On Downlink Capacity of Cellular Data Networks with WLAN/WPAN Relays

Bozidar Radunovic, Alexandre Proutière

*R*INRIA

On Downlink Capacity of Cellular Data Networks with WLAN/WPAN Relays

Božidar Radunović[*]   Alexandre Proutiere[‡]

N°6050

December 2006

Thème COM

R apport de recherche

On Downlink Capacity of Cellular Data Networks with WLAN/WPAN Relays

Božidar Radunović⋆*   Alexandre Proutiere‡†

Thème COM — Systèmes communicants
Projet TREC

Abstract: We consider the downlink of a cellular network supporting data traffic. In addition to the direct traffic from the base-station, each user is equipped with the same type of 802.11-like WLAN or WPAN interface that is used to relay packets to further users and hence to improve the performance of the overall network. We are interested in analyzing what are the design guidelines for such networks and how much capacity improvements can the additional relay layer bring, in comparison to cellular networks. We consider a realistic dynamic setting where users randomly initiate downloads and leave the system upon transfer completion. A first objective is to provide a scheduling/relay strategy that maximizes the network capacity, which is the traffic in bit/s/cell that the network can support. We find that, regardless of the spatial traffic distribution, when the cell approaches saturation (the number of active users is very large), the capacity-achieving strategy divides the cell into two areas: one closer to the base-station where the relay layer is always saturated and some nodes receive traffic through both direct and relay links, and the further one where the relay is never saturated and the direct traffic does not exist. We further show that it is approximately optimal to use fixed link lengths, and we derive this length. We give a simple algorithm to calculate the cell capacity. The obtained capacity is shown to be independent of the cell size (unlike in traditional cellular networks), and it is 20%-60% higher than already proposed relay architectures when the number of users is large. Finally, we provide guidelines for future protocol design.

Key-words: wireless, cellular, relay, modeling, capacity

⋆ *INRIA - ENS, 45 rue d'Ulm, 75005 Paris. Bozidar.Radunovic@ens.fr

† ‡France Telecom R&D, 38-40 rue du Général Leclerc, 92794 Issy-les-Moulineaux, France. Alexandre.Proutiere@ens.fr

Capacité des réseaux cellulaires de données avec relais de types WLAN ou WPAN

Résumé : Nous considérons la voix descendante d'un réseau cellulaire supportant du trafic de données, et dans lequel chaque utilisateur a la possibilité, en plus de celle de recevoir du trafic directement de la station de base, de relayer du trafic en utilisant une interface radio de type WLAN ou WPAN. Cette fonctionnalité additionnelle permet d'augmenter la capacité du réseau et d'en améliorer la performance. Cette étude vise à donner des règles de conception de tels systèmes, mais aussi de quantifier le gain en terme de capacité que la couche de relais permet d'atteindre. Nous considérons un scenario de trafic réaliste ou les utilisateurs génèrent aléatoirement des transferts de données, et quittent le système lorsque le transfert requis s'achéve. Un premier objectif est d'identifier des algorithmes de d'ordonnancement et de relais qui permettent de maximiser la capacité du réseau, définie par le volume de trafic en bit/s/cellule que le réseau peut supporter. Il s'avère que, quelle que soit la répartition géographique du trafic dans la cellule, ces algorithmes optimaux possèdent à forte charge (lorsque le nombre d'utilisateurs actifs devient très grand) la propriété suivante : ils divisent la cellule en deux parties, la première proche de la station de base ou la couche de relais est toujours saturée et ou les utilisateurs reçoivent du trafic à la fois de la station de base et de certains utilisateurs relais, la seconde plus éloignée ou les utilisateurs ne reçoivent de trafic que par l'intermédiaire de relais. Nous montrons de plus qu'il est approximativement optimal d'utiliser des liens de relais de taille fixe, et nous caractérisons cette taille. Nous construisons également un algorithme simple permettant de calculer la capacité de la cellule. La capacité est contrairement au cas des réseaux cellulaires usuels indépendante de la taille de la cellule, et dépasse de 20 à 60% celle obtenue par des stratégies d'ordonnancement et de relais existantes. Nous donnons finalement quelques règles de conception des réseaux cellulaires avec relais.

Mots-clés : wireless, cellular, relay, modeling, capacity

1 Introduction

1.1 Cellular Networks with Relays

Wireless cellular networks operate on expensive licensed frequencies and their bandwidth is a scarce resource limited by regulations. Recently there has been a lot of interest in increasing the capacity of cellular networks using an additional wireless physical layer that operates on an unlicensed frequency band. New generations of mobile devices are already equipped with WLAN (wireless local-area network) or WPAN (wireless personal-area network) interfaces and the question that arises is whether one can use the available relay structure to improve the service of cellular networks.

In this paper we consider such a scenario. We assume each mobile node in a cellular network, and the base-station (BS), is equipped with an additional relay adapter. The BS can communicate with a mobile node using a *direct* link (transmission over the cellular, expensive frequency) or relaying over one or several mobile nodes using *relay* links (over the unlicensed frequency).

A typical relay technology we have in mind is 802.11 WLAN. The physical layer of 802.11 allows a source and a destination to adapt their communication rate. A source selects an appropriate rate to transmit a packet depending on the link quality and the level of interference at the receiver. If the link quality degrades during the packet transmission, the packet is lost and has to be retransmitted. There are no advanced coding techniques, like incremental-redundancy, which can recover some information from a collided packet. In order to guarantee some link quality, 802.11 MAC introduces the RTS/CTS mechanism that prevents nodes in the neighborhood to interfere with an ongoing transmission. The size of this exclusion area depends on the transmission power of RTS and CTS signaling packets. Many of the existing WLAN and WPAN technologies (e.g. 802.11, 802.15.4) are based on the design principles described above. In this work we shall consider such technologies that use these design principles. Furthermore, we assume all nodes posses the same type of relay interface.

The key component of the system is the scheduler. Since all nodes in the system are in a direct control of the BS, one can design a centralized scheduler which will control direct and relay transmissions. The problem of defining an efficient scheduler can be divided into two subproblems: the routing subproblem that decides which routes shall be used for relay traffic, and the scheduling subproblem that decides which links (relay and direct) will transmit in each slot.

We focus on the downlink data traffic. We assume the uplink traffic has much lower volume, and can be handled by direct transmissions only. Similarly, signaling is done directly, to reduce the latency. Our goal is to find a scheduling/routing policy that maximizes the network capacity, and to characterize how the capacity depends on the choice of relay interface, cell size, etc.

1.2 Related work

Augmenting a cellular network with relays is not a novel concept. Some of the first papers that proposed this kind of architecture are [1, 2, 3]. In [1], the authors suppose that mobile nodes cannot relay and introduce dedicated relays which use unlicensed frequencies in order to improve the capacity. In [2], the authors assume mobile nodes themselves dispose of WLAN interfaces, and provide a routing protocol that finds and maintains relay routes. In [3, 4], small networks with 1-hop relays are considered.

Scheduling algorithms for relay networks are discussed in [5, 6]. In [5], the authors discuss several simple scheduling schemes. More advanced scheduling techniques are considered in [6]. There, as opposed to the other related work, it is supposed that the BS and the relays use the same frequency band. Consequently, the BS transmits only to the nearest nodes, and the others receive relay traffic only.

What is common for all the proposed relay protocols is that none of them is based on the objective to maximize a certain network-wide performance criterion. Instead, they are based on a simple local heuristic that considers relaying only for those nodes whose direct communication with the BS is of very low quality. That way one node will never receive traffic from *both* relay and direct links. Typically, closer nodes will receive traffic *only* directly, and distant nodes *only* over relay links.

1.3 Problem Statement and Our Contributions

We wish to find the optimal scheduling and routing policies that maximizes the throughput (capacity) of a densely populated cellular network with relays. We are further interesting in dimensioning such a network, that is calculating its capacity given a certain spatial traffic density.

We find the optimal scheduling and routing policies using different heuristics. We show that it is approximately optimal for each node to use as a relay a node that maximizes *transport capacity* (the rate of a relay link times its length). When a network is highly populated, this policy results in equal-size relay links.

The optimal scheduling divides a cell into two regions. The first region, around the BS, is such that the relay channel is fully saturated. Nodes in this region may receive traffic *both* from relays and directly from the BS, which contrasts with previously proposed relay protocols [6, 4, 2, 3]. In the other region, the relay channel is never saturated, and there is no direct traffic to users in this region.

The routing heuristic we proposed can be directly used in a real protocol design. On a contrary, our scheduling scheme is centralized, and is difficult to implement in a realistic environment. We give several guidelines towards practical implementations, but leave a precise definition of the scheduling scheme for future work.

We further present a simple algorithm to approximately calculate the capacity of a cell with relays, which can be easily used for dimensioning purposes. We show that the capacity remains constant, independent of the cell size, which is in contrast with cellular networks with no relay where the capacity decreases exponentially. We also show a significant improvement in the capacity as compared to when the direct traffic is scheduled only to the nodes nearest to the BS [6]. This improvement is even larger when a low-rate WPAN is used instead of WLAN as a relay network.

1.4 Organization of the Paper

In the following section, we precisely define the modeling assumptions and the performance objectives. In Section 3, we evaluate the capacity of networks with relays. We first derive the results for 1D networks with fixed relay link lengths, then we present a heuristic for defining optimal routing, and finally we generalize the findings to 2D networks. In Section 4, we give guidelines on how to design the optimal routing and scheduling policies. In Section 5, we present extensive numerical results. Finally in Section 6, we conclude.

2 Models and Objectives

We consider the downlink of a single cell whose transmission resources (power and bandwidth) are shared by a dynamic population of data flows. Flows are randomly generated by users and leave the network once the corresponding data transfer has been completed. Flows are characterized by their sizes but also by the position of the corresponding users in the cell. We assume here that users remain still during the entire duration of the data flows. Throughout the paper, we will consider both 1D linear or 2D cells.

2.1 Traffic Characteristics

We denote by $\mathcal{C}$ the set of locations in the cell (this set might be discrete or continuous). The traffic model may be very general. We just assume that data flows are generated at location $x \in \mathcal{C}$ according to a stationary ergodic process of intensity $\lambda(x)$. These flows have arbitrary distributed sizes of mean $\sigma(x)$. The traffic intensity at location x in then defined by $\rho(x) = \lambda(x) \times \sigma(x)$ (in bit/s). We further define $p(x)$ as the proportion of traffic generated at location x: $p(x) = \rho(x)/\rho$, where ρ is the total traffic generated in the network, i.e., $\rho = \int_{\mathcal{C}} \rho(x) \, \mathrm{d}x$.

2.2 Radio Resources - Scheduling and Relay Policies

We next describe the two radio interfaces we shall use.

2.2.1 Direct Transmissions from the BS

We assume that the BS transmits at full power and serves only one user at a time. The service rate of a user at location x is denoted by $C_d(x)$. This rate is a function of the SINR at the receiver and can be well-approximated by Shannon formula:

$$C_d(x) = W_1 \log_2(1 + \mathrm{SINR}(x)). \tag{1}$$

This assumption is quite realistic (up to a multiplicative factor) for example in the case of CDMA 1Ev-Do or UMTS/HSDPA systems. Denote by $t(x) = 1/C_d(x)$. We also assume there is a direct uplink channel, also used for signaling (scheduling, acknowledgments, etc.).

2.2.2 Relay Capabilities

We assume the relay channel is based on the design principles of 802.11 MAC/PHY (nevertheless, our model is valid for most of other WLAN and WPAN physical layers that are designed on the same principles). It supports variable transmission rates. If a signal, coded for a given rate, is received at an SINR below the threshold, the packet is lost. We assume that more advanced techniques, like incremental redundancy coding, are not available at the physical layer.

In order to control the interference at a receiver, we use the idea of the RTS/CTS signaling. When a node wants to receive a packet, it sends a CTS signal. All nodes in an area around a destination that hear the CTS signal have to refrain from sending. We denote the radius of this area, assumed circular, by D, and it depends on a power of the transmitted signal and the attenuation function. Note that in our case there is no need to exclude nodes that hear the RTS signal. This is typically done to ensure ACK reception. However, the signaling consumes a small fraction of bandwidth and we suppose that it is performed over the direct link, thus by omitting RTS, we avoid unnecessary exclusions and increase the capacity of the network. To simplify the analysis and the practical implementation of relay policies, we consider that the exclusion area D will be the same for all relay nodes. This assumption will be discussed in Section 3.3.

Furthermore, we assume that relay nodes transmit at full power (denote by P^{RELAY}) when transmitting. P^{RELAY} is assumed to be identical for all nodes. The choice of full power has been extensively justified in the literature on rate-adaptive, multi-hop networks, see e.g. [7], [8].

For a given link we need to choose coding rate C_r as a function of link length l. Loosing and resending a packet is expensive hence it is important to choose a rate sufficiently low to avoid packet errors. We will choose

$$C_r(l) = W_2 \log_2 \left(1 + \frac{P^{\mathrm{RELAY}} l^{-\alpha}}{N_2 + k P^{\mathrm{RELAY}} D^{-\alpha}} \right), \tag{2}$$

where k is a security factor guaranteeing low packet error rate. This factor basically quantifies the maximum interference generated by other active relay nodes such that the packet error rate on the considered link is negligible.

Finally, note that our system is *fully centralized*, unlike ad-hoc WLAN/WPAN networks. This means that the exclusion principles can be enforced and acknowledgments can be transmitted by the BS, using the direct channel. These packets are short, unlike data packets, hence the overhead is low. Consequently, there is no need for distributed protocols (like sending RTS/CTS); we can emulate the principles of such protocols in a centralized scheduler on the BS. In this work we will thus focus on centralized solutions which are typically much more efficient.

2.3 Scheduling and Relay Policies

We now provide a model to describe how radio resources can be shared by active users. We fix the number of active users and their positions in the cell. Denote by $\mathcal{N} \subset \mathcal{C}$ the finite set of locations of active users and by $\mathcal{L} \subseteq \mathcal{N}^2$ a set of possible relay links (those whose rate is larger than some minimum). Without loss of generality we may assume that two users cannot be at the same position.

2.3.1 Scheduling BS resources

The BS shares its power in time between active users. We denote by $\tau(x)$ the proportion of time the BS serves a user at position $x \in \mathcal{N}$. For example, in the Proportional Fair Scheduler of the CDMA 1Ev-DO standard, $\tau(x)$ is inversely proportional to the feasible rate at position x, $C_d(x)$. A feasible scheduling policy is such that:

$$\sum_{x \in \mathcal{N}} \tau(x) \leq 1. \tag{3}$$

For simplicity of the model, we do not explicitly consider fast fading. However, it can easily be incorporated in (3).

2.3.2 Relay policies

To describe a relay policy, one first has to define a notion of transmission profile. A profile j is a set of simultaneously active relay links: $j = \{(s_1, d_1), \ldots, (s_p, d_p)\}$. Profile j is feasible if and only if the distance between any pair of positions (s_m, d_n) is greater than D for all $m \neq n$. Denote by $\mathcal{J}$ is the set of all possible profiles. A relay policy consists in activating the links from profile $j \in \mathcal{J}$ for transmission a proportion of time $\tau_r(j)$. The relay constraint then reads:

$$\sum_{j \in \mathcal{J}} \tau_r(j) \leq 1. \tag{4}$$

Here again, we do not consider fast fading in the model for simplicity. It can be incorporated as explained in [6].

Unfortunately the number of possible profiles explodes when the number of active users grows, and it then becomes difficult to identify optimal relay policies. Instead, we use the notion of cliques, see e.g. [9], [10].

Definition 1 *A clique is any set of links such that no two links from a clique are allowed to transmit at the same time, and that no other link can be added, satisfying these constraints.*

Now let $\tau_r(s,d), ((s,d) \in \mathcal{L})$ be the proportion of time node s sends relay traffic for node d, $\tau_r(s,d) = \sum_{j \in \mathcal{J}:(s,d) \in j} \tau_r(j)$. Denote by $\{Q_e\}_{e \in \mathcal{E}}, Q_e \subseteq \mathcal{L}$ the set of all cliques. As demonstrated in [9], any feasible relay policy can be characterized by the following set of constraints:

$$\sum_{(s,d) \in Q_e} \tau_r(d,s) \leq \gamma, \quad \forall e \in \mathcal{E}. \tag{5}$$

For $\gamma = 1$ we obtain an upper-bound, that is any feasible policy has to satisfy the constraints (but not any policy that satisfies the constraints is feasible). For $\gamma = 2/3$ we obtain a lower-bound: if the constraints are satisfied, the policy is feasible.

We will first define an optimal relay policy satisfying the constraints relative to the cliques (5), for the upper-bound $\gamma = 1$. We will then prove that this optimal policy corresponds to an actual policy, i.e., that it also satisfies constraints (4).

Let us also define $\phi_d(x) = C_d(x)\tau(x)$ to be the rate of traffic directly sent from the BS to users at position x, and $\phi_r(s,d) = C_r(d-s)\tau_r(s,d)$ to be the rate of traffic sent from users at position s to users at position d. Finally denote by $\phi(x)$ the rate at which a user at position x is served. Then a feasible scheduling/relay policy has to satisfy the following flow conservation constraint:

$$\sum_{s:(s,x) \in \mathcal{L}} \phi_r(s,x) + \phi_d(x) = \phi(x) + \sum_{d:(x,d) \in \mathcal{L}} \phi_r(x,d). \tag{6}$$

2.3.3 Existing scheduling and relay policies

we now define two existing scheduling and relay policies that we shall use as a comparison with our capacity-achieving policy. The first one is the *direct* policy for which no relaying is allowed ($\tau_r(s,d) = 0$ for all s,d). The second one is the *relay* policy, as defined in [6]. It assumes that only the nodes that cannot relay are supplied with the direct traffic ($\tau(d) = 0$ if there exists $s, (s,d) \in \mathcal{L}$).

2.4 Performance Objectives and Cell Capacity

2.4.1 Cell capacity

In data networks, users perceive performance through the time to download a document, i.e., through the flow duration. Under some given scheduling and relay policies, the number of active data flows at the different locations in the cell evolves according to some stochastic process.

A necessary and sufficient condition for the time to download documents to remain finite is the stability of the process of the number of active data-flows. Thus stability is the first indicator for quality of service. For a fixed spatial traffic distribution $\{p(x)\}_{x\in\mathcal{C}}$, an optimal scheduling and relay scheme should maximize the total traffic intensity compatible with stability. This maximum is called the cell capacity, and denoted by $\rho^{\star}$. In the next two sections, we fix the spatial traffic distribution and compute the cell capacity. To this end, we construct an optimal scheduling/relay scheme that ensures stability whenever $\rho < \rho^{\star}$. Such a scheme is said to *achieve maximum stability*. Here, since we only consider stability as a design objective, we can design the scheme when the system is saturated only. Hence we may assume that there are users/relays everywhere in the cell, $\mathcal{N} = \mathcal{C}$.

The cell capacity in absence of relays was introduced and extensively studied in [11]. It is the maximum traffic intensity ρ such that there exists a stable scheduling policy, i.e., such that there exists $(\tau(x), x \in \mathcal{C})$ with $\int_{\mathcal{C}} \tau(x) \leq 1$ and for all $x \in \mathcal{C}$, $\rho p(x) < \tau(x) C_d(x)$. It is straightforward to verify that the cell capacity is given by $\rho^{\star} = \left(\int_{\mathcal{C}} t(x)p(x)dx\right)^{-1}$. Furthermore, any work-conserving scheduling policy achieves maximum stability.

In our setting with relays, a capacity achieving scheme solves the following linear program:

$$\max \quad \rho \tag{7}$$

$$\text{s.t.} \quad \int_{x\in\mathcal{C}} \tau(x)\,\mathrm{d}x \leq 1, \tag{8}$$

$$\int_{(s,d)\in Q_e} \tau_r(s,d)\,\mathrm{d}s\,\mathrm{d}d \leq 1, \forall e \in \mathcal{E}, \tag{9}$$

$$\rho p(x) < \phi_d(x) + \int_d \phi_r(x,d)\,\mathrm{d}d - \int_s \phi_r(s,x)\,\mathrm{d}s, \forall x. \tag{10}$$

We use the continuous setting here just to emphasize that the system is saturated, and that there are users at each position in the cell. Note that the constraint (10) ensures stability. Also note that the solution of the above problem remains the same if we put equalities in constraints (10) and (8).

2.4.2 Optimal scheduling/relay policies

As mentioned above, to define capacity achieving policies we may assume that there are users/relays everywhere in the cell. In practice however, most of the time, the number of active users is limited. In Section 4, we complete the design of optimal policies considering scenarios with limited number of users. Denote by $\mathcal{N}$ the set of positions of active users. Usual policies are based on optimization techniques [12]. For example, a policy can result in rates $\phi(x), x \in \mathcal{N}$ that maximize some system *utility*[1] subject to constraints (3)-(5)-(6). When the system is saturated these constraints are identical to those of linear program (7)-(10). In that case, any Pareto-optimal, and then any utility-based, policy is the solution of the linear program (7)-(10) for some particular $\{p(x)\}_{x \in \mathcal{N}}$, and if $\phi(x) = p(x)\rho$. Identifying $\{p(x)\}_{x \in \mathcal{N}}$ that leads to a given Pareto-optimal policy is usually not easy. However to realize a max-min fair policy, we may choose a uniform $p(x)$.

These observations suggest that we can design efficient scheduling/relay policies exploiting the characteristics of the solution of the linear program (7)-(10), and for different distributions $\{p(x)\}_{x \in \mathcal{C}}$. A policy based on the solution when p is uniform will be close to max-min fairness, whereas for example, if we choose $p(x)$ inversely proportional to the distance from x to the BS, the obtained policy will realize a better trade-off between efficiency and fairness.

3   Capacity under Optimal Relay Schemes

In this section, we evaluate the traffic capacity of a cell with relays in different scenarios. We first consider the case of the linear, one-dimensional cell where users are located in $[0, R]$, and where the BS is located at 0. Later, we will extend the analysis to two-dimensional cells.

3.1   Fixed Relay-link Sizes

Consider first the case where the relay-links are of fixed size D_r and the rate of these links is $C_r = C_r(D_r)$. Later, in Section 3.2, we will show that for a certain D_r this solution is indeed close to optimal. We also assume that users at distance $x < D_r$ from the BS may receive relay traffic from the BS (thus, close to the BS, the relay link sizes can be smaller than D_r).

We are interested in evaluating the maximum traffic intensity ρ such that there exists a scheduling and relay strategy stabilizing the network. The cell capacity is the solution of

---

[1]For instance, to realize a Proportional Fair policy, the utility is chosen as $\sum_x \log \phi(x)$.

the linear program (7). The constraint limiting the BS transmissions is given by:

$$\int_0^R \tau(x)\,\mathrm{d}x \le 1. \tag{11}$$

Here the cliques are easy to identify: for all $x < R - D - D_r$, the segment $[x, x + D + D_r]$ is a clique. Hence the constraints relative to the cliques are given by:

$$\int_x^{x+D+D_r} \tau_r(u)\,\mathrm{d}u \le 1, \quad \forall x \in [0, R - D - D_r]. \tag{12}$$

The flow conservation constraints are:

$$\rho p(x) < C_d(x)\tau(x) + C_r(\tau_r(x) - \tau_r(x + D_r)1_{\{x + D_r \le R\}}). \tag{13}$$

To simplify the notation, we define $\tau(x) = 0 = \tau_r(x)$ for all $x < 0$ and $x > R$.

We now define a scheduling/relay scheme that will be shown to maximize the cell capacity. The idea of this scheme is that since the cell capacity is strongly limited by users at the cell boundaries, these users should be served by relays only. Formally this scheme is defined as follows.

The MaxRelay scheme Assume that the cell capacity $\rho^\star$ is known and define X_r by:

$$X_r = \inf(x : \forall y > x, \int_y^{y+D+D_r} \tau_r'(u)\,\mathrm{d}u < 1), \tag{14}$$

where

$$\tau_r'(x) = \frac{\rho^\star}{C_r} \sum_{i \ge 0} p(x + iD_r)1_{\{x + iD_r \le R\}}.$$

$\tau_r'(x)$ may be interpreted as the proportion of time users at location x should receive relay traffic (from relay users located at $x - D_r$) destined for itself and all its downstream relay users at locations $x + iD_r$, $i \ge 0$ (as if there were no direct traffic). The MaxRelay scheme is defined by:

$$\tau_r^\star(x) = \begin{cases} \tau_r'(x), & \text{if } x > X_r, \\ \tau_r(x + D + D_r), & \text{if } x \le X_r. \end{cases} \tag{15}$$

$$\tau^\star(x) = \begin{cases} 0, & \text{if } x > X_r, \\ t(x)(\rho^\star + C_r(\tau_r^\star(x + D + D_r) - \tau_r^\star(x))), \\ \qquad \text{if } x \le X_r. \end{cases} \tag{16}$$

To prove the optimality of this scheme, we need the following assumption.

Assumption 1 *The function $w(x) = t(x) - t(x - D_r)$ is increasing.*

This assumption is exact when the distance to the BS is not too small (say less than 100m with usual propagation scenario). Furthermore, the MaxRelay scheme proves to be almost optimal even in absence of this assumption.

Theorem 1 *When the spatial traffic distribution p is fixed, and under Assumption 1, the MaxRelay scheme achieves maximum stability.*

Proof. Denote by $\rho^\star$ the cell capacity, i.e., the maximum traffic intensity ρ compatible with constraints (11)-(13). It is straightforward to prove that the schedules $\tau^\star$ and $\tau_r^\star$ achieving this maximum are such that the constraints (11) and (13) are saturated, which just means that the BS always transmits and that the service rate at a given location is equal to the traffic intensity generated at this location. Then, we have:

$$1 = \rho^\star \int_0^R p(x)t(x)\,\mathrm{d}x + C_r \int_0^R t(x)(\tau_r^\star(x + D_r) - \tau_r^\star(x)\,\mathrm{d}x).$$

Define $t(x) = 0$ if $x < 0$. Now assuming $\rho^\star$ is known, $\tau_r^\star$ is the solution of the following linear program:

$$\mathbf{LP1} : \max \quad \int_0^R \tau_r(x)(t(x) - t(x - D_r))\,\mathrm{d}x \tag{17}$$

$$\text{s.t.} \quad \int_x^{x+D+D_r} \tau_r(x)\,\mathrm{d}x \leq 1, \tag{18}$$

$$C_r(\tau_r(x) - \tau_r(x + D_r)) \leq \rho^\star, \tag{19}$$

$$\tau_r(x) = 0, \text{for } x > R, \tag{20}$$

$$\tau_r(x) \geq 0, \forall x. \tag{21}$$

Assume first that we know the optimal relay scheme $\tau_r^\star(x)$ for all $x > X_r$. Then consider the following linear program:

$$\mathbf{LP2} : \max \quad \int_0^{X_r} \tau_r(x)(t(x) - t(x - D_r))\,\mathrm{d}x \tag{22}$$

$$\text{s.t.} \quad \int_x^{x+D+D_r} \tau_r(x)\,\mathrm{d}x \leq 1, \tag{23}$$

$$\tau_r(x) = \tau_r^\star(x), \text{ for } x > X_r. \tag{24}$$

If the solution of LP2 satisfies the constraints of LP1, then it will also be the solution of LP1. Denote by $\lambda(x)$ the Lagrange multiplier associated with constraint (23), for $x \in [0, X_r]$.

We now identify the term in front of $\tau_r(x)$ in the Langrangian of LP2, $t(x) - t(x - D_r) - \int_{\max(x-D-Dr,0)}^{x} \lambda(u)\,\mathrm{d}u$ (we write $t(x) = 0, x \leq 0$). This term must be null when τ_r is the solution of LP2. Since the function $t(x) - t(x - D_r)$ is increasing in x we deduce that $\lambda(x) > 0$.

From KKT optimality conditions we conclude that for all $x < X_r$, $\int_x^{x+D+D_r} \tau_r(x)\,\mathrm{d}x = 1$, which further implies that: $\tau_r(x) = \tau_r(x + D + D_r)$ for all $x < X_r$. The obtained solution satisfies constraints of LP1, so it must be the solution of LP1. Hence we have proved that LP1 is equivalent to the following linear program.

LP3 :
$$\max \quad \int_0^R \tau_r(x)(t(x) - t(x - D_r))\,\mathrm{d}x \tag{25}$$
$$\text{s.t.} \quad \tau_r(x) = \tau_r(x + D + D_r), \forall x < X_r, \tag{26}$$
$$C_r(\tau_r(x) - \tau_r(x + D_r)) \leq \rho^\star, \tag{27}$$
$$\tau_r(x) = 0, \text{for } x > R, \tag{28}$$
$$\tau_r(x) \geq 0, \forall x. \tag{29}$$

Now one can easily verify that the solution of LP3 satisfies $\tau_r^\star(x) = \tau_r'(x)$ for all $x \geq X_r$, which concludes the proof.

Finally we need to show that the MaxRelay scheme can actually be realized, since constraints relative to the cliques provide an upper bound on the feasible rate region. However, it is easy to see that a simple symmetric schedule, where nodes $D + D_r$ far apart are scheduled at the same time can implement the MaxRelay scheme, which concludes the proof. $\square$

We illustrate the MaxRelay scheme in Figure 1. We see that there exist two regions. The first one is for $x < X_r$, and in this region the relay PHY is fully saturated ($\int_x^{x+D+D_r} \tau_r(x)\,\mathrm{d}x = 1$), and some nodes in the region receive both direct and relay traffic. The second region is beyond X_r. For $x > X_r$ the relay PHY is never saturated ($\int_x^{x+D+D_r} \tau_r(x)\,\mathrm{d}x < 1$) and there is no need for direct traffic as it is expensive.

An other interesting observation is that these two regions always exist regardless of the spatial traffic distribution $p(x)$, although $p(x)$ does influence the values of X_r and $\tau_r'(x)$.

3.2 Variable Relay Link Lengths and Rates

Next we relax the restriction on fixed link lengths. We allow each node to relay over multiple nodes, and we assume that the rate of each relay link depends on its length, as explained in Section 2.2.2. Our goal is to derive the optimal scheduling strategy and, in particular, the optimal relay routing strategy.
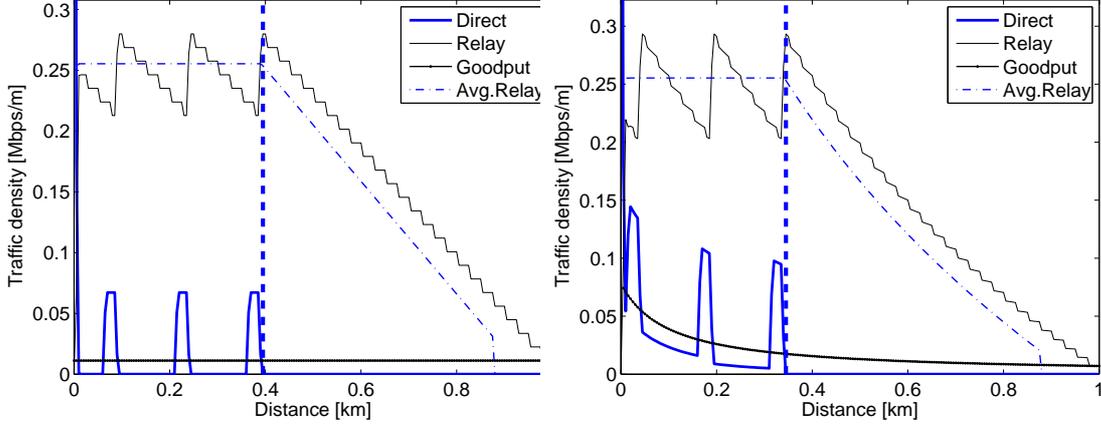
Figure 1: Examples of optimal scheduling/relay schemes for fixed link lengths ($D_r = 30$m, $D = 100$m). We consider 1D cell of 1km and different traffic distributions $p(x) = 1$ (left) and $p(x) = 1/(0.1+x)$ (right). The achieved goodput density is $\rho^\star = 12.2$kbps/m. Average relay traffic is averaged over each clique $[x, x + D + D_r]$, and it is saturated for $x < X_r$.

We will proceed as in Section 3.1. First, we will identify a region $[0, X_r]$ in which all cliques are saturated and show that in the remaining area $(X_r, R]$ no cliques are saturated. Then we will show that a relay routing using links of a certain fixed length is close to optimal. We will also specify this optimal length.

We cannot theoretically prove the results in this section due to the high complexity of the problem. Instead, we demonstrate them using numerical simulations. We calculate the optimal solution by solving the discrete version of linear program (7)-(10) for 200 equidistant nodes and for different values of network radius R, the exclusion area radius D and spatial traffic distribution $p(x)$. We then compare this optimal result with our proposed heuristic, and verify the results presented in this subsection.

Before presenting results, we first need to describe the cliques in the variable link length setting. Let us denote with D_r^{MAX} the maximum allowed relay link length.

Lemma 1 *The only cliques that exist in this networks are $Q(x) = \{(s,d) \in [0,R]^2 \,|\, (s \le x + D$ or $d \ge x), 0 \le d - s \le D_r^{MAX}\}$, for all $x \in [0, R - D]$.*
Proof. We first have to show that $Q(x)$ is a clique, that is, that every two links in $Q(x)$ block each other and that no other such link can be added. It is easy to see that for any two links $(s_1, d_1), (s_2, d_2) \in Q(x)$ we have that $\min(s_1 - d_2, s_2 - d_1) \le D$. Furthermore, we need to verify that if a link (s_1, d_1) does not belong to $Q(x)$ then it is not blocked by all links from $Q(x)$. If $s_1 \le d_1 < x$ then link (s_1, d_1) is not blocked by $(x + D, d), d > x + D$, nor the other way around. If $x + D < s_1 \le d_1$ then link $(s, x), s \le x$ is not blocked by (s_1, d_1), nor the other way around.

Finally, we have to show that there exist no other set $Q' \neq Q(x)$ for all x, which is a clique. Let $x = \min\{d \,|\, (s,d) \in Q'\}$. Then, $x + D \geq \max\{s \,|\, (s,d) \in Q'\}$ because otherwise the two links would not interfere. But the set of points (s,d) that satisfy constraints $d \geq x, s \leq x + D, s \leq d \leq s + D_r$ is exactly $Q(x)$ hence $Q' \subseteq Q(z)$. □

Next, in Proposition 1 and Proposition 2, we present the main results of this section

Proposition 1 *Let $X_r \in [0, R]$ be the maximum such that $\int_{(s,d)\in Q(x)} \tau_r(s,d)\,\mathrm{d}d\,\mathrm{d}s = 1$. Then for all $x < X_r$ we also have $\int_{(s,d)\in Q(x)} \tau_r(s,d)\,\mathrm{d}d\,\mathrm{d}s = 1$. Furthermore, no node after $X_r + D$ receives direct traffic.*

The first part of the proposition has been verified by simulations. One example is depicted in Figure 2, left.

The second part of the proposition follows immediately. Consider nodes $x, y > X_r + D, y < x$. Both nodes do not belong to any saturated clique. Hence, if $\phi_d(x) > 0$ we can redirect some of the direct traffic to y instead, and forward it from y to x using relay, since it is not saturated. That way we gain some of the BS transmission time, which contradicts with optimality.

Proposition 2 *Let $D_r = \arg\max_l l \cdot C_r(l)$, where $l \cdot C_r(l)$ is the transport capacity of a link of length l, as defined in [13]. It is approximately optimal for all $d < X_r$ to use as a relay node $s = \min(d - D_r, 0)$. The optimal relay link length for all $d > D_r$ is thus D_r that maximizes the transport capacity, and it is independent of the location of node d.*

As above, this proposition is based on a heuristic verified by simulations. Although we were not able to formally prove if, we provide below some intuitive explanations to justify it.

Using a similar transform as in **LP1**, we can rewrite the optimization problem (7)-(10) as

LP4 :

$$\max \quad \int_0^R \int_{d-D_r^{MAX}}^d \tau_r(s,d) w(d,s)\,\mathrm{d}s\,\mathrm{d}d \tag{30}$$

$$\text{s.t.} \quad \int_{(s,d)\in Q(x)} \tau_r(s,d)\,\mathrm{d}d\,\mathrm{d}s \leq 1, \tag{31}$$

$$(\forall x \leq R) \int_{x-D_r}^x C_r(x-s)\tau_r(s,x)\,\mathrm{d}s - \tag{32}$$

$$-\int_x^{x+D_r} C_r(d-x)\tau_r(x,d)\,\mathrm{d}d \leq \rho^\star p(x), \tag{33}$$

$$w(d,s) = C_r(d-s)(t(d)-t(s)), \tag{34}$$
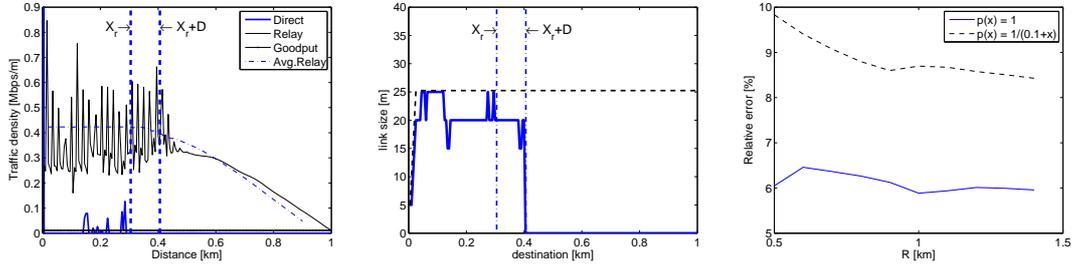
$$\tau_r(s,d) \geq 0. \tag{35}$$

Figure 2: Left: the optimal traffic distribution for variable link lengths. Relay traffic represents the total relay traffic arriving at node d. The rest is as in Figure 1. Middle: the optimal relay link lengths as a function of destination node position. The dashed line represent the optimal routing according to our heuristic. The vertical lines denote X_r and $X_r + D$. Right: relative error of $\rho^\star$ achieved with fixed routing as compared to the optimal routing for different traffic density $p(x)$.

Due to the complex constraints, it is not easy to guess what the solution of this problem is. However, we can see that for the weight associated to link (s, d) in objective function (30) is $w(d, s) = C_r(d - s)(t(d) - t(s))$, and we shall "prefer" links with higher weight.

Let $s(d) = \arg\max_s w(d, s)$, be the relay node with the highest weight with respect to node d. When $d \gg (d - s(d))$, we have $w(d, s) \approx C_r(d - s)t'(d)(d - s)$, and we have $d - s(d) \approx D_r$. However, even when d is of the same order as $d - s(d)$, we verify numerically that $w(d, s(d)) \approx w(d, d - D_r)$.

One can interpret the weight $w(d, s) - t(s)$ as a ratio of time $t(d) - t(s)$ gained on transmitting one bit using the direct link to s instead of transmitting it to d, over the time $1/C_r(d - s)$ needed to relay one bit from s to d. Furthermore, our approximation says that one needs to maximize $(d - s)C_r(d - s)$ which is the rate times the distance. As already mentioned, this is exactly the transport capacity defined in [13], although in [13] it occurs in a different framework (here, it is a result of a performance ratio between the two physical layers).

Finally, we verify our heuristic numerically. We solve problem **LP4** using linear programming and we compare the optimal routing with our routing heuristic. The results are illustrated in Figure 2.

In Figure 2, middle, we see that for $d < X_r$ the optimal routing corresponds well to our heuristic. For $X_r < d < X_r + D$ the optimal link lengths become shorter. This is because the clique $Q(X_r)$ is the last saturated clique, as explained in Proposition 1. Hence for every $s \in Q(X_r)$, it is sufficient to relay data to some node which does not belong to any saturated clique, that is any node $d > X_r + D$. Therefore, link lengths for these nodes tend to be smaller than D_r. Finally, for $d > X_r + D$, relay PHY is not saturated any more hence many routing strategies are possible (including fixed link lengths D_r).

We next show that despite of these discrepancies, fixed-length routing with the optimal D_r is has a comparable performance to the optimal routing. This is illustrated in Figure 2,
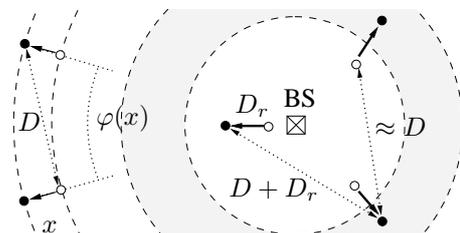
Figure 3: An illustration of the schedule: the white circle denotes nodes $x < (D + D_r)\sqrt{3}$ and the shaded circle denotes nodes $(D + D_r)\sqrt{3} \leq x < D + D_r$.

right, where we compare the achieved traffic density ρ of the optimal routing (found by solving **LP4**) and the routing with fixed link lengths D_r for different cell radii R. We see that the error is less than 10%.

Furthermore, we verified numerically that the same results hold for different traffic density functions $p(x)$. We verify them for typical parameters for WLAN and WPAN physical layers (numerical details are given in Section 5).

Finally, the constraints in linear program **LP4** are formed using cliques which represent an upper bound on the actual performance. On a contrary, the performance of fixed-length routing is exact, as explained in Section 3.1.

3.3   2D Networks

Finally, we consider the case where the cell is a disk of radius R. Again, we first restrict the analysis to the case of fixed relay link sizes, and we discuss variable link length case at the end of the section.

Even with the assumption of fixed link length, deriving the cell capacity is extremely difficult (for example, it proves difficult even to identify cliques). We simplify the problem by the following approximation: we assume only links whose link destinations are on circles of radii $x + k(D + D_r)$, $k \in \mathbb{N}$ may be active at the same time. We next count the maximum number $n_c(x)$ of links that can be simultaneously activated on the circle of radius x. The idea behind the approximation is to map each circle to a node in the 1D case, and to calculate the capacity using the results from Section 3.1.

When x is large enough $n_c(x)$ can be well-approximated by $\lfloor 2\pi/\varphi(x) \rfloor$ where the angle $\varphi(x)$ is characterized by

$$D^2 = x^2 + (x - D_r)^2 - 2x(x - D_r)\cos\varphi(x).$$

We can show that this approximation is tight when $x > (D + D_r)/\sqrt{3}$ (for $x = (D + D_r)/\sqrt{3}$, using the approximation we can have 3 simultaneous relay links with receivers at

distance x from the BS). Now when $x < (D + D_r)/\sqrt{3}$, one can easily prove that if there is one active relay link with receiver at distance x from the BS, one may add two relay links with receivers at distance $y(x)$ from the BS, where

$$y(x) = \sqrt{(D + D_r)^2 + x^2 - \sqrt{3}x(D + D_r)}.$$

All this is illustrated in Figure 3.

It is then reasonable to consider the following approximation for $n_c(x)$:

$$n_x(x) = \begin{cases} 1, & \text{if } x < (D + D_r)/\sqrt{3}, \\ \lfloor 2\pi/\varphi(x) \rfloor + 2, & \text{if } (D + D_r)/\sqrt{3} \leq x < (D + D_r), \\ \lfloor 2\pi/\varphi(x) \rfloor, & \text{if } x \geq (D + D_r). \end{cases}$$

Note that in order to make things tractable we make the additional approximation that the two additional occurrences of circle $y(x)$ are associated with $\tau_r(y(x))$ and not with $\tau_r(x)$.

Now assume that τ is defined so that $\tau(x)dx$ may be interpreted as the proportion of time the BS is serving all users on the ring between distances x and $x + dx$. Similarly, define τ_r so that $\tau_r(x)dx$ represents the proportion of time users located on the ring between distances x and $x + dx$ from the BS simultaneously receive relay traffic. Finally define the traffic distribution p so that $p(x)dx$ is the proportion of traffic generated between distance x and $x + dx$ (note that it does not imply that the traffic distribution is circular symmetric). Then the cell capacity is the maximum traffic ρ such that there exist $(\tau(x), x \in [0, R])$ and $(\tau_r(x), x \in [0, R])$ such that, $\tau(x) \geq 0$, $\tau_r(x) \geq 0$, $\int_0^R \tau(x)\,dx \leq 1$, and for all $x \in [0, R]$ $\int_x^{\min(R, (x+D+D_r))} \tau_r(u)\,du \leq 1$, and $\rho p(x) < \tau(x)C_d(x) + C_r(\tau_r(x)n_c(x) - \tau_r(x + D_r)n_c(x + D_r)1_{\{x + D_r \leq R\}})$.

The MaxRelay scheme   As in one-dimensional cells, we define by $\tau_r'(x)$ as the proportion of time $n_c(x)$ users at distance x should receive relay traffic so as to handle all the traffic to users located at distance $x + iD_r$ from the BS, $i \geq 0$, using relays only,

$$\tau_r'(x) = \frac{\rho^\star}{C_r n_c(x)} \sum_{i \geq 0} p(x + iD_r)1_{\{x + iD_r \leq R\}}.$$

Further define X_r as in (14). The MaxRelay scheme is now defined by (15) and:

$$\tau^\star(x) = \begin{cases} 0, & \text{if } x > X_r, \\ t(x)(\rho^\star p(x) + C_r(n_c(x + D + D_r)\tau_r^\star(x + D + D_r) & (36) \\ \qquad -n_c(x)\tau_r^\star(x))), & \text{if } x \leq X_r. \end{cases}$$

As in the one-dimensional case, the MaxRelay scheme is provably capacity optimal when the function $t(x) - t(x - D_r)$ is increasing. This scheme also provides a tight approximation of the capacity when the latter assumption is violated. The proof of the following result is similar to that of Theorem 1. The result is illustrated in the left of Figure 4.

Theorem 2 *When the spatial traffic distribution p is fixed, and under Assumption 1, the MaxRelay scheme achieves maximum stability.*

Finally, we discuss the variable link length case. Since our problem has circular a symmetric structure, we can assume that all nodes on a circle will use links of identical length. Again, we can construct a similar mapping as in the previous case to map the 2D case to the 1D case. Repeating the same type of analysis as in Section 3.2, we can verify that the choice of the fixed link length maximizing the transport capacity is approximately optimal in this case as well.

This result also provides an intuitive justification why we can assume D independent of x. Since a region between $[0, X_r]$ has a fully saturated relay traffic, it is likely to expect that the same D and D_r will be optimal throughout this saturated region. Formal verification of this assumption is left for future work.

4   Towards Protocol Design

The goal of a scheduler in our system is to define when the BS should transmit and to which node, and also when nodes should receive relay traffic, and from which nodes.

As discussed in Section 2.3, an optimal scheduler maximizes some utility function subject to constraints (3)-(5)-(6). This is a convex optimization problem but with high complexity since the number of possible links (relay source-destination pairs) is high.

In order to reduce the complexity, we divide the problem into a routing and a scheduling subproblem. The routing subproblem chooses what relay links shall be used, and the scheduling subproblem solves the above optimization problem constrained on previously selected routes.

We propose a simple routing algorithm, described in Figure 5, based on the results of Section 3.2. This is a modified Dijkstra algorithm where $\mathcal{N}_c$ is the set of already connected nodes, $\mathcal{N}_n = \mathcal{N} \setminus \mathcal{N}_c$ of nodes to be connected and $\mathcal{L}$ is the set of chosen links. The next node d we connect is the one that minimizes function $C_r(d-s)(d-s)$ to any of the already connected nodes $s \in \mathcal{N}_c$. The candidate source is labeled $s(d)$. If $d - s(d) > D_r^{MAX}$, where D_r^{MAX} is some maximum allowed relay link size, it means that node d does not receive relay traffic. Nevertheless, it is put in $\mathcal{N}_c$ as other nodes may connect to it. The routes are finally defined by $\mathcal{L}$ upon the completion of the above algorithm.

Once the routing algorithm provides us with the set of optimal links $\mathcal{L}$, we solve the scheduling subproblem restricted on links $\mathcal{L}$ (or in other words setting $\tau_r(s, d) = 0$ for all $(s, d) \notin \mathcal{L}$), using classical centralized convex optimization algorithms. The number of variables is now $O(N)$ instead of $O(N^2)$ we had without constraints on routing. An illustration of the result on a sample topology is depicted on Figure 4.
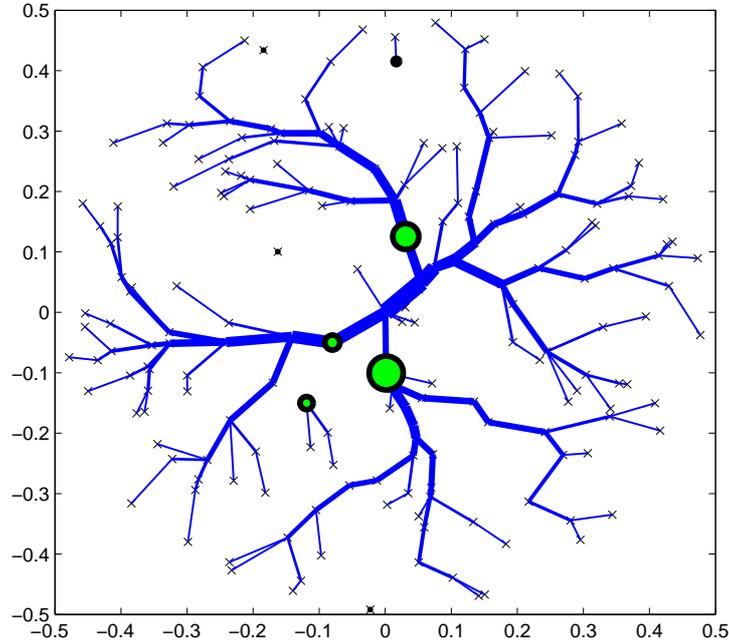
Figure 4: The optimal scheduling and relay scheme for an example of a random topology. We consider 2D cell of radius of 0.5km. We assume $p(x) = 1$ everywhere. The lines represent the relay traffic (the bolder the line is, the higher is the traffic). The circles represent the direct traffic, and the radii correspond to the intensities. The traffic to the disconnected subtrees has to be supplied directly. Also, some of the direct traffic is needed to the areas where the relay interface is congested.

The proposed routing heuristic is simple to implement, for example reusing the routing protocol presented in [2]. It is sufficient to estimate the received signal power in order to have an estimate of transport capacity $C_r(d - s)(d - s)$. On the contrary, the scheduling subproblem is difficult to implement because, in order to formulate the convex optimization, one needs to identify the cliques which is not trivial in a dynamic wireless environment. One possible direction might be to use the scheduling proposed in [6], and to add direct links where needed, to saturate the relay channel. This issue remains as a future work.

$$\mathcal{N}_c = \text{BS}, \mathcal{N}_n = \mathcal{N} \setminus \mathcal{N}_c, \mathcal{L} = \emptyset,$$
$$\text{for } \mathcal{N}_n \neq \emptyset$$
$$\quad s(d) = \arg\max_{s \in \mathcal{N}_c} C_r(d-s)(d-s),$$
$$\quad d = \arg\max_{d \in \mathcal{N}_d} C_r(d-s(d))(d-s(d)),$$
$$\quad \text{if}(d-s(d)) < D_r^{MAX}$$
$$\quad\quad \mathcal{L} = \mathcal{L} \cup \{(s(d), d)\}, \tag{37}$$
$$\quad \text{end}$$
$$\quad \mathcal{N}_c = \mathcal{N}_c \cup \{d\}, \mathcal{N}_n = \mathcal{N}_n \setminus \{d\}$$
$$\text{end}$$

Figure 5: Relay routing algorithm
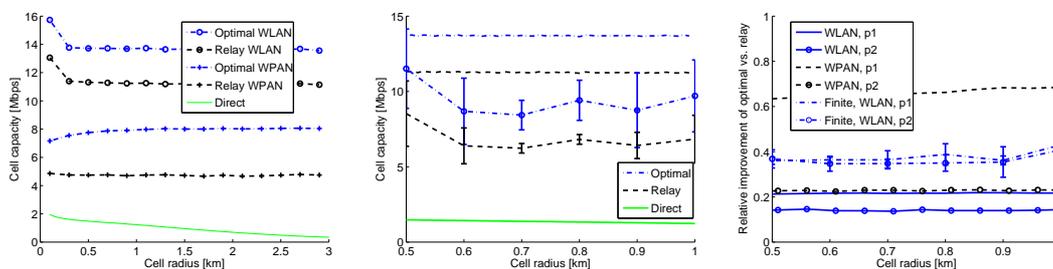
Figure 6: Left: capacities of the optimal, relay and direct policies for the case of WLAN and WPAN relay interfaces. Middle: the maximum throughput random networks with node density of 250 nodes/km^2 compared to the capacity (for the direct policy the two coincides). Right: relative improvements of the optimal routing over the relay routing for WLAN and WPAN interfaces, for both the capacity and the simulated maximum throughput ($p_1(x) = 1, p_2(x) = const \times 1/(0.1 + x)$).

5 Numerical Results

In this section we evaluate the capacity of a single cell network with relays, using the optimal policy derived in the previous sections, and we compare its performance with the direct and relay policies (defined in Section 2.3.3).

We consider two cases of relay networks. One is WLAN relay and we take typical 802.11 parameters (transmission power 100mW, maximum rate 54Mbps). The other one is WPAN relay and we take next generation 804.15.4a parameters (transmission power 1mW, maximum rate 27Mbps). We assume the BS transmits at 20W and its maximal rate is 10Mbps.

We first look at the capacity. In Figure 6, left, we see that the capacity with the direct policy decreases exponentially (as explained in [11]), whereas it stays constant with the

optimal and relay policies. In Figure 6, left, we see that with the optimal policy, we achieve from 20% (in the case of WLAN relay) up to 60% (in the case of WPAN relay) capacity improvement over the relay policy.

WLAN relays offer high throughput hence it can carry almost all traffic itself, and the improvement of the optimal over relay is smaller. WPAN offers lower throughput hence the improvement is higher. Note that in the case of WLAN relays the optimal link length is quite small ($\approx$ 30m). If we choose to use longer links, for example to limit the number of hops, the improvement of the optimal policy over the relay one will be higher.

Next, we consider networks with a finite number of nodes. We use the algorithm described in Section 4 to solve the optimization problem (7)-(10). Note that since this problem is formulated with constraints on cliques, it will give us an approximation of the optimal solution, as discussed in Section 2.3.

The corresponding maximum cell throughputs are compared with the capacities in Figure 6, middle. The maximum cell throughput will be smaller than the capacity since the finite number of nodes implies suboptimal routing (there might be no relay routes to some nodes, and not all links have the optimal lengths). Nevertheless, we see that even though the maximum throughputs are smaller than the capacities, they are relatively close. Hence, our capacity analysis can be used as a performance metric when dimensioning cells with relays. The throughput of the direct policy fits perfectly with the prediction as it does not have routing problems.

In Figure 6, right, we see the relative improvement of the maximum throughput in the case of the optimal and the relay policies. We see that it is larger than the gain obtained when comparing the capacities. This is again due to routing constraints; imperfect routes will additionally limit the performance of the relay layer.

Finally, all the results in this section hold when the node density is large. If it is small, most nodes are disconnected and there is no possibility to relay. This is the reason why we simulated only WLAN interface. Link lengths in the case of WPAN are very short thus require very large node density in order to achieve connectivity. This in turns makes simulations computationally too expensive, even with the heuristics proposed in Section 4.

6   Conclusions and Future Work

In this paper we have derived an approximately optimal relay (routing and scheduling) policy that maximizes the cell capacity. We have shown that it is approximately optimal for relays to use links that maximize the transport capacity. We have also shown that in many cases, a node should receive traffic both from the base station and from a relay, unlike in relay policies proposed by other authors.

We have presented a simple algorithm for calculating the cell capacity. This capacity can be useful for dimensioning purposes. Using this algorithm, we have shown that the

cell capacity with relays stays constant with the cell size, as opposed to the capacity of a cell without relays that rapidly decreases with the cell size. We have also shown that our optimal strategy outperforms other strategies that use direct links only to the nearest node, as proposed in the literature.

We have given guidelines for relay protocol design. The actual protocol implementation remains as a future work. We also plan to consider the impact of a bound on a maximum number of relay hops (e.g. due to delay constraints) and possible inefficiencies of a real schedule on the cell capacity.

References

[1] H. Wu, C. Qiao, S. De, and O. Tonguz, "Integrated cellular and ad hoc relaying systems: iCAR," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 2105–2115, 2001.

[2] H. Luo, R. Ramjee, P. Sinha, L. Li, and S. Lu, "UCAN: a unified cellular and ad-hoc network architecture," in *Proceedings of MobiCom '03*, 2003, pp. 353–367.

[3] S. Mengesha, H. Karl, and A. Wolisz, "Capacity increase of multi-hop cellular wlans exploiting data rate adaptation and frequency recycling," in *Proceedings of MedHoc-Net '04*, 2004.

[4] H.-Y. Wei and R. Gitlin, "Two-hop-relay architecture for next-generation wwan/wlan integration," *IEEE Wireless Communications*, April 2004.

[5] J. Cho and Z. Haas, "On the throughput enhancement of the downstream channel in cellular radio networks through multihop relaying," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 7, pp. 1206–1219, 2004.

[6] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE trans. on Wireless Communications*, vol. 4, no. 5, September 2005.

[7] B. Radunovic and J.-Y. Le Boudec, "Joint scheduling, power control and routing in symmetric, one-dimensional, multi-hop wireless networks," in *Proc. WiOpt*, 2003.

[8] F. Baccelli, N. Bambos, and C. Chan, "Optimal power, throughput and routing for wireless link arrays," in *Proc. INFOCOM*, Barcelona, Spain, April 2006.

[9] B. Hajek and G. Sasaki, "Link scheduling in polynomial time," *IEEE trans. on Information Theory*, vol. 34, no. 5, 1988.

[10] T. Nandagopal, T. Kim, X. Gao, and V. Bharghavan, "Achieving MAC layer fairness in wireless packet networks," in *Mobile Computing and Networking*, 2000, pp. 87–98.

[11] T. Bonald and A. Proutiere, "Wireless downlink data channels: User performance and cell dimensioning," in *Proc. ACM Mobicom*, San Diego, USA, August 2003.

[12] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *Journal of the Operat. Res. Society*, vol. 49, 1998.

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