An Information-Theoretic View of Connectivity in Wireless Sensor Networks

Xin Liu
Computer Science Department
University of California, Davis
Email: liu@cs.ucdavis.edu
R. Srikant
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Email: rsrikant@uiuc.edu

***Abstract*— In this paper, we study the connectivity properties of a wireless sensor network from an information-theoretic viewpoint. We consider both regular linear (one-dimensional) and planar (two-dimensional) networks with unreliable sensor nodes, i.e., each node is inactive/dead with a certain probability. We study the following problems: 1) What is the fundamental limit on the data rate that such a network can support for a single sensor node to the destination under transmission power and network-topology constraints? 2) What are the constraints on the network topology such that any (single) sensor can communicate with the destination at a desired rate? For problem 1), we provide upper and lower bounds on the achievable data rate, and for problem 2), we provide upper and lower bounds on the distance between the nodes required for communication at the desired rate.**

I. Introduction

Recently, large wireless networks have attracted a great deal of research interest [1], [2], [3], [4]. Problems that have been studied include the capacity and scaling law of a large wireless network [5], [6], [7], [8], [9], [10], and its coverage and connectivity properties [11], [12], [13], [14], [15], [16], [17]. A wireless sensor network is a good example of large wireless networks with practical importance.

In this paper, we focus on the connectivity problem of a large wireless sensor network. A widely-used model to study connectivity based on transmission radius is as follows: A node can communicate (directly) with any other node within its transmission/communication radius r (if there is no other transmissions near the receiver). In other words, if the distance between two nodes is large, then these two nodes cannot communicate (directly) with each other. Two nodes are connected if there exists a sequence of relay nodes between them such that the distance at each hop in the route is no larger than r. Another model is based on signal-to-interference-plus-noise ratio (SINR), i.e., if SINR at the receiver is above some threshold, then reliable transmission occurs. Connectivity properties are studied under such models.

In [15], the authors show that $\pi r^2(n)D \sim \log A$ is a necessary and suf£cient condition for coverage and a necessary condition for connectivity in a random network where the nodes are located in a region of area A, according to a Poisson process with density D, and $r(n)$ is the radius of communication.

The authors of [11] consider a unit area with n randomly located nodes. They show that each node should be able to communicate with nodes within a circle of area $\pi r^2(n) = (\log n + c(n))/n$ to guarantee the asymptotic connectivity of a large ad-hoc network if and only if $c(n) \to \infty$.

In [14], the authors study the connectivity and coverage of a regular sensor grid with unreliable nodes in a unit square region. They show that the necessary and suf£cient conditions for the network to cover the region while all active nodes are connected are of the form $p(n)r^2(n) \sim \log(n)/n$, where two nodes within distance $r(n)$ can communicate with each other and $p(n)$ is the probability that a node is active.

It is shown in [17] that each node should be connected to $\Theta(\log n)$ neighbors in order to guarantee asymptotic connectivity for a random ad-hoc network.

Percolation theory has been used to study the connectivity of wireless networks in [12], [13], [15] and references therein. In [12], the authors study both pure ad-hoc and hybrid networks and show that the introduction of a sparse network of bases stations can signi£cantly improve the connectivity. In [13], the authors study the case where other transmissions cause interference, and show that the interference coef£cient is an important factor in connectivity.

A. System Model and Problem Description

While the model based on communication radius or SINR may be able to approximate the operation of current wireless networks, such a model model is overly simpli£ed. In fact, even if two nodes are very far apart, they may still be able to communicate reliably, albeit at a lower data rate. Thus, a more fundamental question is *the rate* at which two nodes can reliably communicate under transmission-power and network-topology constraints.

In addition, the multi-hop operation used in current wireless networks is just one way of communication. When we consider a network of radio nodes, there is a variety of possibilities, as mentioned in [5]. For instance, a (group of) node(s) can help the transmission from the source to the destination by cancelling the interference from other nodes at the destination, and another (group of) node(s) can transmit coherently to boost the desired signal strength at the receiver [18].

To further clarify the idea, let us consider a simple network of three radio nodes, as shown in Figure 1. In the £gure, node S is the source, node D is the destination, and there exists a node R in the vicinity of S. If we consider the transmission radius model, node R cannot help node S to reach node D since it is not closer to D than S. If we consider the interference model, node R's transmission is considered as interference at D, which is "bad" in the model. However, node R can actually help (a lot). Assume that node R is close to node S and can obtain the information that node S wants to communicate to node D easily. Then, node R can transmit its signal coherently to enhance the received signal at D. Assume the transmission power constraint on each node is P and the signal attenuation from S to D and R to D are both α. The signal strength of S and R are both $\sqrt{P}$. In the best case, the signals from S and R combine coherently at the receiver D. Thus, the received signal strength is $2\alpha\sqrt{P}$, and the received power is $4\alpha^2 P$ (as compared with $\alpha^2 P$ if node R does not help, and $2\alpha^2 P$ in a non-coherent case with a total transmission power constraint $2P$). Under the same noise level, the higher the received power, the higher the data rate. Thus, the existence of node R can signi£cantly help the transmission from S to D [18]. This phenomena is not captured in radius-based or SINR-based models, but is captured in information-theoretic models such as the ones in [5], [10]. Note that in a sensor network, especially a randomly deployed one, certain sensor may be close to each other. Coherent communications are possible at least in theory, and thus we do not want to exclude such possibilities.

In this paper, we consider connectivity from an information-theoretic viewpoint. In other words, we do not assume models such as transmission radius, and do not limit the possibilities of collaboration among a network of radio nodes. Instead, we try to understand the ultimate limits on the information transfer rate from a source to the destination in a large network with unreliable nodes. Under such a context, we de£ne connectivity as follows:

The network is connected at rate $R(n)$ if, for a randomly picked source-destination pair, we can guarantee that the source can communicate with the destination at rate $R(n)$, assuming all other nodes act as relay nodes.

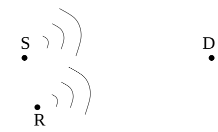

Fig. 1. An example of network cooperation

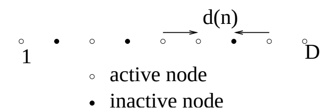

Fig. 2. A regular linear network

In contrast to the traditional de£nition of connectivity, in this paper, connectivity is explicitly associated with a rate which is feasible for any single pair of source and destination. Of course, there may exist source-destination pairs that are in more favorable conditions and could transmit reliably a rate higher than $R(n)$. Consider a network of sensors which collect data and transmit the data to a central location, named the fusion center. If the sensor network is connected at rate $R(n)$, then any single sensor node can communicate with the fusion center at rate $R(n)$.

We consider both regular linear and planar networks as shown in Figure 2 and Figure 3, respectively. There are n nodes in the network. Each node has an average power constraint P_{ind}. The distance between two adjacent nodes is $d(n)$, which can be a function of n. We consider an unreliable sensor network, where a node is inactive with a certain probability, as in [14]. A node may become inactive or dead if it consumes all its power or if it shuts itself off to conserve power. In a practical network, as time passes, the failure probability of a node may increase, which corresponds to the decrease of active probability.

In this paper, we answer the following questions.

- Suppose that the distance between two adjacent sensor nodes is £xed, i.e., $d(n) = d_0$. *At what data rate* can we guarantee that any single sensor node can communicate reliably with the fusion center

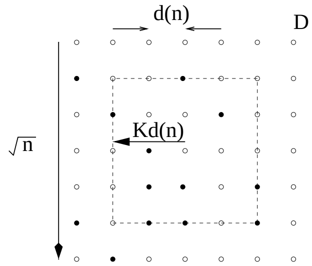

Fig. 3. A regular planar network

with all the other active sensor nodes acting as relays?

- Suppose that we desire a £xed data rate from any active sensor node to the fusion center. With n sensor nodes, how large an area can we "cover", i.e., what is the maximum distance between two adjacent nodes such that the rate constraint is sat-is£ed?

As we will see, the two problems are very closely related. In the paper, we primarily discuss the £rst problem. The solution to the £rst problem automatically provides an answer to the second problem, which is summarized in Section V. In this paper, we study an unreliable network with uniform guarantees for any single sensor node, which is different from [5]. Compared with [10], our network is not limited to a unit area. The distance between nodes plays an importance role in the achievable rate, and thus the results in [10] does not imply our results.

The connectivity result provides a guarantee for the achievable rate under infrequent communication scenarios. For certain sensor-network applications, it is reasonable to expect the communication between sensor nodes and the fusion center to be infrequent. For example, consider a sensor network which is used to monitor faults in a large structure, such as a building or a bridge. A node may only communicate with the fusion center when it detects, say a possible fault, and needs to send the information to the fusion center. In addition, the connectivity result also provides an upper bound on the guaranteed communication rate when there are multiple source-destination pairs transmitting simultaneously.

We use the following notation in the rest of the paper. We assume that a transmitted signal is attenuated by a factor of $1/d^{\delta}$ over a distance d. The attenuation model for very small d is discussed in Section V. Let d_{ij} be the distance between node i and node j. Let p be the probability that a sensor node is active. A node is active independently with probability p. We consider the case where $0 < p \leq 1$ since the case where $p = 0$ is trivial. We consider a white Gaussian noise channel where σ^2 is the variance of the white noise. Let $R(n)$ be the data rate for any single active source to the destination and $P_d(n)$ be the probability of disconnection with respect to rate $R(n)$, i.e., there exists at least one active node that cannot transmit to the destination at rate $R(n)$. The data rate is a function of communication schemes: transmissions from other nodes are harmful if considered as interferences, helpful if acted as coherent relay, or can be cancelled if interference cancellation is preformed. However, as stated later, the data rate is upper-bounded by the ratio of the total received power vs. noise, which provides an upper-bound on the achievable rate *uniformly*. On the other hand, a speci£c interference cancellation transmission scheme can be used to achieve a lower bound uniformly.

We £rst consider the case where $d(n) = d_0$, i.e., the distance between two adjacent nodes is £xed, and study how $R(n)$ scales with n. Then, we use this result to obtain a bound on $d(n)$ if $R(n)$ is required to be no smaller than a £xed constant rate R_{req}.

B. Organization of the Paper

The paper is organized as following. In Section II, we consider a regular linear network where $0 < p < 1$. We provide upper and lower bounds on the achievable data rate when $d(n) = d_0$. In Section III, we consider the case of a regular planar network with $0 < p < 1$. In Section IV, we study the special case where $p = 1$. In Section V, we provide bounds on $d(n)$ when a minimum rate is desired, and conclude the paper in Section VII.

II. Linear Network

Consider a linear network with n nodes where the distance between two adjacent nodes is d_0. The fusion center locates at one end of the network, as shown in Figure 2. We assume that the fusion center is always "on". We show upper and lower bounds on $R(n)$ such that any active sensor node can communicate with the fusion center at rate $R(n)$, provided that all other active nodes act as relays.

Proposition 1: If $\delta > 1/2$, and

$$R(n) \geq \frac{c_1}{d_0^{2\delta}(\log n)^{2\delta-1}},$$

then

$$\lim_{n\to\infty} P_d(n) \geq p - p^2,$$

where

$$c_1 = \frac{(1+\epsilon_0)P_{ind}\left(2\log\frac{1}{1-p}\right)^{2\delta-1}}{(2\delta-1)\sigma^2},$$

and $\epsilon_0 > 0$.

This proposition provides an upper bound on the achievable data rate for all users. In other words, we cannot guarantee that each active node can access the sink node if the desired data rate is larger that the one presented above. Note that it is possible that some nodes can achieve this data rate. The proposition simply claims that this rate cannot be guaranteed for all nodes.

The heuristics of the proof is as follows: Lemma 1 in [5] shows that the achievable data rate is upper bounded by the total received power. Then Lemma 2 shows that as n grows, with a high probability that there exists an isolated node (a node that is far away from other nodes) in the network. Therefore, the total received power at the receiver side is bounded (as a function of the distance of the isolated node to other nodes), and thus limits the *uniform* achievable data rate. Note that in a practical network, such "isolated" nodes may be more important in terms of data gathering. For instance, in a surveillance network where sensors detect its environment, such an "isolated" node may be responsible for a larger area. The events detected by this node may be less likely to be detected by other nodes, and/or the data gathered by this node may be less correlated with data gathered by others. Thus, data transmission of such nodes is not indifferent.

We £rst present the max-¤ow min-cut lemma in [5].

Lemma 1: Let $N_1 \subset N$. A source-destination pair (s_l, d_l) is said to cut N_1 if $d_l \in N_1$ and $s_l \notin N_1$. If $(R_1, \cdots, R_m)$ is a feasible rate vector with sequence of $((2^{TR_1}, \cdots, 2^{TR_m}), T, \lambda_T)$ codes[1] with $\lambda_T \to 0$ as $T \to \infty$, then

$$\sum_{l: d_l \in N_1, s_l \notin N_1} R_l \leq \frac{1}{2\sigma^2} \liminf_{T\to\infty} P_{N_1}^{rec}(T),$$

where $P_{N_1}^{rec}(T)$ is the average power received by N_1, from outside N_1, for code $((2^{TR_1}, \cdots, 2^{TR_m}), T, \lambda_T)$, i.e.,

$$P_{N_1}^{rec}(T) := \frac{1}{T}\sum_{t=1}^{T}\sum_{i\in N_1} E\left(\sum_{j\notin N_1}\frac{X_j(t)}{d_{ij}^{\delta}}\right)^2 .$$

■

The lemma provides an upperbound on the data rate, which is limited by the total received power on average. An analogy is that many people talk to one person at the same time. They can say the same words coherently or say different words individually. The information that can be received by the listener is bounded by the total amplitude of the voices.

Lemma 2: Let $\delta > 1/2$ and $\epsilon_0 > 0$. In a linear network, an active node cannot communicate with rate $R(n)$ to the destination if there are no active neighbors within a radius of Kd_0, where $K \geq 1$ and K satis£es

$$d_0^{2\delta} K^{2\delta-1} \geq \frac{P_{ind}(1+\epsilon_0)}{(2\delta-1)R(n)\sigma^2}. \tag{1}$$

Proof: Consider an active node i and apply Lemma 1. Let $N_1^c = \{i\}$, i.e., $N_1 = \{1, 2, \cdots, i-1, i+1, \cdots, n\}$. The maximum rate at which node i can transmit to the destination, $R_i(n)$, is bounded by the total average received power at all the other nodes, i.e.,

$$R_i(n) \leq \frac{1}{2\sigma^2}\liminf_{T\to\infty} P_{N_1}^{rec}(T),$$

If within radius Kd_0, there are no active neighbors, i.e., nodes $i-K$ to $i+K$ are inactive, then

$$\begin{aligned}
P_{N_1}^{rec}(T) &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{k=K+1}^{\infty} 2E\left(\frac{X_i(t)}{(kd_0)^{\delta}}\right)^2 \\
&\overset{(a)}{\leq} \sum_{k=K+1}^{\infty}\frac{2P_{ind}}{(kd_0)^{2\delta}} \\
&= \frac{2P_{ind}}{d_0^{2\delta}}\sum_{k=K+1}^{\infty}\frac{1}{k^{2\delta}} \\
&\leq \frac{2P_{ind}}{d_0^{2\delta}}\int_K^{\infty}\frac{1}{x^{2\delta}}dx \\
&\overset{(b)}{=} \frac{2P_{ind}}{d_0^{2\delta}(2\delta-1)K^{2\delta-1}},
\end{aligned}$$

where (a) holds because of the average power constraint for each node i,

$$\frac{1}{T}\sum_{t=1}^{T} X_i^2(t) \leq P_{ind},$$

and (b) holds when $\delta > 1/2$. By the hypothesis in (1), we have

$$P_{N_1}^{rec}(T) \leq \frac{2R(n)\sigma^2}{1+\epsilon_0},$$

and thus

$$R_i(n) \leq \frac{R(n)}{1+\epsilon_0}.$$

■

Proof of Prop. 1: We say that a node is *isolated* if it is active and there are no active nodes within radius Kd_0, i.e., the K nearest neighbors on both sides are inactive. Consider the case where the condition in (1) is satis£ed. Thus, if an active node is isolated, then the linear network is disconnected with respect to rate $R(n)$. Recall that $P_d(n)$ is the probability that the linear network is

[1] A $((M_1, \cdots, M_m), T, \lambda_T)$ code refers to a code with M_i codewords between source-destination pair i, with the length of each codeword being T and the probability of decoding error being less than or equal to λ_T.

disconnected with respect to rate $R(n)$, i.e., there exists at least one active node that cannot communicate with the destination at rate $R(n)$. We have

$$
\begin{aligned}
P_d(n) &\geq Pr\{\text{there exists an isolated node}\} \\
&\overset{(a)}{\geq} \sum_{i=3K+1}^{n-K} Pr\{\text{only node } i \text{ is isolated}\} \\
&\geq \sum_{i=3K+1}^{n-K} Pr\{\text{node } i \text{ is isolated}\} \\
&\quad - \sum_{j\neq i} Pr\{\text{nodes } i \text{ and } j \text{ are isolated}\}.
\end{aligned}
$$

In inequality (a), we consider nodes that are at least K hops from node n and $3K$ hops away from the origin. Note that if a node is within K hops from node n, then this node cannot be isolated by the de£nition. On the other hand, we let i be far away from the origin to avoid boundary effects. We have

$$Pr\{\text{node } i \text{ is isolated}\} = p(1-p)^{2K}.$$

Next, we consider the term $Pr\{$ nodes i and j are isolated$\}$. If $K \leq j \leq n-K$, then

$$
\begin{aligned}
&Pr\{\text{nodes } i \text{ and } j \text{ are isolated}\} \\
&= \begin{cases} 0 & \text{if } |i-j| \leq K \\ p^2(1-p)^{4K} & \text{if } |i-j| > 2K \\ p^2(1-p)^{2K+|i-j|} & \text{else} \end{cases}
\end{aligned}
$$

If $j \leq K$, then $|i-j| \geq 2K$. We have

$$Pr\{\text{nodes } i \text{ and } j \text{ are isolated}\} \leq p^2(1-p)^{3K}.$$

If $j \geq n-K$, since j cannot be isolated, we have

$$Pr\{\text{nodes } i \text{ and } j \text{ are isolated}\} = 0.$$

Thus, we have

$$P_d(n) \geq L_1(n) - L_2(n),$$

where

$$
\begin{aligned}
L_1(n) &= (n-4K)p(1-p)^{2K} \\
L_2(n) &= (n-4K)[3Kp^2(1-p)^{3K} \\
&\quad +(n-2K)p^2(1-p)^{4K}].
\end{aligned}
$$

Let

$$2K = \frac{\log n}{-\log(1-p)}. \qquad (2)$$

We have

$$
\begin{aligned}
\log L_1(n) &= \log(n-4K) + 2K\log(1-p) + \log p \\
&= \log(n-4K) + \frac{\log n}{-\log(1-p)}\log(1-p) \\
&\quad + \log p,
\end{aligned}
$$

and thus

$$\lim_{n\to\infty} L_1(n) = p.$$

Similarly,

$$\lim_{n\to\infty} L_2(n) = p^2.$$

Thus, we have

$$\lim_{n\to\infty} P_d(n) \geq p - p^2.$$

Substituting (2) into (1), we have that if

$$R(n) \geq \frac{P_{ind}(1+\epsilon_0)}{(2\delta-1)d_0^{2\delta}\sigma^2\left(\frac{\log n}{-2\log(1-p)}\right)^{2\delta-1}},$$

then

$$\lim_{n\to\infty} P_d(n) \geq p - p^2,$$

which completes the proof. ■

Next, we present a suf£cient condition on $R(n)$ such that the network is connected at rate $R(n)$. The proof is constructive. We develop a strategy such that an active node can communicate with the fusion center at rate $R(n)$. The basic idea is to divide the line segment between node 1 and node n into intervals of length $\rho(n)$, and select one node in each interval to relay packets. We need to show the following: 1) with probability 1 there is at least one active node in each interval of length $\rho(n)$, and 2) $R(n)$ can be supported. Lemma 3 provides a suf£cient condition on $\rho(n)$ for each interval to have at least one active node and Lemma 5 indicates how to relay packets to achieve rate $R(n)$.

Lemma 3: If

$$x(n) \leq \frac{-\log(1-p)}{\log n},$$

then

$$\lim_{n\to\infty} P_x(n) = 1,$$

where $P_x(n)$ is de£ned as

$$P_x(n) = \left(1-(1-p)^{\frac{1}{x(n)}}\right)^{nx(n)}.$$

Proof: We have

$$
\begin{aligned}
P_x(n) &= \left(1-(1-p)^{\frac{1}{x(n)}}\right)^{nx(n)} \\
&= (1-(1-p)^{\frac{1}{x(n)}})^{(1-p)^{-\frac{1}{x(n)}}nx(n)(1-p)^{\frac{1}{x(n)}}}.
\end{aligned}
$$

For small enough y, we have $(1-y)^{1/y} \geq e^{-1-\epsilon}$, where $\epsilon > 0$. Thus, for large enough n,

$$P_x(n) \geq e^{(-1-\epsilon)nx(n)(1-p)^{\frac{1}{x(n)}}}.$$

Further,

$$\begin{aligned} & \log\left(nx(n)(1-p)^{\frac{1}{x(n)}}\right) \\ = \; & \log n + \frac{1}{x(n)}\log(1-p) + \log(x(n)) \\ \leq \; & \log n + \frac{\log n}{-\log(1-p)}\log(1-p) + \log(x(n)) \\ = \; & \log(x(n)). \end{aligned}$$

Thus,

$$\lim_{n\to\infty} \log\left(nx(n)(1-p)^{\frac{1}{x(n)}}\right) = -\infty,$$

and

$$\lim_{n\to\infty} nx(n)(1-p)^{\frac{1}{x(n)}} = 0.$$

Thus,

$$\lim_{n\to\infty} P_x(n) = 1.$$

■

The following lemma is presented in [5], where a coherent-relay-and-interference-cancellation scheme is used. Recall that d_{ij} is the distance between node i and node j.

Lemma 4: Consider the Gaussian multiple relay channel with coherent multi-stage relaying and interference cancellation. Consider $M+1$ nodes, sequentially denoted by $0, 1, \cdots, M$, with 0 be the source, M as the destination, and the other $M-1$ nodes serving as $M-1$ stages of relay. Then any rate R satisfying the following inequality is achievable from 0 to M:

$$R < \min_{1\leq j\leq M} \frac{1}{2}\log\left(1 + \frac{1}{\sigma^2}\left(\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P_{ik}}}{d_{ij}^{\delta}}\right)^2\right)\right)$$

where $P_{ik} \geq 0$ satis£es $\sum_{k=i+1}^{M} P_{ik} \leq P_{ind}$.

Consider a linear relay network where the maximum distance between two adjacent (active) nodes is $2\rho(n)$. The following lemma presents a suf£cient condition for the achievable rate from a single source to the destination.

Lemma 5: Suppose that the maximum distance between two nodes is $2\rho(n)$. If

$$\rho(n) < c_d(n),$$

where

$$c_d(n) = \frac{1}{2}\left(\frac{P_{ind}}{\sigma^2(e^{2R(n)}-1)}\right)^{\frac{1}{2\delta}} \tag{3}$$

then the rate $R(n)$ is achievable from any active node to the destination.

Proof: Let the source be 0 and the destination be M. Let $P_{i,i+1} = P_{ind}$ and $P_{ij} = 0$ for all $j \neq i+1$. We have

$$\begin{aligned} & \min_{1\leq j\leq M} \frac{1}{2}\log\left(1 + \frac{1}{\sigma^2}\left(\sum_{k=1}^{j}\left(\sum_{i=0}^{k-1}\frac{\sqrt{P_{ik}}}{d_{ij}^{\delta}}\right)^2\right)\right) \\ & \geq \min_{1\leq j\leq M} \frac{1}{2}\log\left(1 + \frac{1}{\sigma^2}\left(\sum_{k=1}^{j}\frac{P_{k-1,k}}{d_{k-1,j}{}^{2\delta}}\right)\right) \\ & \geq \min_{1\leq j\leq M} \frac{1}{2}\log\left(1 + \frac{P_{j-1,j}}{\sigma^2 d_{j-1,j}{}^{2\delta}}\right) \\ & \geq \min_{1\leq j\leq M} \frac{1}{2}\log\left(1 + \frac{P_{ind}}{\sigma^2(2\rho(n))^{2\delta}}\right) \qquad (4) \\ & \overset{(a)}{>} \min_{1\leq j\leq M} R(n) \\ & = R(n) \end{aligned}$$

where (a) holds by the hypothesis $\rho(n) < c_d(n)$. Thus, by Lemma 4, $R(n)$ is an achievable rate. ■

To achieve this data rate, the operation is interference cancellation. To be more speci£c, at each stage of relay, the relay node can fully decode the information from the source, and thus it knows all the transmissions in the downstream (i.e., the transmission from j to $j+1$ for all $j \geq i+1$). Hence, it can remove the interference caused by downstream transmissions from the received signal. On the other hand, all upstream transmissions (i.e., the transmission from j to $j+1$ for all $j < i-1$.) can be helpful. However, because the £rst relay node has no upstream transmission, it gives a lower-bound on the rate. This operation is different from the operation mode of current wireless networks, where all other transmissions (upstream and downstream) are considered as interference and no interference cancellation is performed. If we follow the current transmission mode, the achievable rate is lower than the one presented in Prop. 2, but on the same order.

Proposition 2: If

$$R(n) \leq \frac{1}{2}\log\left(1 + \frac{c_2}{(\log n)^{2\delta}}\right),$$

where

$$c_2 = \frac{P_{ind}}{\sigma^2}\left(\frac{(1-\epsilon_0)\log\frac{1}{1-p}}{2d_0}\right)^{2\delta}$$

and $\epsilon_0 > 0$, then the system can support rate $R(n)$ with probability 1 as $n \to \infty$.

Proof: The proof is constructive. Let $\rho(n) = (1-\epsilon_0)c_d(n)$. As explained earlier, the linear network is divided into $nd_0/\rho(n)$ intervals, with each interval having $\rho(n)/d_0$ nodes. The length of each interval is $\rho(n)$. We need to show the following: 1) with probability 1 there is at least one active sensor node in each interval, 2) $R(n)$ can be supported if we select one active node in each interval as a relay node.

We £rst show that with probability 1 that each interval has at least one active node.

$$\begin{aligned} & P_0(n) \\ = \;& Pr\{\text{each interval has at least one active node}\} \\ = \;& (1-(1-p)^{\rho(n)/d_0})^{nd_0/\rho(n)}. \end{aligned}$$

Let $x(n) = d_0/\rho(n)$. By the hypothesis, we have

$$e^{2R(n)} - 1 \leq \frac{c_2}{(\log n)^{2\delta}}.$$

We have

$$\begin{aligned} \rho(n) &= (1-\epsilon_0)c_d(n) = \frac{1-\epsilon_0}{2}\left(\frac{P_{ind}}{\sigma^2(e^{2R(n)}-1)}\right)^{\frac{1}{2\delta}} \\ &\geq \frac{1-\epsilon_0}{2}\left(\frac{P_{ind}}{\sigma^2\left(\frac{c_2}{(\log n)^{2\delta}}\right)}\right)^{\frac{1}{2\delta}} \\ &= \frac{1-\epsilon_0}{2}\frac{2d_0\log n}{(1-\epsilon_0)\log\frac{1}{1-p}} \\ &= \frac{d_0\log n}{\log\frac{1}{1-p}}. \end{aligned}$$

Thus,

$$x(n) = \frac{d_0}{\rho(n)} \leq \frac{-\log(1-p)}{\log n},$$

and

$$\lim_{n\to\infty} P_0(n) = 1$$

by Lemma 3.

Next, we show that $R(n)$ is achievable from any active node to the destination. Consider an active node. Since the previous result guarantees that there is at least one active node within each interval between the active node and the destination, we can randomly pick an active node in each interval as the relay node mentioned in Lemma 4. Because the length of each interval is $\rho(n)$, the maximum distance between two adjacent relay nodes is $2\rho(n)$. Lemma 4 guarantees that $R(n)$ can be supported because $\rho(n) < c_d(n)$. ■

III. Planar Network

The results for the planar network are similar. The distance between two adjacent nodes is d_0. The fusion center is located at the northeast corner of the regular grid, as shown in Figure 3. We present a necessary condition and a suf£cient condition on the rate $R(n)$ for the network to support rate $R(n)$ from any active sensor node to the fusion center with all other sensor nodes acting as relays. Because the ideas of the proofs here are similar to the proofs in the linear case, we omit it here and refer readers to [19] for the complete proof.

Proposition 3: If $\delta > 1$, and

$$R(n) \geq \frac{c_3}{(\log n)^{\delta-1}d_0^{2\delta}},$$

where

$$c_3 = \frac{2P_{ind}(1+\epsilon_0)\left(4\log\frac{1}{1-p}\right)^{\delta-1}}{(\delta-1)\sigma^2}$$

and $\epsilon_0 > 0$, then

$$\lim_{n\to\infty} P_d(n) \geq p - p^2.$$

Proposition 4: If

$$R(n) \leq \frac{1}{2}\log\left(1+\frac{c_4}{(\log n)^{\delta}}\right),$$

where

$$c_4 = \frac{P_{ind}}{\sigma^2}\left(\frac{(1-\epsilon_0)\log\frac{1}{1-p}}{4d_0^2}\right)^{\delta},$$

and $\epsilon_0 > 0$, then the system can support rate $R(n)$ with probability 1 as $n \to \infty$.

IV. The case $p = 1$

Our results in the previous sections assume that $0 < p < 1$. In this section, we consider the special case where all nodes are always active, i.e., $p = 1$, and $d(n) = d_0$. In this case, we show that for both linear and planar networks, the data rate that can be supported from any single node to the destination is $\Theta(1)$.

Proposition 5: For a regular linear network, if $\delta > 1/2$, $\epsilon_0 > 0$, and

$$R_{req} \geq \frac{2(1+\epsilon_0)P_{ind}\delta}{\sigma^2 d_0^{2\delta}(2\delta-1)},$$

then no active user can communicate with the destination node at rate R_{req}.

Proof: The proof is similar to that of Lemma 2. For any node i, let $N_1 = \{i\}^c$. We have

$$\begin{aligned} P_{N_1}^{rec}(T) &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{\infty} 2E\left(\frac{X_i(t)}{(kd_0)^{\delta}}\right)^2 \\ &\leq \frac{4P_{ind}\delta}{d_0^{2\delta}(2\delta-1)} \\ &\leq \frac{2\sigma^2 R_{req}}{1+\epsilon_0}. \end{aligned}$$

Thus, $R_i < R_{req}$ for all i by Lemma 1, and thus completes the proof. ■

Proposition 6: For a regular planar network, if $\delta > 1$, $\epsilon_0 > 0$, and

$$R_{req} \geq \frac{2(1+\epsilon_0)P_{ind}(2\delta-1)}{d_0^{2\delta}\sigma^2(\delta-1)},$$

then no active user can communicate with the destination node at rate R_{req}.

Proposition 7: For both regular linear and planar networks, if

$$R_{req} < \frac{1}{2}\log\left(1+\frac{P_{ind}}{\sigma^2 d_0^{2\delta}}\right),$$

then any single node can communicate with the destination at rate R_{req} regardless of the number of nodes in the system.

Proof: The result follows from the results in [5]. Replace $(2\rho(n))$ with d_0 in (4) in Lemma 5, we obtain the result. ■

V. Topology Constraint under a Minimum Data-Rate Requirement

In the previous sections, we answered the £rst problem proposed in Section I-A. Another question to be answered is the following: suppose we want to support data rate R_{req} for a single active sensor, how large an area can we cover with n sensor nodes, i.e., what is the upper bound on the distance between two adjacent nodes, $d(n)$, such that any single active node can communicate with the fusion center at rate R_{req}?

It is clear that as n increases, the distance between two nodes may become very small when $p < 1$. Thus, we modify the signal attenuation model as follows:

$$\alpha_d = \begin{cases} \frac{1}{d^\delta} & \text{if } d \geq d_{min} \\ f(d) & \text{if } d < d_{min} \end{cases}$$

where $f(d) \geq d_{min}^{-\delta}$.

Note that we only have a lower bound on $f(d)$ and do not assume that we know the function $f(d)$. In other words, we do not exploit the fact that $f(d)$ could be larger than $d_{min}^{-\delta}$ for $d < d_{min}$. This is a realistic assumption since propagation models are not well-de£ned when the distance between the transmitter and receiver is very small [20]. Such a model puts an additional constraint on R_{req}. To be more speci£c, we need

$$R_{req} < R_{max}, \tag{5}$$

where

$$R_{max} = \frac{1}{2}\log\left(1+\frac{P_{ind}}{\sigma^2 (d_{min})^{2\delta}}\right),$$

which is the capacity of a standard additive white Gaussian noise channel.

In the rest of this section, we assume $R_{req} < R_{max}$. Next, we present a lower and an upper bound on $d(n)$ for the system to support data rate R_{req} for any single active node to the fusion center in the linear case and then the planar case.

Corollary 1: If $\delta > 1/2$, $\epsilon_0 > 0$, and

$$d(n) \geq \frac{c_1'}{(\log n)^{1-\frac{1}{2\delta}}},$$

then

$$\lim_{n\to\infty} P_d(n) \geq p - p^2,$$

where

$$c_1' = \left(\frac{(1+\epsilon_0)P_{ind}\left(2\log\frac{1}{1-p}\right)^{2\delta-1}}{(2\delta-1)R_{req}\sigma^2}\right)^{\frac{1}{2\delta}}.$$

On the other hand, if

$$d(n) \leq \frac{c_2'}{\log n},$$

where

$$c_2' = \frac{-(1-\epsilon_0)\log(1-p)}{2}\left(\frac{P_{ind}}{\sigma^2(e^{2R_{req}}-1)}\right)^{\frac{1}{2\delta}},$$

then the system can support rate R_{req} with probability 1 as $n \to \infty$.

The corollary concludes that as n increases, the coverage area of the linear network increase sublinearly and at least as fast as $n\log^{-1} n$.

Proof: By enforcing the constraint $2\rho(n) > d_{min}$ in the proof of Lemma 5, we guarantee that any $R_{req} < R_{max}$ can be achieved.

To obtain an upper bound on the distance $d(n)$ required to achieve a desired rate $R_{req} < R_{max}$, we need

$$Kd(n) \geq d_{min}. \tag{6}$$

However, note that the requirement in (1) in Lemma 2 can be restated as

$$d(n)^{2\delta}K^{2\delta-1} > \frac{P_{ind}(1+\epsilon_0)}{(2\delta-1)R_{req}\sigma^2},$$

and thus a suf£cient condition of (6) is

$$K \geq \frac{d_{min}^{2\delta}(2\delta-1)R_{req}\sigma^2}{P_{ind}(1+\epsilon_0)}.$$

By the de£nition of K in (2), the above condition holds as n increases. The case for the planar network is similar.

The rest of the proof follows from Prop. 1 and Prop. 2. ■

Corollary 2: If $\delta > 1$, $\epsilon_0 > 0$, and

$$d^2(n) \geq \frac{c_3'}{(\log n)^{1-\frac{1}{\delta}}},$$

where

$$c_3' = \left(\frac{2P_{ind}(1+\epsilon_0)\left(4 \log \frac{1}{1-p}\right)^{\delta-1}}{(\delta-1)R_{req}\sigma^2} \right)^{\frac{1}{\delta}}$$

then

$$\lim_{n\to\infty} P_d(n) \geq p - p^2.$$

On the other hand, if

$$d(n)^2 < \frac{c_4'}{\log n},$$

where

$$c_4' = \frac{-(1-\epsilon_0)\log(1-p)}{4} \left(\frac{P_{ind}}{\sigma^2(e^{2R_{req}}-1)} \right)^{\frac{1}{\delta}}$$

then the system can support rate R_{req} with probability 1 as $n \to \infty$. ■.

VI. Simulation

In this section, we simulate a regular unreliable linear network and compare the achievable rate with the developed upper and lower bounds. We choose the linear network because of its numerical simplicity. The simulation setup is as follows: we simulate a linear network of n nodes with the following parameters: $d_0 = 1$, $\sigma^2 = 0.00001$, $P_{ind} = 1$, and $\delta = 3$. Given a value of p, each node is randomly chosen to be active with probability p independently. The achievable rate is limited by the maximum distance between two active nodes, as in Eq. 4. We £rst simulate a linear network with $n = 1000$, shown in Figure 4. In the £gure, the x-axis is p, the probability that a node is active, and y-axis the rate in a logarithm scale. The solid line is the upper bound, dashed line the lower bound, and each star represents an achievable rate of a realization of a random network. The ¤uctuation is due to the randomness in the simulation and the fact that the presented bounds are asymptotic bounds. As p increases, the rate increases and vice versa. The results can be applied to study the chronic performance of a network. As time passes, more nodes fail and thus the achievable rate decreases. In Figure 5, we £x $p = 0.5$ and show the bounds as n increases. In this £gure, the x-axis is the number of nodes in the linear network in a logarithm scale. As n increases, the network size increases linearly, and the achievable rate decreases sub-linearly.

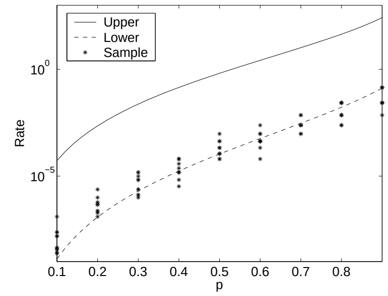

Fig. 4. The effect of p in a linear network.

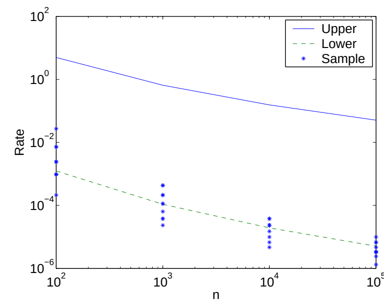

Fig. 5. The effect of n in a linear network.

VII. Conclusion

In the paper, we consider information-theoretic connectivity properties of a regular network with unreliable sensor nodes. We show that when $0 < p < 1$ and $d(n) = d_0$, for a linear network, with a positive probability, there exist active sensor nodes that cannot communicate with the fusion center at rate $R(n)$, if

$$R(n) = \Omega((\log n)^{-2\delta+1}).$$

On the other hand, for large n, we can guarantee that a single active node can communicate with the destination at rate $R(n)$ if

$$R(n) = O((\log n)^{-2\delta}).$$

In addition, for a planar network, with a positive probability, there exists active sensor nodes that cannot communicate with the fusion center rate $R(n)$ if

$$R(n) = \Omega((\log n)^{-\delta+1}).$$

On the other hand, for large n, we can guarantee that each sensor node can communicate with the fusion center if

$$R(n) = O((\log n)^{-\delta}).$$

The conclusion may seem counter-intuitive at the £rst sight: as the number of nodes increases, the achievable rate decreases. Because d_0 is £xed, thus, as n increases, the coverage area increases. To provide uniform guarantees of all active node in a larger area results in a lower rate. Corollaries 1 and 2 conclude the results from a different viewpoint: to guarantee a £xed rate for all nodes, the coverage area grows at least as fast as $O(n/\log n)$, but sub-linearly as n grows.

In the paper, we provide upper and lower bounds on the achievable data rate for any single sensor node to the fusion center given the topology constraints. We also derive the upper and lower bounds on the distance between the sensor nodes to support a required communication rate. The result in the paper can help us understand the communication capability and limit of a large wireless sensor network. We proved the results by picking a £xed destination, the fusion center. However, the proofs can be easily extended to the case where each node has a random (active) destination. First, we consider the achievable results (Prop. 2 and Prop. 4). We prove the results by showing that with probability 1, each square (interval) has at least one active node and we can relay data at rate $R(n)$ from a square to its adjacent squares. Thus, it is clear that we can £nd a route between a source and its randomly picked destination such that rate $R(n)$ is achievable. Next, we consider the converse (Prop. 1 and Prop. 3). The proof shows that there is a positive probability that a node is isolated. If a node is isolated, then it cannot communicate to any node at the desired rate, including its randomly picked destination. Thus, the converse results hold for the case of a random destination. Similar argument holds for Props. 5-7.

There is a big gap between the lower bound and the upper bound proved. It is our conjecture that the lower bound is tight on the order and the upper bound is loose. To provide a tight bound is an interesting research problem. In addition, in this paper, we study the regular grid. The result in this paper should be able to be generalized to the case of a random network where nodes are randomly located within a certain area. The basic approach is similar to the one in this paper while requiring the uniform convergence theorem [21], [22] to bound the number of nodes in a given area uniformly.

REFERENCES

[1] P. R. Kumar, "New technological vistas for systems and control: The example of wireless networks," *IEEE Control Systems Magazine*, pp. 24–37, 2001.

[2] D. Estrin, R. Govindan, J. S. Heidemann, and S. Kumar, "Next century challenges: Scalable coordination in sensor networks," in *Proceedings of ACM Mobicom 1999*, 1999.

[3] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Next century challenges: Mobile networking for smart dust," in *Proceedings of ACM Mobicom 1999*, 1999.

[4] K. Sohrabi, J. Gao, V. Ailawadhi, and G. J. Pottie, "Protocols for self-organization of a wireless sensor network," *IEEE Wireless Communications*, vol. 7, no. 5, pp. 16–27, 2000.

[5] L. L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," 2002, Submitted to IEEE Transactions on Information Theory. URL: http://decision.csl.uiuc.edu/∼prkumar/.

[6] P. Gupta and P. R. Kumar, "Internets in the sky: capacity of 3d wireless networks," in *Proceedings of the 39th IEEE Conference on Decision and Control*, 2000, vol. 3.

[7] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388 –404, 2000.

[8] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *Proceedings of IEEE INFOCOM 2001*, 2001, vol. 3.

[9] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proceedings of IEEE INFOCOM 2003*, 2003.

[10] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: the relay case," in *Proceedings of IEEE INFOCOM 2002*, 2002, vol. 3.

[11] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity," in *Proceedings of the 37th IEEE Conference on Decision and Control*, 1998, vol. 1.

[12] O. Dousse, P. Thiran, and M. Hasler, "Connectivity in ad-hoc and hybrid networks," in *Proceedings of IEEE INFOCOM 2002*, 2002, vol. 2.

[13] O. Dousse, F. Baccelli, and P. Thiran, "Impact of interferences on connectivity in ad hoc networks," in *Proceedings of IEEE INFOCOM 2003*, 2003.

[14] S. Shakkottai, R. Srikant, and N. Shroff, "Unreliable sensor grids: Coverage, connectivity and diameter," in *Proceedings of IEEE INFOCOM 2003*, 2003.

[15] T. K. Philips, S. S. Panwar, and A. N. Tantawi, "Connectivity properties of a packet radio network model," *IEEE Transactions on Information Theory*, vol. 35, no. 5, pp. 1044 –1047, 1989.

[16] P. Santi and D. M. Blough, "The critical transmitting range for connectivity in sparse wireless ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 2, no. 1, pp. 25 –39, 2003.

[17] F. Xue and P. R. Kumar, "The number of neighbors needed for connectivity of wireless networks," To appear Wireless Networks, URL:http://decision.csl.uiuc.edu/∼prkumar/.

[18] B. Schein and R. Gallager, "The gaussian parallel relay network," in *Proceedings of the IEEE International Symposium on Information Theory*, 2000.

[19] X. Liu and R. Srikant, "An information-theoretic view of connectivity in wireless sensor networks," http://www.cs.ucdavis.edu/∼liu/research.html.

[20] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, 2001.

[21] V. N. Vapnik and A. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of probability and its applications*, vol. 16, no. 2, pp. 264–280, 1971.

[22] V. N. Vapnik, *Statistical learning theory*, John Wiley & Sons, 1998.