

Data-Collection Capacity of IEEE 802.11-like Sensor Networks

Chi Pan Chan, Soung Chang Liew

Department of Information Engineering, The Chinese University of Hong Kong
{cpchan4, soung}@ie.cuhk.edu.hk

Abstract – Data-collection is an important application of sensor networks. This paper considers the scenario in which data generated from all sensors are to be forwarded to a single “data center” for processing. Although there have been many studies on this many-to-one communication scenario, it has generally been assumed that the data-collection capacity is upper-bounded by the link capacity L . We show that when the IEEE 802.11 protocol is used, the data-collection capacity has a tighter analytical upper-bound of $3L/4$, and simulated throughput of $0.601L$. In deriving our results, we introduce the notion of “canonical networks”, which is a class of regularly-structured networks whose capacities can be analyzed more easily than unstructured networks. We argue that the capacities of canonical networks serve as a good benchmark for general networks in that the maximum possible capacity of any given network is unlikely to exceed the upper-bound capacity established for canonical networks.

Keywords- sensor networks, 802.11, data collection, data gathering, capacity, wireless multi-hop networks

I. INTRODUCTION

Data collection is an important application of sensor networks. In data collection, there is a “data center” in the sensor network to which data collected by some or all the sensors are to be forwarded. The data center is the place where data collected from the surrounding area are processed. The problem of data collection can be found in many applications. For example, in battle fields, sensors can be used to detect enemies’ intrusion and the aggregated information can be used to deduce the intrusion pattern at the data center.

The achievable rate at which data can be forwarded to the data center is an interesting open issue. This is particularly so in view of the fact that abnormal events detected by individual sensors may occur in synchrony, causing a sudden surge of data to the data center. If the network is not designed well, the data may not reach the data center quickly enough, and may even be lost or discarded in transit, due to traffic overload. Having an understanding of the achievable data-collection rate is important so that networks can be designed to prevent such overloading. For example, the sensors may pace the rate of data forwarding in such a way as to avoid loading the network with traffic above the achievable rate.

There have been many related studies on the capacity of wireless networks. P. Gupta and P.R. Kumar [1] analyzed the capacity in many-to-many situation. It provides the basic model that can be adapted for use in the analysis of the many-to-one communication in data collection. As a loose bound, it can be easily shown that the data-collection capacity is upper-bounded by L [2][3], where L is the single-link capacity.

There is a high probability, however, that the capacity is lower than L for a random network [3]. This paper follows pretty much the approaches used in [1], [2] and [3] in characterizing which nodes can transmit together without packet collisions. The difference is that here we are interested in the capacity under the carrier-sensing (CS) operation of the IEEE 802.11-like distributed MAC protocol [4]. We find that the data-collection capacity with IEEE 802.11-like protocol is at most $3L/4$.

In our analysis, the so-called receiver Restart (RS) Mode is assumed. Ref. [5] has shown that RS is required to remove the hidden-node (HN) problem. This paper focuses on networks that are hidden-node free. The detail of RS will be discussed in Section II of this paper

This rest of this paper is organized as follows. Section II provides the definitions and assumptions used in our analysis. Section III derives the data-collection capacities of IEEE 802.11-like sensor network. Simulation results are presented and discussed in Section IV. Section V concludes this paper.

II. DEFINITIONS AND ASSUMPTIONS

Definition 1: The *source nodes* are nodes that generate data traffic.

Definition 2: The *sink node* is the data center to which the data collected at the source nodes are to be forwarded.

Definition 3: The *relay nodes* relay data traffic from the source nodes to the sink node.

Note that a node can be classified as one of the following: 1) a source node; 2) a sink node; 3) a relay node; or 4) both a source node and a relay node.

Definition 4: Given a network topology, the *data-collection capacity* with respect to a set of source nodes and a sink node is the maximum total rate at which the data from the source nodes can be forwarded to the sink node.

Fig. 1 shows a simple example of a sensor network consisting three nodes. Node 2 is a source node and node 1 is a relay node that forwards packets from node 2 to node 0.

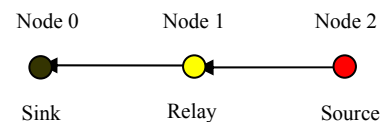


Figure 1. Simple example of data collection

Node 1 does not generate traffic by itself. The data-collection capacity of node 0 with respect to this topology and source traffic is $L/2$, where L is the capacity of one link. This is because node 1 cannot receive and transmit at the same time.

On the other hand, if node 1 is also a source node in addition to being a relay node, then the capacity may reach L . To achieve this, however, only node 1 gets to send its data to the sink node and node 2 must not transmit. If 802.11 MAC protocol is used, node 1 will not enjoy exclusive access to the medium, and the capacity of L in general cannot be achieved.

If both source nodes get to send an equal amount of traffic to node 0, then the capacity is $2L/3$, although the standard 802.11 MAC protocol in which all active nodes have equal opportunity of transmission may not achieve this because achieving this bound requires node 1 to transmit twice as many times as node 2 (since node 1 needs to serve as the relay node for node 2).

Now, if we generalize the above linear network [6] to one consisting $(n+1)$ nodes, in which there are n sources nodes with $(n-1)$ of them also being relay nodes. It is not difficult to see that the capacity is at most $L/3$ for large n (consider that when node i transmits, nodes $(i+1)$ and $(i+2)$ cannot; node $(i+2)$ cannot transmit because the reception at node $(i+1)$ will be corrupted by the transmission by node i). The important observation here is that *in a large network in which all nodes produce the same amount of traffic, the data-collection capacity approaches that in which only the boundary node produces source traffic, with the intermediate nodes being the relay nodes only*. In our linear network example, node n is the boundary node. For large n , it does not matter whether only node n produces traffic or all the n nodes in the linear chain produce traffic as far as the capacity is concerned.

We next consider a general network, such as that in Fig. 2. The capacity bottleneck is likely to be near the sink node because all traffic travels toward the sink nodes. Specifically, nodes near the sink node are responsible for forwarding more traffic and they may contend for access of the wireless medium because they are close to each other. To obtain an idea on the general limit of the data-collection capacity in a large network, it will be worthwhile to consider a “canonical network” in which only the boundary nodes generate traffic to be forwarded to a node in the center, such as that shown in Fig. 3. The capacity of this topology serves as a conservative benchmark for the capacity of a general sensor network in which all nodes generate traffic. We now define the canonical network more specifically (see Fig. 4 for illustration).

Definition 5: A *chain* is formed by a sequence of nodes leading the center sink node. Traffic is forwarded from one node to the next node in the sequence on its way to the sink node. A *linear chain* is a chain which is a straight line.

Definition 6: An *i-hop node* is a node that is i hops away from the sink node in a chain.

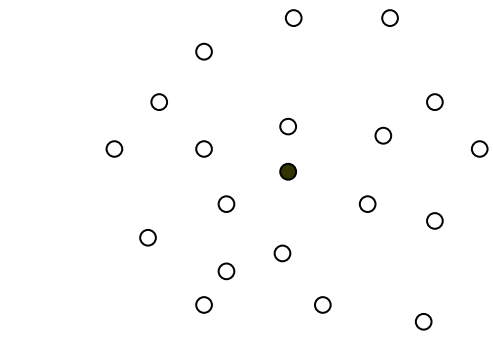


Figure 2. A random sensor network

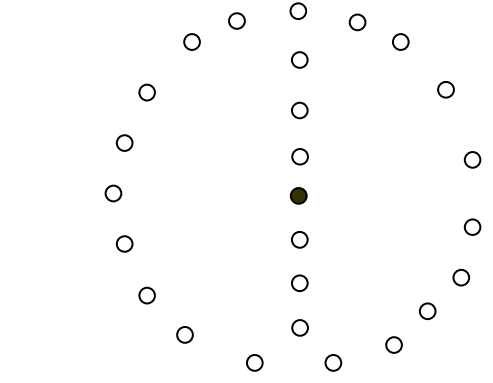


Figure 3. A canonical network with source nodes on the boundary

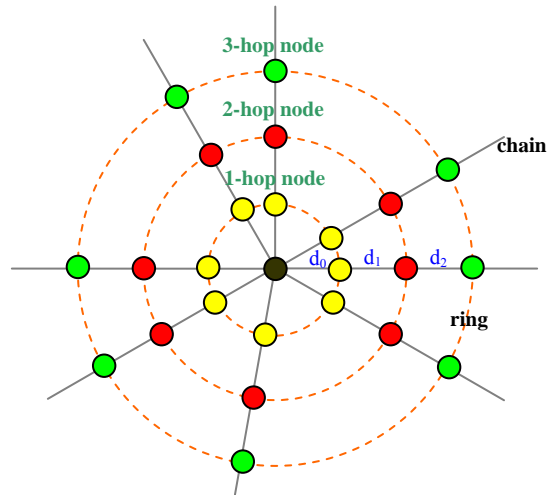


Figure 4. A Canonical Network

Definition 7: A *canonical network* is formed by a number of linear chains leading to a *common* center sink node; the nodes in different chains are distinct except the sink node. In addition, the distance between an i -hop node and an $(i-1)$ -hop node, d_i , is the same for all the linear chains.

Definition 8: A *ring* is a circle centered on the sink node. An i -hop ring consists of all the i -hop nodes of the different linear chains in a canonical network.

Motivation for the Study of Canonical Networks

Canonical networks have regular structures and can be analyzed more easily than general networks. The capacity results obtained are useful as a benchmark (upper bound) for general networks in the following sense. First, in a densely populated network (say, infinitely dense), we may choose to form linear chains from the boundary nodes to the center node for routing purposes. Since the direction of traffic flow is pointed exactly to the center, there is no “wastage” with respect to the case in which the routing direction is at an angle to the center. So, the capacity of the best-structured canonical network is likely to be near optimal, if not optimal. Second, we have defined the class of canonical networks to be quite general in that we do not restrict the number of linear chains in it. Neither do we limit the distance d_i . In deriving the capacity of the canonical network later, we allow for the possibility of an infinite number of linear chains and arbitrarily small d_i . This provides us with a high degree of freedom in identifying the best-structured canonical networks.

We further assume the following:

Assumptions:

- (1) Unless otherwise stated, the nodes and links are homogenous. They are configured similarly, i.e., same transmission power, carrier-sensing range (CSRange), transmission rate, etc.
- (2) ACK is sent by the receiver when a packet is received successfully, as per the 802.11 DCF operation.
- (3) We assume the following constraints on simultaneous transmissions [1][5]. Consider two links (T_1, R_1) and (T_2, R_2) . If the links can transmit at the same time, it must satisfy the following inequalities to avoid collisions:

$$\begin{aligned} |X_{T_2} - X_{R_1}| &> (1+\Delta) |X_{T_1} - X_{R_1}| \\ |X_{R_2} - X_{R_1}| &> (1+\Delta) |X_{T_1} - X_{R_1}| \\ |X_{T_1} - X_{R_2}| &> (1+\Delta) |X_{T_2} - X_{R_2}| \\ |X_{R_1} - X_{R_2}| &> (1+\Delta) |X_{T_2} - X_{R_2}| \end{aligned} \quad (1)$$

where X_i is the location of node i , $|X_i - X_j|$ is the distance between X_i and X_j . $\Delta > 0$ is the distance margin. These are the physical constraints that prevent DATA-DATA, DATA-ACK and ACK-ACK collisions at the receivers.

The received power function can be expressed in the form of

$$P(d) \propto P_t / d^\alpha, \quad (2)$$

where P_t is the transmission power of node t , d is the distance and α is the path-loss exponent, which typically ranges from 2 to 6 according to different environments[7]. Assume $P_{T_1}=P_{T_2}$, $\alpha = 4$, and Signal to Interference Ratio (SIR) requirement of 10dB. Then at R_1 , we require

$$\frac{P(|X_{T_1} - X_{R_1}|)}{P(|X_{T_2} - X_{R_1}|)} > SIR, \quad (3)$$

giving

$$\frac{|X_{T_2} - X_{R_1}|}{|X_{T_1} - X_{R_1}|} > \sqrt[4]{10} = 1.78.$$

In other words, $\Delta = 0.78$. Unless otherwise stated, we assume $\Delta > 0.78$ throughout this work.

(4) We assume Hidden-Node Free Design (HFD). That is, colliding transmissions can be carrier-sensed by transmitters. According to [5], HFD requires

- (i) Use of Receiver Restart (RS) Mode, and
- (ii) Sufficiently large CSRange.

We briefly describe the HFD requirements for understanding of the analysis later. More details can be found in [5]. Fig. 5 is an example showing that no matter how large CSRange is, hidden node (HN) phenomenon can still occur in the absence of an appropriate receiver carrier-sensing operation. In the figure, simultaneous transmissions can occur and the SIR is sufficient at R_1 and R_2 so that no “physical collisions” occur. But HN can still happen, as described below.

Assume T_1 starts first to transmit a DATA packet to R_1 . After the physical-layer preamble of the packet is received by R_2 , R_2 will “capture” the packet and will not attempt to receive another new packet while T_1 ’s DATA is ongoing. If at this time T_2 starts to transmit a DATA to R_2 , R_2 will not receive it and will not reply with an ACK, causing a transmission failure on link (T_2, R_2) . This is the default receiver mode assumed in the NS-2 simulator[8] and most 802.11 commercial products. Note the example in Fig. 5 is independent of the size of CSRange.

This HN problem can be solved with the receiver Restart Mode (RS) which can be enabled in some 802.11 products (e.g., Atheros WiFi chips; however, the default is that this mode is not enabled). With RS, a receiver will switch to receive the stronger packet if its power is C_t times higher than the current packet. The example in Fig. 5 will not give rise to HN with RS if CSRange is sufficiently large.

RS Mode alone, however, cannot prevent HN without sufficiently large CSRange. To see this, consider the example in Fig. 6. Assume T_1 transmits a DATA to R_1 first. During the

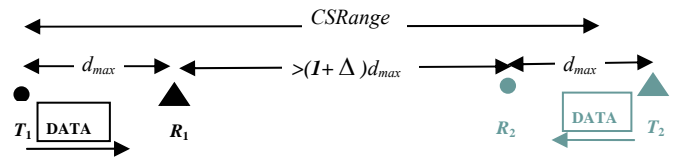


Figure 5. Lack of RS Mode leads to HN no matter how large CSRange and SIR are

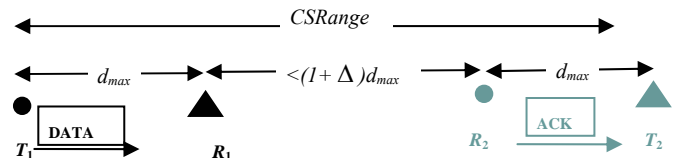


Figure 6. With RS Mode, CSRange not sufficiently large still leads to HN due to insufficient SIR

DATA's period, T_2 starts to send a shorter DATA packet to R_2 . With RS Mode, R_2 switches to receive T_2 's DATA and sends an ACK after the reception. If T_1 's DATA is still in progress, R_2 's ACK will corrupt the DATA at R_1 , since the distance between R_1 and R_2 is within interference range. To prevent T_2 from transmission (hence the collision), the following must be satisfied:

$$|X_{T_1} - X_{T_2}| \leq CSRange. \quad (4)$$

Reference [5] proved that if $CSRange > (3+\Delta) d_{max}$, where d_{max} is the maximum link length, then HN can be prevented in any network in general. However, for a specific network topology, e.g., the canonical network, the required CSRange can be lower.

(5) In the remainder of this paper, unless otherwise stated, the term "collision" refers to collision due to HN (i.e., caused by the failure of carrier-sensing) rather than collision due to simultaneous countdown to zero in the back-off period of the MAC of different transmitters. We assume the latter collisions are negligible (when the network is not overloaded) compared with HN collisions in a large-scale network, a fact which has been borne out by simulations and which can be understood through intuitive reasoning.

Throughout this work, we consider pair-wise interference only. The concept of CSRange and the constraints in (1) rely on this assumption. An analysis which at the outset takes into account the simultaneous interferences from many sources will complicate things significantly. So, given a network topology, our approach is to first identify the HFD solution based on pair-wise interference analysis only, and then verify this solution is still HN-free under multiple interferences. Indeed, it can be shown that for the canonical networks under consideration here, HFD based on pair-wise analysis remains to be valid under multiple interferences. Due to space limitation, however, we omit the multiple-interference verification part in this paper.

III. THEORETICAL ANALYSIS

In this section, we derive the capacities of canonical networks.

A. Fixed Link Distance

We first consider the case where all links have the same length d , i.e., $d_0 = d_1 = \dots = d$. *Theorem 1*, which follows from *Lemma 1* and *Corollary 1* below, proves that the data-collection capacity in this network is upper-bounded by $2L/3$.

Lemma 1: Given three nodes in the circular area of radius d from the center, we can identify two nodes with distance not larger than $(1+\Delta)d$ from each other.

Proof: Assume the contrary that the distance between any two of the three nodes is larger than $(1+\Delta)d$. Then, the triangle whose vertices are these three nodes have sides larger than $(1+\Delta)d$.

Now, consider the *equilateral* triangle inscribed within the circle of radius d , and let t be the length of one side (see Fig. 7). Then

$$t \leq 2d \sin \frac{\pi}{6} = 1.731 d < (1+\Delta)d$$

It is therefore not possible to inscribe another triangle within circle with all sides larger than $(1+\Delta)d$, much less a circle within the triangle. This contradicts the assumption. ■

Corollary 1: At any time, at most *two* 2-hop nodes can transmit at the same time without collision.

Proof: With reference to Fig. 8, suppose that three 2-hop nodes can transmit together. In order that the ACK of any 1-hop node to not interfere with the reception of DATA packet of another transmission, the distances between the three 1-hop nodes must all be larger than $(1+\Delta)d$. By *Lemma 1*, this is not possible. ■

Theorem 1: The data-collection capacity of a canonical network with fixed link distance is upper-bounded by $2L/3$, where L is the link capacity.

Proof: Define "airtime" to be the transmission time of DATA packets as well as the ACK from the receiver [6].

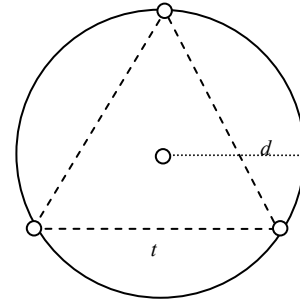


Figure 7. Equilateral triangle inscribed in a circle

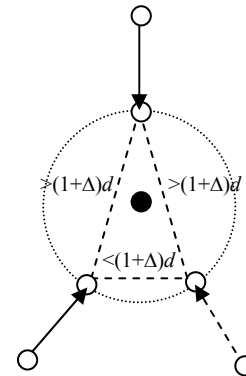


Figure 8. At most two simultaneous transmissions from 2-hop nodes

Let S_{ij} be the airtime occupied by the transmission of the i -hop node on the j -th chain over a long time interval $[0, Time]$.

Let S_1 = the union of airtimes occupied by all 1-hop nodes S_{1j} .

Similarly, S_2 = the union of airtimes occupied by all 2-hop nodes S_{2j} . That is, $S_1 = S_{11} \cup S_{12} \cup \dots \cup S_{1N}$ and

$S_2 = S_{21} \cup S_{22} \cup \dots \cup S_{2N}$. We further define $x_{ij} = |S_{ij}|/Time$.

By definition,

$$|S_1 \cup S_2| \leq Time \quad (5)$$

By the “no collision” assumption in *assumption (3)*, when any 1-hop node transmits, none of the other 1-hop nodes or 2-hop nodes can transmit at the same time if collisions are not to happen. Thus, if carrier-sensing works perfectly and there is no collision, then

$$S_1 \cap S_2 = \emptyset \quad (6)$$

and

$$S_{1i} \cap S_{1j} = \emptyset \quad \text{for } i \neq j. \quad (7)$$

This implies

$$|S_1| + |S_2| = |S_1 \cup S_2| \leq Time \quad (8)$$

and

$$|S_1| = |S_{11}| + |S_{12}| + \dots + |S_{1N}|. \quad (9)$$

By *Corollary 1*,

$$|S_2| \geq \frac{|S_{21}| + |S_{22}| + \dots + |S_{2N}|}{2}. \quad (10)$$

In the canonical network, the 1-hop nodes are relay nodes that do not generate data. All traffic transmitted by 1-hop nodes must come from 2-hop nodes. By the “no collision” assumption, the sum of the airtimes of 1-hop nodes must not be greater than the sum of airtimes of 2-hop nodes. We have

$$|S_{11}| + |S_{12}| + \dots + |S_{1N}| \leq |S_{21}| + |S_{22}| + \dots + |S_{2N}| \quad (11)$$

From (8)-(11),

$$(x_{11} + x_{12} + \dots + x_{1N}) + \frac{(x_{11} + x_{12} + \dots + x_{1N})}{2} \leq 1$$

or $x_{11} + x_{12} + \dots + x_{1N} \leq \frac{2}{3}$

where $(x_{11} + x_{12} + \dots + x_{1N})L$ is the throughput.

■

We now show a specific scheduling example on a 2-chain network which achieves the capacity of $2L/3$. Consider the topology shown in Fig. 9. There are two chains, having link distance d and $CSRange = 3d$ which removes HN. Recall that the general HFD has two requirements, (i) RS mode and (ii) $CSRange > (3+\Delta) d_{max}$ [5]. For the topology in Fig. 9, it turns out that $CSRange = 3d$ is enough.

The numbers shown on the links in Fig. 9 represent a possible transmission schedule. Links with same number transmits at the same time. Following this pattern, capacity of $2L/3$ is “potentially” achievable. Note that whether the capacity is achievable depends on whether the optimal

transmission schedule is realized. Section IV of this paper will show the current IEEE 802.11 protocol cannot achieve this theoretical capacity.

Before going to the next subsection, we note that *Theorem 1* actually applies not just to canonical networks (the proof does not require them), but general networks in which (i) all links are of the same length; and (ii) source nodes are at the boundary. In other words, the chains leading to the data center need not be straight-line linear chains.

B. Variable Link Distance

In this subsection, we consider canonical networks in which the distance between adjacent rings can be varied (i.e., d_0, d_1, \dots may be distinct). With this assumption, the capacity is upper-bounded by $3L/4$. This is proved in *Theorem 2* after *Lemma 2* in the following.

Lemma 2: The number of simultaneous transmissions by 2-hop nodes is at most *three* in a canonical network.

Proof: Assume the contrary that we can have four 2-hop nodes transmitting at the same time. With respect to Fig. 10, consider the four straight lines formed by the four nodes to the center. Four angles are formed between adjacent lines. Let θ be the minimum of the four angles. Four angles are also formed between non-adjacent lines. Let β be the minimum of the four angles.

For simultaneous transmissions, the transmitters should not be able to carrier sense each other. This implies an upper bound for $CSRange$ as follows:

$$CSRange < 2(d_0 + d_1) \sin \frac{\theta}{2}. \quad (12)$$

In addition, by *assumption (4)*, to prevent collision of 1-hop nodes and 2-hop nodes, they should be able to carrier sense each other. This implies a lower bound for $CSRange$. By (4),

$$CSRange \geq \sqrt{(d_0 + d_1)^2 + d_0^2 - 2d_0(d_0 + d_1) \cos \beta}. \quad (13)$$

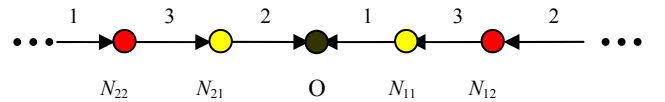


Figure 9. Example of fixed-link-distance topology, $CSRange=3d$

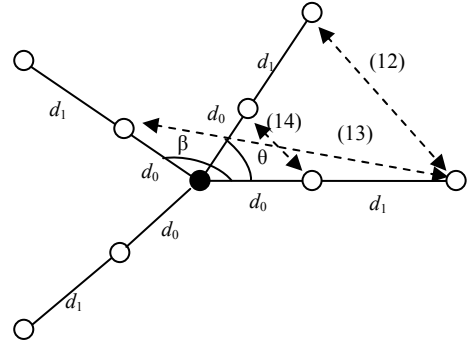


Figure 10. Example of 4-chain canonical network

By *assumption (3)*, the receivers of simultaneous transmissions should not violate the physical constraints. By (1),

$$(1+\Delta)d_1 < 2d_0 \sin \frac{\theta}{2}. \quad (14)$$

Since we assume there are four chains, $\theta \leq 90^\circ$ and $\beta \leq 180^\circ$. From the definitions of θ and β , we have

$$2\theta \leq \beta \leq 180^\circ. \quad (15)$$

From (13) and (15),

$$CSRange \geq \sqrt{(d_0 + d_1)^2 + d_0^2 - 2d_0(d_0 + d_1)\cos(2\theta)}. \quad (16)$$

Let $d_1 = \alpha d_0$. We can form two inequalities from (12), (14) and (16):

$$\alpha < \frac{\sqrt{2(1-\cos\theta)}}{(1+\Delta)}, \quad (17)$$

$$\begin{cases} \alpha > \frac{1-2\cos^2\theta}{1-2\cos\theta} + \frac{\sqrt{(2\cos^2\theta-1)^2 + 1-2\cos\theta}}{1-2\cos\theta} - 1 \\ \text{or} \\ \alpha < \frac{1-2\cos^2\theta}{1-2\cos\theta} - \frac{\sqrt{(2\cos^2\theta-1)^2 + 1-2\cos\theta}}{1-2\cos\theta} - 1 \end{cases} \quad (18)$$

Fig. 11 shows the plot of (17) and (18) when $\Delta = 0.78$. The shadowed region is the area of solution. From the plot,

$$\theta > 1.72912 \text{ radian} = 99.07^\circ.$$

This leads to a contradiction. Thus, there can be at most three simultaneous 2-hop transmissions. ■

Theorem 2: The data-collection capacity of a canonical network is upper-bounded by $3L/4$, where L is the link capacity.

Proof: Similar to the proof of *Theorem 1*, from *Lemma 2*,

$$|S_2| \geq \frac{|S_{21}| + |S_{22}| + \dots + |S_{2N}|}{3}. \quad (19)$$

Hence,

$$\begin{aligned} (x_{11} + x_{12} + \dots + x_{1N}) + \frac{(x_{11} + x_{12} + \dots + x_{1N})}{3} &\leq 1 \\ \text{or} \quad x_{11} + x_{12} + \dots + x_{1N} &\leq \frac{3}{4}, \end{aligned}$$

where $(x_{11} + x_{12} + \dots + x_{1N})L$ is the throughput

Fig. 12 shows an example of a canonical network. The CSRange has to be set larger than $2.62d_0$ and smaller than $3.417d_0$. The numbers on the links show a possible transmission schedule that achieves capacity of $3L/4$. ■

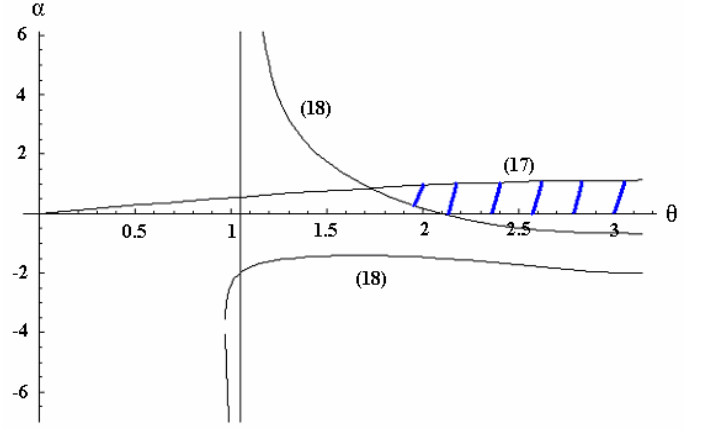


Figure 11. Plot of Inequalities (17) and (18)

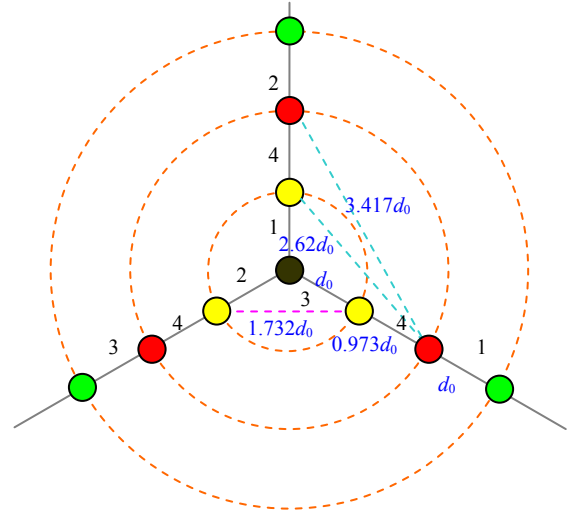


Figure 12. Example of 3-chain canonical network, $CSRange=2.7d$

C. Other topologies

Consider topologies that do not belong to the class of canonical networks. We give a conjecture without a proof:

Conjecture: In a general network where all source nodes are two or more hops away from the center sink node, the data-collection capacity is at most $3L/4$.

For a general network, similar to (5), we have

$$\frac{|S_1| + |S_2| - |S_1 \cap S_2|}{\text{Time}} \leq 1. \quad (20)$$

Similar to (9) and (11),

$$|S_1| = |S_{11}| + |S_{12}| + \dots + |S_{1N_1}| \quad (21)$$

and

$$|S_{11}| + |S_{12}| + \dots + |S_{1N_1}| \leq |S_{21}| + |S_{22}| + \dots + |S_{2N_2}|, \quad (22)$$

where N_1 is the number of 1-hop nodes, N_2 is the number of 2-hop nodes. To obtain a tighter upper-bound capacity than $3L/4$, we need to show that

$$\frac{|S_1| + |S_2| - |S_1 \cap S_2|}{\text{Time}} \leq \frac{3}{4}, \quad (23)$$

which involves finding different possible combinations for $|S_2|$ and $|S_1 \cap S_2|$. Since a general network does not have the regular structure of a canonical network, the analysis is complicated and we leave it as future work as to whether this conjecture is correct.

IV. SIMULATION RESULTS

We use the network simulator NS-2 [8] to simulate the canonical network shown in Fig. 12. As shown in Section III-B, the 3-chain canonical network may achieve capacity $3L/4$ under the right transmission schedule. In the simulation, the IEEE 802.11 protocol is assumed, with RS Mode enabled. Table I shows the details of the simulated configuration. Only nodes at the boundary generate data. Offered load control is applied to the source nodes to prevent them from injecting too much traffic into the network. It has been shown in [6] that offered load control can yield higher throughput.

Fig. 13 shows the simulation result. The x-axis is the number of nodes per chain, including the data center. Given a number of nodes per chain, we vary the offered load in the simulation to identify an offered load that achieves the highest average throughput. When the number of nodes per chain is 3, i.e., the 2-hop nodes are the source nodes, the throughput is 4.43Mbps (0.71L), which is very close to the theoretical capacity $3L/4$, where the link capacity L is around 6.24Mbps as obtained by simulating one single link. But when the number of nodes per chain increases, the throughput drops to 3.75Mbps (0.601L).

An explanation for this phenomenon is that the scheduling scheme of IEEE 802.11 does not result in the optimal transmission schedule discussed in Section II-B. Consider Fig. 12, it is possible for 2-hop and 3-hop nodes to transmit at the same time. To achieve capacity $3L/4$, all the 2-hop nodes must transmit together. However, a 3-hop transmission may prevent this, resulting in only some of the 2-hop nodes transmitting together. That is, there are times when not all 2-hop nodes transmit together, meaning $|S_2|$ cannot reach the lower bound in (19). Meeting the lower bound, however, is essential to achieving the optimal throughput $3L/4$.

Fig. 14 shows the simulation result of canonical networks with different numbers of chains but with fixed link distance. The simulated configuration is shown in Tables II and III. The CSRange for each topology is determined by minimizing its value while preventing HN. The throughput is obtained by varying the offered load and choosing the highest one. From the graph, the highest throughput is 3.63Mbps (0.583L), while the theoretical capacity is $2L/3$. This is likely due to the imperfect scheduling by 802.11, which has been discussed in the previous paragraph.

In Fig. 14, the throughput converges to around 2.1Mbps (0.337L) when the number of chains increases. The

convergence can be explained as follows. From the analysis in Section III, we see that the bottleneck is around the data center. When the number of chains is large, the area near the data center will become dense. The possible transmission patterns are similar in this area, and thus the throughput converges.

The throughput of the network shown in Fig. 12 is higher than the throughput of networks with fixed link distance (compare Fig. 13 and Fig. 14). This matches the theoretical results qualitatively, although the throughput difference is much lower in simulation.

TABLE I. SIMULATION CONFIGURATION FOR VARIABLE-LINK-DISTANCE CANONICAL NETWORKS

Number of chains	3
d_0	250m
d_1	242m
d_i for $i > 1$	250m
Transmission Range	250m
Carrier Sensing Range	675m
Routing Protocol	AODV
Propagation Model	Two Ray Ground
Packet Data Size	1460 bytes

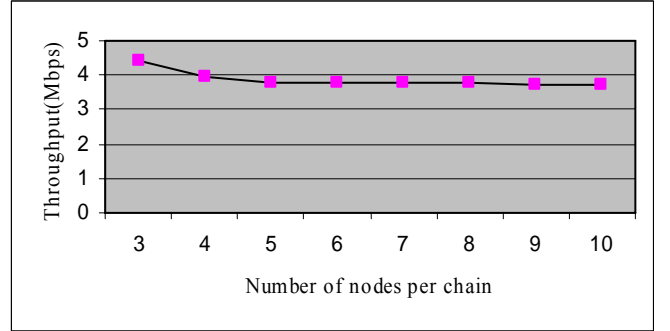


Figure 13. Simulated throughput of a 3-chain canonical network with offered load control

TABLE II. SIMULATION CONFIGURATION FOR FIXED-LINK-DISTANCE CANONICAL NETWORKS

Number of nodes per chain	8
d_i for all i	250m
Transmission Range	250m
Carrier Sensing Range	Refer to Table III
Routing Protocol	AODV
Propagation Model	Two Ray Ground
Packet Data Size	1460 bytes

TABLE III. CARRIER SENSING RANGE FOR FIXED-LINK-DISTANCE CANONICAL NETWORKS

Number of chains	Carrier Sensing Range
2	750m
3	875m
4	750m
5	725m
6	875m
7	800m
8	750m
9	875m
10	825m
>10	900m

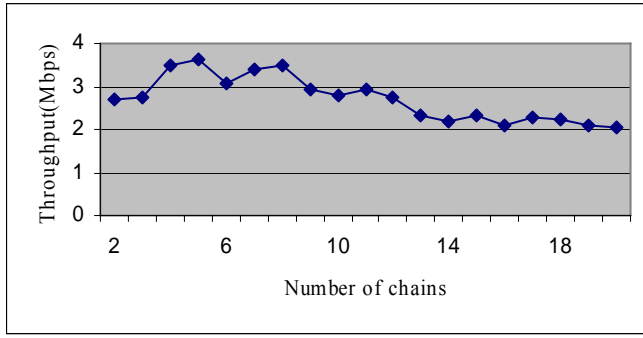


Figure 14. Simulated throughput of fixed-link-distance canonical networks with offered load control

V. CONCLUSION

In this paper, we have studied the data-collection capacity of IEEE 802.11-like sensor networks. We have defined a class of *canonical networks* whose maximum capacity serves as a benchmark for general networks. Specifically, the maximum possible capacity of any network is unlikely to exceed the upper-bound capacity of the canonical networks, which is $3L/4$, where L is the link capacity.

If we restrict our attention to networks in which all links have the same length, the upper-bound capacity is further reduced to $2L/3$. While the $3L/4$ result in the previous paragraph has been established for canonical networks only, the $2L/3$ result applies to general networks so long as (i) source nodes are at least two hops away from the data center; (ii) all links have the same length.

The above results imply that using variable link length is more desirable than using fixed link length. When the network is very dense (say, infinitely dense), if each node chooses a routing path with maximum hop distance in each hop, it may result in an equivalent network with fixed link distance d_{max} , where d_{max} is the maximum hop-distance governed by the transmit power and receiver sensitivity. So, max-hop routing is not optimal in the application of data-collection.

The $3L/4$ and $2L/3$ results above require certain transmission schedules, which may not be achievable by

802.11 exactly. Our 802.11 simulation indicates that when the number of nodes is large, the throughputs are around $0.601L$ (for variable link length) and $0.583L$ (for fixed link length), which does not reach the theoretical capacities of $3L/4$ and $2L/3$, respectively. This implies that there is room for improvement in the design of scheduling schemes, particularly for the variable link distance case.

We have focused on the IEEE 802.11 protocol with Hidden-node Free Design (HFD) here. It will be worthwhile to investigate whether more sophisticated MAC protocols (see, for example, [9]) can further increase the data-collection capacity.

REFERENCES

- [1] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol. IT-46, March 2000.
- [2] D. Marco, E.J. Duarte-Melo, M. Liu, and D.L. Neuhoff, "On the Many-to-One Transport Capacity of a Dense Wireless Sensor Network and the Compressibility of Its Data," *IPSN 2003*, pp. 1-16, April 2003.
- [3] E.J. Duarte-Melo, M. Liu, "Data-Gathering Wireless Sensor Networks: Organization and Capacity," *Computer Networks*, vol. 43, pp.519-537, Nov. 2003.
- [4] IEEE Computer Society LAN MAN Standards Committee, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Std. 802.11, 1997.
- [5] L. Jiang and S. C. Liew, "Removing Hidden Nodes in IEEE 802.11 Wireless Networks," *IEEE VTC*, Sept. 2005.
- [6] P.C. Chung and S.C. Liew, "Offered Load Control in IEEE802.11 Multi-hop Ad-hoc Networks," *IEEE MASS*, Oct. 2004.
- [7] The Institute of Electrical and Electronics Engineers Inc. Press, "Wireless Communications Principles and Practice".
- [8] "The Network Simulator NS-2", <http://www.isi.edu/nsnam/ns>
- [9] P. C. Ng, S. C. Liew, and L. Jiang, "Achieving Scalable Performance in Large-Scale IEEE 802.11 Wireless Networks," *IEEE WCNC*, March 2005.
- [10] W. H. Ho, S. C. Liew, "Achieving Scalable Capacity in Wireless Networks with Adaptive Power Control," *The 5th IEEE Workshop on Wireless Local Networks*, Nov. 2005.
- [11] A. Chakrabarti, A. Sabharwal, and B. Aazhang, "Multi-Hop Communication is Order-Optimal for Homogeneous Sensor Networks," *IPSN 2004*, pp. 178-185, April 2004.
- [12] J. Li, C. Blake et al., "Capacity of Ad Hoc Wireless Networks," *ACM MobiCom*, July 2001.