On Distributed Function Computation in Structure-Free Random Networks

Sudeep Kamath and D. Manjunath
Dept of Electrical Engg
IIT Bombay Mumbai INDIA 400 076
sudeep,dmanju@ee.iitb.ac.in

***Abstract*— We consider in-network computation of MAX in a structure-free random multihop wireless network. Nodes do not know their relative or absolute locations and use the Aloha MAC protocol. For one-shot computation, we describe a protocol in which the MAX value becomes available at the origin in $O(\sqrt{n/\log n})$ slots with high probability. This is within a constant factor of that required by the best coordinated protocol. A minimal structure (knowledge of hop-distance from the sink) is imposed on the network and with this structure, we describe a protocol for pipelined computation of MAX that achieves a rate of $\Omega(1/(\log^2 n))$.**

I. Introduction

Early work on computation of functions of binary data over wireless networks focused on computing over noisy broadcast networks, e.g., [1], [2]. With increasing interest in wireless sensor networks, recent research has concentrated on 'in-network' computation over multihop wireless networks, e.g., [3], [4], [5]. The primary focus of this research has been to define an oblivious protocol that identifies the nodes that are to transmit in every slot. This implies that the nodes have organized themselves into a network and have their clocks synchronized. Both of these require significant effort. In this paper we describe a protocol for in-network computation of MAX in a structure-free network. Nodes transmit using the Aloha protocol. We first describe the One-Shot MAX protocol for one-shot computation of the MAX and its analysis. We show that the sink will have the result in a time that is within a constant factor of that required by a structured network with high probability. We then impose a minimal structure and describe the Pipelined MAX protocol and its analysis. We show that the rate of computing the MAX is $\Omega(\frac{1}{\log^2 n})$.

For pedagogical convenience we will assume slotted-Aloha at the MAC layer. The analysis easily extends to the case of pure Aloha MAC.

II. MAX in Noisefree Multihop Aloha

n nodes are uniformly distributed in $[0,1]^2$. The sink, the node that is to have the value of the MAX is at the origin. The nodes know neither their relative nor their absolute positions but each node knows n. We first assume that the nodes use the s-Aloha MAC protocol.

Spatial reuse is analyzed using the well-known protocol model of interference [6]. For s-Aloha, this model translates to the following. Consider a transmitter T at location X_T transmitting in a slot t. A receiver R, at location X_R, can successfully decode this transmission if and only if the following two conditions are satisfied. (1) $\|X_R - X_T\| < r_n$, and (2) $\|X_R - X_{T_1}\| > (1+\Delta')r_n$ for some constant $\Delta' \geq 0$; T_1 is any other node transmitting in slot t and located at X_{T_1}. r_n is called the transmission radius. A transmission from T in slot t is deemed successful if all nodes within r_n of X_T receive it without collision. The following is a sufficient condition for successful transmission by node T in slot t : $\|X_T - X_{T_1}\| > (1+\Delta)r_n$, $\Delta = 1+\Delta'$, for all transmitters T_1 transmitting in slot t.

A. *One-shot computation of* MAX *using Aloha*

Let Z_i be the value of the one-bit data at Node i and $\mathcal{Z} := \max_{1\leq i\leq n} Z_i$. The protocol One-Shot MAX is as follows. Node i can either receive or transmit in a slot but not both. In slot t, Node i will either transmit, with probability p or listen, with probability $(1-p)$, independently of all the other transmissions in the network. Let $X_i(t)$ be the value of the bit received (correctly decoded in the absence of a collision) by Node i in slot t, $t = 1, 2, \ldots$. If Node i transmits in slot t or if it senses a collision or idle in the slot, then it sets $X_i(t) = 0$. Define $Y_i(0) = Z_i$ and $Y_i(t) := \max\{Y_i(t-1), X_i(t)\}$ for $t = 1, 2, \ldots$. If Node i transmits in slot t, it will transmit $Y_i(t-1)$.

It is easy to see that the correct value of $\mathcal{Z}$ will 'diffuse' in the network in every slot. The performance of the protocol, that is, the diffusion time, depends on p. The choice of p is discussed in Section III.

To study the progress of the diffusion, we will consider a tessellation of the unit square into square cells of side $s_n = \lceil\sqrt{\frac{n}{2.75\log n}}\rceil^{-1}$. This will result in $l_n := \frac{1}{s_n} = \lceil\sqrt{\frac{n}{2.75\log n}}\rceil$ rows (and columns) of cells in $[0,1]^2$. There will be a total of $M_n := \frac{1}{s_n^2} = \lceil\sqrt{\frac{n}{2.75\log n}}\rceil^2$ cells. Let $\mathcal{C}$ denote the set of cells under this tessellation. Let S_c be the set of nodes in Cell c and N_c be the number of nodes in Cell c. Under this tessellation, two cells are said to be *adjacent* if they have a common edge. Let the transmission radius be $r_n = \sqrt{\frac{13.75\log n}{n}} \approx \sqrt{5}s_n$. From [6], the network will be connected with high probability for this value of r_n. The expected number of nodes in a cell is $ns_n^2 \approx 2.75\log n$. Further, from Lemma 3.1 of [7], for our choice of r_n and s_n,

$$\Pr\left(c_1\log n \leq N_c \leq c_2\log n \text{ for } 1\leq c\leq M_n\right) \rightarrow 1 \quad (1)$$

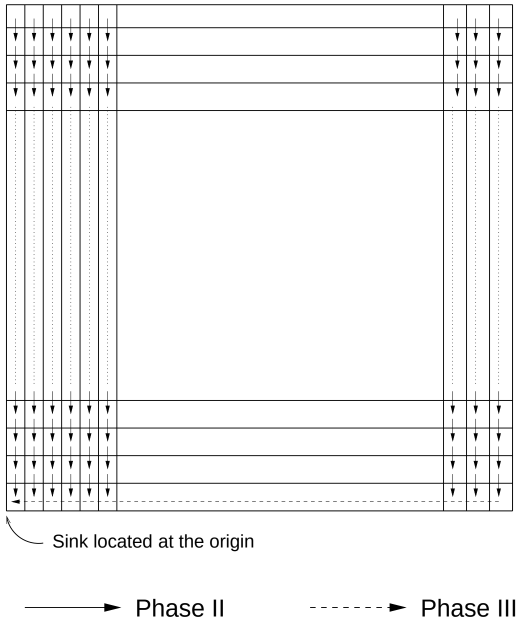

Fig. 1. Direction of diffusion during Phase II and Phase III of protocol One-Shot MAX

where $c_1 = 0.091$ and $c_2 = 5.41$. Our results will hold for networks that are connected and which satisfy (1). As $r_n \geq \sqrt{5}s_n$, a successful transmission by any node from Cell c is correctly decoded by all nodes in Cell c as well as by all nodes in cells adjacent to Cell c.

The value of $\mathcal{Z}$ can reach the sink along any of the many possible trees rooted at the origin. For our analysis, we will divide the progress of the computation into the following three 'phases' and analyze each of the three phases separately.

- *Phase I for data aggregation within each cell.* This phase is completed when every node of the network has transmitted successfully at least once.
- *Phase II for progress to the bottom of the square.* In this phase, the locally computed values of the MAX get diffused into the cells on one side of the unit square as shown in Fig. 1.
- *Phase III for progress into the sink.* In this phase, the value of MAX is transferred to the sink at the origin in the manner shown in Fig. 1.

We show in Section III that Phase I will be completed in $O(\log^2 n)$ slots with high probability, Phase II and Phase III will each be completed in $O\left(\sqrt{\frac{n}{\log n}}\right)$ slots with high probability. We also argue that in another $O\left(\sqrt{\frac{n}{\log n}}\right)$ number of slots, the value of $\mathcal{Z}$ would have diffused to each node of the network. These results can be combined to state the following.

Theorem 1: If all the nodes execute the protocol One-Shot MAX, the maximum of the binary data at the n nodes is available at the sink in $O\left(\sqrt{\frac{n}{\log n}}\right)$ slots with probability at least $\left(1 - \frac{k}{n^\alpha}\right)$ for any positive α and some constant $k > 0$.

Note that the best coordinated protocol, under this choice of r_n, will also require $\Theta\left(\sqrt{\frac{n}{\log n}}\right)$ number of time slots for a one-shot computation of MAX. The bound on the time in Theorem 1 is therefore, tight as well as optimal.

B. *Pipelined computation of* MAX *using Aloha*

If $\mathcal{Z}$ were to be computed continuously using the One-Shot MAX protocol, a throughput of $\Theta\left(\sqrt{\frac{\log n}{n}}\right)$ can be achieved. We believe some structure in the network is necessary to do better. We will assume that all nodes have a transmission range that is exactly r_n. This strict requirement can be easily relaxed but we will keep this assumption for pedagogical convenience.

We impose the following structure in the network. Prior to the computation, each node obtains its minimum hop distance to the sink. Henceforth, we will refer to this as simply the *hop distance* of the node. From (1), each cell in the tessellation is occupied. Since nodes in adjacent cells differ in their hop distance by atmost 1, the largest hop distance of a node in the network is no more than $d := 2\, l_n = 2\lceil\sqrt{\frac{n}{2.75 \log n}}\rceil$.

Let h_i be the hop distance of Node i. Observe that a transmission by Node i can be decoded successfully by Node j only if $|h_i - h_j| \leq 1$. Conversely, if there is a reception by Node i in slot t, then that transmission must have been made by a node with hop distance either $(h_i - 1),\ h_i$, or $(h_i + 1)$. Thus, if a node transmits its hop distance modulo 3 along with its transmitted bit, then every receiver that can decode this transmission successfully, can also, by the receiver's knowledge of its own hop distance, correctly identify the hop distance of the transmitter.

Time is divided into *rounds*, where each round consists of T_0 slots. Minimizing of T_0 maximizes the throughput. We will discuss the choice of T_0 in Section III. Data arrives at each node at the beginning of each round, that is, at the rate of 1 data bit per round. Let the value of the bit at Node i in the round r be $Z_i(r)$. $\mathcal{Z}(r) := \max_{1 \leq i \leq n} Z_i(r)$, for $r = 1, 2, \ldots,$ is to be made available at the sink node, Node s.

The Pipelined MAX protocol is the following. The sink only receives data and does not transmit. The following protocol is executed by the other nodes in the network.

In each slot, Node i either transmits with probability p or listens with probability $(1 - p)$ independently of all other transmissions in the network. The value of p is chosen as in the One-Shot MAX protocol. Each node executes the following protocol for round r.

Transmission: If Node i transmits in slot u of round r, then it transmits three bits $(T_{2,i}(r, u), T_{1,i}(r, u), T_{0,i}(r, u))$ in the slot. Bits $T_{2,i}(r, u)$ and $T_{1,i}(r, u)$ are *identification bits* and are obtained as $(h_i \bmod 3)$. The bit $T_{0,i}(r, u)$ is a *data bit* and is obtained as

$$T_{0,i}(r, u) = \max\{Z_i(r - d + h_i), Y_i(r - 1)\}$$

where, by convention, $Z_i(v) = Y_i(v) = 0$ for $v \leq 0$. Bit $Y_i(r-1)$ is computed from succesful receptions in round $(r-1)$, as described below.

Reception: In round r, Node i maintains $Y_i(r, u)$ for $u = 0, 1, 2, \ldots, T_0$. $Y_i(r, 0)$ is initialized to 0 at the beginning of round r. $Y_i(r, u)$ stores the MAX of the data bits that Node i has decoded from all the slots in round r, upto and including slot u, and which were transmitted by the nodes with hop distance

(h_i+1). In slot u of round r, if Node i successfully receives a transmission from a node with hop distance (h_i+1) (available from the identification bits), then it uses the data bit $X_{0,i}(r,u)$ as follows: $Y_i(r,u) = \max\{Y_i(r,u-1), X_{0,i}(r,u)\}$. Otherwise $Y_i(r,u) = Y_i(r,u-1)$. Let $Y_i(r) = Y_i(r,T_0)$.

The sink node, Node s, obtains the $\mathtt{MAX}$ as $\mathcal{Z}(r-d) = \max\{Z_s(r-d), Y_s(r)\}$, for all $r > d$. The delay of the protocol is d rounds or dT_0 time slots.

Theorem 2: The protocol Pipelined MAX achieves a throughput of $\Omega\left(\frac{1}{\log^2 n}\right)$ with a delay of $O(\sqrt{n\log^3 n})$ time slots. The probability of incorrect computation of $\mathtt{MAX}$ in any round is upper bounded by $\left(\frac{k}{n^\alpha}\right)$ for any positive α and some constant $k > 0$.

The best coordinated protocol for pipelined computation of $\mathtt{MAX}$ can provide a throughput of $\Theta\left(\frac{1}{\log n}\right)$ in the absence of block coding. The penalty for low organization and no coordination is found in the $\log n$ overhead for the length of each "round", T_0, which we have shown in Section III to be $\Theta(\log^2 n)$ time slots. A round in the best coordinated protocol will require $\Theta(\log n)$ slots. Also, for our protocol, Node i, with a hop distance of h_i, requires a memory of $(d-h_i+1)$ bits to store $Z_i(r), Z_i(r-1), \ldots, Z_i(r-d+h_i)$. Thus, the protocol requires each node to have $(d+1)$ bits of memory for storage of past data values.

III. Proofs

A. Preliminaries

1) Bounding the Number of Interfering Neighbors: Define the interfering neighborhood of Node i by $\mathcal{N}_i^{(I)} := \{j : 0 < \|X_i - X_j\| \le (1+\Delta)r_n\}$. As discussed earlier, a transmission from Node i in slot t is deemed successful if all nodes within r_n of Node i can decode this transmission without a collision. A sufficient condition for Node i to be successful in transmitting in slot t is that no node belonging to $\mathcal{N}_i^{(I)}$ must transmit in slot t.

From the protocol model, the choice of s_n and (1), the set of nodes that interfere with a transmission from a node in Cell c, (i.e., $\bigcup_{i\in S_c}\mathcal{N}_i^{(I)}$) is contained within an interference square centered at Cell c. This square contains $k_1 = \left(2\lceil\frac{(1+\Delta)r_n}{s_n}\rceil + 1\right)^2$ cells. From (1),

$$|\mathcal{N}_i^{(I)}| \le k_1 c_2 \log n - 1 \tag{2}$$

Observe that k_1 is a constant for large enough n.

2) Probability of a successful transmission from a cell: Let P_i be the probability that Node i transmits successfully in a slot and $P^{(c)}$, the probability that some node in Cell c transmits successfully in a slot. $P_i \ge p(1-p)^{|\mathcal{N}_i^{(I)}|}$, and from (2), we have $P_i \ge p(1-p)^{k_1c_2\log n - 1}$. Successful transmissions by nodes from Cell c are mutually disjoint events, and hence, $P^{(c)} = \sum_{i\in S_c} P_i \ge N_c p(1-p)^{k_1c_2\log n-1}$. From (1), we have $N_c \ge c_1\log n\ \forall c \in \mathcal{C}$ and hence, $P^{(c)} \ge c_1 \log n\ p(1-p)^{k_1c_2\log n - 1}$. Choosing $p = \frac{1}{k_1c_2\log n}$ maximises the lower bound in this inequality and yields

$$P^{(c)} \ \ge\ \frac{c_1}{k_1c_2}\left(1+\frac{1}{k_1c_2\log n - 1}\right)^{-(k_1c_2\log n - 1)} \ \ge\ \frac{c_1}{k_1c_2e} =: p_S$$

Thus, the probability of successful transmission from a cell is lower bounded by a constant p_S, independent of the number of nodes in the network. This will be crucial to our analysis.

B. Proof of Theorem 1

We will prove Theorem 1 by proving bounds on the total time required by each of phases I, II and III.

1) Phase I: Data aggregation within each cell: Consider Cell c. Let $\mathcal{T}_c$ be the total number of slots required for every node in Cell c to have transmitted successfully at least once. Recall that $p = (k_1c_2\log n)^{-1}$. We will bound $\mathcal{T}_c$ by stochastic domination. Consider a sample space $\mathcal{S}$ containing the set of mutually disjoint events $E_1, E_2, \ldots, E_{N_c}$. Let $\Pr(E_q) = p(1-p)^{k_1c_2\log n-1}$ for $q = 1,2,\ldots,N_c$. Observe that $P_i \ge \Pr(E_q)$, for any Node i in S_c and $q = 1,2,\ldots N_c$. Let $E = \bigcup_{q=1}^{N_c} E_q$. We have $P_E = \Pr(E) = N_c p(1-p)^{k_1c_2\log n - 1}$. Let a sequence of samples be drawn independently from $\mathcal{S}$. The probability of occurence of E in a given sample is P_E and hence, the waiting time in terms of number of samples drawn for the event E to occur is given by the geometrically distributed random variable $\mathrm{Geom}(P_E)$. Let the number of samples required to be drawn from $\mathcal{S}$ so that each of the events $E_q,\ q = 1,2,\ldots N_c$ occurs at least once, be the random variable T_c'. Then, $T_c' = \sum_{j=1}^{R_c'} t_{c,j}'$ where $t_{c,j}' \sim \mathrm{Geom}(P_E)$ and $R_c' \sim \sum_{l=1}^{N_c}\mathrm{Geom}(1-\frac{l-1}{N_c})$. The random variable $t_{c,j}'$ is the waiting time between consecutive occurences of the event E. Now, consider the events of successful occurences of event E. If $(l-1)$ distinct events among $E_1, E_2, \ldots E_q$ have already occured, the probability that the next occurence of E is due to an as yet unoccured event $E_{q'}$ is $(1-\frac{l-1}{N_c})$, as each E_q, for $q = 1,2,\ldots,N_c$ is equally probable. The number of occurences of event E to wait for the occurence of an as yet unoccured event among $E_1, E_2, \ldots, E_q$, when some $(l-1)$ of them have already occurred, is distributed as $\mathrm{Geom}(1-\frac{l-1}{N_c})$. Thus, the random variable T_c' is as obtained earlier.

Now compare the following two events: (1) Event $\mathcal{A}$ defined as the successful transmission from Cell c resulting from a successful transmission by Node i in Cell c and (2) Event $\mathcal{B}$ defined as the occurrence of E in a sample drawn from $\mathcal{S}$ due to the occurence of E_q. Observe that $\Pr(\mathcal{A}) \ge \Pr(\mathcal{B})$. From this comparison, we see that $\mathcal{T}_c$ will be stochastically dominated by T_c' (i.e., $\Pr(\mathcal{T}_c \ge a) \le \Pr(T_c' \ge a),\ a = 1,2,3,\ldots$). Further, T_c' will be stochastically dominated by the random variable $T_c = \sum_{j=1}^{R_c} t_{c,j}$, where $t_{c,j} \sim \mathrm{Geom}(p_S)$ and $R_c \sim \sum_{l=1}^{m}\mathrm{Geom}(1-\frac{l-1}{m})$ with $m = \lceil c_2\log n\rceil$ which is an upper bound on N_c from (1). We therefore, have

$$\Pr(\mathcal{T}_c \ge a) \le \Pr(T_c \ge a),\ a = 1,2,3,\ldots$$

It is convenient to work with the random variable T_c because it is independent of the parameters of Cell c. We will obtain the moment generating functions (mgf) of the distributions of the integer-valued random variables that we analyze. The mgf of random variable X will be denoted by $\mathsf{X}(z) = \sum_{j\in\mathbb{Z}}\Pr(X=j)z^{-j}$. The region of convergence of the mgf is specified in parentheses.

$$
\begin{aligned}
\mathsf{t_{c,j}}(z) &= \frac{p_S z^{-1}}{1-(1-p_S)z^{-1}} := S(z) \quad (|z| > 1-p_S) \\
\mathsf{R_c}(z) &= \Pi_{l=1}^{m} \frac{(1-\frac{l-1}{m})z^{-1}}{1-\frac{l-1}{m}z^{-1}} \\
&= \Pi_{l=1}^{m} \frac{(m-l+1)z^{-1}}{m-(l-1)z^{-1}} \quad \left(|z| > 1-\frac{1}{m}\right) \\
\mathsf{T_c}(z) &= \sum_{r\in\mathbb{N}} \mathsf{Pr}\left(R_c = r\right)[S(z)]^r \\
&= \mathsf{R_c}\left(\frac{1}{S(z)}\right) \\
&= \frac{m! p_S^m}{\Pi_{l=1}^{m}\left(m[z-(1-p_S)]-(l-1)p_S\right)} \\
&\qquad \left(|z| > 1-\frac{p_S}{m}\right)
\end{aligned}
$$

Thus, $\mathbb{E}[e^{sT_c}] = \frac{m!p_S^m}{\Pi_{l=1}^{m}(m[e^{-s}-(1-p_S)]-(l-1)p_S)}$ for $s < \log\left(\frac{1}{1-\frac{p_S}{m}}\right)$. Choose $s_1 = \log\left(\frac{1}{1-\frac{p_S}{2m}}\right)$. After some algebra, we can show the following.

$$
\begin{aligned}
\mathbb{E}[e^{s_1 T_c}] &= \frac{m! p_S^m}{m^m} \Pi_{l=1}^{m}\left(e^{-s_1} - 1 + \frac{m-l+1}{m} p_S\right)^{-1} \\
&= c_m \sqrt{\pi m}.
\end{aligned}
$$

Here $c_m = \frac{2^{2m}}{\binom{2m}{m}\sqrt{\pi m}} \rightarrow 1$ as $m \rightarrow \infty$ by the Stirling approximation. From the Chernoff bound we get $\mathsf{Pr}\left(\mathcal{T}_c \geq V_1\right) \leq \mathsf{Pr}\left(T_c \geq V_1\right) \leq c_m\sqrt{\pi m}\left(1-\frac{p_S}{2m}\right)^{V_1}$. By the union bound, we have

$$
\mathsf{Pr}\left(\max_{c\in\mathcal{C}} \mathcal{T}_c \geq V_1\right) \leq M_n c_m \sqrt{\pi m}\left(1-\frac{p_S}{2m}\right)^{V_1}
$$

To achieve $\mathsf{Pr}\left(\max_{c\in\mathcal{C}} \mathcal{T}_c \geq V_1\right) \leq \frac{k}{n^\alpha}$, it is sufficient to have $(1-\frac{p_S}{2m})^{V_1} \leq \frac{k}{n^\alpha M_n c_m \sqrt{\pi m}}$ or

$$
V_1 \geq \frac{\frac{1}{2}\log m + \log M_n + \alpha \log n - \log k + \frac{1}{2}\log \pi + \log c_m}{-\log(1-\frac{p_S}{2m})}
$$

Here, $m = \lceil c_2 \log n \rceil$, $M_n = \lceil\sqrt{\frac{n}{2.75\log n}}\rceil^2$. Writing $-\log\left(1-\frac{p_S}{2m}\right) = \frac{p_S}{2m} + \frac{p_S^2}{2(2m)^2} + \ldots$, we can see that there exists a choice of $V_1 = O(\log^2 n)$, which is sufficient for Phase I to be complete. That is, every node in every cell of the network would have successfully transmitted at least once in V_1 slots, with probability at least $\left(1-\frac{k}{n^\alpha}\right)$.

2) Phase II: Progress to the bottom of the square: Let the columns of rows as shown in Fig. 1 be numbered $C_1, C_2, \ldots, C_{l_n}$. Each column has l_n cells. Let the cells in each column be numbered from 1 to l_n from top to bottom. In this phase, we are concerned with transmissions in the top $w := l_n - 1$ cells of each column. In Phase I, each node has successfully received the transmissions by every other node in its cell. Hence, Phase II will be completed if the following sequence of events occurs for each column C: A successful transmission by some node in the first cell of the column, followed by a successful transmission by some node in the second cell of the column and so on until a successful transmission by some node in the w-th cell of the column.

Let the number of slots required for this sequence of events be $\mathcal{T}^{(C)}$ for column C. We can see that $\mathcal{T}^{(C)}$ will be stochastically dominated by $T^{(C)} := \sum_{j=1}^{w} t_j^{(C)}$, where $t_j^{(C)} \sim \text{Geom}(p_S)$. We can thus derive the following.

$$
\begin{aligned}
\mathsf{T^{(C)}}(z) &= \frac{p_S^w z^{-w}}{(1-(1-p_S)z^{-1})^w} \\
&\qquad (|z| > 1-p_S) \\
\mathbb{E}[e^{sT^{(C)}}] &= \frac{p_S^w}{(e^{-s}-(1-p_S))^w} \\
&\qquad \text{for } s < \log\left(\frac{1}{1-p_S}\right) \\
\mathsf{Pr}\left(T^{(C)} \geq V_2\right) &\leq \frac{\mathbb{E}[e^{s_2 T^{(C)}}]}{e^{s_2 V_2}} = 2^w(1-\frac{p_S}{2})^{V_2} \\
\mathsf{Pr}\left(\mathcal{T}^{(C)} \geq V_2\right) &\leq 2^w(1-\frac{p_S}{2})^{V_2} \\
\mathsf{Pr}\left(\max_{1\leq j\leq l_n} \mathcal{T}^{(C_j)} q \geq V_2\right) &\leq l_n 2^w(1-\frac{p_S}{2})^{V_2}
\end{aligned}
$$

where we have used $s_2 = \log(\frac{1}{1-\frac{p_S}{2}})$ in the Chernoff bound. Thus, to achieve $\mathsf{Pr}\left(\max_{1\leq j\leq l_n} T^{(C_j)} \geq V_2\right) \leq \frac{k}{n^\alpha}$, it suffices to have $(1-\frac{p_S}{2})^{V_2} \leq \frac{k}{n^\alpha l_n 2^w}$ or

$$
V_2 \geq \frac{\alpha \log n + \log l_n + w\log 2 - \log k}{-\log(1-\frac{p_S}{2})}
$$

Now, $l_n = \lceil\sqrt{\frac{n}{2.75\log n}}\rceil = w+1$, and hence, $V_2 = O\left(\sqrt{\frac{n}{\log n}}\right)$ slots are sufficient for the completion of Phase II with probability at least $\left(1-\frac{k}{n^\alpha}\right)$.

3) Phase III: Progress into the sink: Phase III comprises diffusion of the `MAX` into the cell containing the sink. Let the time required for this to happen be the random variable T_s. It is easily seen from the analysis of the sequence of transmission for Phase II that $\mathsf{Pr}\left(T_s \geq V_3\right) \leq 2^w(1-\frac{p_S}{2})^{V_3}$ where w is as defined before. Calculations similar to those in the analysis for Phase II show that $V_3 = O\left(\sqrt{\frac{n}{\log n}}\right)$ slots are sufficient for completion of this phase with probability at least $\left(1-\frac{k}{n^\alpha}\right)$.

4) : Since each of phases I, II and III get completed in $O\left(\sqrt{\frac{n}{\log n}}\right)$ time slots with probability at least $\left(1-\frac{k'}{n^\alpha}\right)$, for appropriate constants k', the protocol One-Shot MAX achieves computation of the `MAX` at the sink in $O\left(\sqrt{\frac{n}{\log n}}\right)$ number of time slots with probability at least $\left(1-\frac{k}{n^\alpha}\right)$. If the protocol is followed for another V_3 slots, the true `MAX` will diffuse to the complete bottom row, the direction of diffusion being opposite to that in Phase III. In another V_2 slots, the true `MAX` will diffuse out to the complete network by diffusing in the opposite direction to that in Phase II.

C. Obtaining the Hop Distance

The following algorithm Hop Distance Compute obtains the hop distance for each node in the network. $\lceil \log d \rceil$ slots are grouped into a *frame* and $T_0 = \Theta(\log^2 n)$ (as obtained in Phase I earlier) frames form a *superframe.* The algorithm ends after $(d+1)$ superframes.

Let the superframes be denoted by $g_0, g_1, \ldots, g_d$. A node either transmits in every slot of a frame or it does not transmit in any slot of the frame. Each transmission is a number expressed in $\lceil \log d \rceil$ bits. At the beginning of the algorithm, the sink transmits the number 0 expressed in $\lceil \log d \rceil$ bits in each frame of superframe g_0. Each node of the network other than the sink executes the following algorithm. Node i makes no transmission till it has decoded a transmission successfully. Let the first successful reception by Node i happen in a frame belonging to superframe g_i and let the decoded transmission correspond to the number n_i expressed in $\lceil \log d \rceil$ bits. Node i sets its hop distance to $(n_i + 1)$ and ignores other successfully received bits in frames from superframe g_i. During the T_0 frames from superframe g_{i+1}, Node i transmits, in each frame, the number $(n_i + 1)$ expressed in $\lceil \log d \rceil$ bits, with probability p, independently of all the other transmissions in the network and makes no transmission with probability $(1-p)$. After the end of round g_{i+1}, Node i makes no more transmissions. The total number of slots required is $(d+1)T_0\lceil \log d \rceil$.

Lemma 1: The nodes of the network correctly compute their minimum hop distance from the sink, using Hop Distance Compute in $O(\sqrt{n \log^5 n})$ time slots with probability at least $\left(1 - \frac{k}{n^\alpha}\right)$ for any positive α and some constant k.

We omit the proof of this lemma.

D. Proof of Theorem 2

Let the set of nodes at hop distance h be G_h. Let $u_{i,r}$ be the first slot in round r that Node i transmits succesfully in. The number of slots in a round is $T_0 = \Theta(\log^2 n)$ (as obtained in the analysis of Phase I). Every node in the network would have transmitted successfully at least once in each round of T_0 slots with high probability. In the proof, we will assume that each node of the network transmits successfully in each round at least once. We claim that

$$\max_{i \in G_h} T_{0,i}(r, u_{i,r}) = \max_{j \in \bigcup_{h \leq f \leq d} G_f} Z_j(r - d + h)$$

for $0 \leq h \leq d$ and $r > d - h$. The sink being at hop distance 0, proving the claim will complete the proof. Let $h_{\max} \leq d$ be the largest hop distance of a node in the network. The claim is obviously true for $h = h_{\max}$. Assume that the claim is true for $h_0 < h \leq h_{\max}$ for $r > d - h$. We shall show that the claim will then be true for $h = h_0$ and for $r > d - h_0$. Consider transmissions by the nodes at hop distance h_0 in round $(r+1)$.

$$\begin{aligned}\max_{i \in G_{h_0}} T_{0,i}(r+1, u_{i,r+1}) &= \max_{i \in G_{h_0}} \{\max\{Z_i(r+1-d+h_i), \\ &\qquad Y_i(r)\}\} \\ &= \max_{i \in G_{h_0}} \{\max\{Z_i(r+1-d+h_0), Y_i(r)\}\}\end{aligned}$$

Since each node at hop distance (h_0+1) transmits successfully at least once in round r, the transmission of each such node is decoded successfully by some node at hop distance h_0. Hence,

$$\begin{aligned}\max_{i \in G_{h_0}} Y_i(r) &= \max_{j \in G_{h_0+1}} T_{0,j}(r, u_{j,r}) \\ &= \max_{j \in \bigcup_{h_0+1 \leq f \leq d} G_f} Z_j(r - d + h_0 + 1)\end{aligned}$$

where the second equality follows from the induction hypothesis. Hence,

$$\begin{aligned}\max_{i \in G_{h_0}} T_{0,i}(r+1, u_{i,r+1}) &= \max\{\max_{i \in G_{h_0}} Z_i(r+1-d+h_0), \\ &\qquad \max_{j \in \bigcup_{h_0+1 \leq f \leq d} G_f} Z_j(r-d+h_0+1)\} \\ &= \max_{j \in \bigcup_{h_0 \leq f \leq d} G_f} Z_j(r-d+h_0+1)\}\end{aligned}$$

which proves the claim for hop distance h_0 for round $(r+1)$. The claim is therefore, true for hop distance h_0 for $r > d - h_0$. By induction, the claim is true for each h and each round $r > d - h$. The sink Node s computes $Y_s(r)$ and correctly sets $\mathcal{Z}(r-d) = \max\{Z_s(r-d), Y_s(r)\}$. The delay of the protocol is dT_0 slots.

The computation of $\mathcal{Z}(r-d)$ at the end of round r would be unsuccessful only if there exists a node, Node i which does not transmit successfully at all in round $(r - h_i)$. As transmissions by different nodes are independent, the analysis in the diffusion of phase I of One-Shot MAX carries over. The probability that the computed value of $\mathcal{Z}(r)$ is incorrect for any given round is upper bounded by $\frac{k}{n^\alpha}$ for any positive constant α and constant $k > 0$.

IV. Discussion

The total number of transmissions (successful as well as unsuccessful) in one execution of One-Shot MAX is $\Theta(\frac{n^{3/2}}{\log^{3/2} n})$. In Pipelined MAX, a total of $\Theta(n \log n)$ transmissions are made per round. Note that the corresponding number is $\Theta(n)$ with a coordinated protocol for both cases.

The analysis that we provided can be extended to the case where the nodes use pure Aloha as the MAC. We need to use a transmission rate rather than a transmission probability. The success probabilities are calculated similarly except that we now have a collision window that is twice the packet length. All calculations are analogous.

It is fairly straightforward to show that in a noiseless, structure-free broadcast network, the histogram can be computed in $\Theta(n)$ slots with probability at least $\left(1 - \frac{k}{n}\right)$. In the noisy broadcast network, by a simple modification of the protocol of [1], we can show that the histogram can be computed in $\Theta(n \log \log n)$ slots with high probability.

References

[1] R. G. Gallager, "Finding parity in simple broadcast networks," *IEEE Trans. on Info. Theory*, vol. 34, pp. 176–180, 1988.

[2] E. Kushilevitz and Y. Mansour, "Computation in noisy radio networks," in *Proc. of SODA*, 1998, pp. 236–243.

[3] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 755–764, April 2005.

[4] N. Khude, A. Kumar, and A. Karnik, "Time and energy complexity of distributed computation in wireless sensor networks," in *Proceedings of IEEE INFOCOM*, 2005, pp. 2625–2637.

[5] Y. Kanoria and D. Manjunath, "On distributed computation in noisy random planar networks," in *Proc. of IEEE ISIT*, Nice, France, June 2007.

[6] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming*, W. M. McEneaney, G. Yin, and Q. Zhang, Eds. Birkhauser, Boston, 1998.

[7] F. Xue and P. Kumar, "The number of neighbors needed for connectivity of wireless networks," *Wireless Networks*, vol. 10, no. 2, pp. 169–181, March 2004.