Relay Routing and Scheduling for Capacity Improvement in Cellular WLANs

Seble Mengesha, Holger Karl

Relay Routing and Scheduling for Capacity Improvement in Cellular WLANs

Seble Mengesha, Holger Karl*
Fachgebiet Telekommunikationsnetze
Technische Universität Berlin, Sekr. FT 5-2, Einsteinufer 25, 10587 Berlin
mengesha|karl@ee.tu-berlin.de

Abstract

In classical infrastructure-based wireless systems, all mobile terminals communicate directly with the access point. In such systems, the transmission capacity, i.e., the amount of traffic successfully communicated to/from the access point, is important. This paper studies the possibility to increase this capacity by using mobile terminals as intermediate relays. These relays reduce the communication distance and hence mobile terminals can send their traffic at a maximum data rate while maintaining a target packet error rate and transmission power limit.

We present algorithms that solve the relaying problem combining routing, scheduling and adaptive modulation. A crucial aspect is the fair sharing of resources in both direct and relaying systems; two different modes of fairness are identified. For one fairness mode, relaying can improve capacity by up to 30%. In addition, relaying allows to add additional resources, namely frequencies, to a cell, resulting in almost a doubling of capacity.

Keywords: Relaying, capacity, fairness, WLAN, HiperLAN/2

.

1 Introduction

In wireless communication systems, two communicating terminals can be too far apart to reasonably allow direct communication at high data rates. Using intermediate terminals to relay their traffic is the solution used in "ad-hoc" networks. However, relaying can also be beneficial in infrastructure-based systems where many wireless terminals want to communicate only with an access point; typical examples for such scenarios are "hot spots".

Relaying benefits would be due to the reduction of distances over which communication takes place: No longer do all terminals have to communicate with a potentially far-away access point (AP), but could rather communicate with an intermediate relay that is perhaps halfway towards the access point. This reduced distance allows the mobile terminals to reduce the power they use for transmission. This power reduction can improve energy efficiency or electromagnetic immissions; it could also be used to increase the overall transmission capacity of such a system (defined as the amount of data successfully sent/received by the access point per unit time).

Capacity can be improved because smaller radiated power also means smaller interference between or within cells [7]. Additionally, for a single cell, shorter distances also allow faster modulations, resulting in more transmitted data per unit time; this paper concentrates on this second effect. In addition, relaying could also open the way to add additional resources, i.e., frequencies, to a cell.

The contribution of this paper is to show how ad-hoc relaying via mobile terminals can be leveraged to increase transmission capacity. This goal is achieved by deciding which terminal is relaying for which other terminal (a routing problem) and how these communication relationships are organized in time (a scheduling problem). In addition, jointly optimizing transmission power/modulation selection and fairly sharing system resources among terminals when doing relaying are crucial issues. This problem is studied over a wide parameter range; one pivotal result is that relaying improvements are the largest in highly loaded cells, as is a typical scenario for hot spots.

The remainder of this paper is organized as follows. Section 2 describes the model assumptions. Section 3 and 4 describe our optimization algorithms and their performance results, which are contrasted with related work in Section 5. Finally, Section 6 concludes the paper.

2 System model

As underlying technology, we use HiperLAN/2 [8] as a case study as it allows an easy control of relaying relationships and provides a number of modulations to choose from; additionally, it is quite amenable to a relaying extension [5]. In HiperLAN/2, the access point is responsible for computing a communication schedule for a MAC frame 2 ms long. In a frame, each mobile terminal is assigned time slots in which it is allowed to send or receive data to or from the access point or other mobile terminals. This schedule stipulates the transmission power a mobile uses and which one out of seven different standardized modula-

*Part of this work has been supported by the Germany Ministry for Education and Research (BMBF) under the HyperNET/IBMS2 project.

tion types is used within a time slot. Each slot is used to transmit a data packet of constant size; the length of such a slot hence depends on the chosen modulation.

We use this system model to develop routing/scheduling algorithms for relaying that are based on estimates of the channel gains between terminals and access point; these estimates are in fact provided by HiperLAN/2's so-called "radio map." For simplicity, we assume that these estimates are known at the access point at no cost. As we do not yet consider mobility, disregarding estimation errors is an acceptable simplification, but their impact on the algorithm performance is an important issue for future work. For the simulation-based performance evaluation, we identify the channel gains with the relative positions of the terminals. As we do not consider any obstacles, this is also an acceptable simplification.

Based on the distance/the channel gain, we can compute the signal-to-noise ratio at a receiver: The noise is constant; there is no interference as HiperLAN/2 is strictly TDMA-based and we only consider a single cell in this paper; the transmission power can be obtained and a simple pathloss model yields the received signal strength. Given the signal-to-noise ratio and the chosen modulation, the packet error rate (PER) for a data packet can be computed. Details on the approximation of this error function can be found in [7, 8].

3 Optimizing capacity

Relaying can be applied in a number of different modes, depending on the way of sharing resources and on the frequencies used within a cell. Based on these modes, a routing and scheduling problem has to be solved.

3.1 System modes

In a TDMA-based cellular wireless system like HiperLAN/2, mobile terminals need to send their traffic to the desired destination in a given time slot. When attempting to maximize the capacity of such a cell, the best way is to assign all time slots to the terminals that are closest to the AP as they will realize the highest throughput. Evidently, this is not intended by optimization and, hence, some fairness characteristics have to be demanded.

How to fairly divide the system resources (i.e, the time during which the AP can receive) between terminals—even in the direct case—gives raise to at least two different ways of contemplating the system: Fairness among terminals can be maintained by scheduling the communication such that either all terminals obtain an equal share of the total frame time (the "uniform slot size" scheme) or all terminals are allowed to send a uniform amount of data in slots of varying length depending on their modulation and, ultimately, their distance from the access point (the "uniform traffic size" scheme).

The first scheme will result in a higher total goodput as terminals far from the AP are given a relatively smaller weight compared to near terminals that can use fast modulation. In the second scheme, near terminals have to sacrifice some of their time to allow the far terminals to compensate for their low modulations, but it corresponds better to user expectations.

These two fairness schemes also extend to the relaying case: either all terminals are assigned the same amount of time, or transmissions are arranged such that all terminals receive the same effective goodput, no matter whether they are relayed or not. We investigated both these fairness schemes here to see how well they can be supported by relaying.

3.2 Adding resources

Relaying also opens the possibility to add additional resources to a cell. Let us consider a simple cellular network with three terminals A, B, and C as shown in Figure 1; assume that B is a potential relayer for A and C is communicating only directly with the AP. Basically, all these transmissions have to take place sequentially. However, C's sending to the AP and A's sending to B involve different entities and could—in principle—be scheduled concurrently. Doing so on the same frequency is possible and beneficial in large cells, but to avoid interference within a cell, another frequency could be used for the relaying step. Such two-frequency relaying would require at least one entity, e.g. the relay terminal, to switch between different frequencies, preferably within a single MAC frame. This is indeed technically feasible as modern hardware has sufficiently fast switching times.

Adding a second frequency for relaying could be seen as unfair: why not do so in the direct case as well and increase the bandwidth in the direct case correspondingly? First, this would require changes in the hardware of existing terminals: doubling the bandwidth can not easily be done with existing, commodity radio equipment. Of course, a new system could be constructed that does use twice the original's system bandwidth for direct communication. But that would limit the amount of spatial reuse as now a cell *always* uses a larger share of the total bandwidth available to the system. In this sense, adding a second frequency to a cell by relaying allows to do so only when it is needed and beneficial, using existing terminals—essentially, a legacy/cost argument. Therefore, being able to add a frequency via relaying gives a level of flexibility in assigning system resources that is not possible by simply increasing the bandwidth that is assigned to a system; we investigate in this paper whether this increased flexibility does pay off at all.

But, one might ask, when adding a frequency by relaying, is then the spatial reuse pattern not completely turned upside down as this second frequency would have to be borrowed from a neighboring cell and creates interference in these cells (destroying the improved signal-to-noise ratio which was gained by reducing the distance)? Not necessarily: it is in principle possible to use this second frequency only close to the center at a cell for the communication

Figure 1: A simple cellular network scenario

between access point and relaying terminals. The relayed terminals, which are placed far away from the cell's center and responsible for most of the cell's interference with other cells, still use the cell's primary frequency. Hence, the second frequency would likely not be "noticed" by other cells at all, not disturbing the frequency reuse patterns. However, as the present paper only evaluates a single cell, we cannot currently prove this claim, but it is subject of ongoing work.

To sum up: adding a second frequency only in the relaying case might not be a fair comparison with the direct case. But adding frequencies by relaying is practical, whereas in the direct case, frequencies can not be added easily—RF hardware that can adapt its bandwidth consumption is expensive and not usually available, no current communication system is able to adapt the bandwidth that is used over a given link. In this sense, (two frequency) relaying is a "poor man's solution" that improves capacity with minimal modifications to an existing system.

Figure 2 summarizes the different fairness schemes for direct, one- and two-frequency relaying cases. The example is too small to show the effect of keeping the second frequency close to the access point, nonetheless the overall picture should be clear.

What is missing are now routing algorithms that decide when a terminal should rely its traffic via another terminal and scheduling algorithms that determine the time share that is assigned to a particular transmission.

3.3 Routing

The potential benefit of relaying depends on the data rates that can be realized between relayed terminal, relaying terminal, and access point. Hence, a joint optimization of routing and scheduling is necessary that decides which terminal to use for relaying and that selects modulation and transmission power. Currently, the algorithm optimizes only the uplink case and considers a single intermediate relay.

The effective data rate between any two terminals can be determined based on their distance and a target packet error rate which allows to compute, for each modulation, the required transmission power: Based on results in [8], the relationship between signal-to-noise and packet error rate is known (although no closed-form expressions exist, only approximations). Solving for the required signal-to-noise ratio gives an input for the pathloss model, which can in turn be solved for the required transmission power.

Any modulation that requires more than a maximum allowable power (here, 200 mW based on European legal restrictions) or that does not match minimal required receiver sensitivity is ruled out. As a result, the optimal modulation for this pair of terminals is obtained. The transmission power is then adjusted to the smallest value that still meets the target PER for this modulation. An example of such an effective-data-rate selection is shown in Figure 3 for 1% PER.

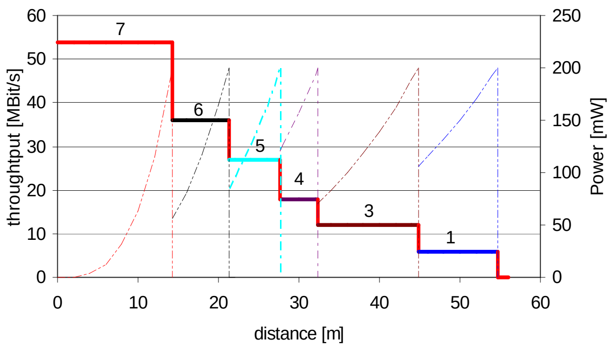

Figure 3: Effective data rate and transmission power with respect to distance for 1 % PER and $\alpha = 3.2$

The shape of this Figure 3 is the primary justification for hoping that relaying will actually improve capacity: *For many distances, the achievable data rate over half the distance is **more than twice** the data rate over the full distance.* For example, over 50 meters, only about 9 MBit/s are achievable, but over 25 meters, about 27 MBit/s can be realized: Assume that a terminal 50 m away from the AP transmits directly for an entire second, transferring 9 MBit. Using a relay terminal in the middle of AP and far terminal, the far terminal can transmit for half a second to the relay, transmitting 13.5 MBit, which are then forwarded in the next half second to the AP, resulting in a total of 13.5 MBit transmitted using relaying. Hence, the total throughput is increased even though the time slot for the direct communication has now to be split in two to accommodate two transmissions.

The routing then proceeds as follows: for a mobile terminal X, its data rate in the direct and relaying case is computed for all possible relaying terminals; the effective data rate for a relayed terminal is the minimum rate of both involved links (as in general, the relaying terminal will not be located precisely in the middle of the two other termi-

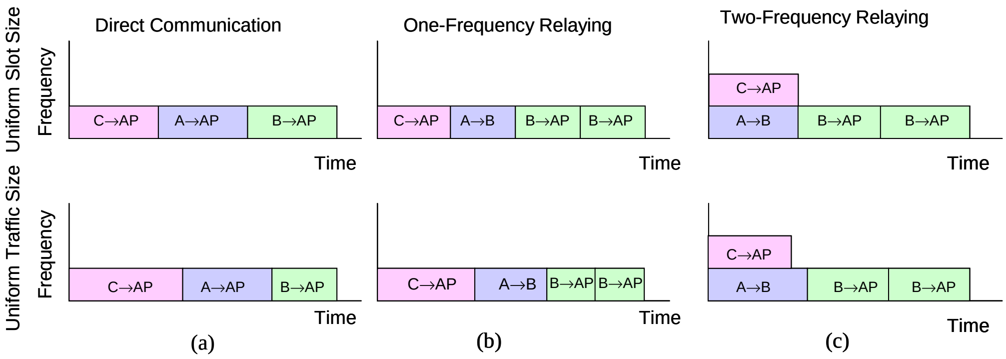

Figure 2: Time-freqency representation of the example scenario

nals). X selects that terminal as a potential relayer that results in the largest data rate for X and that exceeds the direct-case data rate; if no such relaying terminal exists, X will continue to use direct transmission to the AP. Relaying terminals always communicate directly with the AP.

3.4 Scheduling mechanism

For both uniform slot size and uniform traffic size fairness schemes, relaying schedules are computed on the basis of the routing decisions. Let us first consider the uniform slot size case. In direct communication, all terminals are assigned an equal share of transmission time. Using one-frequency relaying, the time slot of a relayed terminal is split in two: one sub slot is used for relayed terminal to relaying terminal communication, the other slot for relaying terminal to AP communication; other slots are unaffected by this operation. For each of these links, the maximum data rate is used.

For the uniform slot size mode, it would also make sense to require that the minimum data rate in the relaying case is at least twice that of the direct case. As only half the effective time is available for a relayed terminal to send out traffic, this would obviously ensure that the capacity of this communication is really increased. However, this is a rather strict and only rarely met condition. In simulative investigations, using the simulation model discussed in Section 4, it turned out that this way of redistributing slot sizes obviously never deteriorated the cell throughput, but also that there were only rarely any benefits. Section 4 will therefore concentrate on showing the ramifications of not making this requirement for the one-frequency, uniform-slot-size case — as it will turn out, this is not beneficial unless the path loss coefficient is very high. In future work, we will also provide more detailed numeric results when the requirement of doubling the data rate over a relaying connection is actually made.

When using two-frequency relaying for the uniform slot size fairness mode, communication from the relayed terminals to the relaying terminals can overlap with other communication. The total number of time slots depends on this overlap; and again, all slots are of uniform length.

In the case of uniform traffic size schedules, the optimal traffic size per slot depends on the modulations chosen in the routing phase. The schedule construction is in principle similar to the uniform slot size case (for both one- and two-frequency relaying), but here the slot length is individually varied for each terminal such that all terminals' goodput at the AP results in the same value, taking into account the need to relay traffic in two slots. Here, the same consideration whether or not to actually use relaying applies: the relay connection could be required to simply increase the direct case data rate or to actually double it. Again, we look at the case when the relay connection is only required to increase the data rate. As it will turn out, this is also not beneficial in general, however, it starts to be useful already for smaller path loss coefficients than in the uniform slots size schedules.

4 Result

For the model described in Section 2, the goodput achieved by capacity-oriented schedules for direct, one-frequency, and two-frequency relaying is evaluated by simulations. For any combination of parameters, the achieved capacity of a single cell is averaged over 55 different random placements of terminals on a square area of 50 m x 50 m, 70 m x 70 m, and 100 m x 100 m (access point in the middle). The path loss coefficient α is 3.2 for most of the simulations. Confidence intervals are not shown as they are too small to be discernible.

To compute modulation and transmission power for a given communication, maximum target PERs of 1%, 3%, and 5% are used. The resulting average total goodputs are shown, for fifteen entities and varying pathloss coefficient α, for both uniform slot size (Figure 4, 5, 8, and 10) and uniform traffic size schedules (Figure 6, 7, and 9). Not all parameter combinations are shown here but are presented in a technical report [10].

For small α, all approaches are capable of fully utilizing the access point's maximum goodput. As α increases, the range of communication over which the target PER condition can be met at a fixed modulation and limited transmission power decreases. Hence, the observed goodput at the access point also goes down. For terminals randomly placed on a square area of 70m x 70m, the target PER can be met even with only the slowest modulation up to $\alpha = 3.2$. But we also considered larger areas to see how often and how far the communication can proceed with the target PER. For such cases, whenever terminals are beyond the range of communication, they send data with the slowest modulation and adjust their PER up to an acceptable rate. But this happens rarely and it does not affect the assumption of uniform maximum target PER within the network.

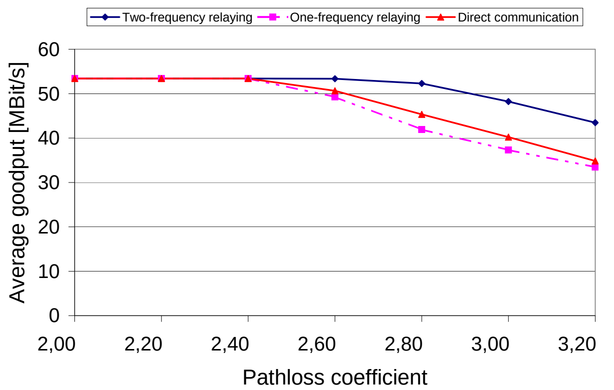

Figure 4: Average goodput as a function of α for uniform slot size fairness in a 50x50 area for 1% PER

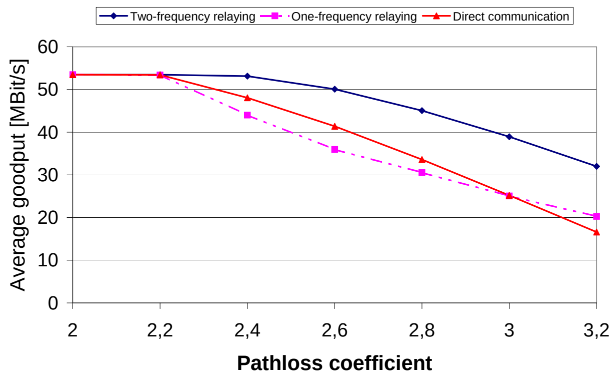

Figure 6: Average goodput as a function of α for uniform traffic size fairness in a 70x70 area for 1% PER

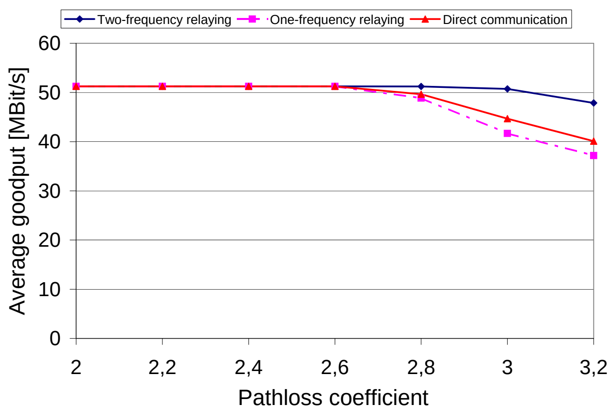

Figure 5: Average goodput as a function of α for uniform slot size fairness in a 50x50 area for 5% PER

For both scheduling types and for terminals randomly placed in smaller area (50m x 50m), the goodput achieved by one-frequency relaying is smaller than that of direct communication as the terminals can have a data rate for direct cases comparable to the relaying case and hence they do not select a relaying terminal. As terminals are placed in a wider area and as the target PER gets smaller, the goodput achieved by one-frequency relaying becomes better than that of direct communication due to the fact that smaller target PERs reduce the range of communication which then forces the terminals to send their data via relaying rather than directly to the AP for a better goodput. In general, we observed that the overall goodput achieved for a smaller target PER is less than that of relatively larger target PER.

For all fairness schemes, two-frequency relaying considerably improves the total performance. In extreme examples, it almost doubles the direct case's goodput. Also as the number of terminals increase, the goodput also increases. This dependency on the number of terminals is essentially a stochastic effect: the far terminals have a higher probability of finding a relay terminal, more often enabling faster modulations. Hence, relaying does generate more capacity when it is most sourly needed.

The performance results for one-frequency relaying also show that the local heuristic used to determine relaying terminals needs to be extended for the uniform slot size

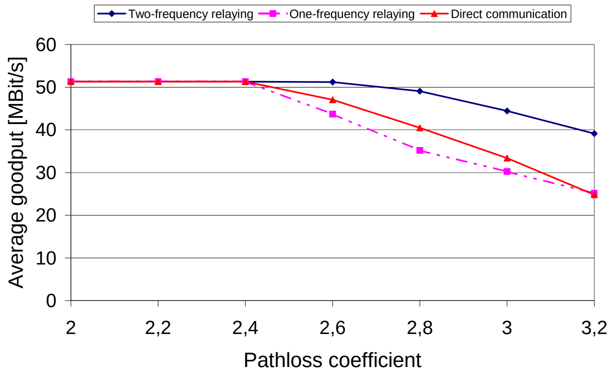

Figure 7: Average goodput as a function of α for uniform traffic size fairness in a 70x70 area for 5% PER

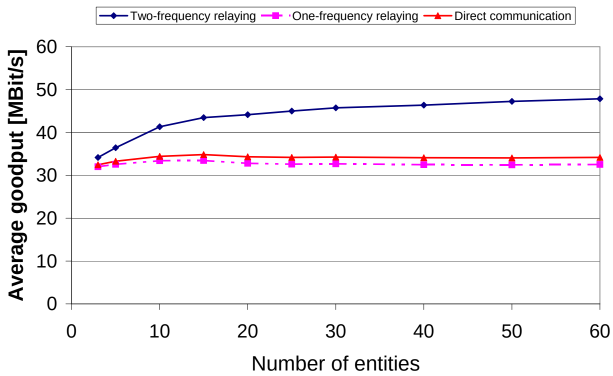

Figure 8: Average goodput as a function of entities for uniform slot size in a 50x50 area for 1% PER and $\alpha = 3.2$

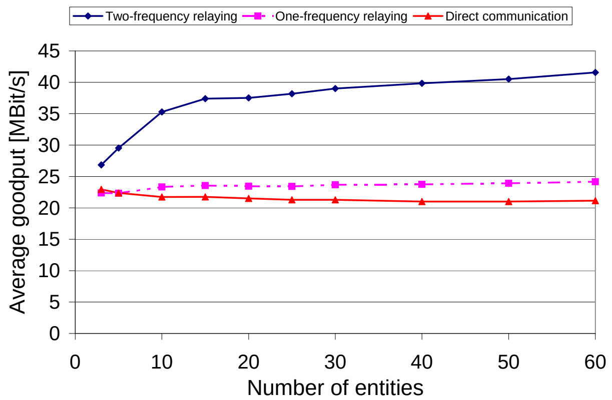

Figure 9: Average goodput as a function of entities for uniform traffic size in a 70x70 area for 3% PER and $\alpha = 3.2$

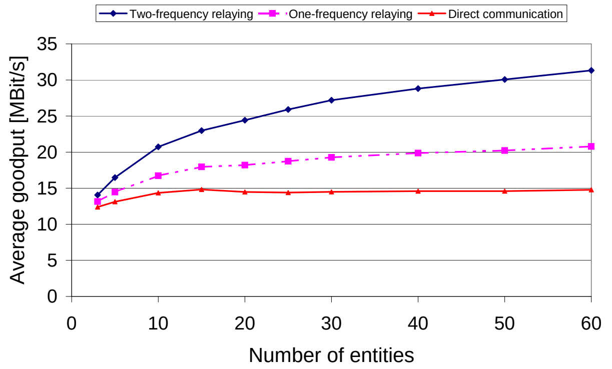

Figure 10: Average goodput as a function of entities for uniform slot size in 100x100 area for 1% PER and $\alpha = 3.2$

scheduling case in order to also result in a *global* improvement of goodput. This is a challenging issue for future work.

5 Related work

The notion of using relaying to improve capacity in wireless networks has been discussed in several contexts. GUPTA and KUMAR [4] have studied the capacity of randomly located ad-hoc networks. Their result shows that for n identical, randomly located nodes, the maximum achievable throughput under optimal circumstances is $\frac{1}{\sqrt{n}}$ times the transmission rate and as the number of nodes n per unit area increases, the throughput decreases accordingly. Using the same physical model of wireless network but applying arbitrarily complex network coding such as multi-access and broadcast codes as opposed to point-to-point coding, GASPER and VERTTERLI [2] showed that the capacity of wireless network with n nodes under a relay traffic pattern behaves like $\log n$ bits per second. GROSSGLAUSER and D.TSE [3] also investigated a similar scenario but introduced mobility and showed how the capacity of this wireless ad-hoc network can be increased using a single mobile relay node. However, the increase of the capacity is at the cost of considerable delay.

Relaying in CDMA networks is studied in [1] with results in principle comparable to ours. RAPP [11] used the HiperLAN/2 system with co-channel interference to show how its throughput and QoS can be increased. He exploited the silent periods in a transmission frame by placing them asymmetrically so as to improve the interference situation in certain areas of the frame for both up- and downlink in co-channel radio cells. Though he considered an environment similar to ours, the approach for improving the throughput of the network is entirely different.

Relaying in cellular and Ad-hoc systems has also been studied by many authors. For e.g., relaying as a means to enhance coverage in cellular radio systems is investigated in [12]. The paper describes methods of selecting a relay node and proposes reusing of channels from the adjacent cells for relaying purpose. Power control is also used to minimize the co-channel interference. In [9] a new wireless network model called the multi-hop cellular network (MCN) is proposed. The model involves mobile devices farther away from the base station communicating with the base station using a multi-hop path consisting of other mobile devices in the cell. The authors focus mainly on system throughput of both MCN and SCN (the conventional singl-hop cellular network) and showed how the throughput of MCN is superior to that of SCN due to the possibility of simultaneous transmissions of multiple packets with in a cell. But since these multiple packet transmissions may consume considerable bandwidth, the throughput improvement is limited.WU et al. [13] studied the performance of iCAR, a new wireless architecture based on the integration of cellular and modern Ad-hoc relaying technology. In their paper, they showed how Ad-hoc relay stations can relay traffic from one cell to another dynamically and lower system-wide call blocking probability.

Compared to these relaying related papers, our work is unique in that it combines relaying with adaptive modulation selection and transmission power control to improve the cellular wireless system under consideration.

6 Conclusions and outlook

Relaying is a viable means to improve the operations of an infrastructure-based wireless communication system as it considerably increases the cell capacity. We described different algorithms which improve the system capacity and maintain fairness among terminals. In particular when a fairness scheme where are terminals in a cell should obtain the same throughput, relaying is a viable means of improving the total capacity of a wireless cell; improvements by up to 30 % over the direct case are possible. Relaying also allows to borrow system resources, i.e., frequency bands, from neighboring cells, considerably increasing capacity in a way that is compatible with existing hardware but not otherwise possible without relaying.

Moreover, the algorithms are practical as they are not computationally intensive, they can be implemented as iterative online algorithms, and are based on information that can be provided by real systems with acceptable overhead. As no additional infrastructure is necessary and also the requirements on the individual terminals are quite modest, our relaying approach can provide a simple and cheap solution to add capacity to a wireless system, particularly in highly loaded networks.

Currently, we are studying the impact of mobility and channel fading and the ensuing errors in channel gain estimates. In addition, the results of this paper will be combined with previous research on the interference reduction due to relaying. The most important question is no doubt the understanding of two-frequency relaying in a multi-cell scenario: do the benefits still hold when interference to other cells is considered, does it work out to "hide" the borrowed frequency band in the interior of a cell?

We are also currently implementing (together with partners from the HyperNET/IBMS2 project [6]) relaying in a real HiperLAN/2 testbed and will study its performance.

References

[1] M. Bronzel, W. Rave, P. Herhold, and G. Fettweis. Interference Reduction in Single Hop Relay Networks. In *Proc. of 11th Virginia Tech Symposium on Wireless Personal Communications*, Blacksburg, VA, June 2001.

[2] Michael Gasper and Martin Vertterli. On the Capacity of Wireless Networks: The Relay Case. *IEEE*, 2002.

[3] M. Grossglauser and D.Tse. Mobility Increases the Capacity of Ad-hoc Wireless Networks. *IEEE/ACM Trans. on Networking*, 2002. to appear, `http://degas.eecs.berkeley.edu:80/~dtse/ton_mob_subm_3.ps`.

[4] P. Gupta and P. R. Kumar. The Capacity of Wireless Networks. *IEEE Trans. on Information Theory*, November 1998. `http://black.csl.uiuc.edu/~prkumar/ps_files/capacity_final.ps`.

[5] D. Hollos and H. Karl. A protocol extension to HiperLAN/2 to support single-relay networks. In *Proc. of 1st German Workshop on Ad-Hoc Networks*, pages 91–108, Ulm, Germany, March 2002. `http://www-tkn.ee.tu-berlin.de/publications/papers/frame2.pdf`.

[6] Webpage of the IBMS^2 project. http://www.ibms-2.de.

[7] H. Karl and S. Mengesha. Analyzing Capacity Improvements in Wireless Networks by Relaying. In *Proc. IEEE Intl. Conf. Wireless LANs and Home Networks*, pages 339–348, Singapore, December 2001.

[8] J. Khun-Jush, G. Malmgren, P. Schramm, and J. Torsner. HIPERLAN type 2 for broadband wireless communication. *Ericsson Review*, 2:108–119, 2000. `http://www.ericsson.com/review/2000_02/files/2000026.pdf`.

[9] Y.-D. Lin and Y.-C. Hsu. Multihop Cellular: A New Architecture for Wireless Communications. In *Proc. INFOCOM*, Tel-Aviv, Israel, March 2000. `http://www.ieee-infocom.org/2000/papers/381.ps`.

[10] Seble Mengesha and Holger Karl. Routing and scheduling for capacity improvement in WLANs. Technical Report TKN-02-012, Techniche Universität Berlin, Germany, August 2002. `http://www-tkn.ee.tu-berlin.de/publications/tknrreports.html`.

[11] J. Rapp. Increasing Throughput and QoS in a HIPERLAN/2 System with Co-Channel Interference. In *IEEE Intl. Conf. on Networking (ICN)*, pages 727–736 (Part I), Colmar, France, July 2001.

[12] V. Sreng, H. Yanikomeroglu, and D. Falconer. Coverage Enhancement through Two-hop Relaying in Cellular Radio Systems. In *Proc. IEEE Wireless Communications and Networking Conf.*, Orlando, FL, March 2002.

[13] H. Wu, C. Qiao, and O. Tonguz. Performance Analysis Of iCAR (Integrated Cellular and Ad hoc Relaying System). In *Proc. ICC*, volume 2, pages 450–455, 2001. `http://www.ieeexplore.ieee.org/iel5/7452/20262/00936981.pdf`.