On the Scalability of Hierarchical Hybrid Wireless Networks

Suli Zhao and Dipankar Raychaudhuri
WINLAB, ECE Department, Rutgers University
671 Route 1 South, North Brunswick, NJ 08902
{sulizhao, ray}@winlab.rutgers.edu

***Abstract*— This paper presents an analysis of the scaling properties of a three-tier hierarchical hybrid wireless network. The network under consideration, which arises in mesh networking scenarios such as 802.11s, aims to achieve better capacity than ad hoc networks without infrastructure support, and also reduces the investment on wired infrastructure. In particular, the hierarchical hybrid network has three tiers consisting of mobile nodes, radio forwarding nodes and wired access points. For a three-level network of n_1 access points, n_2 forwarding nodes and n_3 mobile nodes, we analyze throughput in terms of two tiers of packets: those transmitted by mobile nodes (low-tier) and those transmitted by forwarding nodes and access points (high-tier). It is shown that low-tier capacity increases linearly with n_2, and high-tier capacity increases linearly with n_1 when n_1 grows asymptotically faster than $\sqrt{n_2}$. These results, which are consistent with earlier simulation studies, demonstrate the value of adding radio forwarding nodes to improve scaling behavior and reduce the required number of wired access points. In order to model the capacity of the proposed network, we also study the capacity and traffic distribution of *Random Aggregate Networks*.**

***Index Terms*— Multi-hop wireless networks, hybrid network, ad hoc mesh, throughput, capacity, scalability.**

I. Introduction

Ad hoc ("mesh") wireless networks offer important benefits including rapid deployment and low cost. However, it is well known from Gupta and Kumar [1], that the capacity of a traditional "flat" multi-hop network does not scale well, i.e. throughput per node decreases as the number of nodes in the network, n, becomes large. This motivates consideration of more scalable ad hoc network architectures, possibly based on hierarchical approaches. When considering scalability, it is also important to note that most applications involve traffic flows to and from the Internet in addition to peer-to-peer communication between radio nodes. Also, adding infrastructure nodes to ad hoc networks can effectively reduce the mean number of hops from source to destination, and help produce better performance than flat ad hoc networks [2]. These results show that ad hoc mesh networks benefit from a hierarchical "hybrid" wired/wireless architecture both in terms of scalability and effective integration with the Internet.

Recent results have shown that the asymptotic capacity of two-tier hybrid networks can be improved through deployment of wired infrastructure nodes [2–4]. Infrastructure support increases the network capacity, but wired infrastructure cost can be high, especially for dense networks.

Based on the above considerations, we have proposed a new class of self-organizing hierarchical hybrid wireless networks [5, 6]. As shown in Fig. 1, the proposed network architecture is based on three tiers of wireless devices: low-power "mobile nodes (MN's)" with limited functionality (e.g. no routing capability), higher-power "radio forwarding nodes (FN's)" that route packets between radio links, and "access points (AP's)" that route packets between radio links and the wired infrastructure.

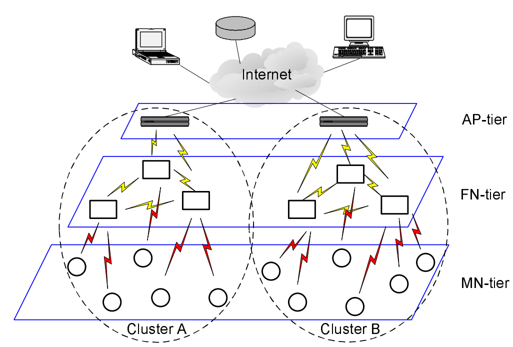

Fig. 1. Hierarchical hybrid wireless network architecture

The scaling properties of the proposed hierarchical architecture have been studied using simulations in [5]. It was shown that the hierarchical network capacity can scale well with a mix of several (lower-cost) FN's and just a few AP's. The simulation results in [5] seemed to show a rough square-law relationship between the density of FN's and AP's for scalability to be maintained. In this paper, we develop a general analytical model for the asymptotic capacity and scaling properties of three-tier hierarchical ad hoc wireless networks.

The rest of this paper is organized as follows. In Section II we discuss related work. Next we describe the three-tier hierarchical network model, separating packets in the network into low-tier and high-tier transmissions. In Section IV, we define and study *Random Aggregate Networks* in order to model the capacity of the hierarchical system under consideration. In Section V, we present the analytical results of aggregate

Research supported part by NJCST Grant #03-2042-007-12 and part by NSF NRT Grant #ANI0335244.

capacity of the three-tier hierarchical hybrid network and also provide a discussion. Section VI summarizes our conclusions.

II. Related Work

In [1], Gupta and Kumar obtain the capacity of multi-hop wireless networks with n identical randomly located nodes, each capable of transmitting at W bits per second, using a fixed range and under a noninterference protocol, which is $\Theta(\frac{W}{\sqrt{n \log n}})$ bits per second per node for randomly chosen destinations.

Liu and Towsley extended the work to two-level hybrid wireless networks [3]. In their model, there are n ad hoc nodes which are randomly distributed in the network, and m base stations (BS's) which are placed on a regular grid and construct hexagonal cells. In the *1-nearest-cell* routing strategy, the infrastructure is used for inter-cell transmissions while ad hoc mode is used for intra-cell transmissions. W_1 and W_2 are the bandwidth allocated to infrastructure and ad hoc mode transmissions. The per cell capacity contributed by infrastructure mode transmissions is $\Theta(W_1)$. The aggregate throughput capacity contributed by ad hoc mode transmissions has different forms for different scaling regimes: if $m = o(\sqrt{n})$, the corresponding aggregate capacity is $\Theta\left(W_2\sqrt{\frac{n}{\log{(n/m^2)}}}\right)$; if $m = \Omega(\sqrt{n})$, the corresponding aggregate capacity is $O(W_2\sqrt{n})$. When $m = \Omega(\sqrt{n})$, the maximum aggregate capacity is achieved by allocating all the bandwidth to carry inter-cell transmissions, and increases linearly with m.

The work of [2] studies the capacity when both ad hoc nodes and AP's are randomly distributed. Suppose that n is the number of ad hoc nodes and W is the wireless bandwidth, it gives the asymptotic per node capacity to be $\Theta(W/\log(n))$ bits per second, provided that the number of AP's scales linearly with n. The work of [4] investigates the asymptotic per user throughput for different scaling regimes for a random hybrid network with arbitrarily placed BS's.

Although providing different bounds for different scaling regimes, the priori work discussed before shows that, in order to obtain a significant improvement in capacity for hybrid networks, infrastructure investments need to be high. The priori work also shows that the scaling regime is related to the number of ad hoc nodes. In our proposed architecture, the middle-tier (FN's) aggregates the traffic of ad hoc nodes, and provides more economical scaling regimes for the grow of the number of infrastructure nodes (AP's) in terms of the number of FN's rather than ad hoc nodes (MN's). Thus the infrastructure cost can be greatly reduced.

In [7], the authors study the achievable throughput of a randomly deployed flat network using multi-hop transmission for any-to-one communication, which has the same asymptotical expression as our result of Random Aggregate Networks in Section IV. In our analytical approach, we obtain not only the asymptotical throughput capacity of the Random Aggregate Network, but also the traffic distribution among the network, which could be helpful in designing scheduling algorithms for such network.

In next sections, we will show that the significant capacity improvement can be obtained with our proposed three-tier hybrid network with less AP's than two-tier hybrid networks. Our analysis uses some results of [1] and [3].

III. System modeling

In the three-tier hierarchical network under consideration, MN's are in the lowest tier and perform end-user functions such as mobile computing or sensing/actuation. MN's do not forward packets for others but send out and receive their own packets. FN's only forward packets for other nodes (i.e. do not act as data sources/sinks), and work as the intermediate tier between the MN- and AP-tiers. AP's are interconnected through the wired network infrastructure. We assume that all packets between MN's and AP's must go through the FN-tier, even when the MN is only one hop away from the AP-tier.

There are two frequencies, f_1 and f_2, used in the network. f_1 is used for transmissions to and from MN's; f_2 is used for transmissions not involving MN's. Therefore, there is no interference between transmissions involving MN's (denoted low-tier transmissions) and transmissions to/from higher-tier FN's and AP's (denoted high-tier transmissions). Each FN is equipped with two radios, and can thus participate in both low- and high-tier transmissions using different radios.

All nodes in the network are assumed to have the same transmission range. We do not consider the mobility of MN's, and also do not take into account the capacity improvement that could be brought in by mobility as discussed in [8]. Through the ad hoc network discovery procedure, each MN is associated with the nearest FN, and we further assume that there is always at least one FN within its range. Each FN is associated with its nearest AP via one-hop or multi-hop transmissions. Thus, low-tier transmissions are all one-hop; while high-tier transmissions are possibly multi-hop using a mix of wireless links and wired infrastructure paths.

We suppose there are n_1 AP's, n_2 FN's and n_3 MN's ($n_1 < n_2 < n_3$) in a disk of area 1 square meter on the plane. FN's and MN's are independently and uniformly distributed, while AP's are placed in a regular pattern like in [3]. The disk is divided into clusters. Each cluster consists of one AP, its associated FN's and their associated MN's. We assume the node on frequency f_1 and f_2 can transmit at W_1 and W_2 bits per second, respectively.

A TDMA (time division multiple access) scheme is used for data transmissions over wireless channels. Time is divided into slots of fixed durations. In each cluster, only one node can transmit data in each time slot on each frequency. Since there is interference between clusters on each frequency, the spatial and temporal transmission schedule has to be deployed on each frequency. The existence of the transmission schedule has been proved in [1] and [3].

A. Protocol model

The Protocol Model from [1] is used for successful reception of a transmission over one hop. All nodes employ the same transmission range r. A transmission from node X_i is successfully received by node X_j if

- The distance between X_i and X_j is no more than r, i.e., $|X_i - X_j| \leq r$.
- For every other node X_k simultaneously transmitting over the same channel $|X_k - X_j| \geq (1 + \Delta)r$.

where $\Delta > 0$ defines the size of the guard zone.

B. Traffic pattern

There are three kinds of traffic in the wireless network: (1) traffic from MN's to AP's (Internet uplink), (2) traffic from AP's to MN's (Internet downlink), and (3) traffic between MN's (local traffic).

Since low- and high-tier transmissions use different frequencies, and there is no interference between them, we can separate the traffic into two parts: one is in the low-tier and the other in the high-tier. We analyze capacity separately for these two parts. Before doing this, we first define and study *Random Aggregate Networks*, and derive their capacity.

IV. Random Aggregate Networks

In a random aggregate network scenario, n nodes are independently and uniformly distributed in a disk of area 1 square meter on the plane. All nodes have a unique destination at the center of the disk, to which each of them wishes to transmit packets at the rate of $\lambda(n)$ bits per second. Note that in the random network scenario [1], destinations are randomly chosen. We further assume that all nodes, including n source nodes and one destination node, use the same and fixed transmission power, and are capable of transmitting at W bits per second. Also we use the Protocol Model for interference.

As in previous work, a throughput of $\lambda(n)$ bits per second for each source node is *feasible* if there exists a spatial and temporal scheme of transmissions, such that each source node can transmit $\lambda(n)$ bits per second on average to its destination node.

The *throughput capacity* of Random Aggregate Networks is of order $\Theta(f(n))$ bits per second if there are deterministic constants $c_1 > 0$ and $c_2 < +\infty$ such that

$$\lim_{n\to\infty} \text{Prob } (\lambda(n) = c_1 f(n) \text{ is feasible }) = 1$$
$$\lim_{n\to\infty} \text{Prob } (\lambda(n) = c_2 f(n) \text{ is feasible }) < 1.$$

Theorem 1: For Random Aggregate Networks on a planar disk in the Protocol Model, the order of aggregate throughput capacity is

$$T = \Theta(W) \quad \text{bits per second.} \tag{1}$$

We use a Voronoi tessellation $\mathcal{V}_n$ of the planar disk to prove *Theorem 1*. The edge effects are ignored. It can be shown that two special properties of *Lemma 4.1* in [1] hold for the planar disk, rewritten as follows:

Lemma 1: For every $\epsilon > 0$, there is a Voronoi tessellation of the disk on the plane with the property that every Voronoi cell contains a disk of radius ϵ and is contained in a disk of radius 2ϵ.

Let $R_v = \sqrt{\frac{100 \log n}{\pi n}}$. Each Voronoi cell $V \in \mathcal{V}_n$ contains a disk of radius R_v and is contained in a disk of radius $2R_v$. Let the transmission range $r = 8R_v$.

We denote by L_i the straight-line segment connecting a source node, X_i, $i = 1, \ldots, n$, to the only destination node, O, which is the center of the disk. As in [1], the routes are chosen to approximate the straight-line segments. The straight-line segments intersect cells in $\mathcal{V}_n$. The packet is relayed from the source cell to the center cell (the cell contains the destination node O, denoted C_O) in a sequence of hops. In each hop, the packet is transferred from one cell to another in the order it intersects the straight-line segment (since there is always direct communication between adjacent cells, according to *Lemma 4.2* of [1]). At last, the packet will be sent to the destination when it arrives at the cell which is adjacent to the center cell. It has been proved in [1] that the multi-hop relaying scheme as described works with high probability. It can also be shown that in the planar disk, this routing is efficient and the load is balanced among the sectors of the disk.

Below we compute the mean number of routes served by each cell. First we consider all cells except the center cell.

A. Cell $V \in \mathcal{V}_n \setminus \{C_O\}$

We bound the probability that a segment L_i intersects a given cell V by computing the probability that L_i intersects the outer disk which contains V (denoted D_V).

Lemma 2: For segment L_i and cell $V \in \mathcal{V}_n \setminus \{C_O\}$,

$$\text{Prob}(L_i \text{ intersects } V) \leq \frac{c_3}{x}\sqrt{\frac{\log n}{n}}(1 - \pi x^2) \tag{2}$$

where x is the distance of the center of D_V from O.

Proof:

Suppose cell V lies at a distance x from O. The angle α subtended at O by D_V, which is of radius $2R_v$, is no more than $\frac{c_4}{x}\sqrt{\frac{\log n}{n}}$. The area of the sector formed is no more than $\frac{c_5 \alpha}{2\pi}$. D_V is inscribed in this sector and divides the sector into three parts: one is the disk itself, the second is the area close to O, and the third is the area to the other side of O. If X_i lies in the third part of the sector, the segment L_i which connects X_i and O will intersect D_V, as shown in Fig.2.

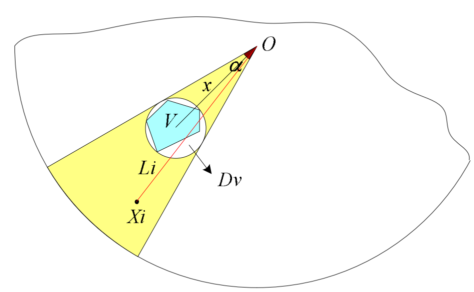

Fig. 2. Voronoi cell V

The sum of the areas of the first two parts are greater than the area of a sector of radius x. Thus the area of the third part is no more than the area of the whole sector subtracted by the area of the sector of radius x, i.e., $\frac{c_5\alpha}{2\pi}(1-\pi x^2)$. Therefore, it gives (2).

■

There are n lines $\{L_i\}_{i=1}^n$, each of which connects X_i with O. Then the mean number of lines passing through a cell which is at a distance x from O and not the center cell is bounded as

$$\begin{aligned} &E[\text{Number of lines in } \{L_i\}_{i=1}^n \text{ intersecting V}] \\ &\leq \frac{c_3}{x}\sqrt{n\log n}(1-\pi x^2) \end{aligned} \tag{3}$$

It is observed that the mean number of lines intersecting a given cell is the function of the distance of this cell from the center, and the lines passing through get denser when the cell is closer to the center. The traffic handled by a cell is proportional to the number of lines passing through it.

For cells close to the center, $x \ll (1/\sqrt{\pi})$ holds, thus $1-\pi x^2 \cong 1$. Then we have, when $x \ll (1/\sqrt{\pi})$,

$$E[\text{Number of lines intersecting V}] \leq \frac{c_3}{x}\sqrt{n\log n}. \tag{4}$$

It is implied from (4) that the number of traffic handled by a cell which is close to the center is proportional to $(1/x)$, where x is the distance from the cell to the center. In order to maximize the capacity of a Random Aggregate Network, it suggests that the resource allocated to each cell has the relationship as (4). Therefore, when the infrastructure is used as a short cut for traffic that needs to traverse a long distance, infrastructure nodes would become traffic hotspots. Equation (4) provides a basis for designing scheduling algorithms to overcome the capacity bottleneck at the hotspots in close proximity to infrastructure nodes.

B. The center cell C_O

For the center cell C_O, since it contains the final destination node, it follows

$$\begin{aligned} \text{Prob}(L_i \text{ intersects } C_O) &= 1 \\ E[\text{Number of lines in } \{L_i\}_{i=1}^n \text{ intersecting } C_O] &= n \end{aligned}$$

Suppose each line L_i carries traffic of rate λ bits per second. The rate at which the center cell needs to transmit is $n\lambda$. Since the center cell has the maximum number of lines passing through, it carries the maximum traffic among all the cells in the disk, i.e.

$$\frac{c_3}{x}\sqrt{n\log n}(1-\pi x^2)\,\lambda < n\,\lambda$$

From *Lemma 4.4* of [1] we know that there exists a schedule for transmitting packets such that in every $(1+c_6)$ slots (c_6 is a constant), each cell gets one slot to transmit (suppose all cells have the same chance to obtain slots to transmit). Thus the rate at which each cell gets to transmit is $W/(1+c_6)$ bits per second. The traffic can be accommodated if it is less than the rate available. Using the even scheduling scheme as described, the traffic handled by the whole network is restricted by the center cell. Therefore it follows

$$\frac{c_3}{x}\sqrt{n\log n}(1-\pi x^2)\,\lambda < n\,\lambda \leq \frac{W}{1+c_6}$$

Thus

$$\lambda \leq \frac{W}{(1+c_6)\,n} \tag{5}$$

Equation (5) gives the lower bound of the per node capacity. For the upper bound, since C_O can only handle data at rate of W bits per second, the aggregate capacity is upper bounded by W. Therefore the order of the aggregate throughput capacity is give by

$$T = \Theta(W) \tag{6}$$

We note that the aggregate capacity of the Random Aggregate Network shown in (6) has the same form as the capacity contributed by the pure infrastructure mode communications in the two-tier hybrid network [3], which implies that the asymptotic capacity of the network having the aggregation traffic pattern is independent of the number of hops required to reach the aggregation node.

V. Capacity of three-tier hierarchical hybrid wireless networks

A. Low-tier transmissions

Low-tier transmissions use frequency f_1 with bandwidth W_1. Each MN communicates with the nearest FN, and there is always at least one FN within its transmission range. Therefore, the traffic is localized and all transmissions are one-hop in this tier.

The disk is divided into sub-clusters, each of which consists of one FN and its associated MN's. Thus there are n_2 sub-clusters in the disk. All low-tier transmissions have to go through the associated FN's, and each FN can only handle data at rate of W_1 bits per second at any time. Therefore the per sub-cluster throughput capacity, T_{l_sc}, is upper bounded by W_1. For the lower bound, since each MN is one-hop away from its associated FN, there is a schedule for each MN to communicate with its associated FN in a round robin fashion, resulting in a throughput of W_1. Hence, it follows

$$T_{l\_sc} = \Theta(W_1)$$

Lemma 1 and *2* of [3] state that in the Protocol Model, there is a spatial schedule that each cell gets one slot to transmit in every constant number of slots, and this constant only depends on Δ. Similarly, there is a scheduling scheme such that each sub-cluster gets a slot to transmit in every constant number of time slots. Therefore, the aggregate throughput capacity contributed by low-tier transmissions, denoted T_l, is given as

$$T_l = n_2\,T_{l\_sc} = \Theta(n_2 W_1) \tag{7}$$

Suppose each MN carries traffic of rate λ bits per second on an average, then

$$T_l = n_3\lambda = \Theta(n_2 W_1)$$

It is observed that the aggregate throughput capacity contributed by low-tier transmissions increases linearly with the number of FN's. The explanation is that FN's work as relay nodes and aggregate the traffic for MN's. It suggests that we can increase the number of FN's to accommodate the traffic of the network. Furthermore, since FN's are not deterministically placed, they can be located where capacity is needed rather than where wired connections are available.

B. High-tier transmissions

After taking away low-tier transmissions, we categorize the remaining part of traffic, which is contributed by high-tier transmissions, into three classes:

- Class 1: traffic from FN's to AP's.
- Class 2: traffic from AP's to FN's.
- Class 3: traffic between FN's.

The bandwidth allocated to Class 1, 2 and 3 traffic are W_3, W_4 and W_5, respectively. Let $W_3 + W_4 + W_5 = W_2$.

1) Class 1 traffic: In the hierarchical network, there are n_1 clusters. Applying Theorem 1, the aggregate throughput capacity contributed by Class 1 traffic is given by

$$T_{h\_1} = \Theta(n_1 W_3) \tag{8}$$

2) Class 2 traffic: Similarly, applying Theorem 1, the aggregate throughput capacity contributed by Class 2 traffic is given as

$$T_{h\_2} = \Theta(n_1 W_4) \tag{9}$$

3) Class 3 traffic: Depending on the locations of the communicating pairs, there are two kinds of Class 3 traffic: intra-cluster and inter-cluster traffic. When the source and destination FN's are in the same cluster, it is intra-cluster traffic. Otherwise, it is inter-cluster traffic. Suppose the bandwidth allocated to inter-cluster and intra-cluster traffic are W_6 and W_7, respectively, where $W_6 + W_7 = W_5$.

a) Inter-cluster traffic: We assume that inter-cluster traffic always goes through the infrastructure (AP's). In particular, the inter-cluster traffic enters the infrastructure at the source FN's associated AP and leaves it at the destination FN's associated AP. When FN's are multiple hops away from their associated AP's, the ad hoc mode is used to route packets between FN's where one end is the source/destination FN and the other is the FN adjacent to the associated AP. Thus, the mix of the infrastructure and ad hoc modes (denoted the mixed mode) might be used for inter-cluster traffic. Note that this may not be the optimal route. For instance, if the source and destination FN's are neighbors, it may be preferable to have direct communication between them via single hop transmission rather than using the infrastructure. However, we use the multi-hop FN-AP-FN path assumption to simplify the analysis.

According to our assumption, one inter-cluster communication can be decomposed into two parts: one is from FN to AP in the source cluster, and the other is from AP to FN in the destination cluster. Only one part is counted in the throughput capacity.

For each cluster, we can apply Theorem 1 and obtain the aggregate throughput capacity contributed by Class 3 inter-cluster traffic as follows

$$T_{h\_3\_inter} = \Theta(n_1 W_6) \tag{10}$$

b) Intra-cluster traffic: AP's do not participate in transferring this kind of traffic. According to the scaling property of n_1 with respect to n_2, there are two cases: $n_1 = o(\sqrt{n_2})$ and $n_1 = \Omega(\sqrt{n_2})$. We apply the results of [3], which have been briefly discussed in Section II, to the analysis below.

(1) $n_1 = o(\sqrt{n_2})$: There are (n_2/n_1) FN's in each cluster and only FN's are involved in transmissions. Applying the results of [3], the aggregate capacity contributed by Class 3 intra-cluster traffic is given as

$$T_{h\_3\_intra} = \Theta\left(W_7\sqrt{\frac{n_2}{\log(n_2/n_1^2)}}\right) \tag{11}$$

Therefore, the aggregate throughput capacity contributed by all Class 3 traffic is given as:

$$T_{h\_3} = \Theta(n_1 W_6) + \Theta\left(W_7\sqrt{\frac{n_2}{\log(n_2/n_1^2)}}\right) \tag{12}$$

when $n_1 = o(\sqrt{n_2})$.

Applying *Corollary 1* of [3], the aggregate throughput capacity contributed by Class 3 traffic is maximized when $W_6/W_5 \to 0$ or equivalently, $W_7/W_5 \to 1$. And the achieved capacity is given as:

$$T_{h\_3\_max} = \Theta\left(W_5\sqrt{\frac{n_2}{\log(n_2/n_1^2)}}\right) \tag{13}$$

when $n_1 = o(\sqrt{n_2})$. Hence in this case it is more beneficial to assign bandwidth to intra-cluster traffic.

(2) $n_1 = \Omega(\sqrt{n_2})$: According to the results of [3], when $n_1 = \Omega(\sqrt{n_2})$, the aggregate throughput capacity contributed by Class 3 intra-cluster traffic is given as

$$T_{h\_3\_intra} = O(\sqrt{n_2}\, W_7) \tag{14}$$

Therefore, the aggregate throughput capacity contributed by all Class 3 traffic is given as:

$$T_{h\_3} = \Theta(n_1 W_6) + O(\sqrt{n_2}\, W_7) \tag{15}$$

when $n_1 = \Omega(\sqrt{n_2})$.

When $W_6/W_5 \to 1$, the aggregate throughput capacity contributed by Class 3 traffic is maximized, and the achieved capacity is

$$T_{h\_3\_max} = \Theta(n_1 W_5) \tag{16}$$

when $n_1 = \Omega(\sqrt{n_2})$. This shows that it is more effective to allocate bandwidth to carry inter-cluster traffic in this case, and the maximum capacity achieved increases linearly with n_1.

4) Maximum capacity of high-tier transmissions: As in Section V-A, it can be shown that there is a scheduling scheme such that each cluster gets a slot to transmit in every constant number of time slots. Then the maximum aggregate throughput capacity contributed by all the three classes of high-tier transmissions, T_{h_max}, can be achieved.

For $n_1 = o(\sqrt{n_2})$,

$$T_{h\_max} = n_1(T_{h\_1} + T_{h\_2} + T_{h\_3\_max}) = \Theta\big[n_1(W_3 + W_4)\big] + \Theta\Big(W_5\sqrt{\frac{n_2}{\log(n_2/n_1^2)}}\Big) \quad (17)$$

For $n_1 = \Omega(\sqrt{n_2})$,

$$T_{h\_max} = \Theta\big[n_1(W_3 + W_4 + W_5)\big] = \Theta(n_1 W_2) \quad (18)$$

Equation (18) reveals that if all the bandwidth of high-tier transmissions are allocated to the communication through the infrastructure network and the number of AP's is the same order of the square root of the number of FN's, the maximum capacity can be achieved, which has the linear relationship with the number of AP's. This capacity is shared among the nodes that are routed through the infrastructure as determined by the deployed routing scheme.

C. Discussions

It is shown that in a three-tier hierarchical hybrid wireless network of n_1 access points, n_2 forwarding nodes and n_3 mobile nodes, the capacity contributed by low-tier transmissions increases linearly with n_2. In addition, the capacity contributed by high-tier transmissions increases linearly with n_1 when n_1 grows asymptotically faster than $\sqrt{n_2}$. Since both low- and high-tier transmissions are involved in each traffic flow, numerically the capacity contributed by low-tier transmissions needs to be the same as that contributed by high-tier transmissions. Therefore, when designing the three-tier hierarchical network, we can adjust n_1 and n_2 such that FN's can accommodate the network traffic and the scaling properties can be achieved by satisfying $n_1 = \Omega(\sqrt{n_2})$. Moreover, since FN's are not deterministically placed, it is easy to position them where capacity is needed rather than where wired connections are available.

It is also observed from (18) that the linear relationship for capacity exists when $n_1 = \Omega(\sqrt{n_2})$, no matter what the fractions of the three kinds of traffic are. In other words, there are no particular rules to allocate bandwidth to different traffic in the system in such condition.

In a two-tier hybrid network of n nodes and m base stations as proposed in [3], the maximum capacity increases linearly with m when $m = \Omega(\sqrt{n})$. In the hybrid network model in [3], the pure infrastructure mode is used for traffic going through BS's, which implies that BS's provide full coverage over the network. In the three-tier hierarchical network, AP's and MN's are data sources/sinks and FN's work as the intermediate tier between them. In this type of three-tier network, AP's do not need to cover the whole area and their number is only required to grow asymptotically faster than the square root of the number of FN's rather than the square root of the number of MN's for scalability to be maintained. Therefore, for serving the same amount of traffic, the number of AP's required can be reduced by adding a new tier of FN's. Since the investment and recurrent wired access cost for AP's or BS's is significantly greater than that for FN's, system costs can be significantly reduced while achieving the same scaling properties.

The above asymptotic capacity and scaling regimes are obtained by deploying a specific routing approach, extended from the work on the two-tier network of [3]. The work in [4] uses a different system model and routing scheme, leading to a different asymptotic capacity and scaling regime. It can be shown that the improvement in capacity and infrastructure cost have similar properties when extending the results of [4] to our proposed three-tier architecture.

VI. Conclusions and Future Work

The scaling properties of a three-tier hierarchical hybrid wireless network have been studied and related throughput capacity bounds have been obtained. The concept of Random Aggregate Networks has been defined and used to determine the capacity of hierarchical hybrid wireless networks. In a three-tier hierarchical network with n_1 access points, n_2 forwarding nodes and n_3 mobile nodes where data sources/sinks are in AP- and MN-tiers, low-tier capacity increases linearly with n_2 and high-tier capacity increases linearly with n_1 when $n_1 = \Omega(\sqrt{n_2})$. The proposed three-tier hybrid network architecture thus provides a means for effectively scaling ad hoc mesh networks while reducing the investment in wired access points relative to the two-tier case. For future work, we plan to work on accurate numerical expressions for estimating the throughput and delay of this class of hierarchical hybrid wireless network.

References

[1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388-404, 2000.

[2] U. C. Kozat and L. Tassiulas, "Throughput capacity of random ad hoc networks with infrastructure support," *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking (MobiCom'03)*, pp. 55-65, 2003.

[3] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," *IEEE INFOCOM 2003*, 2003.

[4] A. Zemlianov and G. de Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE Journal on Selected Areas in Communications*, Vol. 23, No. 3, pp. 657-667, March 2005.

[5] Suli Zhao, Ivan Seskar, and Dipankar Raychaudhuri, "Performance and scalability of self-organizing hierarchical ad hoc wireless networks," *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC 2004)*, 2004.

[6] S. Ganu, S. Zhao, L. Raju, B. Anepu, I. Seskar, and D. Raychaudhuri, "Architecture and prototyping of an 802.11-based self-organizing hierarchical ad-hoc wireless network (SOHAN)," *International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2004)*, September 2004.

[7] E. J. Duarte-Melo and M. Liu, "Data-gathering wireless sensor networks: organization and capacity," *Computer Networks*, pp. 519-537, November 2003.

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