Throughput-Delay Trade-off for Hierarchical Cooperation in Ad Hoc Wireless Networks

Ayfer Özgür, Olivier Lévêque
Faculté Informatique et Communications
Ecole Polytechnique Fédérale de Lausanne
1015 Lausanne, Switzerland
{ayfer.ozgur, olivier.leveque}@epfl.ch

***Abstract*— Hierarchical cooperation has recently been shown to achieve better throughput scaling than classical multihop schemes in static wireless networks. However, the end-to-end delay of this scheme turns out to be significantly larger than those of multihop schemes. A modification of the scheme is proposed here that achieves a throughput-delay trade-off $D(n) = (\log n)^2 T(n)$ for $T(n)$ between $\Theta(\sqrt{n}/\log n)$ and $\Theta(n/\log n)$, where $D(n)$ and $T(n)$ are respectively the average delay per bit and the aggregate throughput in a network of n nodes. This trade-off complements the previous results [1], [2] which show that the throughput-delay trade-off for multihop schemes is given by $D(n) = T(n)$ where $T(n)$ lies between $\Theta(1)$ and $\Theta(\sqrt{n})$.**

***Index Terms*— Ad hoc Wireless Networks, Scaling Laws, Throughput-Delay Trade-off, Hierarchical Cooperation**

I. Introduction

Scaling laws offer a way of studying fundamental trade-offs in wireless networks as well as of highlighting the qualitative and architectural properties of specific designs. Such study has been initiated by the work [3] of Gupta and Kumar in 2000. Their by now familiar model considers n nodes randomly distributed on a unit area, each of which wants to communicate to a random destination at a common rate $R(n)$. They ask what is the maximally achievable scaling of the aggregate throughput $T(n) = nR(n)$ and show that cooperation between nodes can dramatically improve performance. Instead of using the simple non-cooperative scheme of time-sharing between direct transmissions from source nodes to destinations (TDMA), which only achieves aggregate throughput $\Theta(1)$, the nodes can cooperate and relay the packets by multihopping from one node to the next, in which case an aggregate throughput scaling of $\Theta(\sqrt{n})$ is achieved. The price to pay, however, is in terms of delay. In the multi-hop scheme, the packets need to be retransmitted many times before they reach their actual destinations, which results in larger end-to-end delay. More precisely, as shown later in [1], [2], in a multi-hop scheme, bits are delivered to their destinations in $\Theta(\sqrt{n})$ average time after they leave their source nodes, while the average delay for the simple TDMA scheme remains only $\Theta(1)$. Note that this accounts only for on-the-flight delay and the queuing delay at the source node is not considered.

Recently, it has been shown in [4] that under certain assumptions on the channel model, a much better throughput scaling is achievable in wireless networks than the one achieved by classical multi-hop schemes. The authors exhibit a hierarchical cooperation scheme that uses distributed MIMO communication to achieve aggregate throughput scaling arbitrarily close to linear, i.e. $T_h(n) = \Theta(n^{\frac{h}{h+1}})$ for any integer $h > 0$. The parameter h corresponds to the number of hierarchical levels used in the scheme and by increasing h, one can get arbitrarily close to linear scaling. A natural question is whether there is a price to pay for this superior scaling of the throughput. In particular, where is the scheme located on the throughput-delay trade-off discussed earlier? In this paper, we reanalyze the scheme presented in [4] and show that better throughput is achieved at the expense of extremely large bulk-size, where the bulk-size of a scheme is the minimum number of bits that should be communicated between each source-destination pair. More precisely, we show that the bulk-size used by the scheme scales as $B_h(n) = \Theta(n^{\frac{h}{2}})$; in other words, it grows arbitrarily fast as the throughput approaches linear scaling. Note that the bulk-size immediately imposes a lower bound on the end-to-end delay of the communication; even if there is no transmission delay from the source node to the destination node, receiving a bulk of $B(n)$ bits will take at least $\Theta(B(n)/\log n)$ channel uses for a destination node, since a simple application of the cut-set bound upper bounds the rate of reception by (or transmission from) a node with $\log n$ bits per channel use.

In the rest of the paper, we present a modification of the hierarchical cooperation scheme that achieves the same aggregate throughput $T_h(n) = \Theta(n^{\frac{h}{h+1}})$ by using a much smaller bulk-size of $B_h(n) = \Theta(n^{\frac{h}{h+1}})$ bits. The key idea in [4] that yields the hierarchical architecture is to set up the receive and transmit cooperation for the distributed MIMO transmissions as multiple problems of the original kind, that is of communicating between n source-destination pairs in a network of n nodes. Any known solution to the original problem can then be used for cooperation, eventually yielding a better solution. However, if the scheme to begin with uses large bulk-size, using it for cooperation yields a scheme with even larger bulk-size. This is the reason for the increase in bulk size as $\Theta(n^{\frac{h}{2}})$ with increasing number of hierarchical levels h. In this paper, we study the problem of cooperation more carefully. Instead of posing it as multiple unicast problems, we pose it as a network multiple access problem where each of the

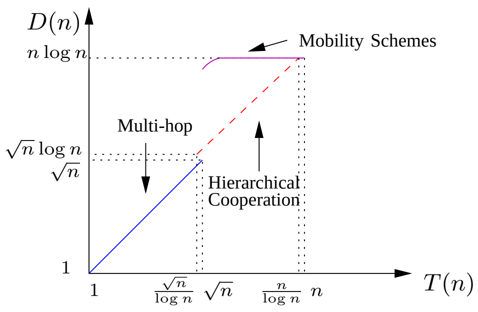

Fig. 1. Throughput-delay performance achieved by hierarchical cooperation together with known results from the literature.

n nodes in the network is interested in conveying independent information, say L bits, to each of the other nodes in the network. We propose a two-phase hierarchical scheme that solves this multiple access problem in $\Theta(n^{\frac{h+1}{h}})$ time-slots for any $h > 0$. Using this scheme for cooperation, we achieve a bulk-size scaling $B_h(n) = T_h(n)$ for our original problem. We show that reduced bulk size consequently reduces the delay to $D_h(n) = n$.

We proceed by optimizing scheduling in this scheme to further reduce the end-to-end delay. To do this, we need to consider a generalized version of the multiple access problem where each node in the network is interested in conveying independent information, say again L bits, to each of the nodes in a subset of $A(n)$ nodes, where the $A(n) < n$ nodes are chosen uniformly at random among the n nodes in the network. We show that this task can be accomplished in $\Theta(\frac{A(n)}{n} n^{\frac{h}{h+1}} \log n)$ channel uses for any $h > 0$ if $A(n) \geq n^{\frac{h}{h+1}}$. This allows us to achieve a throughput delay trade-off of $(T(n), D(n)) = (n^b/\log n, n^b \log n)$ for any $0 \leq b < 1$. This new result is depicted in Fig. 1, together with previous results from the literature.

A related line of research (see e.g. [5], [6], [1], [7]) is the characterization of the throughput-delay trade-off for mobile networks, where nodes move over the duration of communication according to a certain mobility pattern. In general, mobility schemes achieve an aggregate throughput scaling comparable to that of hierarchical cooperation (i.e. up to linear in n), but the delay scaling performance of such schemes may vary significantly, depending on the chosen mobility model. For instance, under the classical random walk mobility model considered in [1], the performance is quite poor, as illustrated in Fig. 1. But from the delay point of view, a more prominent disadvantage which is common to all mobility schemes and which does not appear on the graph in Fig. 1, is the constant that precedes the delay scaling law. Roughly speaking, this pre-constant relates to the speed of nodes in the case of mobility schemes, whereas it relates to the speed of light in the case of hierarchical cooperation.

II. Setting and Main Results

There are n nodes uniformly and independently distributed in a square of unit area. Every node is both a source and a destination. The sources and destinations are paired up one-to-one in a random fashion without any consideration on respective node locations. Each source has the same traffic rate $R(n)$ to send to its destination node under an average power constraint P.[1] The channel between a pair of nodes is modeled by a fading coefficient of the form $r^{\alpha/2}e^{j\theta}$ where r is the distance between the two nodes and θ is a random phase rotation, uniformly distributed over $[0, 2\pi)$ and i.i.d across different pairs. The distances and the phases are assumed to be known to all the nodes in the network.

The aggregate throughput of the system is $T(n) = nR(n)$. Following [2], the delay $D(n)$ of a communication scheme for this network is defined as the average time it takes for a bit or packet of constant size to reach its destination node after it leaves its source node, where the averaging is over all bits or packets traveling inside the network. So defined, the delay of a scheme quantifies the average time spent by the bits traveling inside the network while operated under this scheme.

The following theorem is the main result of this paper.

Theorem 2.1: Using a modified version of the hierarchical cooperation scheme, the following points are achievable on the throughput-delay scaling curve,

$$(T(n), D(n)) = \Theta\left(n^b/\log n, n^b \log n\right)$$

where $0 \leq b < 1$ (see Fig. 1).

III. Overview of the Hierarchical Cooperation Scheme

In this section, we give a brief overview of the hierarchical cooperation scheme as presented in [4] and establish its throughput-delay performance. The reader is referred to [4] for a detailed description of the scheme and its performance analysis.

A. The Three Phase Scheme

The network is divided into clusters of M nearby nodes and a particular source node s sends M bits to its destination node d in three steps:

(1) Node s first distributes its M bits among the M nodes in its cluster, one bit for each node;
(2) These nodes together can then form a distributed transmit antenna array, sending the M bits *simultaneously* to the destination cluster where d lies;
(3) Each node in the destination cluster gets one observation from the MIMO transmission in the previous phase; it quantizes the observation into Q bits, with a fixed Q, and ships it to d, which can then do joint MIMO processing of all the quantized observations and decode the M transmitted bits from s.

From the network point of view, all source-destination pairs have to eventually accomplish these three steps. Step 2 is long-range communication and only one source-destination pair can operate at a time. Steps 1 and 3 involve local communication and can be parallelized across clusters. If TDMA is used to distribute bits and collect MIMO observations in phases 1 and 3 respectively, we need:

[1] In the rest of the paper, we sometimes refer to this traffic pattern as the unicast problem in order to distinguish it from the multicast problem that is discussed in Section IV.

- M^2 time slots to complete phase 1 all over the network since there are a total of $M(M-1) \sim M^2$ bits that needs to be exchanged inside each cluster;
- n time-slots to complete the successive MIMO transmissions for the n source-destination pairs in the network;
- QM^2 time slots to complete phase 3 all over the network since there are a total of $QM(M-1) \sim QM^2$ bits that needs to be exchanged inside each cluster.

In [4], it is shown that each destination node is able to decode the transmitted bits from its source node from the M quantized signals it gathers by the end of Phase 3. The aggregate throughput achieved by the scheme can be calculated as follows: each source node is able to transmit M bits to its destination node, hence nM bits in total are delivered to their destinations in $M^2 + n + QM^2$ time slots, yielding an aggregate throughput of

$$\frac{nM}{M^2 + n + QM^2}$$

bits per time-slot. Choosing $M = \sqrt{n}$ to maximize this expression yields an aggregate throughput $T(n) = \frac{1}{2+Q}\sqrt{n}$.

Note that as opposed to multihop, this three phase scheme allows only bulk transmission between any source-destination pair in the network; i.e. one can not arbitrarily communicate one bit (or L bits with L constant) using the three-phase scheme, but at least $M = \sqrt{n}$ bits should be communicated between all source-destination pairs with each use of the scheme. We say the bulk-size of the scheme is $\sqrt{n}$.

The end-to-end delay of this scheme is simply the total time for the three phases, since the bits are leaving their source nodes at the beginning of the first phase and are only decoded by their respective destination nodes at the end of the third phase. With the choice $M = \sqrt{n}$, we see that the delay of the three phase scheme is $D(n) = (2 + Q)n$. Note that this delay scaling is much worse when compared to the delay of the multi-hop scheme achieving same aggregate throughput.

B. The Hierarchical Cooperation Scheme

Higher aggregate throughput scaling is achieved by using a better network communication scheme than TDMA to establish the transmit and receive cooperations in phases 1 and 3. The traffic demand of exchanging M^2 bits in phase 1 (or QM^2 bits in phase 3) can be handled by setting up M sub-phases, and assigning M pairs in each sub-phase to communicate their 1 bit (or Q bits). The traffic to be handled at each sub-phase now looks similar to the original network communication problem (the unicast network problem defined in Section II), with M users instead of n. Any scheme suggesting a good solution for the original problem can now be used inside the sub-phases as an alternative to TDMA; the multi-hop scheme and the three-phase scheme given in Section III-A would be two alternatives both achieving an aggregate throughput scaling $\Theta(\sqrt{M})$ (in a network of size M) as opposed to the $\Theta(1)$ scaling achieved by TDMA. In general, if a scheme achieving aggregate throughput scaling M^b is used to handle the traffic in each sub-phase, the total completion time for phases 1 and 3 become $M \times M^{1-b}$ and $M \times QM^{1-b}$ respectively. This in turn yields an aggregate throughput

$$\frac{nM}{M^{2-b} + n + QM^{2-b}}$$

bits per time-slot, which is maximized by the choice $M = n^{\frac{1}{2-b}}$, yielding $T(n) = \frac{1}{2+Q} n^{\frac{1}{2-b}}$. Starting with $b = 0$ for TDMA and noticing that $\frac{1}{2-b} > b$ for $0 \le b < 1$, applying the same argument recursively h times, we get a scheme achieving aggregate throughput scaling $T_h(n) = n^{\frac{h}{h+1}}$. Note that this recursion builds a hierarchical architecture with h levels.

For deriving the delay performance of the hierarchical scheme, let us first concentrate on the simplest case $h = 2$. The resultant scheme achieving aggregate throughput scaling $n^{2/3}$ first divides the network into clusters of size $M_1 = n^{2/3}$ and uses the three phase scheme inside these clusters for establishing cooperation. More precisely, the traffic of communicating 1 bit (or Q bits) between M_1 source-destination pairs in each sub-phase of phase 1 (or phase 3) is handled by further dividing the cluster into smaller clusters of size $M_2 = \sqrt{M_1} = n^{1/3}$ and using the three phase scheme (TDMA-MIMO-TDMA) given in Section III-A. Note however that the three phase scheme allows only bulk transmissions between source-destination pairs. In this particular case, one will have to communicate M_2 bits between the source-destination pairs assigned at each sub-phase, as opposed to the original requirement of communicating only 1 bit (or Q bits). For the overall scheme, this in turn increases the bulk size to be communicated between every source-destination pair in the network from M_1 bits to $M_1 \times M_2$ bits, resulting also in larger delay. The delay of the two-level hierarchical scheme is given by $M_2 \times n = n^{4/3}$, as opposed to n for the three phase scheme ($h = 1$). Indeed, it can be checked that the aggregate throughput achieved by the two-level scheme is given by the expression

$$\frac{M_2\, M_1\, n}{M_1(M_2^2 + M_1 + QM_2^2) + M_2 n + M_1 Q(M_2^2 + M_1 + QM_2^2)} \tag{1}$$

and the optimal choices of $M_1 = n^{2/3}$ and $M_2 = n^{1/3}$ maximize the aggregate throughput scaling to $T_2(n) = n^{2/3}$, while the denominator dictating the delay of the scheme is of order $D_2(n) = n^{4/3}$. Note that the increase of the communication bulk size does not affect the throughput, since it corresponds to multiplying the numerator and denominator of (1) by the same factor, but it affects the delay.

Extending the argument for larger h and noticing that the cluster size at the k'th level of an h-level hierarchical scheme is given by $M_k = n^{\frac{h+1-k}{h+1}}$, we obtain the bulk-size in an h-level hierarchical scheme as

$$B_h(n) = M_h \times \ldots \times M_1 = n^{\frac{h}{2}}$$

and its end-to-end delay as

$$D_h(n) = M_h \times M_{h-1} \times \cdots \times M_2 \times n = n^{\frac{h^2+h+2}{2(h+1)}}$$

where we observe that for large h, the delay exponent is linear in h. Recall that the aggregate throughput achieved by an h-level hierarchical cooperation scheme is given by $T_h(n) = n^{\frac{h}{h+1}}$.

The results obtained in this section establish the poor delay performance of hierarchical cooperation. Note that the delay is mostly due to the large bulk-size used by the scheme. This is different from multi-hop schemes since their bulk-size is constant ($\Theta(1)$) independent of the throughput achieved. The delay in multihop is rather due to the time spent in relaying the messages inside the network. In the next section, we modify the scheme so that it achieves the same throughput using much smaller bulk-size.

IV. Hierarchical Cooperation with Smaller Bulk-Size

In this section, we treat the problem of cooperation in the three phase scheme more carefully. We start by defining the network multiple access problem to be the following.

Definition 4.1 (The Network Multiple Access Problem): Consider the assumptions on the network and channel model given in Section II. Let each node in the network be interested in communicating independent information to each of the other nodes in the network. In particular, let us assume that each node has an independent 1 bit message (or L independent bits, with L constant) to send to each of the other nodes in the network and the quantity of interest is the smallest time $F(n)$ required to accomplish this task. This problem we refer to be the network multiple access problem.

The following theorem provides an achievable solution to this problem.

Theorem 4.1: For any integer $h > 0$, the network MAC problem can be solved in

$$F(n) \leq K\, n^{\frac{h+1}{h}}$$

time-slots, for some constant $K > 0$ independent of n.

Proof of Theorem 4.1: Let us start by assuming that there exists a scheme that solves the multiple access problem in $F(n) = n^b$ time-slots with $b > 1$. Note that one such scheme is simple TDMA that yields $b = 2$. Using this existing scheme, we will construct a new scheme that yields smaller $F(n)$.

As before, let us start by dividing the network into clusters of M nearby nodes. Let us first focus on one specific cluster S and one node d located outside of this cluster. In particular, all nodes in S have 1 bit to send to d. These bits can be communicated to d in two steps:

(1) The nodes in S *simultaneously* transmit their 1 bit messages destined to d forming a distributed transmit antenna array for MIMO transmission. The nodes in the destination cluster which d belongs to, form a distributed receive antenna array for this MIMO transmission.

(2) Each node in the destination cluster obtains one observation from the MIMO transmission in the previous phase; it quantizes and ships this observation to d, which can do joint MIMO processing of all the observations and decode the M transmitted bits from the nodes in S.

As a first step towards handling the whole network problem, note that these two steps should be accomplished between S and all other nodes in the network. This can again be done in two steps:

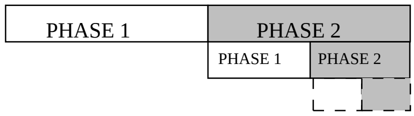

Fig. 2. The figure illustrates the time-division in the hierarchical scheme that solves the network multiple access problem.

Phase 1: MIMO transmissions We perform successive MIMO transmissions between S and all other nodes in the network. In each of the MIMO transmissions, say between S and d, the M nodes in S are simultaneously transmitting the 1 bit messages they would like to communicate to d and the M nodes in the cluster where d lies are observing the MIMO transmission. The MIMO transmissions should be repeated for each node in the network, hence we need at most n time-slots to complete the phase.

Phase 2: Cooperate to decode Clusters work in parallel. Since there are M nodes inside each cluster, each cluster received M MIMO transmissions from S in the previous phase, each transmission intended for a different node in the cluster. Thus, each node in the cluster has M observations, one from each of the MIMO transmissions, and each observation is intended for a different node in the cluster. Each of these observations can be quantized into Q bits, with a fixed Q, which yields exactly the original network multiple access problem, with M nodes instead of n. Using the scheme we assumed to exist in the beginning of the proof, this task can be completed in QM^b time slots.

The total time we have spent during the two phases for handling the traffic originated from cluster S is given by $n + QM^b$. From the network point of view, the above two steps should be completed for all n/M clusters in the network. Thus, the multicasting task can be completed in $\frac{n}{M}(n + QM^b)$ time slots in total. Choosing $M = n^{\frac{1}{b}}$ in order to minimize this quantity yields $F(n) = (1 + Q)n^{2-\frac{1}{b}}$.

Note that $2 - \frac{1}{b} < b$ for $b > 1$. In other words, we have established a solution for the multiple access problem that is better than the one we started with. Indeed, the two phase scheme described above can be used recursively yielding a better scheme at each step of the recursion. In particular, starting with TDMA achieving $b = 2$ and applying the idea recursively h times, one gets a scheme that solves the multiple access problem in $\Theta(n^{\frac{h+1}{h}})$ time slots. The operation of this scheme is illustrated in Fig. 2. □

The interest in the multiple access problem arises from the fact that it exactly models the required traffic for cooperation in the three phase scheme. Recall that this traffic has been handled by decomposing it into a number of unicast network problems in Section III-B. The resultant hierarchical cooperation scheme was optimal in terms of throughput, but not very satisfying in terms of bulk-size. By using the solution to the multiple access problem suggested in this section, one can modify the hierarchical cooperation scheme, so as to achieve the same throughput with smaller bulk-size and consequently smaller delay. Note that the gain is coming from treating the cooperation problem as it is and not necessarily as multiple unicast problems as was previously done in Section III-B.

Corollary 4.1: A modified hierarchical cooperation scheme can achieve an aggregate throughput $T_h(n) \geq K_1 n^{\frac{h}{h+1}}$ with bulk-size $B_h(n) = K_2 n^{\frac{h}{h+1}}$ and delay $D_h(n) \leq K_3 n$, for any integer $h \geq 0$ and some positive constants K_1, K_2, K_3 independent of n.

Proof of Corollary 4.1: Consider the three phase hierarchical scheme described in Section III-A. By Theorem 4.1, the required traffic for transmit and receive cooperation in phase 1 and phase 3 can be handled in $KM^{\frac{h+1}{h}}$ and $KQM^{\frac{h+1}{h}}$ time slots respectively. The expression for the aggregate throughput then becomes

$$\frac{Mn}{KM^{\frac{h+1}{h}} + n + KQM^{\frac{h+1}{h}}}$$

which is maximized by the choice $M = n^{\frac{h}{h+1}}$, yielding aggregate throughput $T_h(n) = \frac{1}{1+K+KQ} n^{\frac{h}{h+1}}$, bulk-size $B_h(n) = M = n^{\frac{h}{h+1}}$ and delay $D_h(n) = (1 + K + KQ)n$. □

V. Hierarchical Cooperation with Better Scheduling

In the previous section, we presented a modified hierarchical scheme that achieves throughput $T_h(n) = \Theta(n^{\frac{h}{h+1}})$ using bulk-size $B_h(n) = \Theta(n^{\frac{h}{h+1}})$. However, the delay of this scheme is still $D_h(n) = \Theta(n)$. In this section, we optimize the scheduling in the scheme to further improve the delay performance to $D_h(n) = \Theta(n^{\frac{h}{h+1}} \log n)$. In the following section, we briefly summarize the scheduling idea for the three phase scheme with $h = 1$ discussed in Section III-A. The scheduling for the modified hierarchical scheme discussed in Section IV is omitted due to space limitations.

A. Better Scheduling for the Three Phase Scheme

Recall the operation of the three phase scheme from the point of view of a single source-destination pair s-d as described in Section III-A: a step (1) where s distributes its M bits among the M nodes in its cluster using TDMA, followed by a step (2) where these M bits are simultaneously transmitted to the destination cluster via MIMO transmission, and a step (3) where the quantized MIMO observations are collected at the destination node d by again using TDMA. These three steps need to be eventually accomplished for each source-destination pair in the network. In this section, we improve the scheduling in accomplishing this task: we organize M successive sessions and allow only n/M source-destination pairs to complete the three steps in each session.

In the beginning of each session we randomly choose *one* source node from each cluster, thus n/M source nodes in total. In general, the n/M destination nodes corresponding to these randomly chosen source nodes can be located anywhere. However, since the source-destination pairs are formed randomly, no more than $\log n$ of these destination nodes are located in the same cluster with high probability. Thus, the three steps can be accomplished for the n/M chosen source-destination pairs in $M + n/M + QM \log n$ time slots.

The operation continues with the next session by choosing a new set of n/M source nodes, one source node per cluster, randomly among the nodes that have not yet accomplished the above three steps. Note that all source-destination pairs will accomplish the these steps in a total of M sessions.

With this rather smoother operation on the network level, we accomplish to serve n/M source-destination pairs in each session, that is transfer $M \times \frac{n}{M}$ bits in total to their destinations in $M + \frac{n}{M} + QM \log n$ time slots yielding aggregate throughput

$$\frac{M \times \frac{n}{M}}{M + \frac{n}{M} + QM \log n} \tag{2}$$

which is maximized by the choice $M = \sqrt{n}$ yielding aggregate throughput $T(n) = \frac{1}{2+Q} \frac{\sqrt{n}}{\log n}$. The delay experienced by each bit is now much less compared to the three phase scheme in Section III-A, since it is again dictated by the total time spent in the three phases (denominator of (2)), which is now less than $D(n) = (2 + Q)\sqrt{n} \log n$.

Note that instead of choosing $M = \sqrt{n}$, which is the optimal choice to maximize the throughput achieved by the scheme, one can choose $M = n^b$ with $0 \leq b \leq 1/2$. In this case, we also restrict the number of source-destination pairs to be served in each session to M, which can now be less than the total number of clusters n/M. Indeed, we operate one source node in each of the $M(\leq n/M)$ clusters and simply keep the remaining clusters inactive. The expression for the aggregate throughput becomes

$$\frac{M \times M}{M + M + QM \log n}$$

which implies that the scheme achieves aggregate throughput $T(n) = n^b / \log n$ and delay $D(n) = n^b \log n$ for any $0 \leq b \leq 1/2$. Note that this throughput-delay trade-off differs only by $\log n$ from the trade-off achieved by multi-hop schemes.

Applying the same kind of a scheduling idea to the modified hierarchical scheme in Section IV, we can show that the delay of this scheme can be reduced to $D_h(n) = \Theta(n^{\frac{h}{h+1}} \log n)$ and that all points on the throughput-delay scaling curve $(T(n), D(n)) = (n^b / \log n, \, n^b \log n)$ for any $0 \leq b < 1$ are achievable. However, the rest of the proof is omitted here due to space limitations.

References

[1] A. El Gamal, J. Mammen, B. Prabhakar, D. Shah, *Optimal Throughput-Delay Scaling in Wireless Networks-Part I: The Fluid Model*, IEEE Trans. on Information Theory 52(6), pp. 2568-2592, 2006.

[2] A. El Gamal, J. Mammen, B. Prabhakar, D. Shah, *Optimal Throughput-Delay Scaling in Wireless Networks-Part II: Constant-Size Packets*, IEEE Trans. on Information Theory 52(11), pp. 5111-5116, 2006.

[3] P. Gupta and P. R. Kumar, *The Capacity of Wireless Networks*, IEEE Trans. on Information Theory 42 (2), pp. 388–404, March 2000.

[4] A. Özgür, O. Lévêque, D. Tse, *Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks*, IEEE Transactions on Information Theory, Vol. 53 (10), pp. 3549-3572, October 2007.

[5] M. Grossglauser and D. Tse, *Mobility Increases the Capacity of Ad-hoc Wireless Networks*, IEEE/ACM Transactions on Networking 10(4), pp. 477-486, 2002.

[6] M. J. Neely and E. Modiano, *Capacity and Delay Tradeoffs for Ad-Hoc Mobile Networks*, IEEE Transactions on Information Theory, Vol. 51 (6), pp. 1917-1937, June 2005.

[7] G. Sharma, R. R. Mazumdar and N. B. Shroff, *Delay and Capacity Trade-offs in Mobile Ad Hoc Networks: A Global Perspective*, 2006 IEEE INFOCOM Conference, Barcelona, Spain, April 2006.