus α was constant. Energy consumed in receiving b bits was $E_r = e_r b$, where e_r also depends on the transceiver used by the node. We did not consider the energy consumed by the sink. We also assumed that the total area of the network was A instead of unit since we examined the effect of different network scales.

We reproduce here the relevant results of [16] to aid our discussion. In a flat network, the energy consumed, E, is:

$$E = xE_t(r) + (x - n)E_r, \tag{31}$$

where is x the total number of transmissions required to deliver one packet from every node to the sink. Details on x and its calculation can be found in [16].

In the above case the network consists entirely of sensing nodes, meaning each node not only relays data, but also generates data. [16] also considered a network with u nodes that generate data, and v nodes that act only as relays, both of them randomly deployed. It was shown that the energy consumed is $E = yE_t + (y-u)E_r$, where y only depends on u, not v. This means that if we have a network with n nodes and to that we add v nodes acting as relays, the minimum

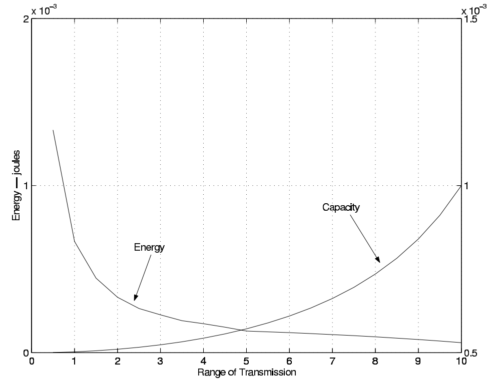

Fig. 5. Flat network, R=10

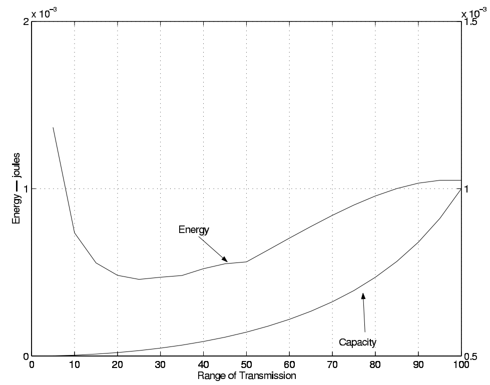

Fig. 6. Flat network, R=100

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. Depending on the size of the network area, this could be beneficial.

In a hierarchical network the energy consumed is:

$$E_H = (x^{'}(E_t) + x^{'}(E_r) + \frac{n}{H}E_t(R))H, \tag{32}$$

where $x^{'}$ is the number of transmissions performed by the nodes in a cluster. Also, $E_{H_1} - E_{H_2} \geq 0$ when $H_1 < H_2$, meaning that as the number of heads increases, the energy consumed decreases. Both results are plotted in [16] for a few different area sizes. Those plots 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 [16] 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.

Figures 5, 6, and 7 show the results for the energy consumption of the flat network (left Y axis) and the capacity that can be achieved in the same network (right Y axis). We see that while energy consumption is affected by the scale of the network, capacity is not (it does change the feasible range of r to ensure connectivity). This means that the relation between energy and capacity changes as the scale of the network changes. In particular, we see that at small scales, the capacity increases as the energy consumption decreases. At large scales, the capacity increases only as the energy consumption increases.

It is important to mention that this set of figures show significant difference in how the energy consumption changes with the transmission range. This is because at smaller scales (network size) e_t is the dominant part of the energy consumption, thus as r increases, the energy consumption decreases (see Fig. 5). At a bigger scale, as r increases the

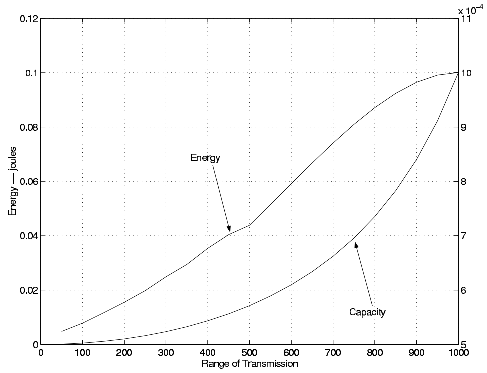

Fig. 7. Flat network, R=1000

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 (see Fig. 7). This observation is important because it is generally accepted that smaller hops are better than large ones when it comes to energy consumption. What this shows is that the answer depends on the scale of the network.

B. In-networking Processing

Our capacity results essentially showed that the many-to-one communication is even less scalable than the random communication case as reported in [5], while many-to-one is the dominant communication in a data-gathering sensor network. This means that for certain applications, the capacity needed may severely limit the feasible size of the network in terms of number of sensors deployed. To make large-scale wireless sensor network a possibility, the need for methods that can reduce the number of bits needed from each sensor has long been recognized and has motivated research in distributed data compression and source coding, distributed signal processing, or more generally, in-network processing of sensor data (see for example [17]). The idea is to increase the amount of data processing within the network so that the total amount of traffic is reduced. In some sense this reflects the need for a dynamic relationship between the amount of processing vs. the amount of communication required to accomplish a task.

In the case of data-gathering, e.g., for the purpose of image reconstruction of the sensing field, scalability may be improved by exploiting the correlation among data from neighboring sensors as we deploy more and more sensors into the field. A sensor can either use distributed data compression schemes or simply compress its own data based on data received from other sensors (to be relayed). This way the amount this sensor transmits is reduced. As data traverse the network via multiple sensors, more data gets compressed. If lossless compression is used, such a scheme will result in sending the same amount of information but with a lower number of bits. This means that the transport capability may be sufficient to transmit the needed information. It also means that less energy are consumed in communication, and likely less energy consumption overall (energy consumed in communication is generally much more than energy consumed in processing). Viewed from a different perspective, compression can also be used to enhance the quality of received data. To see this imagine a given network with a transport capacity of x bits per second per sensor. This throughput capacity allows us to use a quantizer with a certain step size. Now suppose that the same network uses data compression within the network. This will allow the nodes to transmit the same amount of information with less number of bits. It follows then that with the same number of bits each sensor can now deliver data with a smaller quantization error.

It's worth mentioning that in a recent work [18] it was shown that by using source coding the network transport capacity may be sufficient (even in a all-to-all type of communication). However, this result is based on the assumption that the entropy of sampled data from all sensors is on the same order of the rate distortion as the number of sensors goes to infinity. To the best of our knowledge it is not known whether this assumption does hold in the limit.

The above discussion and our results on transport capacity and energy consumption derived in this paper provide further motivation for the exploration of efficient in-network processing methods, which is part of our on-going research.

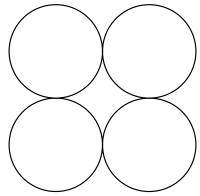

Fig. 8. Non-overlapping Receiving Areas

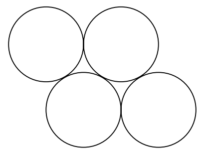

Fig. 9. Non-overlapping Receiving Areas – shifted

C. Practical Implications

In addition to the theoretical contribution of results presented in this paper, there are also practical implications that can be obtained from these results. Our ultimate goal is to apply the understanding of fundamental limits in the design of practical sensor networks. In this section we point out some of the practical issues that arise as a consequence of the results of the previous sections.

First we examine the choice of an efficient MAC scheme. In constructing the two lower bounds on transport capacity we have clearly shown that in the case of many-to-one communication higher per sensor throughput is achieved by having a traffic load-aware MAC. In other words, we need a MAC scheme that will allocate resources proportional to the amount of communication each sensor has to perform. In general contention based MAC may result in resource allocations that reflect communication needs. However, due to collision the actual achievable throughput may be significantly reduced. The development of such a scheme and its distributed implementation is subject to further study. Potential candidates include method proposed in [14].

A second important implication is the organization of a sensor network, i.e., flat vs. hierarchical. As we have seen from our previous discussion, there is no general answer to this. A hierarchical structure can easily achieve the throughput capacity at the expense of more powerful sensors serving as cluster heads. At the same time, this may result in higher energy consumption especially for a network covering a larger area. When clustering is used, the proper placement of cluster heads is of great importance. Throughout the paper 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 allow us to create a more even distribution via selected cluster heads.

VI. CONCLUSION

In this paper we have studied the capacity and scalability issues related to many-to-one communication in a data-gathering wireless sensor network. We showed that overall the transport capacity of such a network is $\Theta(\frac{W}{n})$ per sensor node. We derived an upper bound as well as constructive lower bounds for both the flat and the hierarchical network architecture. Through constructing the achievable lower bounds on capacity we showed that knowledge of the traffic load can double the achievable capacity of a network with multi-hop communications. Using a hierarchical architecture and introducing extra nodes as cluster heads can achieve the ultimate upper bound on throughput capacity. Moreover, the number of clusters needed to reach the capacity of the network is independent of n. Placing a second layer of nodes with higher transmission rate and using clustering can exceed the capacity of the flat network.

APPENDIX

Adopting the same notations we proceed as follows. Any arrangement that minimizes the uncovered area would have the circles making contact with each other. Also such an arrangement would be a regular arrangement since an irregular arrangement would mean that at some parts of the network the uncovered area is bigger than in others, meaning that a

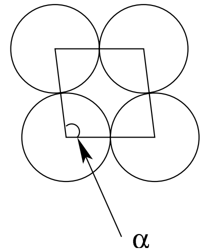

Fig. 10. Uncovered Area

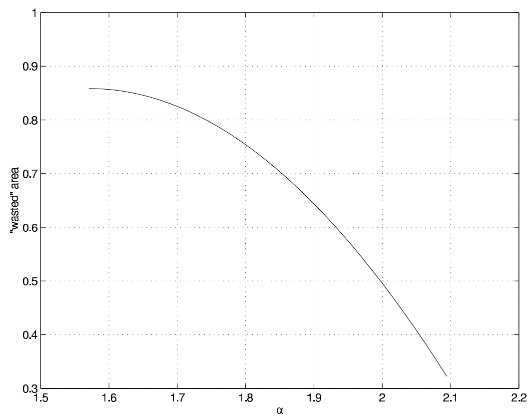

Fig. 11. Y vs. α

better arrangement exists. Based on this consider the arrangement seen in Fig. 8. We have the circles aligned in rows, one on top of the other. Fixing the circles in the top row and shifting the circles in the bottom row we get to the position shown in Fig. 9.

Clearly the uncovered area is minimized in one of the arrangements above or one in between. We create imaginary lines that join the center of the circles as shown in Fig. 10. The sliding of the lower circles can be represented as the change in angle α. Note that $\frac{\pi}{2} \leq \alpha \leq \frac{2\pi}{3}$.

Let Y be the area surrounded by the circles, as a function of α. Simple geometry yields:

$$Y = 8r^2 \left(\cos \frac{a}{2} \sin \frac{a}{2} \right) - \pi r^2 \tag{33}$$

This is a monotonously decreasing function in α and Y is minimized at $\alpha = \frac{2\pi}{3}$, as shown in Fig. 11. Note that a higher α would force the overlapping of circles.

Using such an α, a given circle is surrounded as shown in figure 12. A circle thus "contributes" to six uncovered spaces. Each of those six areas measures $r^2 \left(\sqrt{3} - \frac{\pi}{2} \right)$. Each uncovered space is shared by three circles, thus

$$\mu_m = \frac{6}{3} \cdot \left(r^2 \left(\sqrt{3} - \frac{\pi}{2} \right) \right) = r^2 \left(2\sqrt{3} - \pi \right). \tag{34}$$

Using our approximation of μ_m in Theorem 2 we get:

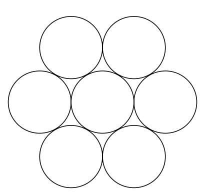

Fig. 12. Minimum Uncovered Area

$$\lambda \leq \frac{W}{1.014n}. \tag{35}$$

REFERENCES

[1] A. Manjeshwar and D. P. Agrawal, "Teen: A routing protocol for enhanced efficiency in wireless sensor networks," in *Proceedings of the 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing in Conjunction with 2001 IPDPS*, 2001.

[2] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration for tracking applications," *IEEE Signal Processing Magazine*, March 2002.

[3] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the Hawaii International Conference on System Sciences*, January 2000.

[4] L. Subramanian and R. H. Katz, "An architecture for building self-configurable systems," in *IEEE/ACM Workshop on Mobile Ad Hoc Networking and Computing (MobiHoc 2000)*, 2000.

[5] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, March 2000.

[6] J. Li, C. Blake, D. S. J. DeCouto, H. I. Lee, and R. Morris, "Capacity of ad hoc wireless networks," in *Mobicom*, 2001.

[7] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *Proceedings of the IEEE Infocom*, 2001.

[8] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proceedings of the IEEE Infocom*, 2002.

[9] P. Gupta and P.R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Application: A Volume in Honor of W. H. Fleming*, March 1998.

[10] F. Xue and P.R. Kumar, "The number of neighbors needed for connectivity of wireless networks," *submitted for publication*, April 2002.

[11] E. J. Duarte-Melo and M. Liu, "Analysis of energy consumption and lifetime of heterogeneous wireless sensor networks," in *IEEE Globecom*, 2002.

[12] G. Chartrand, *Introductory Graph Theory*, Dover Publications, INC, 1985.

[13] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, American Elsevier Publishing CO., INC, 1976.

[14] A. Woo and D. E. Culler, "A transmission control scheme for media access in sensor networks," in *Mobicom*, 2001.

[15] A. Okabe, B. Boots, and K. Sugihara, *Spatial Tessellation Concepts and Applications of Voronoi Diagrams*, New York: Wiley, 1992.

[16] E. J. Duarte-Melo and M. Liu, "Energy efficiency of many-to-one communications in wireless networks," in *IEEE Midwest Symposium on Circuits and Systems*, August 2002.

[17] S. S. Pradhan and K. Ramchandran, "Distributed source coding: Symmetric rates and applications to sensor networks," in *IEEE Data Compression Conference (DCC)*, March 2000.

[18] A. Scaglione and S. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," in *Mobicom*, September 2002.