Throughput Optimization for Hierarchical Cooperation in Ad Hoc Networks

Javad Ghaderi, Liang-Liang Xie, Xuemin (Sherman) Shen
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, N2L 3G1, Canada
{jghaderi, llxie, xshen} @uwaterloo.ca

***Abstract*—For a wireless network with n nodes distributed in an area A, with n source-destination pairs communicating with each other at some common rate, the hierarchical cooperation scheme proposed in [1] is studied and optimized by choosing the number of hierarchical stages and the corresponding cluster sizes that maximize the total throughput. It turns out that increasing the number of stages does not necessarily increase the throughput, and the closed-form solutions for the optimization problem can be explicitly obtained. Based on the expression of the maximum achievable throughput, it is found that the hierarchical scheme achieves a scaling with the exponent depending on n.**

I. Introduction

Wireless networks formed by radio nodes is a subject of much topical interest, and they are found in various applications such as ad hoc networks, mesh networks, sensor networks, etc. For the optimal design and operation of such networks, it is of fundamental importance to determine the information-theoretic capacity of such networks, which, however, is a formidable task, since even for the simple three-node scenario [2], the exact capacity is still undetermined after several decades' effort.

Although the exact capacity is extremely difficult to determine, a lot of insightful upper and lower bounds on the capacity of large wireless networks have been obtained in recent years, e.g., [3]-[10]. The seminal work [3] initiated the study of scaling laws, and discovered a throughput scaling of $\Theta(\sqrt{n})$ under several communication models. Subsequently, a purely information-theoretic approach without any restrictions on the communication schemes was taken in [4], where a more fundamental connection between the total network transmit power and the transport capacity was discovered. As a consequence, the scaling law of $\Theta(\sqrt{n})$ was confirmed in the high signal attenuation regime only, and when the signal attenuation is low, higher scaling laws are shown to be possible for some special relay networks.

Therefore, an interesting question was raised as to what exactly are the scaling laws in the low signal attenuation regime. By incorporating long-range MIMO communications with local cooperations as proposed in [10], a recent work [1] developed a hierarchical architecture which was able to continually increase the scaling by adding more hierarchical stages. Specifically, for a network model where all the nodes are confined in a unit area but still with the far-field signal attenuation, the scaling with h hierarchical stages was claimed to be $\Theta(n^{\frac{h}{h+1}})$. Thus, by letting $h \to \infty$, any scaling of $\Theta(n^{1-\epsilon})$ is achievable, where $\epsilon > 0$ can be arbitrarily small.

However, there is a fundamentally important issue unaddressed in [1], i.e., the pre-constant of the scaling. Actually, the pre-constants of the scalings for different h are different, and they are not even lower bounded from zero. In this paper, we will show that the complete expression for the scaling with h hierarchical stages should be $c(h)n^{\frac{h}{h+1}}$, where, the pre-constant $c(h)$ not only depends on h, but also tends to zero as h goes to infinity.

Since the pre-constant affects the scaling behavior, we will present what can be achieved with the hierarchical scheme by providing an explicit expression of the pre-constant. It will become obvious that adding more stages does not necessarily increase the achievable rate for any fixed n. Actually, for each n, the optimal number of stages to choose is $\sqrt{\log_\beta(n/2)}$, where β is a constant to be defined later, and the corresponding maximum achievable throughput is

$$\frac{\beta R}{\sqrt{\log_\beta(n/2)}}(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}} \tag{1}$$

where R is another constant. Therefore, as shown in (1), the hierarchical scheme actually achieves a scaling with the exponent depending on n.

Generally, a network with area A is distinguished into two categories based on whether $A^{\alpha/2} \leq n$, where $\alpha \geq 2$ is the power path loss exponent. In the case where $A^{\alpha/2} \leq n$, (1) is achievable. In the other case where $A^{\alpha/2} > n$, (1) has to be multiplied by $n/A^{\alpha/2}$ in order to meet the power constraint.

It is worth pointing out that the results in this paper such as (1) apply to finite n. When trying to draw conclusions on scaling laws by taking $n \to \infty$, however, it should be noted that the results for the first case cannot remain valid, since the far-field model would fail to apply after some point.

For clarity, we will first present the results for regular networks. Then the extension to random networks is trivial after we introduce a clustering algorithm that divides the whole network into quadrilateral clusters, each with exactly the number of nodes required for carrying out the hierarchical cooperation scheme.

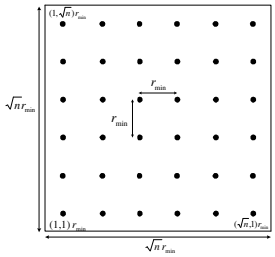

Fig. 1. A regular network with minimum distnce r_{min} and area A.

II. Model of Wireless Network

Consider the following standard additive white Gaussian noise channel model of wireless network.

1) There are a set of n nodes located on a plane.
2) Each node uses a common average power P to transmit.
3) At any time t, each node i transmits the signal $X_i(t) \in \mathbb{C}$, and receives the signal $Y_i(t) \in \mathbb{C}$. The received signal depends on the transmitted signals by all the other nodes as
$$Y_i(t) = \sum_{k \neq i} H_{ik}(t) X_k(t) + Z_i(t)$$
where $Z_i(t)$ is white circularly symmetric Gaussian noise of variance N, and the gain
$$H_{ik}(t) = \sqrt{G} r_{ik}^{-\alpha/2} \exp(j\theta_{ik}(t))$$
where r_{ik} is the distance between nodes i and k, and $\theta_{ik}(t)$ is the phase.

Consider the problem of n source-destination pairs in the network, where each node is a source, with its destination node arbitrarily chosen from the other nodes. For simplicity, assume that each node chooses a different node as its destination, although this requirement can be relaxed to some extend as we can see from the coding strategy described later. Therefore, each node is a source and also a destination for another source. We only consider the case where all pairs communicate at the same rate.

For the simplicity of presentation, and in order to expose the key features of the coding strategy, we will first consider a regular network as depicted in Figure 1, where nodes are located at the grid points $(xr_{\min}, yr_{\min})$ for $1 \leq x, y \leq \sqrt{n}$ in an area $A = nr_{\min}^2$. Then the results can be easily extended to general random networks with high probability, where n nodes are randomly and uniformly distributed inside a square of area A.

III. Hierarchical cooperation in regular networks

A. Double stage cooperation scheme

As a prelude, consider only two stages for the scheme and assume $A = 1$ unit. We basically follow [1], but show what can be achievable by presenting a more transparent description. Divide the regular network into clusters of size M nodes. The *double stage scheme* is based on local transmit and receive cooperation in clusters and MIMO transmissions between clusters. Consider one source node s and its destination node d. The goal of s is to send M sub-blocks of L bits (in overall, ML bits) to d in three steps:

1) s distributes its sub-blocks among the M nodes in its cluster by using TDMA. For this purpose, for each node k in the source cluster, s encodes the k-th sub-block to a codeword of length C_0 chosen from a randomly generated Gaussian codebook $\mathcal{C}_0 \overset{i.i.d}{\sim} \mathcal{N}_{\mathbb{C}}(0, \sigma_0^2)$ where $\sigma_0^2 = P r_{sk}^{\alpha}$.
2) The nodes in this cluster form a distributed array antenna and send the LM bits of information to the destination cluster by MIMO transmissions. To accomplish this step, each node encodes its data sub-block to a codeword of C_1 symbols by using a randomly generated Gaussian codebook $\mathcal{C}_1 \overset{i.i.d}{\sim} \mathcal{N}_{\mathbb{C}}(0, \sigma_1^2)$ where $\sigma_1^2 = P\frac{r_{SD}^{\alpha}}{M}$ and r_{SD} is the distance between the centers of two clusters. Then nodes send their codewords simultaneously to the destination cluster. Therefore, this step needs C_1 time slots to complete. At the end of this step, each node in the destination cluster has accumulated C_1 observations from MIMO transmissions. According to Lemma 4.4 of [1], any rate less than the following is achievable for the second step.
$$L/C_1 \leq \log\left(1 + t\frac{GP}{N + \Delta^2}\right) \frac{(a^2 - t)^2}{2b^4} \qquad (2)$$
for any $0 \leq t \leq a^2$ where $a = (2 - \sqrt{2})^{\alpha/2}$ and $b = (2 + \sqrt{2})^{\alpha/2}$.
3) Each node in the destination cluster quantizes its observations with Q bits per observation to obtain a *quantized observation sub-block* of length $C_1 Q$ bits. ¿From here on, the step is similar to step 1 but in reverse. The cluster nodes send their quantized observation sub-blocks to d which can estimate the observation sub-blocks and decode the transmitted bits.

In the double stage cooperation strategy, the power of each observation must be upper bounded independent of cluster size which leads to quantization with a fixed number of bits for an average distortion Δ^2. When two clusters are neighbor, using the power assignment of $\sigma_1^2 = P r_{SD}^{\alpha}/M$ yields an unbounded received power when the cluster size increases. A simple solution is to divide these clusters into two equal halves. The source node s distributes its sub-blocks among $M/2$ nodes of the half located farther to the border. Then these $M/2$ nodes form a distributed antenna and perform MIMO between the halves located farther away. So the required time for the step 2 is twice the time needed for disjoint clusters, i.e. the required time is $2C_1$ slots. In step 3, $M/2$ nodes take part in delivering the observations to the destination. For source and destination nodes located in the same cluster, we can simply ignore the second step.

Clustering enables spatial reuse in a sense that clusters can work in parallel for local cooperations (step 1 and step 3) provided they locate far enough from each other. This leads

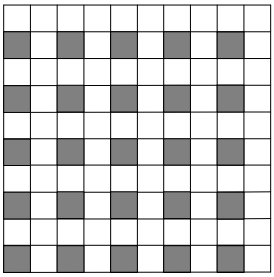

Fig. 2. The operating clusters in one time slot by 4-TDMA

to three phases in the operation of the network:

Phase 1: Setting up transmit cooperation. Clusters work in parallel according to the 4-TDMA scheme in Figure 2 (as opposed to the 9-TDMA in [1]) where each cluster is active in a fraction $1/4$ of the total time of this phase. When a cluster becomes active, its source nodes perform the first step. It is easy to see that the total interference at any receiver due to parallel operating clusters is upper bounded by a finite constant I if $\alpha > 2$. Hence, all communication links can operate at any rate less than the following:

$$L/C_0 \leq \log\left(1 + \frac{GP}{I+N}\right). \tag{3}$$

Each source node needs MC_0 slots. Hence the required time for source nodes of one cluster to exchange their bits is at most M^2C_0 slots. Due to 4-TDMA, the whole phase needs $4M^2C_0$ slots to complete. One can show that the power assignment leads to an overall average power consumption less than P/n.

Phase 2: MIMO transmissions. We perform successive MIMO transmissions according to the step 2, one MIMO for each source-destination pair from source cluster to destination cluster in one time slot, hence we need at most $2nC_1$ slots. Each node encodes the sub-blocks by using a Gaussian code of power σ_1^2 as defined earlier. One can show that this power assignment yields an average power consumption less than P/n.

Phase 3: Cooperate to decode. After the first two phases, each source-destination pair has completed the steps 1 and 2. Each cluster should accomplish the step 3 by conveying the quantized observations to the corresponding destination nodes located in the cluster. This phase is identical to the first phase, except that each node has C_1Q bits to transmit to each node in the same cluster instead of L bits. Therefore, this phase needs $4M^2C_0C_1Q/L$ slots to complete.

In summary, the required time D_2 for the double stage scheme is

$$\begin{aligned} D_2 &= D(phase1) + D(phase2) + D(phase3) \\ &= 4M^2C_0 + 2nC_1 + 4M^2C_0C_1Q/L \\ &= 4M^2C_0(1 + C_1Q/L) + 2nC_1 \end{aligned}$$

For simplicity, all nodes use the same rate for their codewords $\mathcal{C}_0$ and $\mathcal{C}_1$, i.e. $\frac{L}{C_0} = \frac{L}{C_1} = R$ where

$$R \leq \min\{\text{RHS of (2), RHS of (3)}\}$$

Hence, the required time D_2 can be written as

$$D_2(n, M, L) = 4M^2\frac{L}{R}\left(1 + \frac{Q}{R}\right) + 2n\frac{L}{R}$$

We call this quantity delay because each destination can decode its intended bits only after receiving all the corresponding observations , i.e. after the step 3. At the end of this time, each node has delivered ML bits to its destination which yields a total throughput of

$$T_2 = \frac{nML}{D_2}$$

which is maximized by choosing $M = \sqrt{\frac{R}{2(R+Q)}}\sqrt{n}$:

$$T_2^{opt}(n) = \frac{R}{4\sqrt{2(1+Q/R)}}n^{1/2} \tag{4}$$

and the corresponding delay is

$$D_2^{opt}(n, L) = 4\frac{L}{R}n. \tag{5}$$

Obviously, by repeating n times, the double stage scheme can also be used for the problem where each node needs to send different information to all the other nodes. The achievable rate is as the following.

Lemma 3.1: For a regular network of size n, by the double-stage cooperation scheme with clusters of size M, each node can deliver ML different bits to each of the other nodes in a time block of

$$nD_2(n, M, L) = 4nM^2\frac{L}{R}\left(1 + \frac{Q}{R}\right) + 2n^2\frac{L}{R}.$$

Remark 3.1: Note that L denotes the number of bits to be transmitted in a basic time block, and is proportional to the block length for any fixed communication rate. Although for the interest of delay, it is better to choose smaller L as shown in Lemma 3.1, shorter block length leads to higher decoding errors. Hence, there is always a minimum L required to ensure enough reliability.

B. Triple stage cooperation scheme

Recall that in Phase 1 and Phase 3 of the double stage scheme, TDMA was used in each cluster to deliver the bits. Since each cluster is itself a network similar to the original network only with a smaller number of nodes, this implies that one can use the double stage scheme in each cluster to exchange the bits as well. Next, we analyze the throughput and delay of this new *triple stage scheme* when the double stage scheme is used in Phase 1 and Phase 3.

First, divide the whole network into clusters of size M_1. Then divide each cluster of size M_1 into sub-clusters of size M_2, and apply the double stage scheme to the cluster. To avoid the interference from neighboring clusters, use 4-TDMA as before. Hence, according to Lemma 3.1, it takes $M_1D_2(M_1, M_2, L)$ time slots for each node to deliver M_2L bits to each node in the same cluster and this phase needs $4M_1D_2(M_1, M_2, L)$ time slots to complete.

In Phase 2, as before, it takes $2n\frac{M_2L}{R}$ time slots to complete.

In Phase 3, same as phase 1 except that there are $\frac{Q}{R}$ times as many bits to transmit, it takes $4M_1D_2(M_1,M_2,L)\frac{Q}{R}$ time slots to complete.

Totally, with the triple stage scheme, it takes

$$D_3(n,M_1,M_2)=4M_1D_2(M_1,M_2,L)\left(1+\frac{Q}{R}\right)+2n\frac{M_2L}{R}$$

time slots to communicate M_1M_2L bits for each source-destination pair.

C. h-stage hierarchical cooperation scheme

Generally, suppose that with the $(h-1)$-stage hierarchical cooperation scheme with cluster sizes $M_1,M_2,\ldots,M_{h-2}$, it takes $D_{h-1}(n,M_1,M_2,\ldots,M_{h-2})$ time slots to communicate $M_1M_2\cdots M_{h-2}L$ bits for each source-destination pair.

Replacing phase 1 and phase 3 of the double stage scheme with the $(h-1)$-stage scheme, we have the h-stage scheme. Obviously, for the h-stage scheme with cluster sizes $M_1,M_2,\ldots,M_{h-1}$, it takes

$$\begin{aligned}&D_h(n,M_1,M_2,\ldots,M_{h-1})\\&=4M_1D_{h-1}(M_1,M_2,\ldots,M_{h-1})\left(1+\frac{Q}{R}\right)\\&\quad+2n\frac{M_2\cdots M_{h-1}L}{R}\end{aligned}$$

time slots to communicate $M_1M_2\cdots M_{h-1}L$ bits for each source-destination pair.

It can be easily verified that the general formula is

$$\begin{aligned}&D_h(n,M_1,M_2,\ldots,M_{h-1})=M_1M_2\cdots M_{h-1}\frac{L}{R}\times\\&\left\{[4(1+Q/R)]^{h-1}M_{h-1}+2\sum_{i=0}^{h-2}[4(1+Q/R)]^i\frac{M_i}{M_{i+1}}\right\}\end{aligned}$$

Consequently, the throughput is given by

$$T_h(n,M_1,M_2,\ldots,M_{h-1})=\frac{nM_1M_2\cdots M_{h-1}L}{D_h(n,M_1,M_2,\ldots,M_{h-1})}$$

which in general is a function of all the cluster sizes.

We maximize the throughput by using the partial derivatives. Solving $\partial T_h/\partial M_i=0$ for $1\le i\le h-1$ yields

$$M_i^2=\frac{M_{i-1}M_{i+1}}{4(1+Q/R)}$$

where we have defined $M_0=n$ and $M_h=2$. Therefore, the optimal choices of the cluster sizes are

$$M_i=\frac{2(n/2)^{(h-i)/h}}{[4(1+Q/R)]^{i(h-i)/2}}\quad\text{for } 1\le i\le h-1 \tag{6}$$

Next we present one of our main results.

Theorem 3.1: For a regular network of n nodes in the unit area, by the h-stage hierarchical cooperation scheme with the optimal cluster sizes (6), the throughput is given by

$$T_h^{opt}(n)=\frac{R}{h(2\sqrt{1+Q/R})^{h-1}}(n/2)^{1-\frac{1}{h}} \tag{7}$$

and the corresponding delay is

$$D_h^{opt}(n,L)=\frac{h2^{(h+2)(h-1)/(2h)}}{(2\sqrt{1+Q/R})^{(h+3)(h-2)(h-1)/6}}\frac{L}{R}n^{\frac{h-1}{2}+\frac{1}{h}}.$$

For any fixed n, we can find the optimal h to maximize $T_h^{opt}(n)$. Let

$$\frac{dT_h^{opt}(n)}{dh}=0$$

which leads to

$$h^2\ln(2\sqrt{1+Q/R})+h-\ln(n/2)=0.$$

Hence, the optimal number of stages to choose is

$$h^*=\frac{\sqrt{1+4\ln(2\sqrt{1+Q/R})\ln(n/2)}-1}{2\ln(2\sqrt{1+Q/R})}. \tag{8}$$

In order to obtain a simple formula, let

$$\begin{aligned}h^*&=\sqrt{\frac{\ln(n/2)}{\ln(2\sqrt{1+Q/R})}}\\&=\sqrt{\log_\beta(n/2)}\end{aligned} \tag{9}$$

where $\beta:=2\sqrt{1+Q/R}$. Note that

$$\beta^h=\beta^{\log_\beta(n/2)\frac{h}{\log_\beta(n/2)}}=(n/2)^{\frac{h}{\log_\beta(n/2)}}.$$

Therefore,

$$\begin{aligned}T_h^{opt}(n)&=\frac{R}{h(2\sqrt{1+Q/R})^{h-1}}(n/2)^{1-\frac{1}{h}}\\&=\frac{\beta R}{h\beta^h}(n/2)^{1-\frac{1}{h}}\\&=\frac{\beta R}{h}(n/2)^{1-\frac{1}{h}-\frac{h}{\log_\beta(n/2)}}\end{aligned} \tag{10}$$

where letting $h=\sqrt{\log_\beta(n/2)}$, we have

$$T^{opt}(n)=\frac{\beta R}{\sqrt{\log_\beta(n/2)}}(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}} \tag{11}$$

Obviously (11) is a very accurate estimation, although we made some approximation in (9) and h^* should always be an integer.

Theorem 3.2: For a regular network of n nodes in the unit area, by the hierarchical cooperation scheme with the optimal number of stages (8) and the optimal cluster sizes (6), the maximum throughput is approximately given by (11).

Actually, we can provide an exact upper bound of $T^{opt}(n)$ as the following. It follows from (10) that

$$\begin{aligned}T_h^{opt}(n)&\le\beta R(n/2)^{1-\frac{1}{h}-\frac{h}{\log_\beta(n/2)}}\\&\le\beta R(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}}\end{aligned} \tag{12}$$

where, in the last inequality, "=" holds if $h=\sqrt{\log_\beta(n/2)}$.

To check how much different (12) is from the linear scaling law $\Theta(n)$, we take the ratio:

$$\frac{n/2}{(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}}} = (n/2)^{\frac{2}{\sqrt{\log_\beta(n/2)}}}$$
$$= \left(\beta^{\log_\beta(n/2)}\right)^{\frac{2}{\sqrt{\log_\beta(n/2)}}} = \beta^{2\sqrt{\log_\beta(n/2)}} \to \infty.$$

D. Networks with area A

Consider the general model of the regular network with area A. Let us recall that when $A = 1$ unit, running the hierarchy does not need the whole power budget P and the average power consumption is less than P/n per node. We can scale down the general regular network with area A to another regular network with the unit area, but with the power constraint $\frac{P}{(\sqrt{A})^\alpha}$, since the distance between the nodes are reduced by a factor of $\sqrt{A}$. One can dichotomize a general network based on the relation between its area and the number of nodes into two cases:

- Dense network: The network is called dense when $A^{\alpha/2} \leq n$. Then the nodes have enough power to run the hierarchical scheme and get the throughput-delay results as discussed above.
- Sparse network: The network is called sparse when $A^{\alpha/2} > n$. Then the nodes do not have sufficient power to run the hierarchical scheme all the time. Instead, they run the scheme in a fraction $n/A^{\alpha/2}$ of the time with power $PA^{\alpha/2}/n$ and remain silent during the rest of the time. Obviously this bursty modification satisfies the original average power constraint P, and correspondingly, the achieved throughput is modified by a factor of $n/A^{\alpha/2}$, e.g., in (7) and (11).

IV. Extension to random networks

In this section, we extend the results of regular networks to random networks (Refer to [12] for all the details). We first review the extension method of [1]: Consider a random network of unit area with n nodes. Since the average number of nodes in a cluster of area $A_c = \frac{M}{n}$ is M, the hierarchical scheme was applied to this random network by dividing the network into the clusters of area $\frac{M_1}{n}$ and proceeding to clusters of area $\frac{M_{h-1}}{n}$, for the h-stage scheme, and get the throughput-delay of the regular network but with a failure probability. Failure arises from the deviation of number of nodes in each cluster from its average. By a simple Chernoff bound argument, the probability of having large deviations from the average can be bounded (see Lemma 4.1 of [1]). As $n \to \infty$, this probability goes to zero.

The above *clustering* method is not sufficient for the following reasons:

1) The clusters of area $A_c = \frac{M}{n}$ are required to contain *exactly* M nodes to perform the hierarchical scheme. A deviation from the average number of nodes M, even very small, results in failure of the scheme. However, [1] only bounded the probability of large deviation.
2) The probability of having exactly M nodes in a cluster of area $A_c = \frac{M}{n}$ is given by the binomial distribution $\mathfrak{p}(M; n, M/n) = \binom{n}{M}(\frac{M}{n})^M\ (1-\frac{M}{n})^{n-M}$. Using the Stirling's formula to approximate the factorial terms, as $n \to \infty$, yields

$$\mathfrak{p}(M; n, M/n) \approx \frac{M^M}{e^M M!}$$

Recall that for the optimal operation of the scheme, the cluster sizes M are chosen proportional to n^γ where $0 < \gamma < 1$. Hence, the probability of having M nodes is proportional to $\frac{1}{\sqrt{2\pi M}}$ which, in fact, goes to zero.

To resolve the issue of making clusters of exactly M nodes, we will develop a clustering algorithm in this paper. To achieve high probability, we need to consider simultaneously the probabilities of events of the entire class of clusters, which invokes a sort of uniform convergence (in probability) of law of large numbers over the entire class. To resolve this, we will resort to the *Vapnik-Chervonekis* theorem.

A. Choosing an appropriate cluster shape

We use the Vapnik-Chervonekis theorem [11] to find the appropriate cluster shape. An application of this theorem has been already presented in [3] for the set of disks on the plane. In this section, we consider a more general case; we apply the Vapnik-Chervonekis theorem to the set of all the clusters that partition the given random network with n nodes in the unit area. Note that a finite VC-dimension, VC-d($\mathcal{F}$), for the set of clusters $\mathcal{F}$, is a sufficient condition for the uniform convergence in the weak law of large numbers. Assume that this condition is satisfied and the set of clusters has a finite VC-dimension (We will later derive a sufficient condition for the cluster shapes to make the VC-dimension finite). Denote the area of each cluster $c \in \mathcal{F}$ by A_c and its number of nodes with N_c, then we have the following lemma:

Lemma 4.1: For every cluster $c \in \mathcal{F}$ that contains exactly M nodes,

$$\frac{M - \xi \log n}{n} < A_c < \frac{M + \xi \log n}{n} \qquad (13)$$

with probability larger than $1 - \frac{\xi \log n}{n}$ where $\xi = \max\{8\text{VC-}d(\mathcal{F}), 16e\}$.

Note that if a cluster has an area less than $\frac{M-\xi \log n}{n}$, then with high probability it contains less than M nodes. Similarly, if its area is greater than $\frac{M+\xi \log n}{n}$, with high probability, it contains more than M nodes. Next, we choose a right shape for clusters, according to the following Lemma, to make the VC-dimension finite. We have presented the proof in [12].

Lemma 4.2: The VC-dimension of the set of convex r-laterals is finite and upper bounded by $6r \log(3r)$ where r is the number of sides.

We will use a set of quadrilaterals as the clusters. Since the VC-dimension is at most $24 \log 12$, we can apply Lemma 4.1 with $\xi = 800$ to these clusters. Next, we develop an algorithm to make clusters of exactly M nodes.

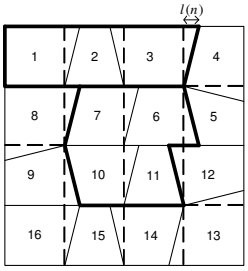

Fig. 3. Clustering of a random network with exactly M nodes in each quadrilateral cluster.

B. Clustering algorithm

Divide the network into squares of area $\frac{M}{n}$, and start from the square located on the top left corner. Depending on how many nodes are within this square, three situations may arise:

1) if the number of nodes in the square is exactly M, ignore this square and go to the next one.
2) if the number of nodes in the square is less than M, make a quadrilateral cluster by expanding the square: Move the top right vertex of the square to the right such that the created quadrilateral cluster contains exactly M nodes.
3) if the number of nodes in the square is more than M, make a quadrilateral by shrinking the square: Move the top right vertex of the square to the left such that the resultant quadrilateral cluster contains exactly M nodes.

After making the first cluster, go to the second cluster on the right side and make it a quadrilateral with exactly M nodes by expanding or shrinking as discussed above. Repeat the procedure for all the squares in the first row. For the top right square, use its bottom right vertex to do expanding/shrinking. For the second row, starting from the right square, move to the left side, and make the quadrilateral clusters of M nodes by expanding-shrinking. Perform the same procedure for all the rows, and we will have a set of quadrilateral clusters; each one contains exactly M nodes. One instance of such a clustering algorithm has been depicted in Figure 3. Note that according to Lemma 4.1, the amount of expanding/shrinking in the areas of the squares is less than $\frac{\xi \log n}{n}$ with high probability.

C. Network operation

The operation of random networks is similar to the operation of the regular networks. The centers of the quadrilateral clusters are defined as the centers of the original squares. Note that the new quadrilateral cluster will include the center of its original square with high probability. To observe this property of our clustering algorithm, consider the combination of the clusters 1, 2, and 3 in Figure 3. This combination gives a larger quadrilateral cluster with $N_c = 3M$, hence the deviation of the area of this cluster from its average $(3M/n)$ must be less than $\frac{\xi \log n}{n}$ and consequently $l(n) \leq \frac{2\xi \log n}{\sqrt{nM}}$. Therefore $l(n)$ is much smaller than the square side $\sqrt{M/n}$ (recall that $M = n^\gamma$ for $0 < \gamma < 1$) and the quadrilaterals are concentrated on the squares. Hence, the hierarchical scheme can be applied to the random networks by using the corresponding quadrilateral of each square instead of original square cluster. By making clusters of M_{h-1} nodes for the bottom stage of the hierarchy using the clustering algorithm, these clusters can be combined to make larger clusters of M_{h-2} nodes for the upper stage. Following the same procedure, make clusters of exactly M_1 nodes for the top stage. It is worth noting that for combined clusters, for example, combination of clusters 6, 7, 10, and 11 in Figure 3, we can define the same deviation factor $l(n)$ as defined for the clusters of the bottom stage.

V. Conclusion

In this paper, the exact achievable throughput of the hierarchical scheme with any number of stages is derived. The optimal cluster sizes for all the stages are found to maximize the total throughput. We also find the optimal number of stages to choose for any network size n. We observe that linear scaling is not achievable via the hierarchical scheme. As one increases the number of stages of the hierarchy to achieve a closer scaling to the linear one, the overhead due to using 4-TDMA scheme for parallel operating clusters and quantizing and re-encoding the observations at different stages, reduces the performance significantly. It also leads to an exponential growth for the delay. Finally, it is worth pointing out that the results presented in this paper provide solid conclusions to networks with finite sizes, not only limiting results.

References

[1] A. Ozgur, O. Leveque, D. Tse, Hierarchical cooperation achieves optimal capacity scaling in Ad Hoc networks, *IEEE Trans. on Information Theory*, vol 53, no. 10, pp. 3549-3572, October 2007.

[2] E. C. van der Meulen, Three-terminal communication channels, *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[3] P. Gupta and P. R. Kumar, The capacity of wireless networks, *IEEE Trans. on Information Theory*, vol. 46, pp. 388- 404, March 2000.

[4] L.-L. Xie and P. R. Kumar, A network information theory for wireless communication: Scaling laws and optimal operation, *IEEE Trans. on Information Theory*, vol. 50, no. 5, pp. 748-767, February 2004.

[5] A. Jovicic, P. Viswanath, and S. R. Kulkarni, Upper bounds to transport capacity of wireless networks, *IEEE Transactions on Information Theory*, vol. 50, pp. 2555–2565, November 2004.

[6] O. Leveque and E. Telatar, Information theoretic upper bounds on the capacity of large extended ad hoc wireless networks, *IEEE Trans. Inform. Theory*, pp. 858–865, vol. 51, no. 3, March 2005.

[7] F. Xue, L.-L. Xie, and P. R. Kumar, The transport capacity of wireless networks over fading channels, *IEEE Trans. Inform. Theory*, 2005. pp. 834–847, vol. 51, no. 3, March 2005.

[8] L. -L. Xie and P. R. Kumar, On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks, *Joint Special Issue of IEEE Trans. on Information Theory and IEEE/ACM Trans. on Networking and Information Theory*, vol. 52, no. 6, pp. 2313-2328, June 2006.

[9] M. Franceschetti, O. Dousse, D. Tse and P. Thiran, Closing the gap in the capacity of wireless networks via percolation theory, *IEEE Trans. on Information Theory*, vol. 53, no. 3, pp. 1009-1018, March 2007.

[10] S. Aeron and V. Saligrama, Wireless ad hoc networks: strategies and scaling laws for the fixed SNR regime, *IEEE Trans. on Information Theory*, vol. 53, no. 6, pp. 2044-2059, June 2007.

[11] V. N. Vapnik and A. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability and its Applications*, vol. 16, no. 2, pp. 264-280, 1971.

[12] J. Ghaderi, L.-L. Xie, and X. Shen, Hierarchical cooperation in Ad Hoc networks: optimal clustering and achievable throughput, submitted to *IEEE Trans. on Information Theory*, February 2008.