Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

HANS-FLORIAN GEERDES, HOLGER KARL[1]

The Potential of Relaying in Cellular Networks

[1] Technische Universität Berlin,Telecommuication Networks Department

ZIB-Report 02-40 (October 2002)

The Potential of Relaying in Cellular Networks

Hans-Florian Geerdes[1], Holger Karl[2]
[1]Konrad-Zuse-Zentrum für Informationstechnik, Berlin
Address : Takustr. 7, D–14195 Berlin
Phone: +49 (30) 8 41 85–278, Fax: +49 (30) 8 41 85–269
Email: geerdes@zib.de
[2]Technical University Berlin, Telecommunication Networks Group
Address: Sekr. FT5-2, Einsteinufer 25, D–10587 Berlin
Phone: +49 (30) 3 14-23826, Fax: +49 (30) 3 14-23818
Email: hkarl@ieee.org

Abstract

Relaying is a protocol extension for cellular wireless computer networks; in order to utilize radio resources more efficiently, several hops are allowed within one cell. This paper investigates the principle potential of relaying by casting transmission scheduling as a mathematical optimization problem, namely, a linear program. We analyze the throughput gains showing that, irrespective of the concrete scheduling algorithm, performance gains of up to 30% on average for concrete example networks are achievable.

Wireless ad hoc networks, Resource Allocation, Integer Programming, Routing, Scheduling, Operation Research, Network Performance

1 Introduction

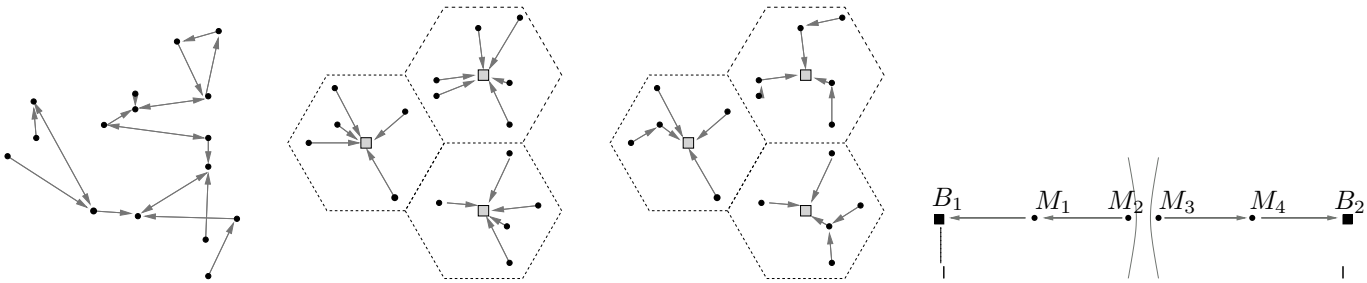
Wireless communication is becoming increasingly popular. The most important asset for delivering wireless services is *throughput*, the amount of data that can be transmitted per time unit. Throughput is limited: only a part of the radio spectrum is available and the presence of thermal noise and interference stemming from other users reduces the possible data rates even further.

1.1 Relaying in Cellular Networks

Telecommunication networks generally fall into two categories: *cellular* networks and *ad hoc* networks. The two principles are contrasted in Figures 1(a) and (b). In ad hoc networks, any two mobile nodes are free to exchange data by radio transmission. In cellular networks, on the other hand, nodes are usually allowed to transmit data only to or from fixed, special nodes called *base stations*.

To use *relaying* means to have mobile nodes acting as intermediaries for data exchange. This is a normal process in many ad hoc networks, but for cellular networks it is not normally used—it contradicts the very principle of cellular network organization. In the context of this work, relaying is examined as a way to improve the performance of cellular networks without changing their cellular nature. That is, we use relaying as a means to transport data to (and from) the base station within one cell, as shown in Figure 1(c), while maintaining the central control by base stations.

Using relaying increases the number of data transmissions, since several hops may be taken to the base station instead of a single one. Nevertheless, it can pay out to do so because it becomes increasingly expensive to send data across long distances: Modern radio technology, e.g., wireless LANs, adapt data rate to distance, allowing higher data rates over shorter distances. Overall, it might be more efficient to use several fast, short hops instead of one slow, long one. An extreme case is depicted in Figure 1(d): the two "inner" mobiles M_1 and M_4 close to the respective base stations B_1 and B_2 are in an ideal position to be used as intermediaries by the "outer" mobiles M_2 and M_3 far away from the base stations. Another potential benefit is the reduction of transmission power and, hence, interference, improving the transmission situation. The goal of this work is to quantify the actually achievable benefit in relaying.



(a) Ad hoc network (b) Cellular network (c) Relaying in cellular networks (d) Network with high benefit in relaying

Figure 1: Relaying in cellular networks

1.2 Assessing the Potential of Relaying by Mathematical Optimization

In order to achieve the *potential* benefits of relaying, we can rely on the foundations of cellular networking: we have the possibility to schedule transmissions in time and to prescribe which transmissions are carried out simultaneously—and we can do so across the network. Here lies the main difference of our work to ad hoc networking. In ad hoc networks this is impossible because a) nodes are not synchronized as tightly and b) no central entity has knowledge of the entire network.

Under smart scheduling and combination of transmission parameters, relaying increases the data throughput capacity for mini networks of the type shown in Figure 1(d) by up to 60% [7]. The maximal additional data throughput by using relaying is calculated for different distance parameters in [7]. There are also cases where direct communication is better. On the other hand, for asymptotically large ad hoc networks, the benefits of relaying actually vanish (a vast amount of research exists, e.g., the classical paper by Gupta and Kumar [4]).

Until now, no *general* method has been presented to find good transmission schedules in *arbitrary* cellular networks. Hence, some questions have remained open:

- Is there any substantial, general benefit in introducing relaying into cellular networks for realistically sized networks?
- For any given concrete network, what is the *best possible* throughput that can be achieved?
- Is it possible to quantify a potential relaying benefit *on the average* for set of network conditions? What network conditions particularly benefit from relaying?

These questions are answered in this work by defining the optimization problem of scheduling transmissions and determining transmission parameters for maximal data throughput in cellular networks using relaying, and by algorithms to find good solutions to this optimization problem.

1.3 Related Work

There is an array of information-theoretic work on questions of ad hoc networks, relaying, and capacity; [4] has already been mentioned. On the optimization side, a similar resource allocation problem is posed in [1] for ad hoc networks. The authors develop two mathematical formulations, one of which (called the *Link-Slot Formulation*) is similar to our setting. However, the important impact of different data rates is not modeled, in so far our model extends this approach.

Some work has been done on subproblems. The question of selecting power and transmission control is treated in [9] with a similar mathematical model. However, the authors do not consider routing issues here. In [2] some results and methods for routing relaying communication in cellular networks can be found, but since the authors consider CDMA networks, scheduling of communications in time is not considered.

In [10] a computational analysis on the achievable data throughput ("capacity regions") in ad hoc networks under idealized rate adaption is carried out. The authors study the behavior of the capacity region under varying assumptions, also including a special case of multihop routing in one single cell, in a related mathematical model. In comparison, our work falls on the more practical side, specific technology constraints are considered more accurately, which leads to stronger quantitative results.

2 The Optimization Problem

The problem we consider is to find a feasible transmission schedule for a fixed period of time that maximizes the amount of data that is transmitted to the base station. We limit ourselves to the uplink case and to a single frequency channel.

Underlying Technology. Our problem formulation and our problem instances rely on HiperLAN/2 [8], one of the many wireless network technologies. As is typical for next generation wireless networks, this technology supports rate and transmission power adaptation. It has also been chosen because its organized structure is amenable to a controlled and easy to analyze relaying schedule [5]. Despite its currently waning popularity, the result apply in principle to any current WLAN technology, in particular to the more common IEEE 802.11.

In HiperLAN/2, time is divided into intervals with a fixed length, called *frames*. At the beginning of each frame the base station transmits a transmission schedule to all of its mobiles. This schedule determines which stations may send/receive data units to/from the base station at what time, and it fixes transmission power and data rate for each transmission. We enable relaying by admitting transmissions between mobile stations; we then also have to require that all data units eventually reach the base station. Our goal is to find transmission schedules that maximize the data throughput achieved within one frame.

A transmission schedule consists of *transmission commands*, each command containing the following information: Link to be used (sender and receiver), data rate, transmission power and transmission time. A transmission schedule thus combines information related to several subproblems:

Intra Cell Routing. By specifying sender and receiver for each transmission, we determine which nodes should act as intermediaries and which path is followed by a data unit in order to reach the base station.

Scheduling. By specifying a transmission time for each transmission we coordinate which transmissions take place simultaneously. The combination of concurrent transmissions is crucial since they may interfere.

Rate and Power Adaption. The data rate to be achieved by a specific transmission can be chosen among several physical modes in HiperLAN/2. Transmission power can also be chosen within some range, but has to be adapted to the physical mode.

The transmission commands have to be composed such that no receiver suffers from too much interference from concurrent transmissions. A key figure for measuring the impact of interference is the so-called *Signal-to-Interference-and-Noise Ratio* (SINR). The SINR at a receiving station is the quotient of the strength of the wanted signal and the sum of all other signals plus noise. The interference situation at the receiver is feasible if the SINR stays above a certain threshold:

$$\frac{\text{Received Signal}}{\sum \text{Interfering Signals} + \nu} \geq \xi \tag{SINR}$$

The parameter ν represents the omnipresent thermal background noise. The threshold ξ depends on the data rate; it is higher for higher data rates.[1] When calculating the signal strength at the receiver, attenuation between the stations has to be taken into account.

Fairness Conditions. If one considers all transmission schedules that are feasible in the above sense, the ones that maximize the overall data throughput might lead to undesired results. Consider the example in Figure 2: Because mobiles M_3 and M_4 are at the cells' border one can achieve more throughput by only granting access to the "inner" mobiles M_1 and M_2 to their respective base stations B_1 and B_2. The outer mobiles would starve under a simple maximization scheme. Therefore, additional fairness constraints are required if this is not supposed to happen. This is a matter of interpretation, and different fairness schemes are possible. We have implemented a notion of *total user fairness*: each mobile station should get the same amount of data throughput, and this common throughput is to be maximized. (An alternative fairness model leading to different optimization problems and solutions is presented in [3].)

3 Model and Method

We merely outline the structure of our mathematical model here, a detailed description can be found in [3]. The model is two-tiered. We construct approximative transmission schedules from sets of simultaneously conducted transmissions called *transmission patterns*. In the first tier (see Section 3.1), we successively combine transmission patterns to ever better feasible schedules. We do not consider all possible transmission patterns at once, since there

[1] In more detail, this quotient determines the bit error rate (BER). But as higher-layer protocols usually do not work properly with too high a BER, it is appropriate to fix a certain maximal BER and, hence, minimal ξ.

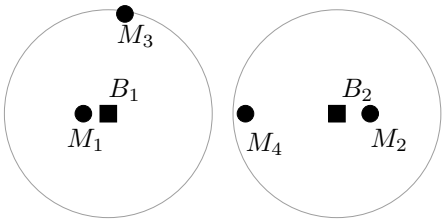


Figure 2: Without fairness conditions, the outer mobiles would starve

are too many of them. Instead of that, the second tier (see Section 3.2) produces transmission patterns that are useful for improving the current solution. The two tiers can be used within a column generation scheme.

3.1 Combining Transmission Patterns

We specify a transmission schedule by expressing for how much time which transmission patterns are executed in the network. This is only possible since we apply a relaxation of the problem: we pretend that data units may be arbitrarily fragmented. Transmission schedules then correspond to *weightings* of all transmission patterns, the sum of all weights being bounded by the available frame time.

We use a matrix Φ having one column for each possible transmission pattern. The row elements of Φ, with one row for each mobile station, contain the data throughput per time slot that the station achieves under the referring transmission pattern. (Since the base station are mere "sinks" in the uplink case, they need not be considered explicitly.) The *absolute value* of a matrix coefficient reflects the transmission data rate: higher data rates result in higher coefficients; a coefficient of zero, on the other hand, means that the referring station is not communicating in the referring transmission pattern. The *sign* of an entry reflects data flow direction: incoming data (the node is used as relay and will have to forward the received data) is counted negative. We group the weights in a vector $\boldsymbol{\eta}$ having as many components as there are transmission patterns. The net throughput per station can now be calculated as the product of Φ and $\boldsymbol{\eta}$. After introducing an auxiliary variable $\bar{\alpha}$ for the minimum throughput, the optimization problem reads:

$$\begin{array}{llcll} \max & \bar{\alpha} & & & \text{Maximize minimum throughput} \\ \text{s.t.} & \mathbf{1}\boldsymbol{\eta} & \leq & T & \text{Available time is bounded} \\ & \Phi\boldsymbol{\eta} & \geq & \bar{\alpha} & \text{Calculate minimum throughput} \\ & \boldsymbol{\eta} & \geq & 0 & \end{array} \tag{LP}$$

Since data units are atomic in reality—they may only be sent entirely or not at all—the fractional solutions $\boldsymbol{\eta}$ to (LP) do not correspond directly to feasible transmission schedules. However, by rounding off all fractional data units we can correct this and obtain feasible transmission schedules.

3.2 Generating Transmission Patterns

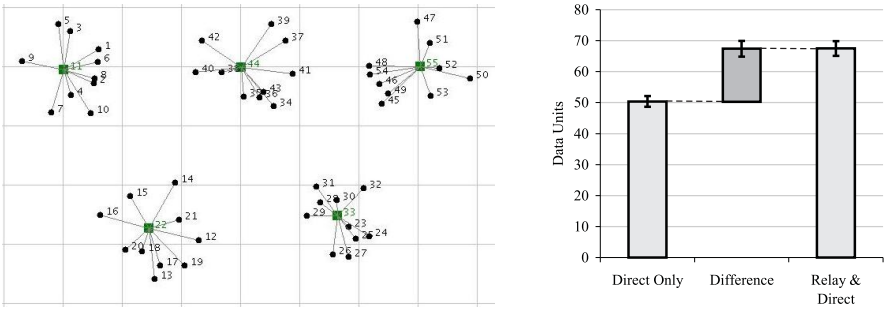
Since the matrix Φ has too many columns to be considered explicitly, the second tier of our model is responsible for generating good transmission patterns, that is, columns of Φ. This corresponds to solving the *pricing problem* for the linear program described in the previous section. We model the set of all possible transmission patterns as a mixed-integer linear program (MIP). All transmissions that might be included in a transmission pattern are denoted by binary variables $x \in \{0, 1\}$, and the associated transmission power is specified in related variables p. If $x = 1$, this means that the referring transmission takes place, otherwise the transmission is not scheduled. We denote the set of all stations by S and the set of available physical modes (corresponding to data rates) by M. The index set of all possible transmissions is then

$$I := M \times S \times S.$$

We have the following variables

$$\begin{aligned} p_{mij} &\in [0, P_{\max}] \;\; \forall\, (m, i, j) \in I \\ x_{mij} &\in \{0, 1\} \;\; \forall\, (m, i, j) \in I \end{aligned}$$

For an optimal solution to a version of (LP) containing only selected columns, let y^* denote the dual variable corresponding to the time constraint and let $\boldsymbol{w}^*$ be the vector of dual variables for the minimum throughput



(a) Example instance network

(b) Additional data throughput per mobile

Figure 3: Computational results for a network class with five cells and 10 mobiles per cell

calculation constraints. Transmission patterns that can improve the current transmission schedule correspond to assignments of values to all x and p variables so that the following inequalities hold:

$$\begin{array}{lll} \sum_{(m,i,j)} \left(\frac{w_i^* - w_j^*}{l_m}\right) x_{mij} > y^* & & \text{Dual Infeasibility in (LP)} \\ \sum\limits_{j \in S} \sum\limits_{m \in M} (x_{mij} + x_{mji}) \ \leq \ 1 & \forall\, i \in S & \text{Transmit } \textit{or} \text{ Receive} \\ P_{\min} x_{mij} \ \leq \ p_{mij} \ \leq \ P_{\max} x_{mij} & \forall\, (m,i,j) \in I & \text{Power Range} \\ \gamma_{ij} p_{mij} + \xi_m SG_\infty (1 - x_{mij}) \geq & & \\ \quad \xi_m \left(\nu + \sum_{(n,s,r) \neq (m,i,j)} \gamma_{si} p_{nsr}\right) & \forall\, (m,i,j) \in I & \text{SINR Feasibility} \end{array} \qquad \text{(PR)}$$

The first constraint changes with the current schedule, since $\boldsymbol{w}^*$ and y^* depend on the current solution to (LP). It ensures that only transmission schedules are produced that can increase the data throughput achieved by the current schedule. The second constraint ensures that any station either transmits or receives or is idle at any time. The semantics between the x and p variables are ensured by the Power Range inequalities. They enforce that $p = 0$ if and only if $x = 0$, otherwise a feasible transmission power has to be chosen. The SINR constraints are the most complex ones. The presented form is a linearized version of (SINR) that uses a number SG_∞ that bounds the sum of all signal and interference values; the values γ represent signal attenuation.

3.3 Algorithm

The following algorithm based on column generation computes an optimal (fractional) schedule for any given network configuration:

1. Initialize (LP) without any transmission patterns.
2. Solve the current (LP) and obtain a solution $\boldsymbol{\eta}^*$ and dual variables $\boldsymbol{w}^*$, y^*.
3. If there are any feasible solutions to (PR),
   insert the corresponding columns into (LP), go back to 2.
4. Otherwise $\boldsymbol{\eta}^*$ is optimal.

The solution $\boldsymbol{\eta}^*$ corresponds to a transmission schedule with fractional data units. We used a simple rounding heuristic (described in [3]) to obtain solutions corresponding to entire data units. For solving the occurring LPs and MIPs we resorted to the optimization software suite Ilog Cplex 7.5 [6]. Relaying can be disabled by ruling out all transmissions to mobile stations as solutions to (PR) in Step 3.

4 Computational Results

We were able to compute good transmission schedules for small to medium-sized networks on a standard PC within a few hors. The biggest class for which the problems could be solved in reasonable time consisted of five cells, each cell containing ten mobile nodes (cf. the example instance in Figure 3(a)). For the given set of five cells we randomly varied the position of ten fixed terminals in each disc-shaped cell. The signal attenuation between the

hosts is calculated using a simple distance-dependent path loss model. Note that the cells do not cover the whole area, they are separated by approximately one cell diameter. This is because adjacent cells would use different frequencies to enable efficient frequency reuse; these cells are neglected here.

The resulting number of data units that each mobile terminal could transmit with and without relaying were considered samples of a random variable. We then proceeded taking samples until the respective confidence intervals for the average throughput became satisfyingly small. The results for the chosen network with five cells are depicted in Figure 3(b): relaying can increase data throughput by about 30% *on the average* (assuming uniform distribution of mobiles in the cells). The error bars represent 5%-confidence intervals. Further experiments showed that relaying pays off more as channel conditions degrade, and that the average benefit of relaying increases under the presence of more mobile hosts.

5 Conclusions

We have addressed open questions related to relaying in wireless cellular networks. Based on realistic parameters, we have shown for an example class of networks that the *average* additional data throughput by using relaying is 30% in the considered scenarios. This average benefit is measured in terms of additional data throughput per time unit; the average is taken under the assumption of uniform mobile host placement. This result was obtained by developing a mathematical programming problem and deriving algorithmic tools for its solution. Some practical issues have not been considered in full detail yet, but the results represents a strong incentive to carry out the engineering task of realizing the disclosed potential. The presented method can serve as an engineering benchmark.

Acknowledgments

We thank Andreas Eisenblätter, Thorsten Koch, and Arie Koster for extensive mathematical support. Seble Mengesha has helped in engineering and technical questions.

References

[1] P. Björklund, P. Värbrand, and D. Yuan. Resource optimization of spatial TDMA in ad hoc radio networks: A column generation approach. In *Proc. IEEE INFOCOM*, San Francisco, CA, March 2003.

[2] M. Bronzel, W. Rave, P. Herhold, and G. Fettweis. Interference reduction in single-hop relay networks. In *Proc. 11th Virginia Tech/MPRG Symposium on Wireless Personal Communications*, pages 49–60, June 2001.

[3] Hans-Florian Geerdes. Capacity improvements in TDMA-based cellular networks by relaying and flexible transmission scheduling. Master's thesis, TU Berlin, 2002. Available online at `http://www.zib.de/groetschel/students/thesis_relaying.ps.zip`.

[4] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388–404, March 2000.

[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.

[6] ILOG, S.A., Paris, France. *ILOG CPLEX 7.5 User's Manual*, November 2001.

[7] H. Karl and S. Mengesha. Analysing capacity improvements in wireless networks by relaying. In *Proc. of IEEE Intl. Conf. on 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.

[9] S. Kim, Z. Rosberg, and J. Zander. Combined power control and tranmission rate selection in cellular networks. In *Proc. of the 49th IEEE Vehicular Technology Conference (Fall)*, pages 1653–1657, Amsterdam, The Netherlands, September 1999.

[10] Stavros Toumpis and Andrea Goldsmith. Capacity regions for wireless ad hoc networks. In *Proc. Intl. Conf. on Communications (ICC)*, 2002.