Gain of a wireless ad hoc hybrid network over the classical pure ad hoc network

A. Puig, Y. Bar-ness and A. Laufer

Index

Index of figures

Acknowledgements

When a thesis and a research is written there normally take part countless of people that directly or indirectly contribute with information and support to the researcher. In most of the cases, when one wants to mention if not all, the majority of the people who take part in one way or another, the list results too long and anyhow there will always be unforgivable omissions. For this reason, therefore, I prefer to begin by being grateful to these people that although they are not mentioned in here, they helped and supported me before, during and after the research.

I also want to mention *Fundación Vodafone España* that, thanks to the grant they provided to me, made possible this research and my stay in the USA.

I am very grateful to the *CWCSPR* that offered me the opportunity of my life when they accepted my application to doing my Master Thesis in the *New Jersey Institute of Technology*. To Osvaldo Simeone, Fabio Ianello, Tariq and William, the group of the *Thurday's seminar*; every two weeks we all met together discussing, presenting and exposing our own research. This meeting was of great aid since – apart from learning new things and acquiring new knowledge from the other member's researches – when your exposing time arrived you had to have clear all the points of your research and, as well, that gets your nerves out to future exposition of the thesis; all the unclear things were brought to light, and then, those were the points which had to be improved on *a posteriori*. Thanks to them I have always had in mind what was what I was investigating and what the aim of the thesis was.

To my friends. To those who always had had a great time with me y with whom I always had had a great time. For keep being my friends despite the bad moments we've gone through and for keep also being my friends in the good times; for making me go through such wonderful moments. As the saying goes: "*every friend is a treasure*" and is graspable since without them life won't be so marvelous. I don't mention them in here because surely the list would be longer than I imagine and for sure that I would forget someone and neither him nor I would forgive this, so, to all of you.

To whom, as well, have shared with me the university life, undoubtedly the best period of life. For all the above written and for all we have lived together for the last 6 years; for making classes more enjoyable, for all the notes they left me and for all the notes they had taken for me, for explaining me the things when I didn't understand, anyway, for all those things they had made for me during this wonderful period of my life.

To Alfons Martínez, my brother-in-law, who made possible the interfamiliar communications during my stay far from home. Thanks to these, my stay was more bearable and easier despite the distance.

To Maria Dolors Vilar, my aunt, and to Joaquim Vilar and Manuela Payà, my grandparents, who were always next to me and treated and took care of me as what I am: their only nephew and grandson. Thanks for all.

To Laia Puig, my sister, because she has been that and much more. Friend and confident we have shared all in this life. Because nobody better than her keeps the secrets and because nobody better than her comprehends me and knows what I feel.

To Olga Vilar, my mother, and again to my sister, who have always been by my side whenever I needed it and have always listened to me when I needed somebody to talk to. For all their support, comprehension and help in all the moments of my life.

And especially to Jordi Puig, my father; if he would have not been as he is I would neither have been who I am nor I would have not become what I have become. Because although he never recognizes it in our presence I know that he has always been very proud of his daughter and his son.

Finally, but the most important to me, I am very very grateful to my girlfriend, Mireia Bossy, who has always been by my side, whenever I needed and when I didn't need it, who has always given support in everything to me and who was one of the people that supported me the most when I decided to embark on this adventure; if she had ask me to stay God only knows that I would have stayed by her side. I want to thank her – and I will always do it and will always bear this in mind – for being as she is and for always being by my side. To her this thesis is dedicated.

1. Introduction

It is obvious that wireless networks present many advantages over the wired ones, otherwise recent and future studies would not have been in the way of replacing or quitting wires of our lives.

The main advantage of wireless networks is the fact that connections don't exist over a physical support such as a cable, but they do exist over what we name a shared media, i.e. air. This means that one can, theoretically, connect any pair of nodes without a cable constraint and also that the communication channel (air) might be shared with other pairs of communicating nodes, and confers high flexibility to the network.

Although one can try to have a large set of pairs of nodes transmitting at the same time this way, it must also be taken into account that all transmissions cause interference on the other ones and that a node is just able to either receive or transmit a single transmission at the same time. It is also remarkable that although one might try to establish a direct communication between any pair of nodes, chances are that the channel has a bad quality between a random pair of nodes or that the communication between them is not successful if they want to communicate of transmit between them, they will have to use relay nodes.

If the channel quality is not good enough or a lot of nodes want to transmit at the same time is when the main problem appears. Interference of other nodes transmitting may result in a non-successful transmission between a pair of nodes.

R. Gowaikar, B. Hochwald and B. Hassibi [1] introduced and studied a pure ad hoc network with random connections. M. Palanques, A. Laufer and Y. Barness [4] focused their work on the throughput improvement introduced by the addition of infrastructure nodes to their scheme, the minimum infrastructure requirements, and the scaling of the throughput as a function of the infrastructure size. They also introduced a different scheduling scheme since the one in [3] was very suboptimal.

This hybrid network works in this way: a set of base stations is installed in the network; is in this work that we study which is the minimum number of these infrastructures needed in order to improve the throughput and respect the scheduling imposed. Every infrastructure contains a set of nodes, which can be shared with other base stations, and the network is divided into sub-graphs.

Every of these sub-graphs contain one and only one base stations and the scheduling is made in each sub-graph. A definition and a construction of circles are made by the authors in [4] and the scheduling is then settled.

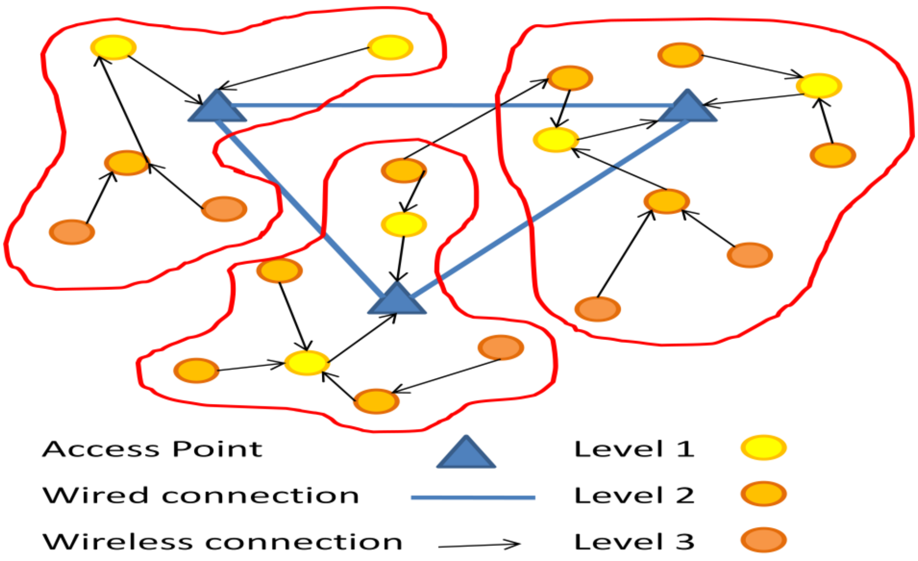

Figure 1. Hybrid network scheme and sub-graph division

In this work we continue the study made in [4] and present a series of simulations and results comparing the hybrid ad hoc network analyzed in [4] with the pure ad hoc network presented in [1].

We compare the scheduling and the achievable throughput of a hybrid ad hoc network over a pure ad hoc network. That is it, what would happen if we introduce base stations in the pure ad hoc network? Will the throughput be increased? Will the interference be reduced or it will be higher? And finally, a new scheduling approach is defined in order to take profit of the fact of having base stations in the network.

The achievable throughput of the network presented in [4] depends basically on the number of base stations included in the network and the average number of hops a message needs to perform in order to reach the intended destination. The maximum achievable throughput is obtained when the number of hops needed to reach the nearest base station is only one. However, this may transform an ad hoc network into a cellular network and we don't want this to happen. What we suppose is that, after the circle's construction is done (as in [4]), if a source wants to transmit a message to a destination placed in another circle then we will use the base station, otherwise the sub-network works as a pure ad hoc network.

In Section 2 we summarize the work in [1] and [4] and use them as an introduction to what our work will be; we also analyze their performance and make some comments on the suitability of their characteristics and what we suppose the gain or the advantage of introducing base stations will result. We first analyze the work in [1] and its main result, then we analyze the work in [4] and also its main result, and finally we introduce how we will compare these two networks in the following sections. In Section 3 we suppose that the probability of having a good connection between two nodes is nothing but a constant. Although this might be an hypothetic case and never been found in a real scenario, it would be a good idea to analyze and see the network's behavior under this supposition seeing how many base station might be needed and what the achievable throughput would be.

After Section 3 we start studying possible real scenarios. In Section 4 we analyze the hybrid ad-hoc network under a shadow fading environment and compare the achievable throughput with the pure ad-hoc one. We will see that this gain is important and that by the addition of some base stations the throughput is highly increased. In Section 5 we analyze the same network but this time under an exponential density scenario. Since there exists and optimal choose of the parameter β we analyze the network in two cases, the first one for a non-optimal value of β and the second one for the optimum β. It can be seen that the throughput is increased for any value of β but the increase will be higher whenever we the value of β is closer to the optimum.

Finally in Section 6 we extract conclusions of our work and give some guidelines for future research.

2. Previous work

In this section all the work and studies made on ad hoc networks is presented. A pure ad hoc network is first analyzed showing bounds for the achievable throughput, the scheduling used, the network interference and operation and some definitions and suppositions.

After analyzing and studying the pure ad hoc network the hybrid ad hoc network is, then, analyzed. That's it, what would happen if we introduce base stations in the pure ad hoc network? Will the throughput be increased? Will the interference be reduced or it will be higher? And finally, a new scheduling approach is defined in order to take profit of the fact of having base stations in the network.

2.1 Communications over a pure ad hoc wireless network

In the paper "Communication over a wireless network with random connections" R. Gowaikar, B. Hochwald and B. Hassibi [1] analyze a network of nodes in which pairs communicate over a shared wireless medium and in which a model based on distance is changed to another one based on randomness. They assume that adopting the premise that randomness can have a first-order effect on the behavior of a network and assuming that channels between nodes are drawn independently from an identical distribution, they study the achievable throughput of a pure ad hoc network with independently and identically drawn random connection strengths between nodes instead of that other studies in which the strength of the connections are governed by geometry and a decay-versus-distance law.

The authors in [1] state that whenever the first order event that governs the signal strength at a receiving node is a random event[1] their model would be the most appropriate. Besides, they affirm that certain forms of randomness could also be helpful for the expected aggregate data traffic in a wireless network.

[1] It is understood by a random event an event that happens in a non-deterministic manner. In other words, they can occur at any time and any place. Such an event could be, for instance, the existence of an obstacle.

2.1.1 Approach

It is generally known that for a message to reach its destination in a classical ad hoc network[2] a node may communicate that message to nearby neighbors so that in a certain number of hops it finally reaches the intended destination. In other words, the source node sends its message to the nearest node; this node would act as relay node sending the message to its nearest neighbors and so on until the message reaches the intended destination. This idea of "near neighbor", in which a node may directly communicate with another node if they are close enough, is now changed by introducing the equivalent idea of "*good path*": it is understood by *good path* those connection strengths between two nodes greater than a chosen threshold β_n. Now transmissions to relays and destinations do only occur along these good paths and a node is a relay to another one if their path is good. Now that the definition of good path is known, it is time to see what kind of scheduling is necessary to achieve this objective.

In the schedule used in [1] transmissions are performed by finding vertex-disjoint paths, that is, paths that do not share any node during all the time needed to cross them. To that purpose, authors use the algorithm and results introduced in [2] and they draw a graph whose vertices are all the nodes in the network but whose edges are only the good paths. A specific random graph model $G(n,p)$ is obtained, where an edge between any pair of the n nodes exists with probability p (p is the probability that a connection strength exceeds β_n).

Let us consider a graph model $G(n,p)$ for our network. The network is formed by n nodes, so these will be the vertexes of the graph and each connection, or edge, exists with probability p. Since the connection strengths do not change in time, the graph will also remain the same. The reader can see an example in the following figure:

[2] A classical ad hoc network is that one in which the model of the wireless network is based on a decay-versus-distance law.

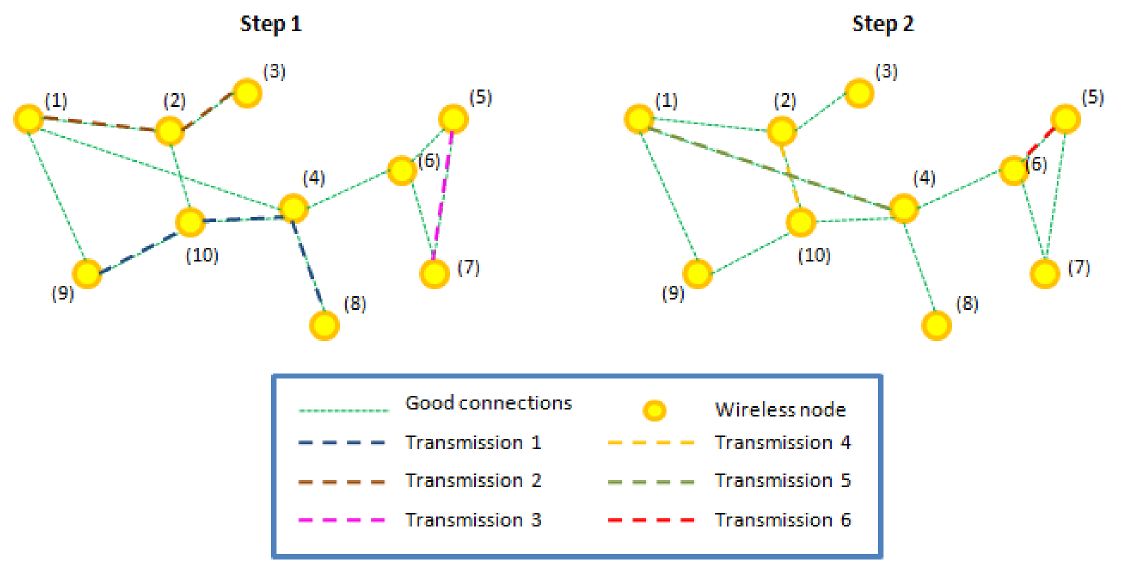

Figure 2. Vertex-disjoint paths on a pure ad hoc network

Figure 1 shows how the scheduling works. The transmission of every message is done by using a sequence of relay nodes, performing h hops, and therefore needs h time slots to be completed; note that in figure 1 nodes (2), (4), (8) and (10) act as relay nodes. First, a set of source-destination pairs, with the related vertex-disjoint paths, is found. Those paths are simultaneously used within step 1, which lasts 3 time slots (the duration of the longest path). After that, a new set of source-destination pairs and paths is found and it is used along step 2.

One can see that the duration of a step is always the duration of the longest path performed during it, which is upper-bounded by the network's diameter[3].

It is known that for such a graph whenever $p > \frac{\log(n)}{n}$ the probability of connectivity goes to one rapidly, the network does not have any isolated node[4]. However it is also stated that if $p < \frac{\log(n)}{n}$ the network can be under-connected, and that is not always good because there may be some isolated nodes, this is nodes not connected to any other one. Thus, although this last does not condition a value of p the first one does and so it is necessarily assumed.

[3] The diameter of the network is the highest among the minimum distances, in terms of hops, between any pair of nodes in the network: the maximum possible distance between a source node and its respective destination

[4] An isolated node is a node not connected to any other one. This is a node that cannot transmit to any other one because no possible connection exists between this and the others.

With this condition, the diameter of the graph behaves like $\frac{\log(n)}{\log(np)}$. A bound for the number of hops that a message needs to perform is also found: One can see that the total number of hops that a message would need to perform is, at most, the diameter of the network and, asymptotically, all the paths have lengths that grow no faster than $h = \frac{\log(n)}{\alpha \cdot \log(np)}$.[5]

2.1.2 Model

As said before, it is supposed in [1] that the connection strengths between the *n* nodes are drawn independently and identically distributed (i.i.d.) according to an arbitrary distribution $f(\cdot)$.

Being the network formed by *n* nodes, then, there are $\binom{n}{2}$ pairs of nodes taken as channel random variables. Thus, if $h_{i,j}$ is the connection between nodes *i* and *j*, then the channel strengths, $\gamma_{i,j} = |h_{i,j}|^2$ are i.i.d. random variables with marginal distribution $f(\gamma_{i,j})$. For maximum generality they allow $f(\gamma) = f_n(\gamma)$ to be a function of the number of nodes *n*. Once drawn, all these channel variables are assumed not to change with time.

The concept of good connection is defined in [1] as a connection between two nodes *i* and *j* with a channel strength greater than a chosen threshold β_n. This step is done without taking interference into account; a connection is good regardless of whether the interference is strong or low. Indeed, this edge between any pair of the *n* nodes exists with probability $p = p_n = P(\gamma_{i,j} \geq \beta_n) = Q_n(\beta_n)$, i.e. the probability that a connection is good generally will depend on the number of nodes in the network (*n*) and the chosen threshold (β_n).

However, although a pair of nodes is in good connection that does not mean that the transmission between them will be successful, or even possible. Assuming that *k* nodes $i_1, i_2, \ldots, i_k$ are simultaneously transmitting signals $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ of power *P*, respectively and the channel has additive white Gaussian noise (with mean 0 and variance σ^2), supposing that

[5] A proof for this can be found in [1] page 2862.

only node i_1 wishes to communicate with node j, then the signal-to-interference-plus-noise ratio (SINR) for node j is given by:

$$\rho_j = \frac{P \cdot \gamma_{i_1,j}}{\sigma^2 + P \cdot \sum_{l=2}^{k} \gamma_{i_l,j}} \tag{1}$$

The reader can see that the existence of interference is not considered in the definition of the threshold for a good connection. It is said in [1] that a transmission is possible when the SINR exceeds some threshold ρ_0. If this SINR is less than a given ρ_0, they say that transmission is not possible. The minimum SINR is defined by the authors in [1] as:

$$\rho_0 = \frac{a_n \cdot \beta_n}{\frac{\sigma^2}{p} + (k-1)\mu_\gamma} \tag{2}$$

Where a_n is a positive constant less than 1 and μ_γ is the mean of γ .

2.1.3 Network operation and throughput

A group of k nodes, denoted by $s_1, s_2, \ldots, s_k$, are randomly chosen as sources. For every source s_i a destination node d_i is chosen at random, there will therefore exist k source-destination pairs. The reader can see that these $2k$ nodes are distinct since a schedule of vertex-disjoint paths is used and that $k \leq n/2$.

The authors in [1] desire to see how many of these source-destination pairs may communicate simultaneously. To that purpose they set that sources can either talk directly to the destination or decide to communicate through a series of relay nodes and therefore reaching the destination in a certain number of hops. It is assumed in [1] that the transmission of every message is done by using a sequence of relay nodes, i.e. a message would need to perform h hops to reach its intended destination from the source, and it will therefore need h time-slots for the transmission to be completed.

Supposing that source s_i whishes to transmit a message M_i to destination d_i the transmission rate for this link is $\log(1+\rho_j)$, which is higher or equal to $\log(1+\rho_0)$.[6] In order to simplify the analysis, authors in [1] use the lower bound $\log(1+\rho_0)$, instead of $\log(1+\rho_j)$.

Assuming that there is allowed a scheduling of *k* simultaneous transmissions during *h* time-slots with some probability of error (ε) in the transmission of every message; the achievable throughput is defined as:

$$T = (1-\varepsilon)\cdot\frac{k}{h}\cdot\log(1+\rho_0) \tag{3}$$

The resulting throughput depends on the number of nodes and so the authors usually add subscripts to the variables involved. The mentioned probability of error is considered by the authors as the probability that a particular message fails to reach its intended destination. Note that, as said before, destination fails to receive a message if the SINR falls below the threshold ρ_0 at any of the *h* hops performed by the message. Thus, an upper bound to this probability of error is given by

$$\varepsilon \le \frac{\log(n)}{\alpha\log(np)}\frac{\sigma_\gamma^{\ 2}}{(k-1)\left(\dfrac{P\cdot\beta_n-\rho_0\cdot\sigma^2}{(k-1)\cdot P\cdot\rho_0}+\mu_\gamma\right)^2} \tag{4}$$

This probability of error is required to go to zero as the number of nodes increase[7].

2.1.4 Main result

The main result in [1] is summarized in what authors call *Theorem 1*[8] which states that:

Considered a network with *n* nodes whose edge strengths are drawn i.i.d. from a certain probability distribution function $f_n(\gamma)$ and with a complementary cumulative distribution function $Q_n(\gamma)$ they choose any βn such that $Q_n(\beta_n)=\dfrac{\log(n)+\omega_n}{n}$, where ω_n is a function that goes to infinite as *n* goes also to infinite. Then there exists a positive constant α such that a throughput of

[6] This is the sustainable throughput per user if the users do not collide.
[7] One can find a development to achieve this result in [1] page 2862
[8] See page 14 in [1] for a detailed explanation.

$$T = (1-\varepsilon)\cdot \alpha \cdot K_n(\beta_n)\frac{\log(nQ_n(\beta_n))}{\log(n)}\cdot \log\left(1+\frac{a_n\beta_n}{\frac{\sigma^2}{P}+(K_n(\beta_n)-1)\mu_\gamma}\right) \tag{5}$$

Is achievable for any positive $a_n < 1$ and any k that satisfies the conditions:

1. $k \le \alpha \cdot n\frac{\log(nQ_n(\beta_n))}{\log(n)}$ (6)

2. $\varepsilon \le \frac{a_n{}^2}{\alpha\cdot(1-a_n)^2}\frac{(k-1)\sigma_\gamma{}^2}{\left(\frac{\sigma^2}{P}+(k-1)\mu_\gamma\right)^2}\frac{\log(n)}{\log(nQ_n(\beta_n))}\to 0$ (7)

They also state that whenever $\frac{\sigma^2}{P}-\mu_\gamma \ge 0$ the throughput is maximized by choosing k as large as possible. Hence, the equality is chosen and therefore $k = \alpha \cdot n\frac{\log(nQ_n(\beta_n))}{\log(n)}$.

2.2 Communications over a hybrid ad hoc wireless network

Although wireless network is a good solution in order to avoid the need of a physical support, such as cable, one must also take into account that all transmissions cause interference on the other ones and that a node is just able to either receive or transmit a message at a certain time. In actual communications, a transmission may often need a connection to a wired network (such as the internet or the plain telephone network) so there may be one or several nodes connected to the wired network while the remaining nodes that wish to reach this wired network will have to use those wired-connected nodes as a relay.

Despite wireless networks having the advantage of being independent and more flexible, meaning that the network can be created by any set of nodes without need of an infrastructure (access point) and any node may enter or leave the network anytime if needed, these networks, due to the lack of a coordinator, have the drawback of a more complicated scheduling than the wired ones.

The simplest and most common approach is the wireless network with infrastructure and a hierarchical structure where one has an access point that manages the transmissions of the wireless nodes and the access of them to the wired network. In such networks every transmitter uses the access point as a relay node, meaning that the source transmits directly to the access point and this one transmits directly to the destination.

In the thesis "*Communication over a hybrid ad hoc wireless network*" M. Palanques, A. Laufer and Y. Bar-ness [4] focus their study on a network that is halfway through those two models, i.e. a **hybrid ad hoc network** (an ad hoc network with infrastructure support). This is, an ad hoc network where some of the nodes are fixed and connected between them by wired links while the others are connected wirelessly and not fixed. In this hybrid network, nodes can communicate across the infrastructure network, through a path of relay nodes or even use the infrastructure network after using relay nodes to reach an access point (which is called a hybrid path).

Based on the premises and results obtained in [1] authors in [4] obtain their own results and conclusions in a hybrid scenario.

2.2.1 Proposed hybrid ad hoc network

The proposed network by the authors in [4] consists of n wireless nodes and m infrastructure nodes (access points). The access points are wired and strongly connected with high capacity links, and thus they consider that the probability of any pair of access points to be connected is 1. The wireless connections between nodes and also between any pair node/infrastructure are randomly independent and identically distributed as in [3], with probability distribution function $f_n(\gamma)$, cumulative distribution $F_n(\gamma)$ and complementary cumulative distribution $Q_n(\gamma)$.

The reader can notice that, while the probability of having a good connection between base stations is deterministic and always 1, the probability of good connection between any pair of nodes or even between a node and a base station is random and so the total number of random variables in the scenario is given by $\binom{n+m}{2}-\binom{m}{2}$. Note that the first term is the total number of nodes in the network while the second one is the total number of wired nodes or infrastructure nodes.

In their proposed network scheme, the overall probability of having a good connection considers all the links in the graph and this probability (dubbed q by the authors) is greater than the p found in [1] thanks to the deterministic probability of the connection between access points. This probability is given by

$$q = \frac{\binom{m}{2}}{\binom{n+m}{2}} \cdot 1 + \frac{\left(\binom{n+m}{2} - \binom{m}{2}\right)}{\binom{n+m}{2}} \cdot p = p + \frac{\binom{m}{2}}{\binom{n+m}{2}}(1-p) \tag{8}$$

Since the second term is always positive, it is easy to see that the connectivity q is always greater than p, which is the probability of having a good connection in a pure ad hoc network. However, it can also be noticed that if the number of nodes is great enough ($n \to \infty$) the probability becomes the same.

Hence, and since the number of nodes in the new network is *n+m*, the new lower bound for the probability of having a good link in order to ensure that there are no isolated nodes in the network is

$$q > \frac{\log(n+m)}{n+m} \tag{9}$$

Since the denominator grows faster than the numerator, this condition is always less restrictive than the one found for the pure ad hoc network. The reader can notice that the condition (9) is applied to q, which is larger than p by the existence of the wired links as showed in (8), and therefore they reach a double improvement. However, in order to simplify the calculations, they use a worst case and require

$$p > \frac{\log(n)}{n} \tag{10}$$

to be accomplished in the study.

2.2.2 Scheduling

Scheduling is a crucial issue in a wireless network and it becomes even more important in a network where traffic is not managed by an infrastructure since collisions are the most important thing to be avoided; if a lot of collisions occur in a network the throughput is much lower than in the case they don't exist. The aim of scheduling is to optimize the usage of resources allowing the nodes to be as operative and efficient as possible.

Since in a wireless network all messages are transmitted over the same channel, the interference has a significant effect on the performance. It is necessary to recall that it is assumed that nodes are not capable of receiving two transmissions at the same time. If two nodes are transmitting towards the same receiver at the same time, both messages will be lost, which has a strongly negative impact on the throughput. The chosen scheduling will therefore have to establish some criteria that allow nodes to transmit knowing that the message won't collide with any other one.

The design of the scheduling establishes a bound to the number of simultaneous transmissions that the network is able to perform in a time slot. Given this bound, it is to be chosen the number of simultaneous transmissions that maximizes the throughput but maintaining, at the same time, a compromise between this number of simultaneous transmissions and the amount of interference this number might cause. This means that, in some cases, it will be desirable, in terms of throughput, to decrement the number of simultaneous transmissions in order to decrease the impact of the interference.

The authors in [4] suppose that all infrastructures are strongly connected, meaning that from any access point any other one can be reached by using just one time slot. They also assume that there are enough infrastructure nodes to guarantee that almost every ad hoc node is capable of reaching any infrastructure node in h_{max} hops or less. To this purpose the nodes are to be divided within what they call sub-graphs, where every node is at most at h_{max} hops from the access point.

It would be easier for the reader to see this with an example.

Let us consider a hops metric instead of the geometrical layout of the network, one can imagine the infrastructure node as the center of a set of concentric circumferences with radius 1, 2, 3, ..., h_{max} hops and in which each one contains the nodes that can reach the center by performing 1, 2, 3, ..., h_{max} hops respectively. The infrastructure node is also the center of an imaginary circle of radius h_{max} hops that contains every node within h_{max} hops of the access point, i. e. every node in the sub-graph. In this case the diameter of every sub-graph is $2 \cdot h_{max}$, as a matter of fact, a message would need to perform h_{max} hops to reach the access point and another h_{max} hops to reach the intended destination. Note that the circumference's notion is not related to the geometrical placement of the nodes; it is related, however, to the probability of having or not a good connection.

Since the number of hops is necessarily an integer, the nodes of the sub-graph are placed in concentric circumferences with radius 1 to h_{max}. Then a set of variables are defined by the authors in [4]. They dub C_i^j the circle centered at infrastructure node *j* with radius *i*, which contains all nodes within *i* hops from infrastructure node *j*; c_i^j is the circumference (also dubbed ring) centered at the *j*th infrastructure node with radius *i*, and which contains the nodes that are exactly *i* hops away from the infrastructure node *j*; finally N_i^j and n_i^j are the number of nodes within the circle and its circumference respectively. A node is said to be in the *i*th level if the shortest path from that node to the closest access point takes exactly *i* hops.

From random graphs theory[9], it is known that the number of adjacent vertices to a certain node (vertices with a good link connection to a certain node) is close to np, the number of vertices/nodes at distance 2 is close to $(np)^2$. Hence, the number of nodes at distance d is approximately $(np)^d$. Due to this theory, the authors in [4] suppose that any base station is connected to np nodes, so, if a set of m base station is there in the network then a total number of mnp nodes will exist in the first ring. It is easy to see that each node in the first ring will have np connections. Thus, the number of nodes in the second ring will be $m(np)^2$.

However, one must also take into account the fact that a node might be contained in several circles (in a different level and/or around another infrastructure node). This means that the

[9] Chapter VII in [5] for a reference

number of nodes at the first level is smaller than mnp since some nodes might be in more than one circle of the first level.

Even more, in the next level (second ring) there may be two sources of error by considering $m(np)^2$. The first one is the overlap between the circles in which a node can be contained in more than one ring, and even in more than one sub-graph. The second one is due to the error committed in the estimation of the number of nodes in the first level. As the reader may see, the error increases in every level, leading to an over-estimation in the number of needed infrastructure nodes and in the number of nodes in every circle and sub-graph.

To solve this problem, the authors in [4] introduce a mechanism to construct disjointed sub-graphs and calculate the expected number of nodes in every sub-graph.

a) Circle's construction

Starting at any access point, the authors build a ring c_1^1 containing all the nodes adjacent to it. Next, they move to another infrastructure node and build the first circle around it c_1^2 taking all the immediate neighbors (nodes directly connected to it) that aren't already contained in c_1^1. They continue performing this operation for every access point *j*, including in c_1^j all nodes directly connected to the infrastructure node *j* and not included in the previous circles.

Although every infrastructure node is expected to be directly connected to np nodes, it is taken into account that a node can be directly connected to more than one access point, thus the number of nodes directly connected to the infrastructure is smaller than mnp. The mechanism explained above constructs the first level assuring that no pair of circles do share any node. Once the number of nodes in the first ring is found then the number of nodes in the following rings is calculated in the same way. These are the results that can be found in [4]:

The authors find out that the number of nodes in the first ring is with high probability:

$$n_1 > n - 5n(1-p)^m + 4n(1-p)^{2m} \tag{11}$$

This means that the number of nodes not contained in this first ring is the difference between the total number of nodes in the network and the total number of nodes in the first ring. This is:

$$n - n_1 < 5n(1-p)^m - 4n(1-p)^{2m} \tag{12}$$

It is demonstrated then that there is a minimum probability p for which all the nodes in the network are contained in the first ring::

$$p > 1 - \sqrt[m]{\frac{5n - \sqrt{(5n)^2 - 16n}}{8n}} \tag{13}$$

And then, the minimum number of base stations needed is given by:

$$m > \frac{\log\left(\frac{5n - \sqrt{(5n)^2 - 16n}}{8n}\right)}{\log(1-p)} \tag{14}$$

Sometimes, due to space or economical reasons, it may not be possible to insert this minimum number of base stations required in order to assure just one hop, which maximizes the achievable throughput. If that was the case, the authors find the minimum probability that assures a maximum number of hops. So, the number of nodes not contained in the first i levels is upper-bounded by:

$$n - \sum_{j=1}^{i} n_j < 5\left(n - \sum_{j=1}^{i-1} n_j\right)(1-p)^{n_{i-1}} - 4\left(n - \sum_{j=1}^{i-1} n_j\right)(1-p)^{2n_{i-1}} \tag{15}$$

And the minimum probability for which the ith level will be the last one is:

$$p > 1 - \sqrt[n_{i-1}]{\frac{5\left(n - \sum_{j=1}^{i-1} n_j\right) - \sqrt{\left(5\left(n - \sum_{j=1}^{i-1} n_j\right)\right)^2 - 16\left(n - \sum_{j=1}^{i-1} n_j\right)}}{8\left(n - \sum_{j=1}^{i-1} n_j\right)}} \tag{16}$$

b) Average number of hops

Once the number of nodes that every level contains is known, then the average number of hops between a node and the nearest base station can be recalculated. Of course, as the number of rings increases so does the average number of hops:

$$\tilde{h} = \frac{1}{n}\sum_{i=1}^{h_{\max}} i \cdot n_i \tag{17}$$

c) Throughput

Finally, the authors in [4] find a bound for the achievable throughput of a hybrid ad hoc wireless network. The throughput is given by:

$$T = (1-\varepsilon)\cdot\frac{m}{2}\cdot\log\left(1+\frac{a_n\beta_n}{\frac{\sigma^2}{P}+\left(K_n(\beta_n)-1\right)\mu_\gamma}\right) \tag{18}$$

At this point, the throughput of a pure ad hoc network and the throughput of a hybrid ad hoc network are both known and therefore they can be compared. In the following work we study the throughput of both cases for different types of probability distribution functions and channel estimations and we will try to analyze the gain of the hybrid ad hoc network over the pure one. At the end of the study we will be able to compare both scheduling ideas and see if there is any improvement in the achievable throughput by the addition of infrastructures.

3. Analysis under a constant probability of having a good connection

After we have seen what the differences and the main advantages of a pure and a hybrid ad-hoc network are, we can start by analyzing the network under different scenarios. The first one is not a real nor possible one but it is a good option in order to understand what we will do in future analysis.

Having a constant probability of good connection is impossible since the channel is not defined and interference will also appear even if there are no nodes transmitting at the same time.

In this hypothetical scenario we find a minimum number of base stations needed and the gain in the throughput obtained. It can also be seen that the relation between infrastructures and nodes in the network goes rapidly to zero as the network increases. We start by analyzing the number of base station needed to then compare the throughput between the two networks and obtain the gain of the hybrid one over the pure one.

3.1 Number of base stations needed in the hybrid ad hoc network

Before we start to study the achievable throughput for different channel estimations we have to fix a way to choose how many base stations we need in our network. In [4][10] a minimum number of base stations needed in the network in order to allow nodes to reach any base station in just one hop are found. As seen before, this low-bound is given by

$$m > \frac{\log\left(\frac{5n - \sqrt{(5n)^2 - 16n}}{8n}\right)}{\log(1-p)} \tag{19}$$

First of all, and before doing any further studies, we can see what happens whenever the probability of having a good connection is a constant and does not depend on the number of nodes existent in the network. Although this might be an hypothetic case and never been found in a real scenario, it would be a good idea to analyze and see the network's behavior

[10] See [4] page 61.

under this supposition seeing how many base station might be needed and what the achievable throughput would be.

In this case, one can see that, whenever $n \to \infty$ the number of base stations needed also goes to infinity $(m \to \infty)$. A proof of this is shown and demonstrated in appendix A.

Although the number of base stations goes to infinite it does it very slowly. The growth behaves like $m \to \alpha \log\left(a/n\right)$ with α a negative constant and a a positive constant. In the figure below one can see how many base station the network needs for a probability of having a good connection $p = 0.015$.

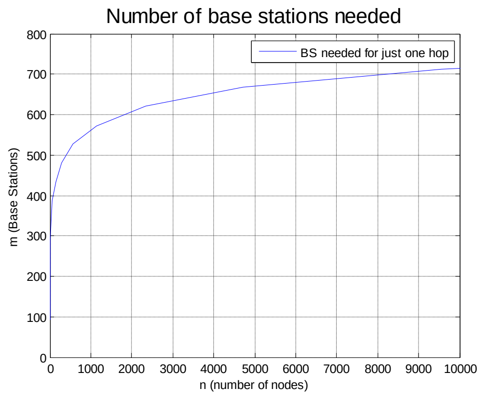

Figure 3. Number of base stations needed for a probability *p* = 0.015

Although the number of base stations is very high, one can see that the relation between the minimum number of infrastructures needed and the total number of nodes in the network goes to zero as the number of nodes increase. This means that for a large enough network, the relation between the number of base stations needed (m) and the total number of nodes in the network (n) is practically equal to zero. We can see a fact of this in the next figure where, for a probability of 0.015, a relation between m and n is represented.

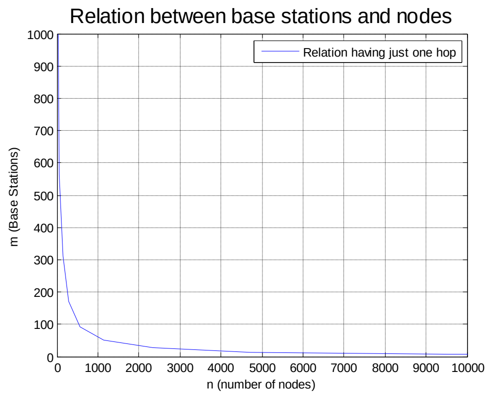

Figure 4. Relation between the number of base stations needed and the number of nodes in the network

As commented before, one can see how, with a constant probability, this relation goes to zero; this means that the larger the network is the less base stations in relation we will need to use.

However, this is the minimum number of infrastructures required to reach any of them with just one hop. We can see now what would happen if we allow the nodes to reach their nearest base station within three hops, this is one, two or at most three hops. It can be demonstrated that reducing the minimum number of infrastructures to $^{3}/_{1000}$ of the total number of nodes, then the number of hops increases to 3. The reader can see that although m continues to go to infinity with n, the number of base stations is drastically reduced.

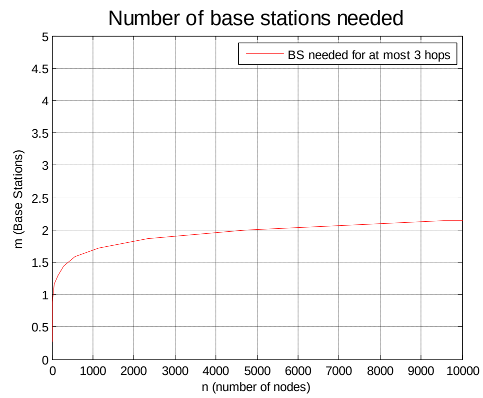

Figure 5. Number of base stations needed if it is allowed up to 3 hops in the network

It is obvious to see that the relation between the number of nodes and the number of access points needed goes to zero much faster than it did before.

In this case as the number of infrastructures needed is not so high in respect of the number of nodes, we can study the gain in the throughput when forcing the nodes to reach the base station within just one hop.

3.2 Throughput of the hybrid and pure ad hoc networks

Let us consider two different cases to this purpose. The first case will be that in which the probability of having a good connection between nodes is $p = 1$ while the second one the probability p will be a constant different than 1.

<u>CASE 1</u> $(p = 1)$

It is easy to see that the inclusion of infrastructures in such a network is nonsense and a waste of money and resources. Obviously, if the probability of having a good connection is 1 it means that all the nodes are well-connected and any node can reach any other node and communicate with it without any problem and satisfying all the conditions.

We can see this mathematically: we know that the minimum number of base stations needed satisfies the condition $m > \dfrac{\log\left(\dfrac{5n-\sqrt{(5n)^2-16n}}{8n}\right)}{\log(1-p)}$. One can see that whenever the probability is 1, the denominator goes to infinity and, therefore, the number of base stations needed is 0. The throughput in this case is no greater than a constant and there would be no gain since no infrastructures are added in the hybrid network.

<u>CASE 2</u> (p = *constant different than 1*)

We can first analyze a pure ad hoc network to compare later with the hybrid one we are studying. For the optimization of k, and whenever the number of nodes n goes to infinity, the maximum achievable throughput becomes constant, again, and it is given by

$$T_{pn} = \frac{(1-\varepsilon)\cdot\alpha\cdot a_n}{p} \tag{20}$$

One can see that if in this case $p = 1$ we get the same throughput as in [3] with probability 1. This is because every node is correctly connected to any other node in the network.

If one takes a look at the achievable throughput of a hybrid network, it is possible to see that as $m \to \infty$ the throughput becomes also constant and the bound for this achievable throughput is given, now, by

$$T_{hn} = \frac{(1-\varepsilon)\cdot a_n}{2\cdot \tilde{h} \cdot p} \tag{21}$$

A proof of this can be found in appendix B.

3.3 Gain of the hybrid ad hoc network over the pure ad hoc network

Now that the bounds for the achievable throughput in the two cases[11] are known, it is possible to calculate the gain in the throughput we may obtain by the addition of base stations in the network. And this gain is

$$G = \frac{T_{hn}}{T_{pn}} = \frac{\frac{(1-\varepsilon)\cdot a_n}{2\cdot \tilde{h}}}{(1-\varepsilon)\cdot a_n \cdot \alpha} = \frac{1}{2\cdot \tilde{h} \cdot \alpha} \tag{22}$$

And this gain is greater than 1 whenever $\tilde{h} < \frac{1}{2\cdot \alpha}$. So, we can conclude that this gain will be greater than the unit always that the average number of hops in the hybrid network is less than half the total possible number of hops of a pure network.

Recall from [3] that this α is a constant that depends on the maximum number of hops that a message would need to perform to reach its intended destination in a pure ad hoc network. In other words, as the number of hops increase the value of this constant α decreases:

$\alpha = \frac{\log(n)}{h\cdot \log(np)}$, and it becomes a constant $\left(\alpha = \frac{1}{h}\right)$ whenever the number of nodes n goes to infinity.

Let us see how this gain behaves with a numerical example. Figure 3 is a graphic of the expected gain we may achieve when adding infrastructures in a network with probability of

[11] We know the achievable throughput for a pure ad hoc network (T_{pn}) and for a hybrid ad hoc network (T_{hn}).

good connection $p = 0.015$. One can see that as the number of nodes increases the achievable gain goes to a constant.

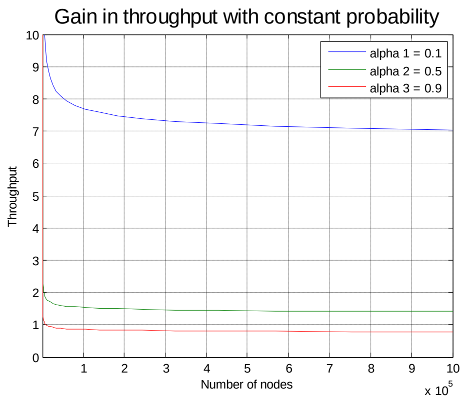

Figure 6. Expected gain for a probability p = 0.015

It is possible to see, on this graphic, the gain for different values of the constant α. Obviously, the gain is less than 1 when the value of alpha is near to 1 because any node in the pure ad hoc network will reach its destination with just one hop and, therefore, the inclusion of base stations is nonsense.

The problem, nevertheless, is that the probability of having a good connection is never a constant but a function of the number of nodes in the network and also of the channel distribution. On the following chapters a series of different channel estimations are studied.

4. Analysis under a shadow fading environment

In wireless communications, the presence of reflectors in the environment surrounding a transmitter and receiver create multiple paths that a transmitted signal can traverse. As a result, the receiver sees the superposition of multiple copies of the transmitted signal, each traversing a different path. Each signal copy will experience differences in attenuation, delay and phase shift while travelling from the source to the receiver. This can result in either constructive or destructive interference, amplifying or attenuating the signal power seen at the receiver. Strong destructive interference is frequently referred to as a deep fade and may result in temporary failure of communication due to a severe drop in the channel signal-to-noise ratio.

The pdf that models the signal strength where strong shadow fading is present is:

$$f_n(\gamma) = (1 - p_n)\delta(\gamma) + p_n\delta(\gamma - 1) \tag{23}$$

where the signal power is 0 in the presence of an obstruction and 1 otherwise. It is easy, now, to find the cumulative distribution function:

$$F_n(\gamma) = \int_{-\infty}^{\gamma} f_n(x)dx = \int_{-\infty}^{\gamma} [(1 - p_n)\delta(\gamma) + p_n\delta(\gamma - 1)]dx = (1 - p_n)u(\gamma) + p_n u(\gamma - 1) \tag{24}$$

and the complementary cumulative distribution function:

$$Q_n(\gamma) = 1 - F_n(\gamma) = 1 - (1 - p_n)u(\gamma) + p_n u(\gamma - 1) \tag{25}$$

A natural choice for the goodness threshold β_n is 1[12] which gives

$$Q_n(\beta_n) = 1 - (1 - p_n)u(\beta_n) + p_n u(\beta_n - 1) = p \tag{26}$$

Is at this point that we can study thoroughly what happens when the number of nodes increases.

[12] Remind that the goodness threshold is chosen supposing no interference in the scenario. So, choosing βn as 1 means that there may exist an obstruction, fading causes a complete loss.

4.1 Throughput of a pure ad hoc network with a shadow fading model

Again, is at this point that we can study thoroughly what happens when the number of nodes increases; in other words, we can study what happens whenever the probability of having a good connection goes to zero[13]. It is said in [3] that according to *Corollary 1*[14] the maximum possible k achieves maximum throughput. Therefore, for large n and $p \to 0$ the throughput is maximized for $p = \frac{\log(n) + \omega_n}{n}$ and is given by:

$$T = \left(1 - \frac{a_n^{\ 2}}{\alpha^2 \cdot (1-a_n)^2} \frac{\log^2(n)}{\log^2(\log(n)+\omega_n)} \frac{1}{n}\right) \cdot a \cdot \alpha \frac{\log(\log(n)+\omega_n)}{\log(n) \cdot (\log(n)+\omega_n)} \cdot n \tag{27}$$

where ω_n is a function that goes to infinity slowly with n[15]. If the probability p is lower than the above value then there will exist isolated nodes while if it is greater the network might be over connected.

4.2 Throughput of a hybrid ad hoc network with a shadow fading model

It is time now to see what happens in a hybrid network. In other words, what would happen with the addition of base stations? Will the throughput increase? If this is the case, which will the gain be?

It is known from [4] that the condition that the probability of error tends to zero as n goes to infinity is necessary to be satisfied.

$$\varepsilon \le \frac{a_n^{\ 2}}{(1-a_n)^2} \cdot 2 \cdot h_{\max} \frac{\left(m\tilde{h} - 1\right)\sigma_\gamma^{\ 2}}{\left(\frac{\sigma^2}{P} + \left(m\tilde{h} - 1\right)\mu_\gamma\right)^2} \to 0 \tag{28}$$

---

[13] We are only interested in the case where p goes to zero. Otherwise the probability would be high enough to consider the inclusion of base stations as nonsense.

[14] One can find a proof of this in R. Gowaikar, B. Hochwald and B. Hassibi "*Communication over a wireless network with random connections*" IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 52, NO. 7, JULY 2006 page 2860

[15] We may therefore choose ω_n to be log(n).

For sufficient large m the condition $\frac{\sigma^2}{P}-\mu_\gamma \geq 0$ is always satisfied and for large n and $p \to 0$ the throughput is maximized, as in [3], by getting $p=\frac{\log(n)+\omega_n}{n}$ and is given by:

$$T=\left(1-\frac{a_n{}^2}{(1-a_n)^2}\cdot 2\cdot h_{\max}\cdot\frac{\left(m\tilde{h}-1\right)\sigma_\gamma{}^2}{\left(\frac{\sigma^2}{P}+\left(m\tilde{h}-1\right)\mu_\gamma\right)^2}\right)\cdot\frac{m}{2}\cdot\frac{a_n\beta_n}{\frac{\sigma^2}{P}+\left(m\tilde{h}-1\right)\mu_\gamma} \tag{29}$$

If (29) is developed, a throughput almost linear in n is obtained for $m\to\infty$:

$$T=\left(1-\frac{a_n{}^2}{(1-a_n)^2}\cdot 2\cdot h_{\max}\cdot\frac{1}{m\tilde{h}}\right)\cdot\frac{a_n}{4\cdot\tilde{h}\cdot\log(n)}\cdot n^{16} \tag{30}$$

The throughput is almost linear in n and requires the network to be sparsely connected.

4.3 Gain of the hybrid ad hoc network over the pure ad hoc network

Now that the throughputs of a pure and a hybrid wireless network are known it is possible to compare both of them and calculate the gain it might be obtained by the addition of infrastructures in the network. Hence, whenever n goes to infinity the expected gain is

$$G=\frac{(1-\varepsilon_M)\cdot\frac{a_n}{4\cdot\tilde{h}\cdot\log(n)}\cdot n}{(1-\varepsilon_H)\cdot a_n\cdot\alpha\cdot\frac{\log(2\cdot\log(n))}{2\cdot\log^2(n)}\cdot n}=\frac{1}{4\cdot\tilde{h}\cdot\alpha}\cdot\frac{2\cdot\log^2(n)}{\log(n)\log(2\cdot\log(n))}=\frac{1}{2\cdot\tilde{h}\cdot\alpha}\cdot\frac{\log(n)}{\log(2\cdot\log(n))} \tag{31}$$

For a very large enough n it is easy to see that the gain goes also to infinity as a function of $\log(n)$.

$$G\to\frac{1}{2\cdot\tilde{h}\cdot\alpha}\log(n) \tag{32}$$

16 Again under the supposition of $p=\frac{\log(n)+\omega_n}{n}\equiv\frac{2\log(n)}{n}$.

4.4 Number of base stations needed

As in the previous section we may start by deciding how many base stations are needed to be established in a network with shadow fading. Since ω_n is a function tending to infinity very slowly, we can therefore suppose p to be $p = \frac{2\log(n)}{n}$. The condition for the minimum number of base stations needed responds (14). However, in this case we have that the denominator depends also on n. One can see that, whenever $n \to \infty$ the number of base stations needed also goes to infinity $(m \to \infty)$. Since p goes to zero as n goes to infinity we can approximate the expression $\log(1-p) = \log(1+(-p)) \approx -p$. Hence, we calculate the limit for m as

$$\lim_{n\to\infty} m = \lim_{n\to\infty} \frac{\log\left(\frac{5n-\sqrt{(5n)^2-16n}}{8n}\right)}{\log(1-p)} \approx \lim_{n\to\infty} \frac{\log\left(\frac{5n-\sqrt{(5n)^2-16n}}{8n}\right)}{-p} = \lim_{n\to\infty} \frac{\log\left(\frac{5n-\sqrt{(5n)^2-16n}}{8n}\right)}{-\frac{2\log(n)}{n}} =$$

$$= \lim_{n\to\infty} \frac{n}{2} \frac{\log\left(\frac{5n-\sqrt{(5n)^2-16n}}{8n}\right)}{-\log(n)} = \lim_{n\to\infty} \frac{n}{2} \cdot \lim_{n\to\infty} \frac{\log\left(\frac{5n-\sqrt{(5n)^2-16n}}{8n}\right)}{\log\left(\frac{1}{n}\right)}$$

It is easy to see that the first term goes to infinity with n. However, it is not so easy to see that the second limit goes to 1 whenever n goes to infinity. We can plot the second limit to see that this is true:

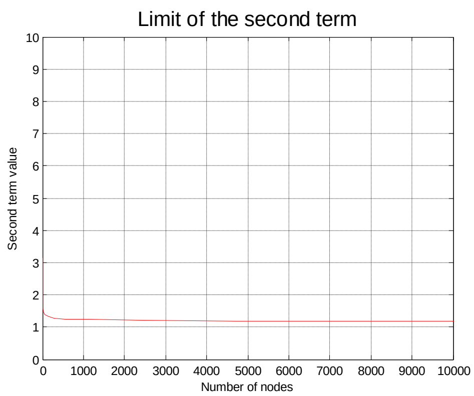

Figure 7. Limit of the second term expression in the above polynomial

We can conclude that for a probability $p = \frac{2\log(n)}{n}$ the number of base stations needed in the network increases, approximately, as $m \approx \frac{n}{2}$ and therefore the choice of m increases almost linear in n (number of nodes).

In the given network the minimum number of base stations needed would be

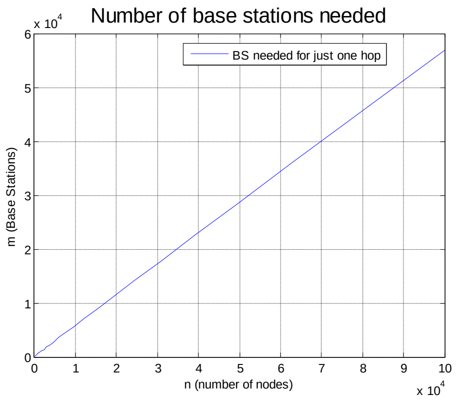

Figure 8. Number of base stations needed for a one-hop hybrid network

One can see that if the number of nodes in the network is 100,000 we will need approximately 57,000 base stations to have, with high probability, just one hop for every node. The relation between the number of base stations and the total number of nodes in the network is nevertheless a 56.98%, approximately. This is a very high percentage.

The problem in here is that this is the minimum m needed for which, with high probability, all of the nodes reach the nearest base station with just one hop. The number of infrastructures could nevertheless be very high and so not very economical to introduce them in the network. One option is reducing this number of base stations so that, in the worst case[17] they are at most the 1% of the nodes in the network.

Let us reduce the number of infrastructures to a $3/1000$ of the total. We will need now, with very high probability, at most 3 hops to reach the nearest base station, but the total number of infrastructures would be now about 200, which is much less than the 1% of the total number of nodes in the network. In this case we increase the number of hops needed to

[17] The worst case in terms of number of infrastructures needed is the maximum number of base stations allowed in the network

reach the nearest base station, but we reduce very much the number of needed infrastructures.

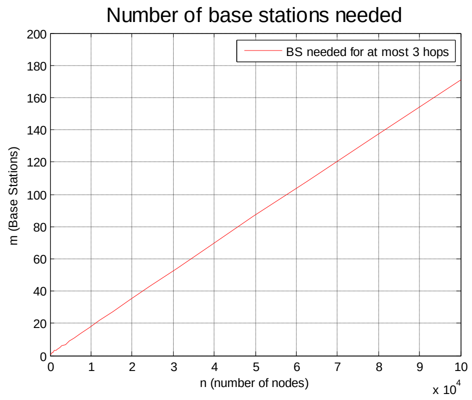

Figure 9. Number of base stations needed for a three-hops hybrid network

One can see that if the number of nodes in the network is 100,000 we will need approximately 170 base stations to have, with high probability, at most three hops. The relation between the number of base stations and the total number of nodes in the network is decreased to a 0.17%, approximately, which is a very interesting percentage. Note that if for any reason the probability p decreases then the number of base stations needed will increase and so will do the ratio between m and n. This is another reason why is preferable to do more than 1 hop (2 or at most 3) rather than just one.

Different from the case in which p is a constant, in this, if we study the relation between the number of nodes in the network and the number of base stations needed when the number of nodes goes to infinity in both cases, the relation tends to a constant. This is

$\lim_{n\to\infty} \frac{m(n)}{n} \to \frac{n/2}{n} = 0.5$. As we said before, if we want to assure that any node will reach its nearest access point with just one hop we will need almost one base station for every two nodes. This is the reason why we have found the second ratio.

$\lim_{n\to\infty} \frac{m(n)}{n} \to 0.001561$ and this means that we will need just one access point for every thousand nodes.

The following graphics show this:

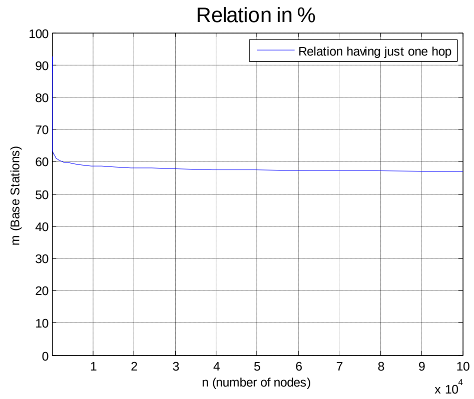

Figure 10. Relation between the number of base stations needed and the total number of nodes in a one-hop hybrid network

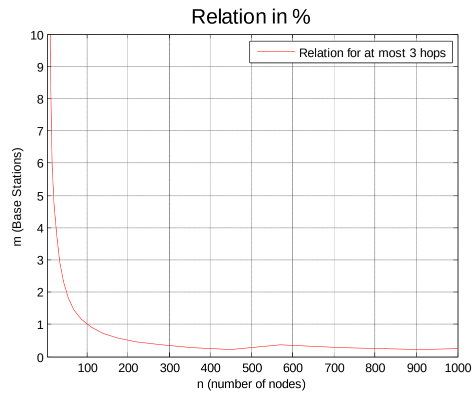

Figure 11. Relation between the number of base stations needed and the total number of nodes in a three-hop hybrid network

Once the number of infrastructures to be put in the hybrid network is decided it is possible, then, to see how many nodes will every circle c_i contain. (Note: recall from [4] that a circle is the number of nodes that reach their nearest base station with a certain number of hops. For instance, c_1 is circle in which the nodes reach their nearest base station with just one hop, c_2 with two, and so on.)

We can see this with an example.

4.5 Numeric example

Supposed a network with $n = 1000$nodes and $p = \frac{2\log(n)}{n}$ [18] the number of base stations needed in this network would be $m = 2$. So, the number of nodes in the first ring would be, with high probability, up to 27; this means that every base station is connected to 13 of the nodes in the first ring. The number of nodes in the second rind ring would be up to 308, which means that, on average, every node of the first ring is connected to 11 of the second. And finally, the number of node in the third ring would be at most 654, which means that now, again on average, every node of the second ring is connected to just 2 nodes of the third one.

With these numbers, we are able now to calculate the average number of hops that a message needs to perform in such a network. This is

$$\tilde{h} = \frac{1}{n}\sum_{i=1}^{h_{\max}} i \cdot n_i = \frac{1}{1000}\sum_{i=1}^{3} i \cdot n_i = \frac{1 \cdot 27 + 2 \cdot 308 + 3 \cdot 654}{1000} \approx 2.6 hops$$

At this point, once the network and scheduling is performed, it is easy to study, calculate and compare the gain in the achievable throughput of hybrid ad hoc network over the pure one, i.e. the gain in the throughput it might be obtained by the addition of base stations in the network.

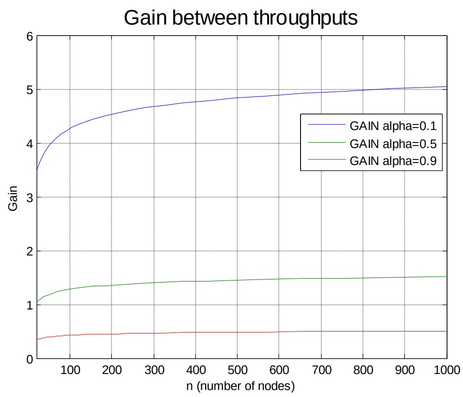

Figure 12. Gain of the average throughput in a hybrid network over a pure ad-hoc network for different values of the parameter α

Like in the previous section, the gain in the throughput depends on the value of the constant α. One can see that if α is near to one there would be no gain in the throughput

[18] for similarity with the simulations made by authors in [3]

since the addition of base stations is nonsense. It is also obvious that as more hops a message need to perform in a pure ad hoc network the less the value of α will be and the greater the gain it will be obtained.

We can therefore conclude that if the channel model is a *Shadow fading model* the addition of base stations implies a great improvement in the throughput.

5. Analysis in a channel with exponential density

An exponential distribution model describes the times in which events occur continuously and independently at a constant average rate. The pdf that models this situation is:

$$f_n(\gamma) = \lambda \cdot e^{-\lambda\gamma} \tag{34}$$

where $\lambda > 0$ is the parameter of the distribution, often called the *rate parameter*. It is easy, now, to find the cumulative distribution function and its complementary:

$$F_n(\gamma) = \int_{-\infty}^{\gamma} f_n(x)dx = \int_{-\infty}^{\gamma} \lambda \cdot e^{-\lambda x}dx = 1 - e^{-\lambda\gamma} \tag{35}$$

$$Q_n(\gamma) = e^{-\lambda\gamma} \tag{36}$$

Although the value of λ can be anyone depending on the rate of the arrivals, a choice of $\lambda = 1$ will make the mean and the variance be the same and equal to one.

The choice for the goodness threshold β_n is nevertheless not as easy to find as in the previous section. However, since the complementary distribution function is, in terms of goodness threshold, $Q_n(\beta_n) = e^{-\beta_n} = p$ a bound for this threshold is found

$$\beta_n \leq \log(n) \tag{37}$$

Authors in [3] find only the throughput for an optimum choice of β_n for the pure ad hoc network. We, instead, will compare both scheduling for this optimum probability and for the optimum probability for our hybrid network and conclude if the addition of base stations means great throughput improvement. That's it; we will first analyze the throughput for the optimum β of the pure ad hoc network and compare it with the throughput we obtain in our hybrid network. Then we find the optimum β of the new hybrid ad hoc network and compare the throughputs we obtain in both networks. Finally we extract conclusions. It will be seen that a gain in the throughput is obtained whatever the value of β is. Obviously, the gain will be higher as the value of β is near to the hybrid optimum value.

a) Optimum probability for pure ad-hoc network

Authors in [3] find out that the optimum value of β_n in order to obtain a maximum throughput is

$$\beta_n = \frac{\log(n)}{2} \tag{38}$$

Which means that the optimum probability of good connection is $p = \frac{1}{\sqrt{n}}$. Note that in this case the probability also goes to zero as n increases and it is easy to see that condition (7) is always satisfied. A bound for the achievable throughput becomes now as shown in [3]:

$$T = \left(1 - \frac{a^2}{\alpha^2 \cdot (1-a)^2} \frac{4}{n}\right) \cdot \frac{a \cdot \alpha \cdot \log(n)}{4} \tag{39}$$

A random network dominated by an exponential pdf has a throughput that scales only logarithmically with n. This network has good connectivity since the number of hops is small, only 2 hops, but is also unfortunately dominated by interference. Thus, only few transmissions can occur simultaneously. In this point is that the introduction of base stations and hybrid scheduling will provide the best improvement.

Let us start by comparing both networks with optimum probability above commented. The reader can notice that this optimum probability for a pure ad hoc network is no longer the optimum in a hybrid ad hoc network. Once this comparison is studied then a set of other values of the probability can be studied in order to see how it behaves.

The first step is, again, to find the achievable throughput obtained in a hybrid network for the probability above commented. This is

$$T = \left(1 - \frac{a^2}{(1-a)^2} \cdot 2 \cdot h_{\max} \cdot \frac{1}{m\tilde{h}}\right) \cdot \frac{a_n \cdot \beta_n}{2 \cdot \tilde{h} \cdot \mu_p} \tag{40}$$

We have previously calculated, and so we know, that $\beta_n = \frac{\log(n)}{2} = \log(\sqrt{n})$ and that the mean of the probability of having a good connection is $\mu_p = 1$. So the average throughput obtained for our hybrid ad-hoc network is bounded by

$$T = \left(1 - \frac{a^2}{(1-a)^2} \cdot 2 \cdot h_{\max} \cdot \frac{1}{m\tilde{h}}\right) \cdot \frac{a \log\left(\sqrt{n}\right)}{2 \cdot \tilde{h}} \tag{41}$$

Now that the throughputs of a pure and a hybrid wireless network are known it is possible to compare both of them and calculate the gain it might be obtained by the addition of infrastructures in the network. Hence, whenever n goes to infinity the expected gain is

$$G = \frac{2}{\tilde{h} \cdot \alpha} \cdot \frac{\log\left(\sqrt{n}\right)}{\log(n)} = \frac{1}{\tilde{h} \cdot \alpha} \tag{42}$$

$$G \approx \frac{h}{2 \cdot \tilde{h}} \tag{43}$$

For the optimum probability the minimum number of base stations necessary to install in the network whenever the number of nodes, n, goes to infinity is

$$\lim_{n\to\infty} m = \lim_{n\to\infty} \frac{\log\left(\frac{5n - \sqrt{(5n)^2 - 16n}}{8n}\right)}{\log(1-p)} \approx \lim_{n\to\infty} \frac{\log\left(\frac{5n - \sqrt{(5n)^2 - 16n}}{8n}\right)}{-p} = \lim_{n\to\infty} \frac{\log\left(\frac{5n - \sqrt{(5n)^2 - 16n}}{8n}\right)}{-\frac{1}{\sqrt{n}}} =$$

$$= \lim_{n\to\infty} C \cdot \sqrt{n}$$

Where C is a constant and, therefore, it can be concluded that for a probability $p = \frac{1}{\sqrt{n}}$ the number of base stations needed in the network $m \propto \sqrt{n}$.

In the given network the minimum number of base stations needed would be

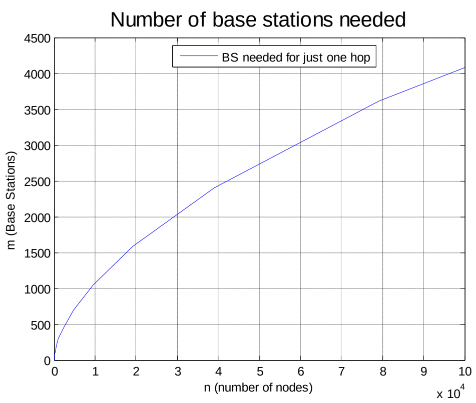

Figure 13. Number of base stations needed in an exponential density scenario

The reader can see that the number of base stations goes to infinity with n, yet the relation between m and n is much lower than in shadow fading model and goes to zero as n increases. For instance, if there is a set of 100,000 nodes in the network, the minimum number of base stations needed represent just a 4.14% and this value decreases as the number of nodes increases.

Different from the shadow fading model in here the relation is not so high and therefore it can be imposed to allow just one hop. The number of infrastructures is not so high and so economically acceptable to introduce them in the network. However, this number of access points might be reduced to a $2/100$ of the total. We will need now, with very high probability, at most 2 hops to reach the nearest base station, and the total number of infrastructures would be now about 85, which is less than the 0.1% of the total number of nodes in the network. In this case the number of hops needed to reach the nearest base station is increased in one, but the number of needed infrastructures is very much reduced.

If the relation between the number of nodes in the network and the number of base stations needed is now studied when the number of nodes goes to infinity in both cases, the relation tends to zero. This is

$$\lim_{n\to\infty}\frac{m(n)}{n}\to\lim_{n\to\infty}\frac{C\cdot\sqrt{n}}{n}=\lim_{n\to\infty}\frac{C}{\sqrt{n}}\to 0 \tag{44}$$

And if the relation is plotted the graphics shows this fact

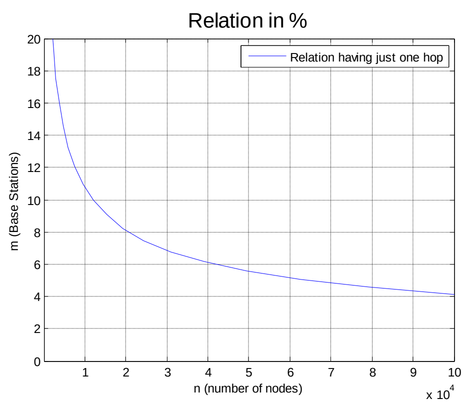

Figure 14. Relation between the number of base stations needed and the total number of nodes in a one-hop hybrid network

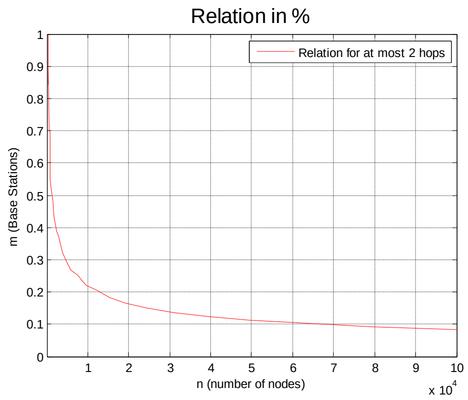

Figure 15. Relation between the number of base stations needed and the total number of nodes in a three-hop hybrid network

Once the number of infrastructures to be put in the hybrid network is decided it is possible, then, to see how many nodes will every circle c_i contain. It can be seen with an example and then comparing the difference between allowing just one hop and allowing at most 2 hops.

Every node reaches its nearest base station in just one hop

Supposed a network with $n = 100{,}000$ nodes and $p = \frac{1}{\sqrt{n}}$, the number of base stations needed in this network so that, with high probability, all of the nodes are in the first ring would be $m = 4{,}144$. There's no need to calculate the average number of hops because, obviously, it will be just one.

At this point, once the network and scheduling is performed, it is easy to study, calculate and compare the gain in the achievable throughput of hybrid ad hoc network over the pure one, i.e. the gain in the throughput it might be obtained by the addition of base stations in the network. Authors in [3] find out that for the optimum β, the number of hops in the network is $h = 2/\alpha$. It is possible to calculate this gain for different values of alpha and it is represented in the following figure.

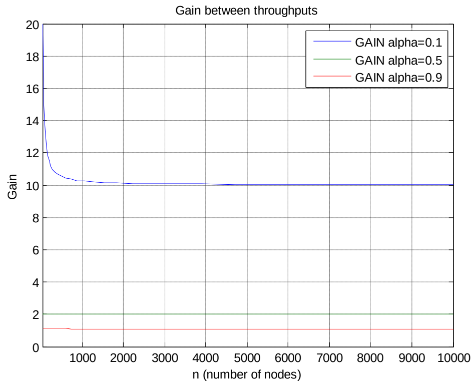

Figure 16. Gain of the average throughput in a hybrid network over a pure ad-hoc network for different values of the parameter α

As always, the gain in the throughput depends on the value of the constant α.

One can see that for this value of β if the number of hops in a pure network is 2 then α is one (recall that the number of hops in the network is $h = 2/\alpha$) and there would be no gain in the throughput since the addition of base stations is, again, nonsense. However, if the number of hops is 4 the throughput is doubled. Thus, it is obvious that as more hops a message need to perform in a pure ad hoc network the less the value of α will be and the greater the gain it will be obtained. In this case the gain is almost the inverse of the value of α or, what is the same, the gain behaves as $h/2$.

Every node reaches its nearest base station in at most two hops

Supposed the same network as above, with $n = 100{,}000$ nodes and $p = \frac{1}{\sqrt{n}}$, the number of base stations needed in this network so that, with high probability, all of the nodes are within the second ring would be $m = 83$. So, the number of nodes in the first ring would be, with high probability, up to 23,117; this means that every base station is connected to 278 of the nodes in the first ring. The number of nodes in the second rind ring would be up to 76,883, which means that, on average, every node of the first ring is connected to 3 of the second.

With these numbers, we are able now to calculate the average number of hops that a message needs to perform in such a network. This is

$$\tilde{h} = \frac{1}{n}\sum_{i=1}^{h_{\max}} i \cdot n_i = \frac{1}{100000}\sum_{i=1}^{2} i \cdot n_i = \frac{1 \cdot 23{,}117 + 2 \cdot 76{,}883}{100000} \approx 1.76 hops \tag{45}$$

At this point, once the network and scheduling is performed, it is easy to study, calculate and compare the gain in the achievable throughput of hybrid ad hoc network over the pure one, i.e. the gain in the throughput it might be obtained by the addition of base stations in the network. In this case is obvious that for α greater than 0.5 the inclusion of base stations is nonsense since the messages of a pure ad hoc network will reach their intended destinations in less or equal than 2 hops.

However, for different values of α lower than 0.5 it might be obtained the following gain graphic:

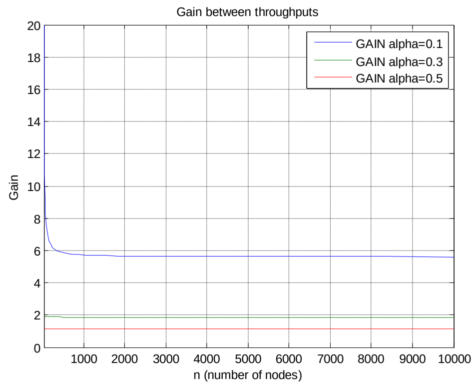

Figure 17. Gain of the average throughput in a hybrid network over a pure ad-hoc network for different values of the parameter α

One might expect the gain to be 1 for $\alpha = 0.5$ as the average number of hops is less than 2 (recall from (44) that it is actually 1.76), but the gain obtained is really a bit more than 1; concretely 1.12. In this case the gain is almost the inverse of the double of the value of α, so the less the value is the greater the gain will be.

So, it has been studied the gain for the optimum value of β in a pure ad hoc network. Let us see which would be the optimum value of this β in a hybrid network and do again the same comparison.

b) Optimum probability for the hybrid network

The optimum value of β for a hybrid ad hoc network with exponential distribution (in terms of throughput) would be a value of $\beta_n \to \infty$. For instance, the value of β may be $\beta_n = \log(n) - \log\log(n^2)$, so that the probability is, as in a shadow fading model $p = \frac{2\log(n)}{n}$.

Note that in this case the probability also goes to zero as n increases and it is easy to see that condition (7) is also satisfied. A bound for the achievable throughput in a pure ad hoc network becomes now:

$$T = \left(1 - \varepsilon_{[3]}\right) a_n \cdot \alpha \cdot \left(1 + \log\log(n^2) - \frac{2\log\log(n^2)}{\log(n)}\right) \tag{46}$$

Again, the first step is to find the achievable throughput obtained in a hybrid network for the probability above commented. We know that

$$T = \left(1 - \frac{a^2}{(1-a)^2} \cdot 2 \cdot h_{\max} \cdot \frac{1}{m\tilde{h}}\right) \cdot \frac{a_n \cdot \beta_n}{2 \cdot \tilde{h} \cdot \mu_p} \tag{47}$$

We have previously calculated, and so we know, that $\beta_n = \frac{\log(n)}{n}$ and that the mean of the probability of having a good connection is $\mu_p = 1$. So the achievable throughput obtained in a hybrid network for the probability above commented is

$$T = \left(1 - \varepsilon_{[4]}\right) \cdot \frac{a_n}{2 \cdot \tilde{h}} \cdot \log\left(\frac{n}{2\log(n)}\right) \tag{48}$$

Now that the throughputs of a pure and a hybrid wireless network are known it is possible to compare both of them and calculate the gain it might be obtained by the addition of infrastructures in the network. Hence, whenever n goes to infinity the expected gain is

$$G = \frac{1}{2 \cdot \tilde{h} \cdot \alpha} \cdot \frac{\log(n) \cdot \left(\log(n) - \log\log(n^2)\right)}{\log(n) + \log\log(n^2) \cdot \left(\log(n) - 2\right)} \tag{49}$$

$$G \approx \frac{1}{2 \cdot \tilde{h} \cdot \alpha} \cdot \left(\log(n) - 1\right) \tag{50}$$

For this probability the minimum number of base stations necessary to install in the network whenever the number of nodes, *n*, goes to infinity is

$$\lim_{n\to\infty} m \approx \frac{n}{2} \tag{51}$$

In the given network the minimum number of base stations needed would be the same as in a shadow fading scenario. Let us see know how the gain behaves in such a scenario with the same example we have been doing so far.

Supposed a network with $n = 100{,}000$ nodes and $p = \frac{2\log(n)}{n}$, the number of base stations needed in this network would be $m = 114$. So, the number of nodes in the first ring would be, with high probability, up to 2,592; this means that every base station is connected to 22 of the nodes in the first ring. The number of nodes in the second rind ring would be up to 43,784, which means that, on average, every node of the first ring is connected to 16 of the second. And finally, the number of node in the third ring would be at most 53,621, which means that now, again on average, every node of the second ring is connected to only 1 nodes of the third one.

With these numbers, we are able now to calculate the average number of hops that a message needs to perform in such a network. This is

$$\tilde{h} = \frac{1}{n}\sum_{i=1}^{h_{\max}} i \cdot n_i = \frac{1}{100000}\sum_{i=1}^{3} i \cdot n_i = \frac{1 \cdot 2592 + 2 \cdot 43784 + 3 \cdot 53621}{100000} \approx 2.51 hops \tag{52}$$

At this point, once the network and scheduling is performed, it is easy to study, calculate and compare the gain in the achievable throughput of hybrid ad hoc network over the pure one, i.e. the gain in the throughput it might be obtained by the addition of base stations in the network.

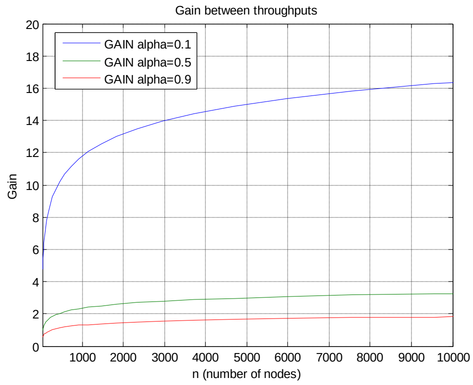

Figure 18. Gain of the average throughput in a hybrid network over a pure ad-hoc network for different values of the parameter α

Different from everything we have studied so far, in this case and with this probability, we always obtain a gain greater than one and, besides, this gain increases with the number of nodes in the network.

Hence, we can conclude that if the channel model has an *Exponential density* the addition of base stations implies a great improvement in the throughput. This improvement would be much greater if the election of the goodness threshold β is not the optimum.

6. Conclusions

First of all recall that we have based all the work under the premises taken in [4], especially the ones obtained in Section VI, which is the scheduling we use to compute the results in our work. The figures and equations shown in this work are basically the simulations of the results previously obtained in [4] and later corroborated in this work.

We have seen that the achievable gain depends always on the parameter α and that if this parameter is high enough the inclusion of infrastructures in a new hybrid network will be nonsense, and sometimes a waste of money.

We have also seen that to achieve the maximum possible throughput all the nodes must reach its nearest base station within one hop, but if we do such a thing we would transform an ad hoc network into a cellular system and we don't want to do that. What our scheduling does is that if the number of relays we need to reach a destination is greater than a certain number of hops then we send the message to the nearest base station which sends the message to nearest infrastructure to the destination.

We have seen that with this scheduling (the one studied in [4] section VI) we obtain an important improvement in the aggregate throughput with the inclusion of base stations. So the transformation of a pure ad-hoc network into a hybrid ad-hoc network will entail to an important gain in the average throughput.

It remains to study, as further work, a new a better scheduling. The scheduling found in [4] and used later in this work is not the optimum, so a better scheduling will probably entail to a better gain in the average throughput.

Furthermore, it can also be studied a new hybrid network which combines the classical geometry and decay-versus-distance law with the goodness probability studied in [1] and [4].

This work is intended to be presented in future conferences and a paper has been also written to present this work to the electrical and computering community.

Comment also that a paper based on [4] and on this work has been written by the authors together, and has been already presented in a Globalcom conference and is intended to be presented in any other conference related on the topic.

It remains me to thank again all the people who have aided and given support to myself while I was doing and writing this thesis and to all the people who were next to me.

Appendices

Appendix A

Claim: Whenever the number of nodes goes to infinity $(n \to \infty)$ the number of base stations needed also goes to infinity $(m \to \infty)$.

Proof:

Recall that the probability of having a good connection is supposed to be a constant and that the number of base stations obeys the relation

$$m > \frac{\log\left(\dfrac{5n - \sqrt{(5n)^2 - 16n}}{8n}\right)}{\log(1-p)} \tag{19}$$

Since p is a constant greater than zero (and obviously less than 1) we can obviate the denominator by just putting a minus sign in front, and the relation between m and n is

$$m > -\log\left(\frac{5n - \sqrt{(5n)^2 - 16n}}{8n}\right) \tag{19.1}$$

$$m > -\log\left(\frac{5}{8} - \sqrt{\frac{25n^2}{64n^2} - \frac{16n}{64n^2}}\right) \xrightarrow[n\to\infty]{} -\log\left(\frac{5}{8} - \frac{5}{8}\right) \tag{19.2}$$

And therefore

$$m > -\log(0) \to \infty \tag{19.3}$$

As we wanted to demonstrate.

Appendix B

Claim: the bound for the achievable throughput of a hybrid network is given, now, by

$$T_{hn} = \frac{(1-\varepsilon)\cdot a_n}{2\cdot \tilde{h}\cdot p} \tag{21}$$

Proof:

From (18) we know that the throughput is bounded by

$$T = (1-\varepsilon)\cdot\frac{m}{2}\cdot\log\left(1+\frac{a_n\beta_n}{\frac{\sigma^2}{P}+(K_n(\beta_n)-1)\mu_\gamma}\right) \tag{18}$$

Where $K_n(\beta_n) = m\cdot\tilde{h}$ so

$$\lim_{n\to\infty} T = \lim_{n\to\infty}(1-\varepsilon)\cdot\frac{m}{2}\cdot\log\left(1+\frac{a_n}{\frac{\sigma^2}{P}+\left(m\cdot\tilde{h}-1\right)\mu_\gamma}\right) \approx \lim_{n\to\infty}(1-\varepsilon)\cdot\frac{m}{2}\cdot\frac{a_n}{\left(m\cdot\tilde{h}\right)p} = (1-\varepsilon)\cdot\frac{a_n}{2\cdot\tilde{h}\cdot p}$$

The average throughput, as we wanted to demonstrate, is bounded by

$$T_{hn} = \frac{(1-\varepsilon)\cdot a_n}{2\cdot \tilde{h}\cdot p} \tag{21}$$

References

[1] R.Gowaikar, B. Hochwald and B. Hasabi "Communication Over a Wireless Network With Random Connections," *IEEE Transactions on Information Theory,* vol. 52, no. 7 pp. 2857-2871, Jul. 2006.

[2] A. Z. Broder, A. M. Frieze, S. Suent and E. Upfals "An Efficient Algorithm for the Vertex-Disjoint Paths Problem in Random Graphs", *Proc. 7th Symp. Discrete Algorithms*, pp. 261-268, 1996.

[3] B. Bollobás, "Random Graphs", *Cambridge University Press*, 2nd edition, Jan. 2001.

[4] M. Palanques, Y. Bar-ness and A. Laufer "Communication over a hybrid ad hoc wireless network" Master Thesis – New Jersey Institute of Technology Feb. 2009

[5] O. Dousse, P. Thiran and M. Hasler, "Connectivity in ad-hoc and hybrid networks," *IEEE Infocom 2002*, pp 1079-1088.