On the Optimal Design of Wireless Relay Networks

Vahid Pourahmadi†, Shervan Fashandi†, Aladdin Saleh‡, and Amir K. Khandani†

† University of Waterloo, Waterloo, ON, Canada,{vpourahm,sfashand, khandani}@uwaterloo.ca

‡Bell Canada, Mississauga, ON, Canada, aladdin.saleh@bell.ca

Abstract

It is known that the achievable data rate per user can be increased when relays are deployed in wireless networks. However, the drawback of this solution is that some of the network resources should be allocated to the relays. In this paper, we consider a two-tier network where all users should send/receive data in two hops. Applying vector quantization, we approximate the best relays' locations. The relays' locations are also computed analytically when the number of relays is less than six. Having the relays' locations, the network average transmission rate is evaluated in terms of a set of network parameters. Then, in the space of these network parameters, we introduce the concept of *neutrality-surface*. The *neutrality-surface* is defined such that the performance of any relay network operating below this surface is inferior to that of a no-relay network with the same parameters. Finally, we study the relative relaying gain for different network configurations.

Index Terms: Resource-limited networks, Relay networks, Relay location, Resource assignment

I. Introduction

RELAY networks combine the advantageous of two radically different design strategies: hierarchical structures and ad hoc networks. Each of these two architectures has its own pros and cons. For instance, Ad hoc networks show good reliability and scalability. However, they usually require complicated routing algorithms, and are vulnerable to security flaws. Complicated base stations and high signaling overhead are among the disadvantages of hierarchical structures. On the positive side, hierarchical networks are able to work with inexpensive and simple terminals. Relay networks provide a balance between these two extremes.

Relay nodes are used in wireless networks to extend the coverage or enhance the channel quality. Since relay deployment is a cost-effective solution for many applications, several standardization committees, including IEEE 802.16j, work on adding the relaying functionality to their current standard. For instance, IEEE 802.16j adopts a two-hop network connection, i.e., Source→Relay→Destination, without a direct link between the source and the destination [1], [2]. IEEE 802.16j relaying schemes are designed such that the relay stations stay transparent to the end users, and let the conventional users connect to the relay network without any modification in their MAC/PHY layers.

Several studies have been conducted to improve the network performance using relays. Cooperative data relaying is studied in [3]. A multihop relay network is introduced in [4], while [5] discusses intra-cell handover management schemes. Furthermore, many researchers have investigated the efficient scheduling and node assignment techniques in relay networks [2], [6]. Reference [7] addresses the problem of optimal relay placement in WLAN networks using the Lagrangian relaxation technique.

Two opposite effects can be identified as relays are deployed in the network. On the one hand, average user transmission rate increases since users transmit to a closer destination and typically experience higher Signal to Noise Ratio (*SNR*) levels. On the other hand, relays must retransmit the aggregated users' traffic to the Base-Station (BS) which takes some of the network resources.

In this work, the network average transmission rate is defined as the metric of performance. Relying on this metric, performance of a relay network depends on the following six parameters: network range (cell size), network scheduler, ratio of the network resources assigned to the users, number of relays, locations of the relays, and users' transmission power. Given the cell size and number of relays, we propose a (suboptimum) algorithm based on vector quantization to compute the best relays' locations. Having the relays' locations and assuming a fair network scheduler, performance of the network is analyzed with respect to the four remaining parameters. Furthermore, for the cases where the relay count is less than six, this analysis provides a mathematical tool for the network design. Using this tool, the designer can find the proper values of these four parameters based on the design constraints. For instance, these constraints can impose conditions on the cell size, user power, and ratio of the network resources assigned to the users.

This paper is organized as follows. First, section II describes the network model and the mathematical notations. Then, in section III, we propose a scheme to determine the best relays' locations. Section IV presents evaluation of the average transmission rate for both of the relay and no-relay networks. Numerical results and network performance analysis is presented in section V. Finally, section VI concludes the paper.

II. NETWORK MODEL

A. *Network Topology and Channel Model*

This paper assumes a two-tier network [8], where each of the user stations transmits to one of the relays in the relay tier. Then, the relay retransmits the data to the BS. In other words, it is assumed that there is no direct connection between the users and the BS. The coverage area of the network is assumed to have a circular shape with radius R. We also assumed that user-stations are placed uniformly in the network [9]. Thus, using continuous approximation, we can consider a density of 'η' for the users such that the total number of users in the network would be $\eta\pi R^2$.

Consider the user u_i which is assigned to the relay r_j. Let $d_{u_i r_j}$ denote the distance between u_i and r_j. The channel gain between u_i and r_j is affected by a deterministic path loss and a lognormal shadowing factor. Thus, the channel gain, $G_{u_i r_j}$, can be written as $G_{u_i r_j} = S\frac{1}{d_{u_i r_j}^{\alpha}}$. Here, α is the path loss exponent, where generally

$2 \leq \alpha \leq 5$ [10]. S denotes the shadowing factor which has a lognormal distribution , i.e., $S = 10^{\frac{Z}{10}}$ where Z is Gaussian with mean and standard deviation of zero and 8dB, respectively [11].

Since the shadowing factor varies very slowly, we can safely assume that the channel gain remains fixed during one Uplink frame. This model is similar to the ergodic quasi-static channel model which is widely used in Information Theory, [3] and [12]. For such a case, the channel between u_i and r_j has a well-defined capacity. This capacity per the unit of bandwidth is given by Shannon's formula

$$C(d_{u_i r_j}) = \log_2(1 + \frac{P\ G_{u_i r_j}}{N_0}), \tag{1}$$

where P and N_0 are the users' transmission power and noise spectral density, respectively.

Obviously, practical data transmission schemes never archive the exact capacity. However, the data transmission rate can get very close to the capacity if the system is designed properly. To this end, firstly, the transmitter has to have a relatively accurate estimation of the channel gain, $G_{u_i r_j}$. For example, in WiMAX networks, the channel is estimated by sending pilot signals and the assumption that the Uplink and Downlink channels are reciprocal. Secondly, the transmitter coding scheme should be adopted based on the channel SNR. For instance, the transmitter can utilize the LDPC codes in [13], [14] which achieve the Shannon capacity within less than 1dB gap with relatively short block lengths. In other words, the achievable data rate can be approximated by the Shannon capacity if the power is shifted by less than 1dB. Similarly, IEEE 802.16 standard has the option to use LDPC codes. The standard also adjusts the code rate and constellation size according to the estimated received SNR [15].

Additionally, it should be noted that in this work, we *are not interested in the exact value* of data transmission rate for a specific power. Instead, *our goal* is *to compare* the achievable data rate in two cases of relay and no-relay networks. Thus, the final results are less affected by the small gap between the Shannon capacity and the practical achievable rate.

Based on the above arguments, we approximate the data transmission rate in each frame with equation (1). Similarly, many references have adopted the Shannon formula as an approximation of the achievable data rate between two nodes [16]–[20]. Having $C(d_{u_i r_j})$ in each frame, our objective is to maximize the time average of the transmission rate over multiple frames, $\overline{C}(d_{u_i r_j})$. $\overline{C}(d_{u_i r_j})$ can be written as

$$\overline{C}(d_{u_i r_j}) = \left\langle \log_2(1 + \frac{P\ G_{u_i r_j}}{N_0}) \right\rangle \overset{(a)}{=} \mathbb{E}_{G_{u_i r_j}} \left\{ \log_2(1 + \frac{P\ G_{u_i r_j}}{N_0}) \right\} = \mathbb{E}_S \left\{ \log_2(1 + \frac{P\ S/d^{\alpha}_{u_i r_j}}{N_0}) \right\}, \tag{2}$$

where $<\ \ >$ denotes the time average operator, and (a) follows from the assumption that the shadowing process is ergodic. In the case of a no-relay network, equation (2) reduces to

$$\overline{C}(d_{u_i B}) = \mathbb{E}_S \left\{ \log_2(1 + \frac{P\ S/d^{\alpha}_{u_i B}}{N_0}) \right\}, \tag{3}$$

where S and $d_{u_i B}$ represent the shadowing factor and the distance between the i^{th} user and the BS, respectively. In the remainder of the paper, we assume that the path loss exponent α is equal to 4.

B. *Frame Structure and Mathematical Representation*

In this study, we assume that nodes (users or relays) are not allowed to transmit on the same frequency band simultaneously. Thus, different nodes make no interference on each other. As an example, the network can operate in Time Division Duplex (TDD) mode and use Orthogonal Frequency Division Multiple Access (OFDMA) as the network medium access sharing mechanism [1]. This frame structure is similar to the frame structure proposed for the IEEE 802.16j standard [1]. This frame structure does not allow simulations transmission over one OFDM subcarrier. It should be noted that our analysis is not limited by the TDD-OFDMA frame structure. Indeed, all of the results in the paper remain valid as long as different users are not allowed to transmit on the same frequency band simultaneously.

Focusing on the Uplink transmission, T denotes the total duration of the uplink frame. Indeed, the Uplink frame of each transmission block, say k, is divided into two parts. Figure 1 shows the Up-Link portion of a frame. In the first $T_1(k)$ seconds (**UL Access Zone**), users transmit to the relay nodes, and for the next $T_2(k)$ seconds (**UL Relay Zone**), relays retransmit the aggregated data to the BS. Both of these zones, furthermore, are partitioned into blocks of bandwidth-time which will be assigned to the users (denoted by $\xi_{u_i}(k)$) or to the network relays (denoted by $\xi_{r_j}(k)$), see Fig. 1. For more detailed explanation of the frame structure, please refer to [1].

In fact, depending on the scheduler's decision, the values of $T_1(k)$ and $T_2(k)$ change in different frames. However, in each frame, say k, the *Access/Relay Zones'* durations should satisfy $T_1(k)+T_2(k)=T$. Furthermore, assuming no frequency reuse, it is clear that $\sum_{i=1}^{|U_\mathcal{T}|}\xi_{u_i}(k)=W\ T_1(k)$, and $\sum_{j=1}^{N_r}\xi_{r_j}(k)=W\ T_2(k)$. Here, W and N_r show the available bandwidth and the total number of relays in the cell, respectively. Moreover, $\mathcal{T}$ is defined as the whole area of the network, and

$$U_\mathcal{T}=\{u_i\text{: User `}i\text{' is located in the region defined by }\mathcal{T}\}. \tag{4}$$

$|U_\mathcal{T}|$ denotes the cardinality of $U_\mathcal{T}$ which equals the total number of users in the cell. Note that, in the case of no-relay networks, $T_2(k)$ would be equal to zero and all the uplink duration would be shared by the users, i.e., $T_1(k)=T$. Thus, $\sum_{i=1}^{|U_\mathcal{T}|}\xi_{u_i}(k)=W\,T$.

Given the above notations, let us define $\widehat{\omega}_\mathcal{T}(k)$ and $\omega_\mathcal{T}(k)$ as the average rate that the users transmit in frame k in the cases of relay and no-relay networks, respectively. $\widehat{\omega}_\mathcal{T}(k)$ and $\omega_\mathcal{T}(k)$ can be evaluated as

$$\widehat{\omega}_\mathcal{T}(k)=\frac{1}{T}\sum_{u_i\in U_\mathcal{T}}\overline{C}(d_{u_ir_j})\xi_{u_i}(k),\ \ \omega_\mathcal{T}(k)=\frac{1}{T}\sum_{u_i\in U_\mathcal{T}}\overline{C}(d_{u_iB})\xi_{u_i}(k), \tag{5}$$

where T is the frame duration.

Furthermore, to capture the average performance of the network, we need to average $\widehat{\omega}_\mathcal{T}(k)$ and $\omega_\mathcal{T}(k)$ over different frames. Thus, we define $\widehat{\Omega}_\mathcal{T}$ and $\Omega_\mathcal{T}$ as the average transmission rate over L (tends to infinity) frames for relay and no-relay networks, respectively. Therefore, we have

$$\widehat{\Omega}_{\mathcal{T}} = \lim_{L\to\infty} \frac{1}{L}\sum_{k=1}^{L} \widehat{\omega}_{\mathcal{T}}(k), \;\; \Omega_{\mathcal{T}} = \lim_{L\to\infty} \frac{1}{L}\sum_{k=1}^{L} \omega_{\mathcal{T}}(k). \tag{6}$$

C. *Network Scheduling*

To complete the network model, the network scheduling algorithm should be defined. For example, in some applications, the scheduler tries to maximize the network throughput. Therefore, in each frame, it allocates all of the network resources to the user with the best channel condition. In another class of scheduling techniques, the objective is to allocate the network resources to all users in a fair manner.

There are several definitions for fairness in the literature [21]. In this work, similar to reference [22], fairness is defined such that $\lim_{L\to\infty}\frac{1}{L}\sum_{k=1}^{L}\xi_{u_i}(k)$ is the same for all users. $\frac{1}{L}\sum_{k=1}^{L}\xi_{u_i}(k)$ denotes the average amount of the network resources (bandwidth-time blocks) assigned to u_i. This definition of fairness implies that

$$\lim_{L\to\infty}\frac{1}{L}\sum_{k=1}^{L}\xi_{u_i}(k) = \frac{W\ \overline{T}_1}{|U_{\mathcal{T}}|}, \quad \forall i \tag{7}$$

where $\overline{T}_1$ is the average duration of the uplink access zone, and $|U_{\mathcal{T}}|$ denotes the total number of users in the cell. If this condition is satisfied, all users will have an equal chance to transmit their data.

There might be many schedulers which satisfy the fairness condition in (7). One possible implementation of such schedulers is described here. In each frame k, and for each user u_i, the scheduler computes $\sum_{j=k-\mathcal{L}}^{k-1}\xi_{u_i}(j)$. This value is a measure of the accumulated amount of the network resources (bandwidth-time blocks) assigned to u_i in the last $\mathcal{L}$ frames. Having this metric for all users, the scheduler partitions the resources of the k'th frame among the users as follows

$$\xi_{u_i}(k) = \left(\Lambda - \sum_{j=k-\mathcal{L}}^{k-1}\xi_{u_i}(j)\right)^{+} \tag{8}$$

where $(x)^{+} = x$ if $x \geq 0$, and $(x)^{+} = 0$ if $x < 0$. Λ is determined such that $\sum_{i=1}^{|U_{\mathcal{T}}|}\xi_{u_i}(k) = W\ T_1(k)$. Intuitively speaking, this scheduling algorithm applies a waterfilling technique to give priority to the users which have had less resources in the last $\mathcal{L}$ frames. A similar technique is used in reference [23] to guarantee fair scheduling of data streams in a single server queuing system.

Due to practical limitations, it is not possible to assign a fraction of one bandwidth-time block to a user. Therefore, the values of $\xi_{u_i}(k)$ in (8) should be rounded to the closest valid number of blocks. It can be seen that if $L \gg \mathcal{L}$, this rounding error becomes negligible and the condition in (7) is satisfied.

It should be noted that the described scheduler is just an example of the class of schedulers which are fair in the sense of condition (7). Our analysis in this work remains valid for any scheduler of this type. In case the scheduler does not satisfy the fairness condition in (7), a different analysis is required which is not addressed in this paper.

III. Optimal Relay Location Estimation

Performance of relay networks is strongly affected by the relays' locations. In fact, there is a coupling between the relays' locations and the amount of resources that the network should assign to the relays. This fact implies that the resource allocation and relay placement problems should be addressed jointly. In fact, finding the optimal relays' locations is a very complex problem. In this study, we propose a suboptimum scheme where the goal is to place the relays such that the average transmission rate, as given in (9), is maximized. From (6) we have,

$$\begin{aligned}\widehat{\Omega}_{\mathcal{T}} &= \lim_{L\to\infty}\frac{1}{L}\sum_{u_i\in U_{\mathcal{T}}}\frac{1}{T}\sum_{k=1}^{L}\overline{C}(d_{u_ir_j})\xi_{u_i}(k) \overset{(a)}{=} \frac{W\overline{T}_1}{T|U_{\mathcal{T}}|}\sum_{u_i\in U_{\mathcal{T}}}\overline{C}(d_{u_ir_j}) \\ &\overset{(b)}{\geq} \frac{W\overline{T}_1}{T}\mathbb{E}_S\left\{\log_2\Big(1+\frac{P\ S/N_0}{(\frac{1}{|U_{\mathcal{T}}|}\sum\limits_{u_i\in U_{\mathcal{T}}}d^2_{u_ir_j})^{\alpha/2}}\Big)\right\}.\end{aligned} \tag{9}$$

where $\overline{C}(d_{u_ir_j}) = \mathbb{E}_S\{\log_2(1+\frac{P\ S/d^{\alpha}_{u_ir_j}}{N_0})\}$, as in (2). In equation (9), (a) follows from the fair resource allocation assumption (equation (7)) and the lower bound in (b) is based on applying the Jensen's inequality and the fact that $\mathbb{E}_S\{\log_2(1+\frac{P\ S/d^{\alpha}_{u_ir_j}}{N_0})\}$ is a convex function of d^2 when $\alpha \geq 2$. Note that, to find the optimal relays' locations, one should maximize $\widehat{\Omega}_{\mathcal{T}}$, which unfortunately does not have a close form solution. Thus, to approximate the optimal relays' locations analytically, in this study, instead of *maximizing* $\widehat{\Omega}_{\mathcal{T}}$, we *maximize the lowerbound in* (9). Indeed, this would be a suboptimal solution for the original problem. It is also obvious that *lowerbound maximization* is equivalent to *minimizing the average user to relay distance square* ($D^2 = \frac{1}{|U_{\mathcal{T}}|}\sum\limits_{u_i\in U_{\mathcal{T}}}d^2_{u_ir_j}$).

Given this approximation, we are now able to formulate the relay placement as *a vector quantization problem*. The users' locations are considered as two-dimensional input points, and the outputs of the vector quantization algorithm would be 'N_r' two-dimensional vectors representing the relays' locations. For instance, the '*K-means*' algorithm with distance square metric inputs users' locations as a series of source samples ($\mathbf{x}$) and computes 'N_r' quantization points ($\widehat{\mathbf{x}}$) as well as the appropriate mapping between x_i and $\widehat{x}_i$.

Applying this algorithm, for the cases of $N_r \in [2, 6)$, it can be observed that relays are placed on a circle centered on the BS. In other words, the relays partition the cell into N_r sectors with the central angle of $\dfrac{2\pi}{N_r}$ each. Users in the j^{th} sector are assigned to the relay at (x_j, y_j) placed on the corresponding bisector. This simple topology is no longer valid when $N_r \geq 6$. For such cases, relays are placed on two or more circles. Figure 2 shows the results of the '*K-means*' algorithm for the cases of $N_r = 3$ and $N_r = 9$. Moreover, to see the accuracy of the lower bound maximization, we compare the final placement of the relays computed by two schemes. First, we use the '*K-means*' algorithm to maximize the exact value of the average transmission rate, $\widehat{\Omega}_{\mathcal{T}}$. Then, as a second method, the '*K-means*' procedure is applied to maximize the lower bound of the equation (9). The results are depicted in Fig. 3 and show that this change of optimization function leads to a small degradation in the final solutions.

Although the '*K-means*' algorithm is able to approximate the optimal relays' locations numerically, for the case of $N_r < 6$ (symmetrical structures), we can calculate them analytically as well. For instance, the location of the first relay, (x_1, y_1), is the point which minimizes the average square distance (D^2) of all users inside the R_1 relaying region. For convenience, in the following, the set of users which are associated to the relay R_j is denoted by $U_{\mathcal{R}_j}$. Using this notation, D^2 can be written as

$$D^2 = \int_{U_{\mathcal{R}_1}} [(r\cos\theta - x_1)^2 + (r\sin\theta - y_1)^2] r d\theta dr, \tag{10}$$

where $y_1 = \tan(\frac{\pi}{N_r})x_1$. Setting $\frac{\partial D^2}{\partial x_1} = 0$, (x_1, y_1) is computed as follows

$$(x_1, y_1) = \begin{cases} x_1 = \frac{N_r}{3\pi}\sin(\frac{2\pi}{N_r})R,\ y_1 = \tan(\frac{\pi}{N_r})x_1 & 2 < N_r < 6 \\ x_1 = 0;\ y_1 = 0.426R & N_r = 2 \end{cases}. \tag{11}$$

Similarly, the location of the j^{th} relay can be computed by rotating (x_1, y_1) by $\frac{(j-1)2\pi}{N_r}$ radians.

IV. Network Average Transmission Rate

According to section III, the optimal location of relays can be computed for different number of relays and any cell size. Having the relays' locations and assuming a fair network scheduler, the average transmission rate can be computed with respect to the following four parameters: cell size, ratio of the network resources assigned to the users, number of relays, and users' transmission power.

The ratio of the network resources which is assigned to the users is identified by the parameter Γ defined as $\Gamma = \frac{\bar{T}_1}{T}$. The path loss exponent is assumed to be $\alpha = 4$ in the rest of the paper.

A. *Average Transmission Rate: No-Relay Network ($\Omega_{\mathcal{T}}$)*

For networks with no relay, the average transmission rate $(\Omega_{\mathcal{T}})$ can be computed as follows

$$\begin{aligned} \Omega_{\mathcal{T}} &\overset{(a)}{=} \lim_{L\to\infty} \frac{1}{L}\sum_{k=1}^{L}\frac{1}{T}\sum_{u_i\in U_{\mathcal{T}}} \xi_{u_i}(k)\overline{C}(d_{u_iB}) = \lim_{L\to\infty}\frac{1}{T}\sum_{u_i\in U_{\mathcal{T}}}\overline{C}(d_{u_iB})\frac{1}{L}\sum_{k=1}^{L}\xi_{u_i}(k) \overset{(b)}{=} \frac{W}{|U_{\mathcal{T}}|}\sum_{u_i\in U_{\mathcal{T}}}\overline{C}(d_{u_iB}) \\ &\overset{(c)}{=} W\int_{r=0}^{R}\overline{C}(r)f(r)\,dr = \int_{r=0}^{R}\mathbb{E}_S\left\{W\log_2(1+\frac{PS/r^4}{N_0})\right\}\frac{2}{R^2}r\,dr \\ &= \mathbb{E}_S\left\{W\log_2(1+\frac{P\ S/R^4}{N_0}) + \frac{2W}{\ln 2}\frac{\sqrt{\frac{PS}{N_0}}}{R^2}\arctan(\frac{R^2}{\sqrt{\frac{PS}{N_0}}})\right\}, \end{aligned} \tag{12}$$

where $\overline{C}(d_{u_iB})$ is given in (3). $f(r)$ is defined as the probability density function (pdf) of d_{u_iB}. For a uniform distribution of users, we have $f(r) = \frac{2\pi r\eta}{\pi R^2\eta} = \frac{2r}{R^2}$. In equation (12), (a) follows from equations (5) and (6). (b) results from equation (7), and (c) relies on the continuous approximation.

B. *Average Transmission Rate: Relay Network ($\widehat{\Omega}_{\mathcal{T}}$)*

As described in section II, all stations transmit their data to the BS via one of the relay nodes. Considering a non-cooperative structure, each user is associated with one of the N_r relays. Consequently, the users can be partitioned into N_r non-overlapping groups, each sending data to a particular relay. Thus, $\widehat{\Omega}_{\mathcal{T}}$ can be computed as

$$\widehat{\Omega}_{\mathcal{T}} \stackrel{(a)}{=} \lim_{L\to\infty} \frac{1}{L}\sum_{k=1}^{L}\sum_{j=1}^{N_r}\frac{1}{T}\sum_{u_i\in U_{\mathcal{R}_j}} \xi_{u_i}(k)\overline{C}(d_{u_i r_j}) = \lim_{L\to\infty}\sum_{j=1}^{N_r}\frac{1}{T}\sum_{u_i\in U_{\mathcal{R}_j}}\overline{C}(d_{u_i r_j})\frac{1}{L}\sum_{k=1}^{L}\xi_{u_i}(k) \tag{13}$$
$$\stackrel{(b)}{=} \frac{W\ \overline{T}_1}{T}\sum_{j=1}^{N_r}\frac{1}{|U_{\mathcal{T}}|}\sum_{u_i\in U_{\mathcal{R}_j}}\overline{C}(d_{u_i r_j})$$

where $U_{\mathcal{R}_j}$ is the set of users associated with the relay j, and $\overline{C}(d_{u_i r_j})$ is given in (2). In equation (13), (a) follows from equations (5) and (6), and (b) results from equation (7).

To analyze (13) mathematically, we can simplify it by considering networks whose $N_r \in [2, 6)$. As mentioned in section III, in such conditions, the relays should be placed symmetrically in the network. Thus, $U_{\mathcal{R}_j}$ becomes identical for all $j \in \{1, 2, \cdots, N_r\}$, and therefore, we can write

$$\widehat{\Omega}_{\mathcal{T}} = \frac{W\ \overline{T}_1}{T}\frac{N_r}{|U_{\mathcal{T}}|}\sum_{u_i\in U_{\mathcal{R}_1}}\overline{C}(d_{u_i r_1}) \stackrel{(a)}{=} W\ \Gamma\frac{1}{|U_{\mathcal{R}_1}|}\sum_{u_i\in U_{\mathcal{R}_1}}\overline{C}(d_{u_i r_1}) \tag{14}$$
$$\stackrel{(b)}{=} W\ \Gamma\int_{U_{\mathcal{R}_1}}\overline{C}(l)\,\frac{2l}{R^2}d\theta dl = W\ \Gamma\int_{U_{\mathcal{R}_1}}\mathbb{E}_S\left\{\log_2(1+\frac{P\ S/l^4}{N_0})\ \frac{2l}{R^2}d\theta dl\right\},$$

where $\overline{C}(l)$ is defined as $\overline{C}(l) = \mathbb{E}_S\{\log_2(1+\frac{P\ S/l^4}{N_0})\}$, and $l^2 = (r\cos\theta - x_1)^2 + (r\sin\theta - y_1)^2$. In equation (14), (a) is based on the assumption of uniform distribution for users, i.e., $\frac{N_r}{|U_{\mathcal{T}}|} = \frac{1}{|U_{\mathcal{R}_1}|}$, and (b) relies on the continuous approximation.

V. Numerical Results

Using (12) and (14), we can compute the network performance for different combinations of *network range* (R), *number of relay nodes* (N_r), *ratio of the network resources allocated to the users* (Γ), and *users' transmission power* (P). In this section, the performance of the network is evaluated for different values of (R, N_r, Γ) and for the cases that $P = 21$ and 27 dBm and the noise floor is set to $N_0 = -93$ dBm [24]. The no-relay network performance is considered as a benchmark for comparison.

A. Network Neutrality-Surface and Neutrality-Curve

In this subsection, the average transmission rates of relay and no-relay networks are compared. For a fixed value of P, the relay network is determined by the parameters (R, N_r, Γ). To have a fair comparison, it is assumed that both networks have the same cell size and users' transmission power. As mentioned in section IV, there is a tradeoff between the performance improvement achieved by relaying and the degradation caused by the loss of resources allocated to the relays. In other words, a relay network with parameters (R, N_r, Γ) is able to compensate the loss in resources due to relaying *iff*

$$\widehat{\Omega}_{\mathcal{T}} \geq \Omega_{\mathcal{T}}, \tag{15}$$

In the *3-dimensional space* of (R, N_r, Γ), the set of points satisfying (15) with equality form a *surface*, called the *neutrality-surface*. In fact, all the triplets below the *neutrality-surface*, i.e., $\widehat{\Omega}_{\mathcal{T}} < \Omega_{\mathcal{T}}$, have worse performance as compared to the corresponding no-relay network. To illustrate the *neutrality-surface*, we project it over its *2-dimensional* subspaces. To this end, the value of Γ is fixed at Γ_1. Then, the *neutrality-curve* is defined as the intersection of the *neutrality-surface* and the plane corresponding to $\Gamma = \Gamma_1$.

Figure 4 depicts the *neutrality-curves* for the two cases of $\overline{T}_1 = \frac{4}{3}\overline{T}_2$ $(\Gamma = 0.57)$ and $\overline{T}_1 = \overline{T}_2$ $(\Gamma = 0.5)$. In each case, the *neutrality-curve* is plotted for two different values of transmission power (P). To ensure that the circular placement of relays is optimal, we have limited the number of relay to $1 < N_r < 6$.

As Fig. 4 shows, for a fixed Γ, the minimum required number of relays decreases when the network range (cell size) increases. This can be justified noting that for smaller network sizes, the received SNR at the BS is high enough for successful decoding. On the contrary, in larger networks, the received SNR reduces because of the path loss. Consequently, even a small number of relays can improve the average transmission rate.

Moreover, as Fig. 4 suggests, increasing the transmission power shifts the *neutrality-curve* to the right. In other words, for a fixed cell size and higher values of $\frac{P}{N_0}$, more relays are required to satisfy $\widehat{\Omega}_{\mathcal{T}} = \Omega_{\mathcal{T}}$. Thus, the effect of increasing $\frac{P}{N_0}$ is similar to the effect of reducing the cell size. They both enhance the received SNR and reduce the necessity of relaying.

All of the above observations confirm the intuitive result that relaying improves the performance significantly only in the low SNR region. In the high SNR region, the resource loss due to the relaying outweighs the gain achieved by increasing the received SNR.

Finally, comparing figures 4(a) and 4(b), it can be concluded that for a fixed R, if Γ is reduced, we need more relays to stay on the *neutrality-surface*. In fact, the lower is Γ, the lower will be the fraction of resources allocated to the users. Thus, to keep $\widehat{\Omega}_{\mathcal{T}} = \Omega_{\mathcal{T}}$, the user to relay connection has to work in a higher SNR, which implies deploying more relays in the network.

*B. **Relative Relaying Gain***

In this section, two performance metrics are defined to characterize the performance of relay networks. It is shown that these metrics are related to the *neutrality-curve* defined in subsection V-A.

First, let us define the *relative relaying gain* as the ratio of the average transmission rate with and without relays ($\frac{\widehat{\Omega}_{\mathcal{T}}}{\Omega_{\mathcal{T}}}$). As an example, we have evaluated the relative relaying gain for a typical network where $\Gamma = 0.5$ and $P = 21$ dBm. The results are depicted in Fig. 5. Since the figure is plotted for a fixed transmission power, the larger values of R correspond to the low SNR region.

Figure 5 shows that for fixed values of R and Γ, increasing the number of relays always increases the relative relaying gain. Since $\Omega_{\mathcal{T}}$ is independent of N_r, this result means that $\widehat{\Omega}_{\mathcal{T}}$ is an increasing function of N_r. Moreover, it can be seen that in low SNR values (large cell sizes), the impact of adding relays is more considerable as compared to the high SNR region (small cell sizes). For instance, doubling the relay count from $N_r = 2$ to $N_r = 4$ in a network of cell radius $R = 1$Km improves the relative relaying gain by about 60%, while this improvement is around 90% when the cell radius is $R = 3$Km. Clearly, this observation is consistent with the intuition that relaying is more advantageous when the network operates in the low SNR region.

It is also interesting to note that the dashed line in Fig. 5 shows the unity gain, i.e., $\widehat{\Omega}_{\mathcal{T}} = \Omega_{\mathcal{T}}$, and is another representation of the *neutrality-curve* in Fig. 4.

VI. Conclusion

In this paper, a tradeoff associated with relaying in wireless networks is studied. Relaying results in higher SNR values and transmission rates. On the other hand, it potentially degrades the performance by consuming some of the network resources to retransmit the aggregated data to the BS. Considering a two-tier scenario, we have discussed how vector quantization can be used to approximate the optimal relays' locations. Then, in the space of network parameters, the *neutrality-surface* is introduced. This surface characterizes the points of balance in the tradeoff, i.e., where the two opposite effects of relaying cancel each other. Finally, introducing relative relaying gains as a performance metric, different network configurations are compared.

References

[1] IEEE 802.16j Task Group, "IEEE 802.16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems: Multihop Relay Specification," *IEEE Standardization Committee*, 2007.

[2] V. Pourahmadi, A. Khandani, W. Tong, P. Zhu, "Effective Node Assignment in 2- Hop Fixed Relay Networks," in *IEEE 802 Plenary Session*, Dallas, 12- 17 November 2006.

[3] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative Strategies and Capacity Theorems for Relay Networks," *IEEE Transactions on Information Theory*, vol. 51, pp. 3037– 3063, 2005.

[4] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "Multihop diversity in wireless relaying channels," *IEEE Transactions on Communication*, vol. 52, pp. 1820– 1830, 2004.

[5] Shengjie Zhao et. al, "Macro Diversity Handover and Fast Access Station Switching for MMR Network," in *IEEE 802 Session 47*, London, UK, January 2007.

[6] Guoqing Li and Hui Liu, "Resource Allocation for OFDMA Relay Networks With Fairness Constraints," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 2061– 2069, 2006.

[7] A. So and B. Liang, "A Lagrangian Approach for the Optimal Placement of Wireless Relay Nodes in Wireless Local Area Networks," *LECTURE NOTES IN COMPUTER SCIENCE*, vol. 3976, pp. 160, 2006.

[8] YT Hou, Y Shi, HD Sherali, and SF Midkiff, "On Energy Provisioning and Relay Node Placement for Wireless Sensor Networks," *Wireless Communications, IEEE Transactions on*, vol. 4, pp. 2579– 2590, 2005.

[9] M. Gastpar and M. Vetterli, "On the Capacity of Wireless Networks: The Relay Case," in *INFOCOM* , 2002.

[10] A. Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.

[11] S. Toumpis, A.J. Goldsmith, "Capacity Regions for Wireless Ad-Hoc networks," *IEEE Transactions on : Wireless Communications*, vol. 2, pp. 736– 748, 2003.

[12] S. Shamai and A. Steiner, "A Broadcast Approach for a Single- User Slowly Fading MIMO Channel," *Information Theory, IEEE Transactions on*, vol. 49, pp. 2617– 2635, Oct. 2003.

[13] I. Sason and S. Shamai, "Improved upper bounds on the ensemble performance of ml decoded low density parity check codes," *Communications Letters, IEEE*, vol. 4, no. 3, pp. 89–91, Mar 2000.

[14] B. Myhre, V. Markhus, and G.E. Øien, "LDPC-coded adaptive modulation on slowly varying Nakagami-fading channels," in *Proc. European Wireless*, 2002, pp. 822–828.

[15] IEEE, "IEEE 802.16 Standard: Air Interface for Fixed Broadband Wireless Access Systems," 2004.

[16] H. Viswanathan and S. Mukherjee, "Performance of Cellular Networks with Relays and Centralized Scheduling," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2318– 2328, 2005.

[17] H. Viswanathan and S. Mukherjee, "Throughput-range tradeoff of wireless mesh backhaul networks," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 3, pp. 593–602, March 2006.

[18] Yingbo Hua, Yi Huang, and J.J. Garcia-Luna-Aceves, "Maximizing the throughput of large ad hoc wireless networks," *Signal Processing Magazine, IEEE*, vol. 23, no. 5, pp. 84–94, Sept. 2006.

[19] M. Saad and S.A. Muhaureq, "Joint routing and radio resource management in multihop cellular networks using particle swarm optimization," *National Radio Science Conference, 2008. NRSC 2008*, pp. 1–9, March 2008.

[20] J.K. Sundararajan and H. Viswanathan, "Comparison of schemes for streaming multicast in cellular networks with relays," *Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE*, vol. 2, pp. 1178–1183, 0-0 2006.

[21] H. J. Kushner, and P. A. Whiting, "Convergence of Proportional-Fair Sharing Algorithms Under General Conditions," *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, vol. 3, pp. 1250– 1259, July 2004.

[22] B Sadeghi, V Kanodia, A Sabharwal, and E Knightly, "Opportunistic Media Access for Multirate Ad Hoc Networks," in *International Conference on Mobile Computing and Networking*, Atlanta, Georgia, USA, 2002.

[23] H. Yousefi'zadeh, "A neuro-forecastwater-filling scheme of server scheduling," *Systems and Networks Communications, 2006. ICSNC '06. International Conference on*, pp. 9–9, Oct. 2006.

[24] IEEE 802.16 Task Group, "IEEE 802.16m Evaluation Methodology Document," Tech. Rep., [Online] Available: `http://wirelessman.org/tgm/`, 2008.

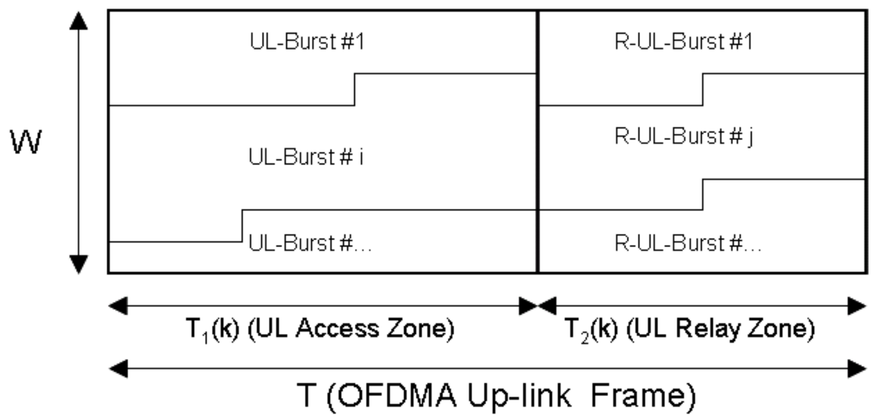


Fig. 1. OFDMA Up-link Frame Structure

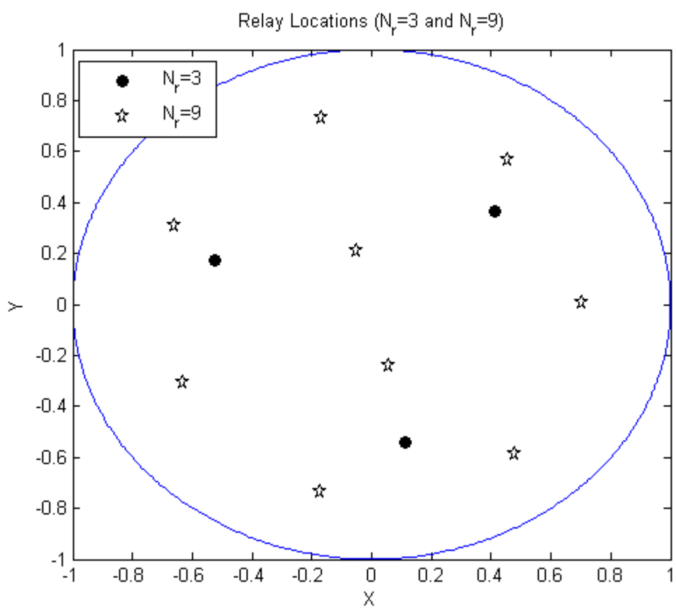


Fig. 2. Optimal relays' locations determined by the *K-means* algorithm

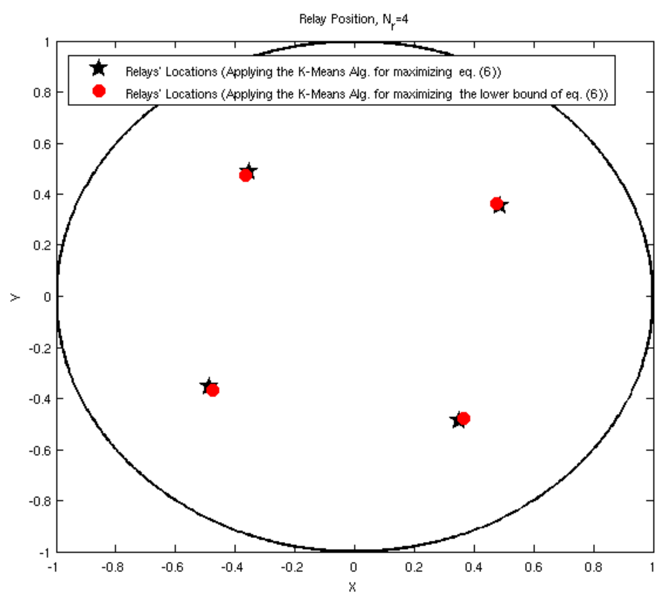
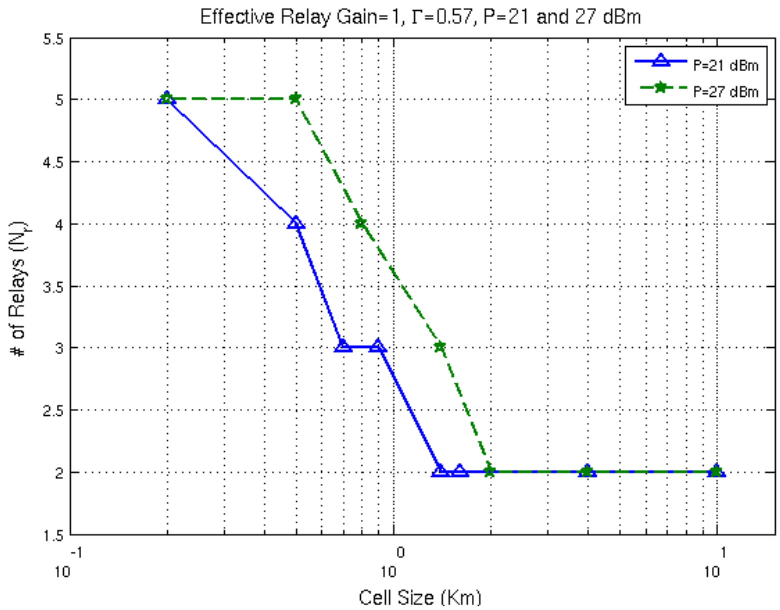
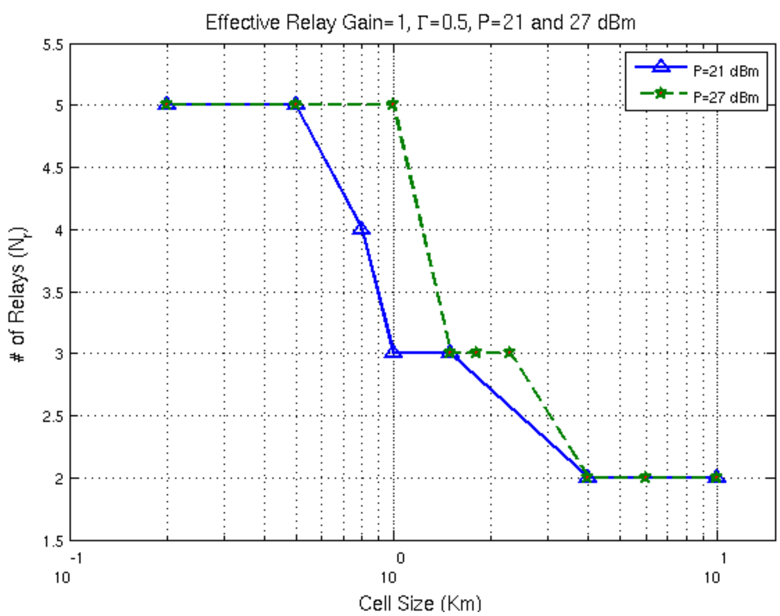


Fig. 3. Comparison between the optimal relays' locations when (a) The *K-means* algorithm is used to maximize the exact value of the average transmission rate (equation (6)) and (b) The *K-means* algorithm is applied to maximize the lower bound of equation (6).



(a) $\Gamma = 0.57$



(b) $\Gamma = 0.5$

Fig. 4. Network Neutrality-Curves

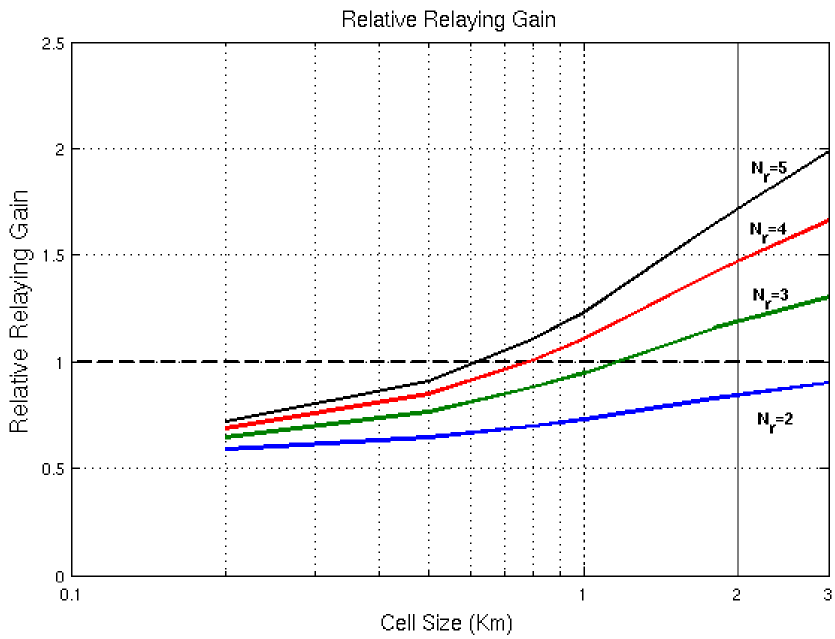


Fig. 5. Relative Relaying Gain as a function of cell size for $P = 21$dBm, $\Gamma = 0.5$, and different number of relays