Numerical Limits for Data Gathering in Wireless Networks

Mohammad Nozari Zarmehri

Faculdade de Engenharia da Universidade do Porto (FEUP)

INESC TEC

Rua Dr. Roberto Frias, s/n 4200-465 Porto Portugal

Email: mnz@inescporto.pt

Ana Aguiar

Faculdade de Engenharia da Universidade do Porto (FEUP)

Instituto de Telecomunicações (IT)

Rua Dr. Roberto Frias, s/n 4200-465 Porto Portugal

Email: ana.aguiar@fe.up.pt

Abstract—In our previous work, we proposed to use a vehicle network for data gathering, i.e. as an urban sensor. In this paper, we aim at understanding the theoretical limits of data gathering in a time slotted wireless network in terms of maximum service rate per node and end to end packet delivery ratio. The capacity of wireless networks has been widely studied and boundaries for that capacity expressed in Bachmann-Landau notation [1]. But these asymptotic limits do not clarify the numeric limits on data packets that can be carried by a wireless network. In this paper, we calculate the maximum data that each node can generate before saturating the network. The expected number of collision and its effect of the PDR% and service rate are investigated. The results quantify the trade off between packet delivery rate and service rate. Finally, we verify our analytical results by simulating the same scenario.

I. Introduction

Data collection is a major application of wireless sensor networks. Typically, each node in the network captures information about its environment through sensors and sends it towards the sink node, which has all resources and functionalities to store and/or process the data. All nodes are also relays, helping data from other nodes get to the sink. In this scenario, it is crucial to know the maximum amount of data that each node can produce without causing the network to overload.

In this paper, we provide initial numeric results on the limit of wireless data gathering along a chain with respect to the probability of collisions due to hidden nodes. The scenario for this calculation is many-to-one communication from all nodes in the chain to the sink. To the best of our knowledge, there are no previous results that enable the calculation of a numeric limit for the data generated by each node as a function of the network parameters. The lack of such a calculation motivates us to show the limitation of the network in a chain of nodes with some numerical results. In this study we focus on the maximum service rate at each node for achieving a certain packet delivery rate at the sink. For example for a chain of 15 nodes, for having guaranteed 90% packet delivery for each node at the sink, the service rate must be less than 25%.

The rest of the paper is organized as follows. We summarize the previous calculation about the capacity of wireless sensor networks for different network models in the next section. The network model which has been used for our calculation is illustrated in Section III. The main contribution for limitation of the data gathering is shown in sections IV and V. Section VI shows the results for two well-known performance metrics: packet delivery rate (PDR) and service rate. Results are discussed in section VII. In Section VIII, we verify our calculation by simulating the same scenario. Finally, we conclude the paper in Section IX.

II. Related Work

Many studies have been done for data gathering capacity in wireless networks. Usually they introduce a boundary for the theoretical throughput. Firstly, Gupta and Kumar [2] study the capacity of wireless networks. They present $\Theta(\frac{W}{\sqrt{n \log n}})$ as throughput per node in a network under a noninterference protocol, where W is the channel bit-rate and n is the total number of nodes in the network.

The capacity has also been calculated under different communication scenarios in random networks. The most interesting scenarios are:

- One-to-One communication where there is one sender and one receiver [3], [4], [5].
- Some-to-Some communication, where some nodes select their destinations randomly [6], [7], [8], [9], [10]
- Many-to-One communication, where all nodes are data sources and there is a sink [11], [12], [13], [14], [15]. This scenario is the scenario which is used for data gathering. [11] showed that the bottleneck is the data collector and the maximum data that can be collected is bounded by $\Theta(\frac{1}{N})$ where N is the total number of time slots which is needed for receiving a complete snapshot from the sensors. Duarte-Melo et al [12] showed that the maximum per node throughput in a many-to-one wireless network scenario is upper bounded by $\Theta(\frac{W}{n})$. $\frac{W}{n}$ is achieved when the sink is 100% busy shared by n nodes. [13] Showed that the data collection capacity of a canonical network with fixed link distance is upper-bounded by $\frac{2W}{3}$ and for variable link distance is upper-bounded by $\frac{3W}{4}$.

In these studies, the boundary of the capacity is introduced by Bachmann–Landau notation [1], which does not measure important metrics like PDR% or service rate for concrete networks with different parameters like number of nodes, packet size and channel bandwidth. Hence, we introduce a

Fig. 1. Network model

way that enables us to understand the overall performance of the network with respect to the concrete network parameters.

III. NETWORK MODEL

According to our previous study for an urban scenario [16], each source passes its packets through a chain of nodes to the sink. So we consider a network model deployed as a flat chain of n nodes, which are all data sources. Also, there is one sink node, located at the end of chain, which we will consider to be at the right side without loss of generality. Each node is not only a source of data packets, but also a relay for the data coming from all its left side neighbors on its way towards the sink. The network model used for the calculations is illustrated in Figure 1.

Each node has at most two neighbors located at the end of its transmission range. Moreover, nodes forward all packets received from the left side and ignore all packets received from the right side (implied ranking). The channel transmission rate is W [bps] and the packet transmission duration is τ, and can be calculated as $\tau = L/W$, where L is the packet size.

IV. MAXIMUM SERVICE RATE WITHOUT COLLISIONS

In this section, we calculate the maximum service rate as the maximum amount of data that can be generated at each node. The actual generation rate should be less than the maximum achievable service rate to avoid overload. We make the simplifying assumption that nodes send their packets in a coordinated way from left to right, starting at the farthest node from the sink, and each node sends one packet of its own and forwards one packet from each node in its left, adding up to i packets at node i. All nodes use the same amount of resources, time, for sending each data packet, τ, so node i needs $i \cdot \tau$ to transmit its packets. We also assume that each node requires a certain time to access the channel for each transmission, the channel access time T_{CA}, which is on average equal for all nodes.

Assuming a total available time of T, the first node on the left has $T - T_{CA}$ time for its transmission. But the second node cannot send during the transmission of the first and third nodes to avoid collisions, where the second node has $T - T_{CA} - 2\tau$ time to send its data packets, which include his own data packet and the data packet forwarded from node one. This calculation can be continued to the end of the chain. Although node n is the last node in the chain, it is not the bottleneck because it must not wait for any transmission on its right side, since there is the sink. Hence, the bottleneck of this scenario is node $n-1$ because it is the last node that must avoid collision with neighbors on both sides.

The total waiting time for the bottleneck in this scenario is given by the sum of the transmission times of the neighbor nodes, $n-2$ and n, with the average waiting time: $T_{CA} + (n-2)\tau + n\tau$. To accommodate all transmissions from all nodes without collisions, the total available time T should be at least equal to the waiting time plus the time needed for its own transmissions at the bottleneck:

$$T \geq T_{CA} + (n-2)\tau + n\tau + (n-1)\tau \tag{1}$$

From here, we can calculate the maximum duration of a single packet, τ, depending on the other network parameters as follows:
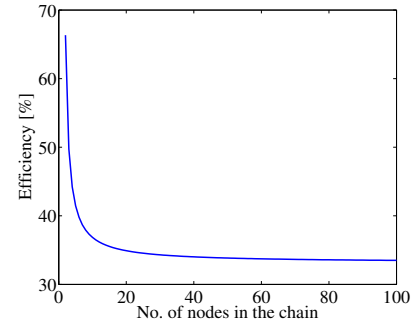
$$\tau \leq \frac{T - T_{CA}}{(3n - 3)} \ [seconds], \tag{2}$$

And the maximum amount of data that can be generated by each node within each period T without causing congestion is:

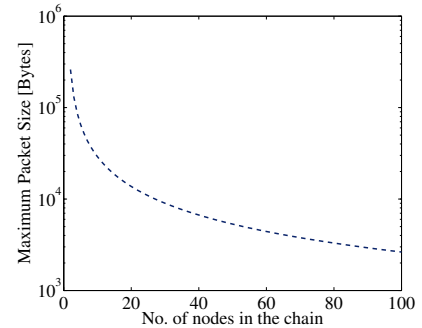$$L \leq \frac{(T - T_{CA})W}{(3n - 3)} \ [bits] \tag{3}$$

We also calculate the best-case resource usage efficiency, ρ, which can only be obtained through perfect scheduling, by dividing the maximum amount of data that can transmitted by all nodes per unit time by the available channel bit rate (W):

$$\rho = \frac{L \cdot n}{T \cdot W} \tag{4}$$

(a) Efficiency

(b) Maximum packet size

Fig. 2. Performance of a simple Chain

Figure 2(a) shows numeric values for the resource efficiency and Figure 2(b) the maximum amount of data per node for T=1 s, T_{CA}=5 ms, and available bandwidth W=6 Mbps. Both values decrease with increasing number of nodes in the chain,

as expected. The bandwidth efficiency is below 70% for very short chains and stabilizes at around 35%, and we have not yet considered hidden node collisions.

V. INTERFERENCE MODEL

In the previous section we assumed that the transmission starts were fully coordinated and node i would only start to transmit after it received all packets from node $i-1$, in which case no collisions could occur. In this section, we extend the results to a more realistic scenario, in which transmission starts are not coordinated and packet receptions at any node can be impaired by collisions caused by simultaneous transmissions from both neighbor nodes, which cannot hear each other (hidden nodes). This only impairs the packet delivery rate from the left side neighbor, since the data flows from left to right. To estimate the amount of collisions and its influence on the system performance we assume a time slotted system, in which the total time T is divided into N equal time slots.

The minimum number of time slots can be calculated by considering the bottleneck in the network according to Eq 5. Node $n-1$ should wait during the transmissions of nodes $n-2$ and n to avoid collision, and needs $n-1$ time slots for own transmission.

$$N = (n-2) + (n-1) + n = 3n - 3 \tag{5}$$

Even with this number of time slots, collision can still happen when two hidden nodes select the same time slot for a transmission. For example, nodes 2 waits for node 1 and 3 but cannot hear transmission of node 4 and so it is possible that nodes 2 and 4 select the same time slots for their transmission, causing a collision that impairs the reception of the packet from node 2 at node 3. So, nodes in the chain do not receive all packets generated to their left side, as part of those collide as a consequence of the hidden node problem. Equation 6 expresses the total received packets at node $n-1$, R_{n-1}, obtained by subtracting the total number of collisions ($E[Col]$) from node 2 to node $n-1$ from the total packets generated to the left of the node.
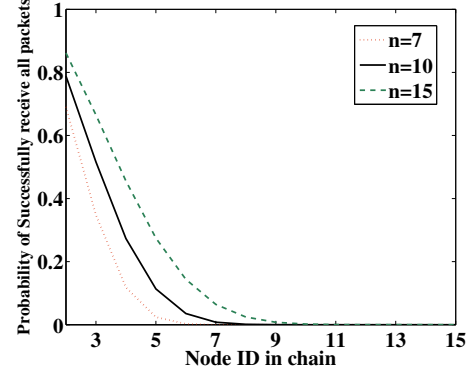
$$R_{n-1} = \sum_{i=1}^{n-2} k_i - \sum_{j=2}^{n-1} E_j[Col] \tag{6}$$

The best case scenario is when time slot selection is coordinated and no collisions occur, meaning that nodes receive correctly all packets generated at their left side. Therefore, the probability of receiving all packet successfully in node $n-1$ can be calculated by Equation 7. Node $n-1$ receives successfully all the packets only if it selects its time slots for transmission from all time slots excluding the ones chosen by nodes $n-2$ and n.
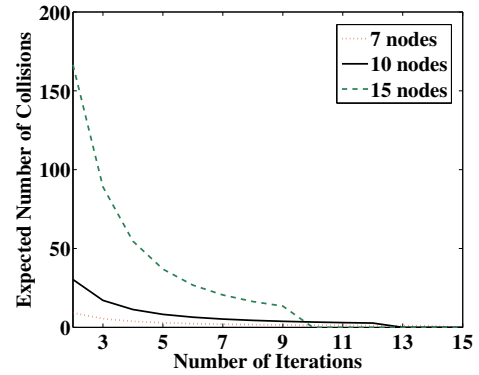
$$P_{n-1}[Successful] = \frac{\binom{N-k_{n-2}-k_n}{k_{n-1}}}{\binom{N}{k_{n-1}}} \tag{7}$$

This calculation is plotted in Figure 3(a) for each node in the chain and for three different chain lengths. The probability of successfully receiving all packets from all nodes decreases as the number of source nodes increases, and goes to zero before reaching to the sink for all chain lengths, meaning that it is not possible to receive all packets at the sink without collision.

(a) Probability of receiving packets successfully at each node in a chain

(b) Effect of increasing N on the expected number of collision

Fig. 3. Chain Performance

For considering possible collisions in Equation 7, which is more realistic, this equation has been extended. We use Y_i as a random variable that indicates the number of collisions at node i. For example $F_{Y_2}(0) = P(Y_2 = 0) = P_2[Successful]$ or $F_{Y_2}(1) = P(Y_2 = 1)$ is the probability of having one collision at node two. The probability of having y collision at node $(n-1)$ is:

$$F_{Y_{n-1}}(y) = P(Y_{n-1} = y) = \frac{\binom{N-k_{n-2}-k_n+y}{k_{n-1}-y} \times \binom{k_{n-1}}{y}}{\binom{N}{k_{n-1}}} \tag{8}$$

In Equation 8, the numerator has two factors. The first factor shows that node $n-1$ as a receiver has y collisions. So this node has y common time slots with its neighbors. The second part just shows the selection of these collisions from all received packets.

So, the expected number of collision can be calculated for node $(n-1)$ from Equation 9:

$$E_{n-1}[Col] = \sum_{i=1}^{kn-2} i \times F_{Y_{n-1}}(i) \qquad (9)$$

In the presence of collisions, some of the time slots are used for unsuccessful transmissions. So, to evaluate the effect of increasing the number of time slots on system performance, we iteratively add the total number of expected collisions over all nodes (Equation 10) to the number of time slots (N). Thus, we do an iterative search for the number of total time slots required to transmit one packet from each node with high packet delivery rate at the sink.

$$E_{Total}[Col] = \sum_{i=2}^{n} E_i[Col] \qquad (10)$$

$$N_{Updated} = 3n - 3 + E_{Total}[Col] \qquad (11)$$

Figure 3-b shows how the expected number of collision is improved by increasing the number of time slots (N). In this figure, each node generates one packet. By increasing N, the expected number of collision decreases (Figure 3-b). However, it causes also a decrease in the service rate.

VI. RESULTS

A. End-2-End PDR%

Following our previous work [17], the most interesting metric for an urban data collector (UDC) protocols is the number of packets received at the sink (PDR%). Moreover, this packet delivery ratio is considered between each source and the sink (End-to-End) because we assume that the sensing application needs information from each source to create a macro vision about the city. According to Equation 6, collisions are the main reason of PDR% reduction at the sink.

Figure 4 shows the PDR% at the sink for a chain with varying number of nodes. The average number of hops from each source node to the sink in our previous study [17] was between 10 to 15 hops, so we focus on this chain length range. The plot shows the end-to-end PDR% for various chain lengths, calculated for the farthest node from the sink, as it is the worst case PDR.

Table VI-A shows the number of time slots in each iteration for each configuration. For 10 nodes, after 12 iteration, there is no collision and the number of time slots remains the same. It also happens for 15 nodes after 10 iterations. Therefore, the curves for 10 and 15 nodes in figure 4 reach to the 100% and cross other curves.

The PDR% with the smallest N is almost zero for the farthest node from the sink. By increasing N, giving more opportunities for transmission, the PDR% is improved to 100% after 12 and 10 iterations for 10 and 15 nodes, respectively.
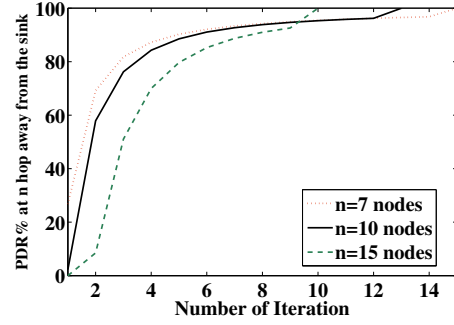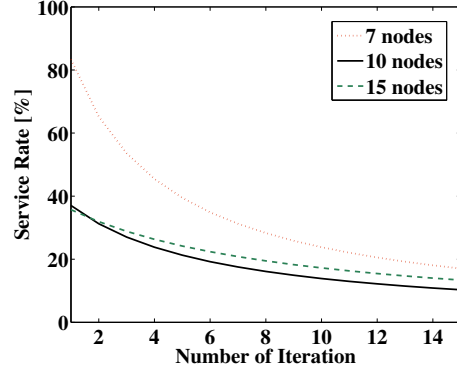

Fig. 4.   PDR% for different chain length

Fig. 5.   Service Rate for different chain length

B. Service Rate

To increase the PDR, we are increasing the amount of time slots necessary to transmit one packet from each node, hence decreasing the service rate and reducing the resource usage efficiency. So, now we evaluate the service rate for this scenario. The service rate is the amount of packets per unit time that each node can generate.

Figure 5 shows the service rate for different chain lengths in different iterations. Clearly, by increasing the number of nodes in the chain, the minimum total number of required time slots increases (see Eq. 5) and so the achievable service rate decreases. The difference in service rate between different chain lengths is due to existence of high number of collision in a longer chain. So, increasing the number of time slots improves the packet delivery rate at the cost of a sharp decrease on the service rate. This trade-off is discussed in the next section.

VII. DISCUSSION

The previous results provide a quantification for the trade-off between the packet delivery rate and service rate. By increasing the number of time slots, the packet delivery rate increases due to the reduced number of collisions and the service rate falls because the total amount of generated packets is fixed. Figure 6 shows this trade-off for better illustration. For the shortest chain, the service rate falls very sharply by

TABLE I
NUMBER OF TIME SLOTS FOR EACH ITERATION

No. of Nodes	Iteration														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
7	18	40	49	54	58	61	63	65	67	69	70	71	72	73	74
10	27	97	127	144	155	163	169	174	179	183	186	189	189	189	189
15	42	422	588	677	731	767	794	814	830	843	843	843	843	843	843

Fig. 6. Trade-off between Service rate and PDR%

(a) 7 nodes

(b) 10 nodes

(c) 15 nodes

Fig. 7. Verification of probability of at least one collision

increasing the number of time slots because there are less time slots in this scenario.

As an example of our goal which is system design, if we have 10 nodes and want to guarantee a PDR of 85% at the sink, then each node should generate three packets per unit time and we should devise 432 time slots for the system (fourth iteration: $3 * 144 = 432$).
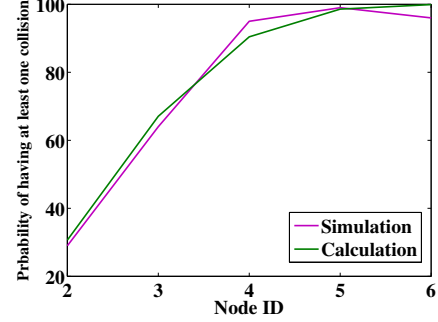
VIII. VERIFICATION OF THE RESULTS

We verify the analytical results by simulating the same scenario. We design a time slotted system with $N = 3n - 3$ slots and each node randomly selects time slots for its transmission, meaning that node one selects one time slot, node two selects two time slots and so on.
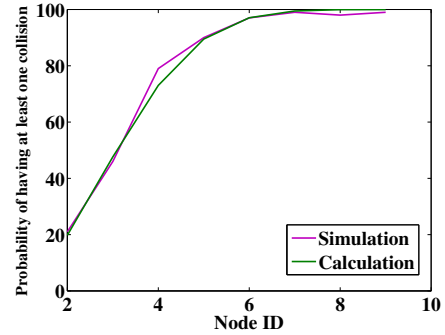
After choosing the time slots for all nodes, we count the number of collisions at each receiver. As a results, at each receiver, the common time slots between the receiver and both its one-hop neighbor in two directions are counted. E.g. in node three, we count the common time slots between nodes (3, 2), (3,4), and (2, 4). It shows the number of collisions at node three (receiver).

We start by verifying the probability of having at least one collision ($1 - P_i[Successful]$). We calculate this probability with both Equation 7 and simulation. In the simulation, we count the number of successful receptions and then reduce it from one. This gives the probability of having at least one collision at each node. Figure 7 shows the results for a chain with different number of nodes and the analytical results match the simulation results.
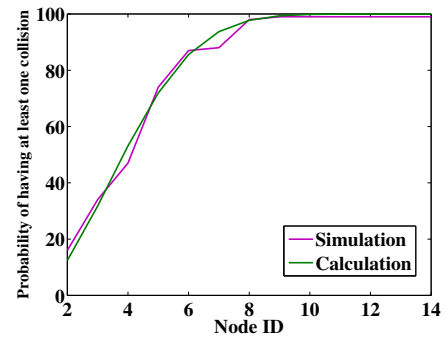
Then, we verify the PDR% at the sink, i.e. the results from Figure 4. Again, the PDR% is calculated for the packet generated at the first node in the chain. To calculate this metric in the simulation, we calculate the probability of having at least one collision at each node and then reduce it from one to obtain PDR_i. Then, we use Equation 12 to calculate the
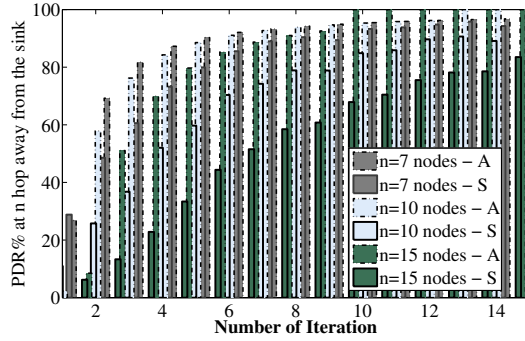
Fig. 8. Comparison of Simulation (S) and analytical (A) results for the end-to-end PDR%

PDR% at the sink for the first node in the chain:

$$PDR_{min} = \prod_{i=2}^{n-1} PDR_i, \tag{12}$$

By comparing the Simulation (S) and analytical (A) results (Figure 8), we can see that the results are almost the same, although there is a small difference. This difference happens because in the simulation, most of the time, there are few time slots that are not selected by any node. So, the obtained results are for a lower effective number of time slots, which decreases the available resources. So, the number of time slots in the simulation is actually less than the number of time slots in the analytical calculation. As a result, the PDR% in the simulation is slightly less than the calculation.

IX. CONCLUSION AND FUTURE WORK

In this paper, we study the limitations of data gathering in the sense of many-to-one scenario and we provide a method to estimate the maximum amount of data that each node can generate to guarantee a pre-defined packet delivery ratio at the sink. Firstly, we calculate the maximum packet size for a scenario without considering the collision. Secondly, we add up collision to our calculation and calculate the end-to-end packet delivery ratio and also the service rate. The results quantify the trade-off between these two metrics. Finally, we verify our analytical results by simulation of the same scenario.

The main line of future work will be to extend these results to more complex network topologies used in data gathering, especially trees. Furthermore, we will model more precisely the impact of contention, and verify the results in experimental settings.

REFERENCES

[1] D. E. Knuth, "Big Omicron and big Omega and big Theta," *ACM SIGACT News*, vol. 8, no. 2, pp. 18–24, Apr. 1976.

[2] P. Gupta and P. Kumar, "Internets in the sky: capacity of 3D wireless networks," in *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No.00CH37187)*, vol. 3. Sydney, Australia: IEEE, 2000, pp. 2290–2295.

[3] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

[4] B. Liu, P. Thiran, and D. Towsley, "Capacity of a wireless ad hoc network with infrastructure," in *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '07*. New York, New York, USA: ACM Press, Sep. 2007, p. 239.

[5] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the Gap in the Capacity of Wireless Networks Via Percolation Theory," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1009–1018, Mar. 2007.

[6] A. Keshavarz-Haddad and R. H. Riedi, "Bounds for the capacity of wireless multihop networks imposed by topology and demand," in *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '07*. New York, New York, USA: ACM Press, Sep. 2007, p. 256.

[7] X.-Y. Li, S.-J. Tang, and O. Frieder, "Multicast capacity for large scale wireless ad hoc networks," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking - MobiCom '07*. New York, New York, USA: ACM Press, Sep. 2007, p. 266.

[8] S. Shakkottai, X. Liu, and R. Srikant, "The Multicast Capacity of Large Multihop Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 6, pp. 1691–1700, Dec. 2010.

[9] B. Liu, D. Towsley, and A. Swami, "Data gathering capacity of large scale multihop wireless networks," in *2008 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems*. IEEE, Sep. 2008, pp. 124–132.

[10] P. Jacquet and G. Rodolakis, "Multicast Scaling Properties in Massively Dense Ad Hoc Networks," in *11th International Conference on Parallel and Distributed Systems (ICPADS'05)*, vol. 2. IEEE, pp. 93–99.

[11] 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," pp. 1–16, Apr. 2003.

[12] E. J. Duarte-Melo and M. Liu, "Data-gathering wireless sensor networks: organization and capacity," *Computer Networks*, vol. 43, no. 4, pp. 519–537, Nov. 2003.

[13] C. Chan and S. Liew, "Data-Collection Capacity of IEEE 802.11-like Sensor Networks," in *2006 IEEE International Conference on Communications*. IEEE, 2006, pp. 3339–3346.

[14] H. Zheng, S. Xiao, X. Wang, and X. Tian, "On the Capacity and Delay of Data Gathering with Compressive Sensing in Wireless Sensor Networks," pp. 1–5, 2011.

[15] S. Chen, M. Huang, S. Tang, and Y. Wang, "Capacity of Data Collection in Arbitrary Wireless Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 1, pp. 52–60, Jan. 2012.

[16] M. Nozari Zarmehri and A. Aguiar, "Data Gathering for Sensing Applications in Vehicular Networks," in *2011 IEEE Vehicular Networking Conference*. Amsterdam: 2011 IEEE Vehicular Networking Conference, 2011, pp. 139–156.

[17] M. Nozari Zarmehri and A. Aguiar, "Supporting Sensing Application in Vehicular Networks," in *ACM MobiCom Workshop on Challenged Networks*. Istanbul: ACM Press, Aug. 2012.