

# Design of Zone Tariff Systems in Public Transportation

Horst W. Hamacher

Department of Mathematics, Universität Kaiserslautern, 67663 Kaiserslautern, Germany, hamacher@mathematik.uni-kl.de

Anita Schöbel

Department of Numerical and Applied Mathematics, University of Göttingen, 37083 Göttingen, Germany,  
schoebel@math.uni-goettingen.de

*Dedicated to our doctor father and grandfather, Prof. Dr. E. Rainer Burkard, on the occasion of his 60th birthday*

Given a public transportation system represented by its stops and direct connections between stops, we present OR models for two problems dealing with the prices for the customers. In the first, the *fare problem*, subsets of stops are already aggregated to zones and “good” tariffs have to be found in the existing zone system. Closed form solutions for the fare problem are presented for three objective functions. The second problem, the *zone problem* includes the design of the zones. In an exemplary way we study this problem for one of the objectives. It is NP hard and we therefore propose three heuristics which prove to be very successful in the redesign of one of Germany’s transportation systems.

*Subject classifications:* integer programming: algorithms, heuristics, applications; transportation: models, traffic.

*Area of review:* Transportation.

*History:* Received May 2001; revisions received September 2002, January 2003, July 2003; accepted July 2003.

## 1. Tariff Systems in Public Transportation

In this paper, we deal with the design of tariff systems in public transportation. This complex real-world problem was brought to our attention by a regional public transportation company several years ago. While working on the design of a fair tariff system we developed a mathematical theory and a visualization tool to evaluate the effects of tariff changes. We present our studies and report on our practical experience when designing zone tariff systems.

When using a bus or a train a passenger usually pays for a trip. There are several possibilities for defining ticket prices in public transportation.

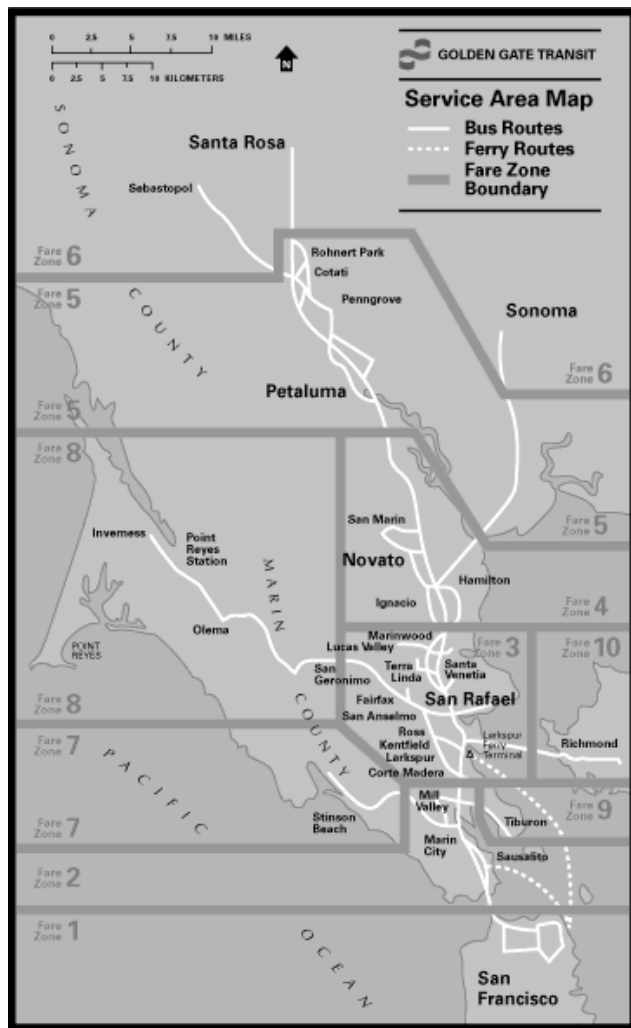
- In a *distance tariff* system, the trip price is dependent on its length. The longer the trip, the higher the fare. This system is generally considered to be fair. To determine the ticket prices one needs the distance between a pair of stations. This makes a distance tariff inconvenient for the public transportation company and for the customers.
- The simplest tariff system is the *unit tariff*. In this case all trips cost the same, independent of their length. A unit tariff is very easy to handle, but the public often does not accept that a short trip between two neighboring stations costs the same as a long trip through the whole system.
- A model between these two tariff systems is a *zone tariff* system. To establish a zone tariff, the area is divided into subregions (the *tariff zones*). The price for a trip in a zone tariff system is dependent only on the trip’s starting and ending zones. If the price can be chosen arbitrarily for each pair of zones, we call the tariff system a *zone tariff with arbitrary prices*. An example of such a tariff system can, for

instance, be found north of San Francisco (Figure 1). The prices are given in form of a matrix (Table 1).

The most popular variant of a zone tariff system is the *counting zone tariff* system. To know his fare in this system a customer counts how many zones his trip will pass and reads off the price assigned to the number of crossed zones. The prices in this system are dependent on the starting and the ending zones of the trip, but trips passing the same number of zones must have the same price. Figure 2 shows a counting zone tariff system south of San Francisco; the corresponding price for a single one-way trip is in Table 2.

Because of their simplicity, zone tariff systems are very popular. In Germany, nearly all tariff associations already have or are introducing zone tariff systems. When a public transportation company wants to change its tariff system to a zone tariff system, it has to design the zones and establish new fares, such that the resulting tariff system is accepted by the customers and does not decrease the income of the company. The goal often is to design the zones in such a way that the new and old prices for most of the trips are as close as possible. This means that neither the public transportation company nor the customers will have major disadvantages when changing the current tariff system to a zone tariff system.

Another goal can be to design fair zones. In this case we do not consider the deviation from old prices, but the deviation from a reference price, that is, one that is considered to be fair such as the distance tariff. In this approach, the public transportation company needs to estimate its new income.

**Figure 1.** Zone tariff system with arbitrary prices.

Source: <http://www.transinfo.org>.

In spite of the importance of this zone design problem there has been limited literature on corresponding operations research (OR) models. The only literature we are aware of deals with the zone design problem with arbitrary prices (Hamacher and Schöbel 1995; Schöbel 1994a, 1996; Babel and Kellerer 2001). For the zone design problem in which we count the number of zones, to the best of our

**Table 1.** Fares (in U.S. dollars) for a one-way trip.

| Zone | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
|------|------|------|------|------|------|------|------|------|------|------|
| 1    | 2.15 |      |      |      |      |      |      |      |      |      |
| 2    | 2.35 | 1.50 |      |      |      |      |      |      |      |      |
| 3    | 2.95 | 1.50 | 1.50 |      |      |      |      |      |      |      |
| 4    | 3.55 | 1.50 | 1.50 | 1.50 |      |      |      |      |      |      |
| 5    | 5.05 | 3.55 | 2.95 | 2.35 | 2.15 |      |      |      |      |      |
| 6    | 5.70 | 4.10 | 3.55 | 2.95 | 2.15 | 2.15 |      |      |      |      |
| 7    | 4.10 | 3.00 | 3.00 | 3.00 | 5.05 | 5.70 | 1.50 |      |      |      |
| 8    | 4.75 | 3.00 | 3.00 | 3.00 | 4.45 | 5.85 | 3.00 | 1.50 |      |      |
| 9    | 2.95 | 1.50 | 1.50 | 1.50 | 4.45 | 5.05 | 3.00 | 3.00 | 1.50 |      |
| 10   | 4.75 | 2.95 | 2.35 | 2.95 | 4.45 | 5.05 | 4.75 | 4.10 | 3.55 | 1.45 |

**Figure 2.** Counting zone tariff system.

Source: <http://www.transinfo.org>.

knowledge, there is no literature dealing with suitable OR models.

In this paper, we present an optimization model for the latter problem. The remainder of this paper is organized as follows. In the next section, we present our model for the zone design problem with counting zones. In §3, we consider the fare problem and show how the fares for each number of passed zones can be calculated easily by closed-term formulas for three different objectives. In §4, we take one of these objectives to study the zone problem. We discuss complexity issues, develop bounds, and propose three solution algorithms for designing good zones. We discuss their numerical behavior in a real-world example in §5. Finally, we draw some conclusions in §6.

## 2. A Model for the Counting Zone Tariff

Let the station graph  $G = (V, E)$  of the public transportation company be given, where  $V$  refers to the set of stops and  $E \subseteq V \times V$  represents the available direct connections without intermediate stops. Furthermore, let  $d_{ij}$  be a reference price for traveling from station  $i \in V$  to station  $j \in V$ .  $d_{ij}$  can be the current ticket price of the public transportation company, or it can be a fair price such as a distance tariff.

If  $L$  denotes the number of planned zones, the zone (planning) problem identifies a partition

$$P = \{V_1, V_2, \dots, V_L\}$$

of  $V$  (i.e.,  $V_i \subseteq V, i = 1, 2, \dots, L$ , pairwise disjoint and  $\bigcup_{i=1}^L V_i = V$ ). In the fare (planning) problem ticket prices

$$c(p), \quad p = 0, 1, 2, \dots,$$

are determined that are only dependent on the number of zones  $p$  in the journey. Here,  $c(p)$  is the price for crossing

**Table 2.** Fares by number of zones in journey for one-way trips.

| Number of zones      | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    |
|----------------------|------|------|------|------|------|------|------|------|------|
| Fare in U.S. dollars | 1.25 | 2.00 | 2.75 | 3.25 | 4.00 | 4.75 | 5.25 | 6.00 | 6.75 |

$p$  zone borders. In particular,  $c(0)$  gives the fare for traveling within any zone,  $c(1)$  is the price for crossing one zone border, i.e., for going from one zone to an adjacent one, and so on.

To evaluate some partition  $P$  with a price vector  $c$ , we define for each pair of stations  $i, j \in V$ ,  $n_{ij}$  as the number of crossed zone borders, when traveling from station  $i$  to station  $j$ . (Adding to the confusion most public transportation companies count the number  $n'_{ij}$  of passed zones on the trip from station  $i$  to station  $j$ , including both the starting and the ending zones, i.e.,  $n'_{ij} = n_{ij} + 1$ . We prefer our denotation for simplicity of our model.) The new ticket price for traveling from  $i$  to  $j$  is then given by

$$z_{ij} = \begin{cases} c(n_{ij}) & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

Given the reference prices  $d_{ij}$  for a trip between stations  $i$  and  $j$ , the absolute deviation in ticket price is calculated by

$$|d_{ij} - z_{ij}| = |d_{ij} - c(n_{ij})| \quad \text{for all } i, j \in V.$$

Let  $w_{ij}$  be the number of customers traveling from station  $i$  to station  $j$  and let  $W = \sum_{i,j \in V} w_{ij}$  be the sum of all customers of the public transportation company. The minimization of the following three objective functions is of interest:

- Maximum absolute deviation:

$$b_{\max} = \max_{i,j \in V} w_{ij} |d_{ij} - z_{ij}|.$$

- Average absolute deviation:

$$b_1 = (1/W) \sum_{i,j \in V} w_{ij} |d_{ij} - z_{ij}|.$$

- Average squared deviation:

$$b_2 = (1/W) \sum_{i,j \in V} w_{ij} (d_{ij} - z_{ij})^2.$$

All three objectives are considered to be good models by practitioners. The first objective function,  $b_{\max}$  with identical weights, models the fact that the greatest deviation of ticket prices in the two different tariffs should be as small as possible. It gives a bound for the largest change in the ticket price for any customer. In the weighted case,  $b_{\max}$  minimizes the maximum deviation in the revenue of the company over all possible trips.  $b_1$  gives the average of all absolute deviations, and  $b_2$  gives the average of all squared deviations in ticket prices. The objective function  $b_2$  leads to a smaller percentage of strongly affected customers than  $b_1$ . Nevertheless, from our experience,  $b_1$  is slightly better accepted by the practitioners than is  $b_2$ . It also should be mentioned that deviations in

price increases and decreases are treated equally, such that the model reflects both the interests of customers and of transportation companies.

We denote two zones  $V_k, V_l$  as *adjacent* if there exist stops  $i \in V_k, j \in V_l$ , such that  $(i, j) \in E$ , i.e., with a direct connection in the station graph  $G$ . To obtain the numbers  $n_{ij}$  a shortest path algorithm (e.g., Floyd 1962, Warshall 1962) can be used according to one of the following models.

**Station Graph Model.** We use the station graph  $G = (V, E)$ , but introduce new weights  $u_{ij}$  for all  $(i, j) \in E$ , defined by

$$u_{ij} = \begin{cases} 0 & \text{if } i \text{ and } j \text{ are in the same zone,} \\ 1 & \text{if } i \text{ and } j \text{ are in adjacent zones.} \end{cases}$$

The length of a shortest path between two stops equals the minimum number of crossed zone borders. This approach will be needed later to update the zone distances in the greedy heuristic in §4.

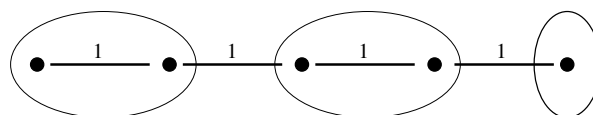
**Zone Graph Model.** To reduce the size of the network we define the zone graph  $G' = (P, E')$  whose node set  $P$  is given by the zones and  $(V_k, V_l) \in E'$  if  $V_k$  and  $V_l$  are adjacent. All edges have weight 1. For  $i \in V_k$  and  $j \in V_l$  we get the minimum number of crossed zone borders  $n_{ij}$  on a trip from  $i$  to  $j$  as the length of a shortest path from  $V_k$  to  $V_l$  in  $G'$ .

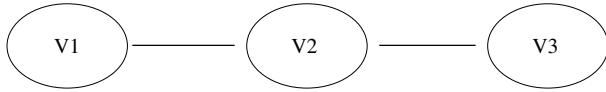
The following example demonstrates the calculation of  $b_{\max}$ ,  $b_1$ , and  $b_2$ . Let a station graph  $G$  with a partition into three zones  $V_1 = \{1, 2\}$ ,  $V_2 = \{3, 4\}$ , and  $V_3 = \{5\}$  be given (see Figure 3).

Suppose that  $w_{ij} = 1$  for all  $i, j \in V, i \neq j$ , i.e.,  $W = 20$ . If we assume that the distance between any adjacent pair of nodes is 1, the matrix  $d_{ij}$  according to the distance tariff system can be

$$D = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{pmatrix}.$$

**Figure 3.** A station network with five stations and three zones.



**Figure 4.** Zone graph with three zones.

The corresponding zone graph  $G'$  consists of three nodes (see Figure 4). The number of crossed zone borders between stations  $i$  and  $j$  is then given by

$$N = \begin{pmatrix} 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 & 2 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 2 & 2 & 1 & 1 & 0 \end{pmatrix}.$$

Suppose the new fares for crossing  $p = 0, 1$ , or  $2$  zone borders are given by

$$c(0) = 0.5, \quad c(1) = 1, \quad c(2) = 1.5.$$

The new ticket prices can be calculated as

$$Z = \begin{pmatrix} 0 & 0.5 & 1 & 1 & 1.5 \\ 0.5 & 0 & 1 & 1 & 1.5 \\ 1 & 1 & 0 & 0.5 & 1 \\ 1 & 1 & 0.5 & 0 & 1 \\ 1.5 & 1.5 & 1 & 1 & 0 \end{pmatrix}.$$

The deviations between the reference prices  $d_{ij}$  and the new ticket prices  $z_{ij}$  are

$$D - Z = \begin{pmatrix} 0 & 0.5 & 1 & 2 & 2.5 \\ 0.5 & 0 & 0 & 1 & 1.5 \\ 1 & 0 & 0 & 0.5 & 1 \\ 2 & 1 & 0.5 & 0 & 0 \\ 2.5 & 1.5 & 1 & 0 & 0 \end{pmatrix}$$

and the objective values can be calculated as

$$b_{\max} = 2.5, \quad b_1 = \frac{20}{20} = 1, \quad b_2 = \frac{32}{20} = 1.6.$$

### 3. Solution of the Fare Problem with Fixed Zones

In this section, we solve the fare problem with respect to a given zone partition. Our first result shows that a closed-form solution is possible for each of the three objectives  $b_{\max}$ ,  $b_1$ , and  $b_2$  introduced in §2.

**THEOREM 1.** Let  $P = \{V_1, V_2, \dots, V_L\}$  be a given zone partition and let  $d_{ij}$  be given reference prices. To minimize  $b_{\max}$ ,  $b_1$ , and  $b_2$  we choose for all  $p = 0, 1, \dots, L$ ,

$$c_{\max}^*(p) := \max_{\substack{i, j \in V: i \neq j, \\ n_{ij} = p}} d_{ij} - \frac{z_p^*}{w_{ij}}, \quad (1)$$

where  $z_p^*$  is defined as

$$z_p^* = \max_{\substack{i_1, j_1, i_2, j_2 \in V: \\ n_{i_1 j_1} = n_{i_2 j_2} = p}} \frac{w_{i_1 j_1} w_{i_2 j_2}}{w_{i_1 j_1} + w_{i_2 j_2}} (d_{i_1 j_1} - d_{i_2 j_2}).$$

$$c_1^*(p) := \text{median} \underbrace{\{d_{ij}, \dots, d_{ij}\}}_{w_{ij} \text{ times}}, \quad (2)$$

$$c_2^*(p) := \frac{1}{W} \sum_{i, j \in V: n_{ij} = p} w_{ij} d_{ij}. \quad (3)$$

**PROOF.** Given the zone partition  $P$  we have to find fares  $c(p) \in \mathbb{R}$  for all  $p = 0, 1, \dots$ , minimizing  $b_{\max}$ ,  $b_1$ , and  $b_2$ , respectively. Define

$$M_p = \{(i, j): i, j \in V \text{ and } n_{ij} = p\}$$

and  $W_p = \sum_{m \in M_p} w_m$  as the sum of all weights belonging to pairs of stations in the set  $M_p$ . First, we note that each of the three objective functions can be separated into at most  $L + 1$  independent subproblems,  $K_{\max}(p)$ ,  $K_1(p)$ , and  $K_2(p)$ , respectively (for  $p = 0, 1, \dots, L$ ).

$$\begin{aligned} b_{\max} &= \max_{i, j \in V} w_{ij} |d_{ij} - z_{ij}| \\ &= \max_{p=0, 1, \dots, L} \max_{m \in M_p} w_m |d_m - c(p)| =: \max_{p=0, 1, \dots, L} K_{\max}(p), \end{aligned}$$

$$\begin{aligned} Wb_1 &= \sum_{i, j \in V} w_{ij} |d_{ij} - z_{ij}| \\ &= \sum_{p=0}^L \sum_{m \in M_p} w_m |d_m - c(p)| =: \sum_{p=0}^L K_1(p), \end{aligned}$$

$$\begin{aligned} Wb_2 &= \sum_{i, j \in V} w_{ij} (d_{ij} - z_{ij})^2 \\ &= \sum_{p=0}^L \sum_{m \in M_p} w_m (d_m - c(p))^2 =: \sum_{p=0}^L K_2(p). \end{aligned}$$

Consequently, to minimize  $b_{\max}$ ,  $b_1$ , and  $b_2$  we determine the optimal fare  $c(p)$  for  $p = 0, 1, \dots, L$  separately, in each of the three objective functions.

• For  $b_{\max}$ : For all  $p = 0, 1, \dots, L$ , the problem of minimizing

$$K_{\max}(p) = \max_{m \in M_p} w_m |d_m - c(p)|$$

is well-known from location theory when locating a point on a line such that the maximum distance to a given set of existing facilities on the same line is minimized. The proof for (1) can therefore be found in the location literature; see e.g., Love et al. (1988), Hamacher (1995). Note that  $K_{\max}(p) = z_p^*$ .

• For  $b_1$ : Because

$$K_1(p) = \min_{m \in M_p} \sum w_m |d_m - c(p)|$$

is a one-dimensional, piecewise linear and convex function, its minimization is known in statistics (see e.g., Hays 1981) and in location theory as the one-dimensional median problem (see, e.g., Hamacher 1995, Plastria 1995). It is shown that the above problem is solved by the so-called weighted median of the set  $\{d_m: m \in M_p\}$ , i.e., by any real number  $c = c_1^*(p)$  which satisfies

$$\sum_{m: d_m < c} w_m \leq \frac{W_p}{2} \quad \text{and} \quad \sum_{m: d_m > c} w_m \leq \frac{W_p}{2}.$$

- For  $b_2$ : Here we have to minimize  $K_2(p)$ , i.e.,

$$\min \sum_{m \in M_p} w_m (d_m - c(p))^2.$$

Using the theorem of Steiner (see e.g., Sarkadi and Vincze 1974) of statistics, we note that the weighted mean of the values in  $\{d_m: m \in M_p\}$  is the unique optimal solution for  $c(p)$ . Q.E.D.

To demonstrate the result of Theorem 1 we continue the example of §2. The optimal values for the zone prices and the resulting values for the objective functions  $b_{\max}$ ,  $b_1$ , and  $b_2$  are listed in Table 3.

Calculating the objective function by using (1), (2), and (3) according to Theorem 1 yields the following corollary.

**COROLLARY 1.** *Given a zone partition  $P = \{V_1, V_2, \dots, V_L\}$  and reference prices  $d_{ij}$ , the optimal values of the objective functions are given as follows:*

$$b_{\max}^* = \max_p z_p^*,$$

$$b_1^* = \sum_p \left( \sum_{(i,j) \in V_p^+} w_{ij} (d_{ij} - c_1^*(p)) + \sum_{(i,j) \in V_p^-} w_{ij} (c_1^*(p) - d_{ij}) \right),$$

where  $V_p^+ := \{(i, j): n_{ij} = p, d_{ij} > c_1^*(p)\}$  and  $V_p^- := \{(i, j): n_{ij} = p, d_{ij} < c_1^*(p)\}$ ,

$$b_2^* = \sum_p \text{Var}\{d_{ij}: i, j \in V, n_{ij} = p\},$$

where  $\text{Var}$  denotes the variance of the set.

**Table 3.**

| Zones      | $c_{\max}$ | $c_1$ | $c_2$ | Example |
|------------|------------|-------|-------|---------|
| 0          | 1          | 1     | 1     | 0.5     |
| 1          | 2          | 2     | 11/6  | 1       |
| 2          | 3.5        | 3     | 3.5   | 1.5     |
| $b_{\max}$ | 1          | 1     | 1.167 | 2.5     |
| $b_1$      | 8          | 8     | 8.667 | 1       |
| $b_2$      | 7          | 8     | 6.722 | 1.6     |

In practice, restrictions are often given on the new fares; sometimes there are even politically desired fares that have to be realized for the number of zones in the journey. With the help of Corollary 1, one can easily calculate the increase of the objective functions when using such given fares instead of the optimal ones. In particular, Corollary 1 shows that for the objective function  $b_{\max}$  the optimal fares  $c_{\max}^*(p)$  are not needed to calculate the optimal objective value for a given zone partition. This will be needed in the next section when we optimize the zone partition with respect to  $b_{\max}$ . If, additionally,  $b_{\max}$  is used in the unweighted case, i.e., with  $w_{ij} = 1$  for all  $i, j \in V$ , we can further simplify Theorem 1 and Corollary 1.

**COROLLARY 2.** *In the case of equal weights, the optimal fares  $c_{\max}^*(p)$  and the corresponding objective value  $b_{\max}$  are given by*

$$c_{\max}^*(p) = \frac{1}{2} \left( \max_{i, j \in V: n_{ij}=p} d_{ij} + \min_{i, j \in V: n_{ij}=p} d_{ij} \right), \quad (4)$$

$$K_{\max}(p) = \frac{1}{2} \left( \max_{i, j \in V: n_{ij}=p} d_{ij} - \min_{i, j \in V: n_{ij}=p} d_{ij} \right), \quad (5)$$

$$b_{\max}^* = \frac{1}{2} \max_{p=1, \dots, L} \left( \max_{i, j \in V: n_{ij}=p} d_{ij} - \min_{i, j \in V: n_{ij}=p} d_{ij} \right). \quad (6)$$

**PROOF.** We calculate  $z_p^*$  as

$$\begin{aligned} z_p^* &= \max_{m_1, m_2 \in M_p} \frac{w_{m_1} w_{m_2}}{w_{m_1} + w_{m_2}} (d_{m_1} - d_{m_2}) \\ &= \frac{1}{2} \left( \max_{m \in M_p} d_m - \min_{m \in M_p} d_m \right), \end{aligned}$$

and consequently,

$$\begin{aligned} c_p^* &= \max_{m \in M_p} \left( d_m - \frac{z_p^*}{w_m} \right) \\ &= \max_{m \in M_p} \left( d_m - \frac{1}{2} \max_{\tilde{m} \in M_p} d_{\tilde{m}} + \frac{1}{2} \min_{\tilde{m} \in M_p} d_{\tilde{m}} \right) \\ &= \frac{1}{2} \left( \max_{m \in M_p} d_m + \min_{m \in M_p} d_m \right). \end{aligned}$$

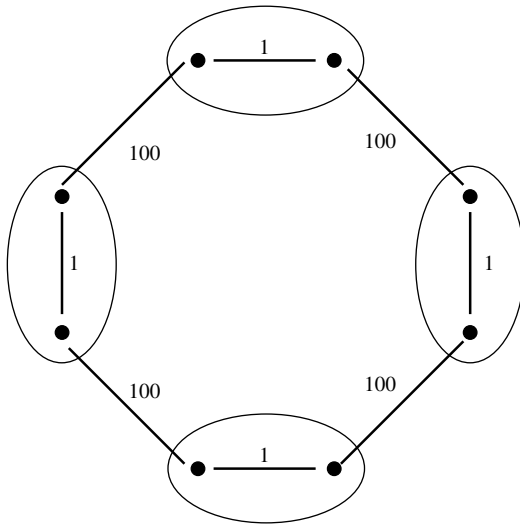
Using Corollary 1 and  $z_p^* = K_{\max}(p)$ , the remaining parts follow immediately. Q.E.D.

We remark that similar results can be derived for the zone design problem with arbitrary prices (see Hamacher and Schöbel 1995, Schöbel 1994b).

## 4. Finding Zone Partitions for the Maximum Deviation Problem

The consequence of the results of §3 is that we can concentrate on finding the zones, because the zone pricing follows easily from the choice of the objective function. We now focus our attention on the maximum deviation problem. Unfortunately, this problem is NP-hard and therefore difficult to solve.

**Figure 5.** A station network where the objective value for  $L = 4$  is better than the objective value for  $L = 5$ .



A first observation deals with the monotonicity of the objective function dependent on the number of planned zones  $L$ . Whereas it is easy to see that for the zone design problem with arbitrary prices all three objectives are monotone in  $L$ , this is not true for the zone design problem with counting zones, as Figure 5 shows. The station network consists of eight nodes, and we assume that  $w_{ij} = 1$  for all pairs of nodes  $i, j$ . The reference prices are given as weights between any two adjacent nodes, as shown in the figure. Between any other pair of nodes the reference prices are given as the sum of the weights along a shortest path connecting the nodes. For the (unweighted) max absolute deviation problem, Corollary 2 shows that any solution with  $L = 5$  zones leads to a strictly higher objective value than the graphed solution with  $L = 4$  and  $b_{\max} = 1$ . We will therefore fix  $L$  in the following.

**THEOREM 2.** *The zone design problem with counting zones and objective function  $b_{\max}$  is NP-hard for all fixed  $L \geq 3$ .*

The proof of Theorem 2 is given in the appendix. Note also that the zone design problem with arbitrary prices is NP-hard (Babel and Kellerer 2001).

To motivate the heuristics of this section, we first present the following two observations for getting upper and lower bounds on the objective value  $b_{\max}$ .

**LEMMA 1.**

$$b_{\max} \leq \frac{1}{2} \left( \max_{i,j \in V} d_{ij} - \min_{i,j \in V} d_{ij} \right).$$

**PROOF.** For any zone partition  $P$  and any integer  $p$ , we have by (5) that

$$\begin{aligned} K_{\max}(p) &= \frac{1}{2} \left( \max_{i,j \in V, n_{ij}=p} d_{ij} - \min_{i,j \in V, n_{ij}=p} d_{ij} \right) \\ &\leq \frac{1}{2} \left( \max_{i,j \in V} d_{ij} - \min_{i,j \in V} d_{ij} \right). \end{aligned}$$

Hence,  $b_{\max} = \max_{p=1, \dots, L} K_{\max}(p)$  also satisfies this inequality. Q.E.D.

**LEMMA 2.** *Given a zone partition  $P$ , let  $\text{INT} \subseteq E$  be the set of edges with both end nodes within the same zone and  $\text{BET} = E \setminus \text{INT}$ . Then, we have*

$$(1) \quad b_{\max} \geq \frac{1}{2} \left( \max_{(i,j) \in \text{INT}} d_{ij} - \min_{(i,j) \in \text{INT}} d_{ij} \right),$$

$$(2) \quad b_{\max} \geq \frac{1}{2} \left( \max_{(i,j) \in \text{BET}} d_{ij} - \min_{(i,j) \in \text{BET}} d_{ij} \right).$$

**PROOF.**

(1) Because

$$\begin{aligned} \min_{i,j \in V, n_{ij}=0} d_{ij} &\leq \min_{(i,j) \in \text{INT}} d_{ij} \text{ and} \\ \max_{i,j \in V, n_{ij}=0} d_{ij} &\geq \max_{(i,j) \in \text{INT}} d_{ij}, \end{aligned}$$

we get

$$\begin{aligned} b_{\max} &\geq K_{\max}(0) \\ &\geq \frac{1}{2} \left( \max_{i,j \in V, n_{ij}=0} d_{ij} - \min_{i,j \in V, n_{ij}=0} d_{ij} \right) \\ &\geq \frac{1}{2} \left( \max_{(i,j) \in \text{INT}} d_{ij} - \min_{(i,j) \in \text{INT}} d_{ij} \right). \end{aligned}$$

(2) Analogously,  $\text{BET} \subseteq \{i, j \in V, n_{ij} = 1\}$ , hence we get the result by using  $b_{\max} \geq K_{\max}(1)$ . Q.E.D.

Lemma 2 suggests a zone design in which edges with high weights are collected in BET and edges with small weights in INT, or vice versa. To be more specific, let  $\text{Diam}$  be the maximal diameter over all zones. Assuming that edge weights along a path are additive, we get, again using (5),

$$\begin{aligned} K_{\max}(p) &= \frac{1}{2} \left( \max_{i,j \in V, n_{ij}=p} d_{ij} - \min_{i,j \in V, n_{ij}=p} d_{ij} \right) \\ &\leq \frac{1}{2} \left( (p+1)\text{Diam} + p \left( \max_{(i,j) \in \text{BET}} d_{ij} - \min_{(i,j) \in \text{BET}} d_{ij} \right) \right), \end{aligned}$$

yielding that the maximal diameter  $\text{Diam}$  should be small, and consequently edges with large weights should be in BET while edges with small weights should be in INT.

Following these considerations, we present three heuristics for solving the zone design problem with counting zones. As input data we need—for any of the following algorithms—a set of  $n$  stations with reference prices  $d_{ij}$  and a number  $L$  of planned zones. The output is then given by a zone partition with  $L$  zones.

### Algorithms Based on Clustering Theory

The first algorithm is based on ideas from clustering theory, and in particular on the sequential agglomerative hierarchical nonoverlapping (SAHN) algorithms (see, e.g., Duran and Odell 1974). The idea is to start with  $n$  zones, each

of which contains a one single station, and to combine in each step the two closest zones to a new one. Depending on the particular definition of the distance between two zones, different algorithms can be obtained. Two of them have been applied to the zone design problem: single linkage and complete linkage.

ALGORITHM 1 (ZONE DESIGN USING SAHN ALGORITHMS).

*Step 1.* Start with a partition  $P$  consisting of  $n$  zones, each of which contains a single station. Let  $d(V_i, V_j) := d_{ij}$  for all zones  $V_i, V_j \in P$ .

*Step 2.* Determine two zones  $V_i \neq V_j \in P$  with minimum distance  $d(V_i, V_j)$ .

*Step 3.* Join  $V_i$  and  $V_j$  to a new zone  $V_k$  and get a new partition  $P$ .

*Step 4.* Calculate the new distances for all  $V \in P$ :  $d(V_k, V) := \frac{1}{2}(d(V_i, V) + d(V_j, V) + c|d(V_i, V) - d(V_j, V)|)$ .

*Step 5.* If the number of planned zones is attained, then Stop, Output  $P$ , else go to Step 2.

The parameter  $c$  in Step 4 determines the formula for calculating the distance between two zones. In the context of the zone design problem, we have used

- $c = -1$  for the single linkage algorithm, and
- $c = 1$  for the complete linkage algorithm.

The interpretation for single linkage is the following: The distance between two zones is defined as the smallest distance between elements of the zones; consequently in each step we join along a shortest edge. Note that in complete linkage the distance between two zones is defined as the maximum distance between their elements, and in each step complete linkage tries to minimize the maximum diameter of the zones.

## Greedy Approach

This approach is a variant of the SAHN algorithms discussed above, but with more emphasis on the specific structure of the zone design problem. Using the basics of Algorithm 1, we calculate for all edges  $(i, j)$  the objective value  $b_{\max}^{ij}$  when contracting  $(i, j)$  of the current zone graph. Finally, we contract the edge with the smallest increase in the objective function. This is rather time consuming, but, as we will show in the next section, leads to very good results in practice. The formulation of the greedy approach is the following:

ALGORITHM 2 (ZONE DESIGN BY GREEDY APPROACH).

*Step 1.* For all edges  $(i, j) \in E$  with  $n_{ij} = 1$ : Contract  $i$  and  $j$  temporarily and calculate  $b_{\max}^{ij}$ .

*Step 2.* Contract the edge  $(i^0, j^0)$  permanently, where

$$b_{\max}^{i^0, j^0} = \min\{(i, j): b_{\max}^{ij}\},$$

and let  $n_{i^0, j^0} = 0$ . If the graph has  $L$  nodes, Stop.

*Step 3.* For all  $i, j \in V$ , recalculate  $n_{ij}$  as shortest distance, and goto Step 1.

## Spanning Tree Approach

The idea of the following heuristic is to determine a set of edges BET that contains mostly edges with high weights.

ALGORITHM 3 (ZONEV DESIGN BY SPANNING TREE APPROACH).

*Step 1.* Find a maximum spanning tree  $T$  in the complete graph with edge weights  $d_{ij}$ .

*Step 2.* Omit the  $L - 1$  largest edges of  $T$  and get a forest with  $L$  components.

*Step 3.* Output: Zones are the connected components.

Note that in trees, the spanning tree approach is equivalent to the single linkage algorithm of clustering theory. In general graphs, it is always possible to find a spanning tree such that omitting its  $L - 1$  largest edges leads to the same result as single linkage. However, if we start with a spanning tree with maximal weight (which performed best in practice), the spanning tree approach differs significantly from single linkage.

## 5. Practical Experiences in Saarland, Germany

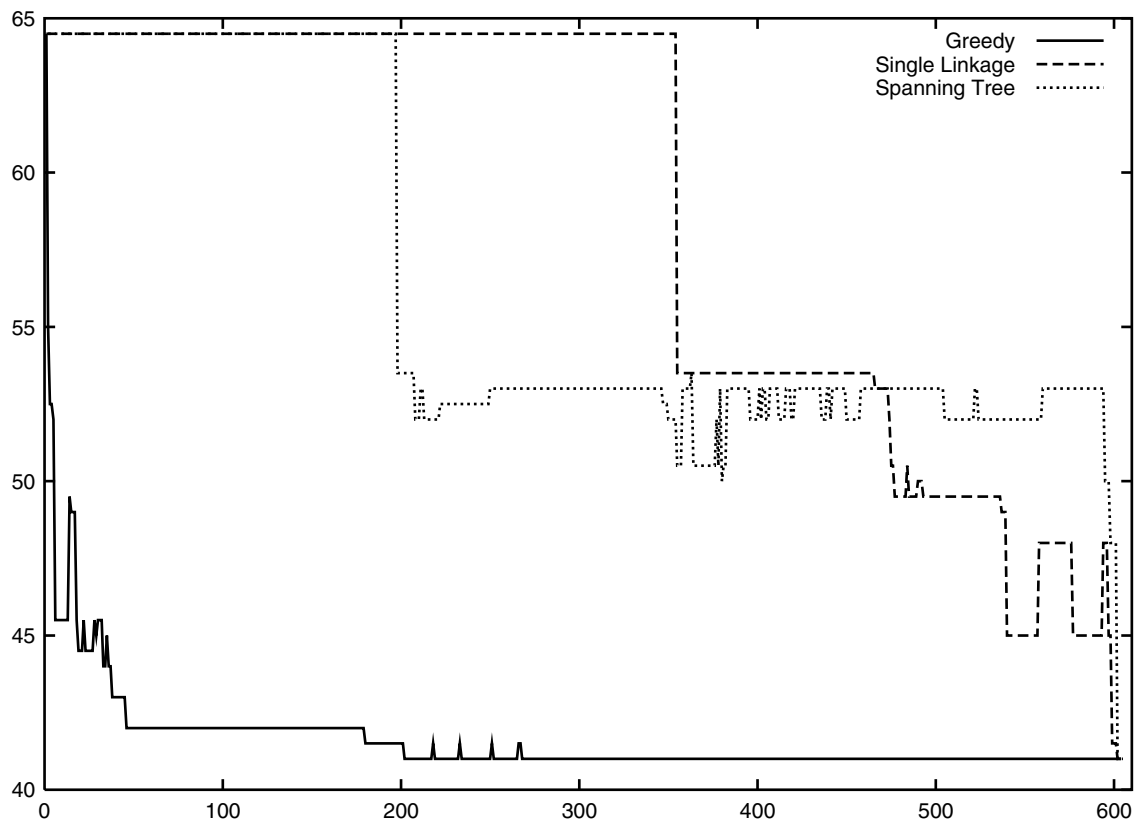
As an example for the practical value of our approach, we consider the situation in the state of Saarland, Germany. Currently, there are six public transportation companies operating in the Saarland, each with its own tariff system.

- Four public transportation companies already use a counting zone tariff system, but their fares for crossing  $p$  zones and the structure of their zones are completely different, although they are partly operating in the same geographical region.

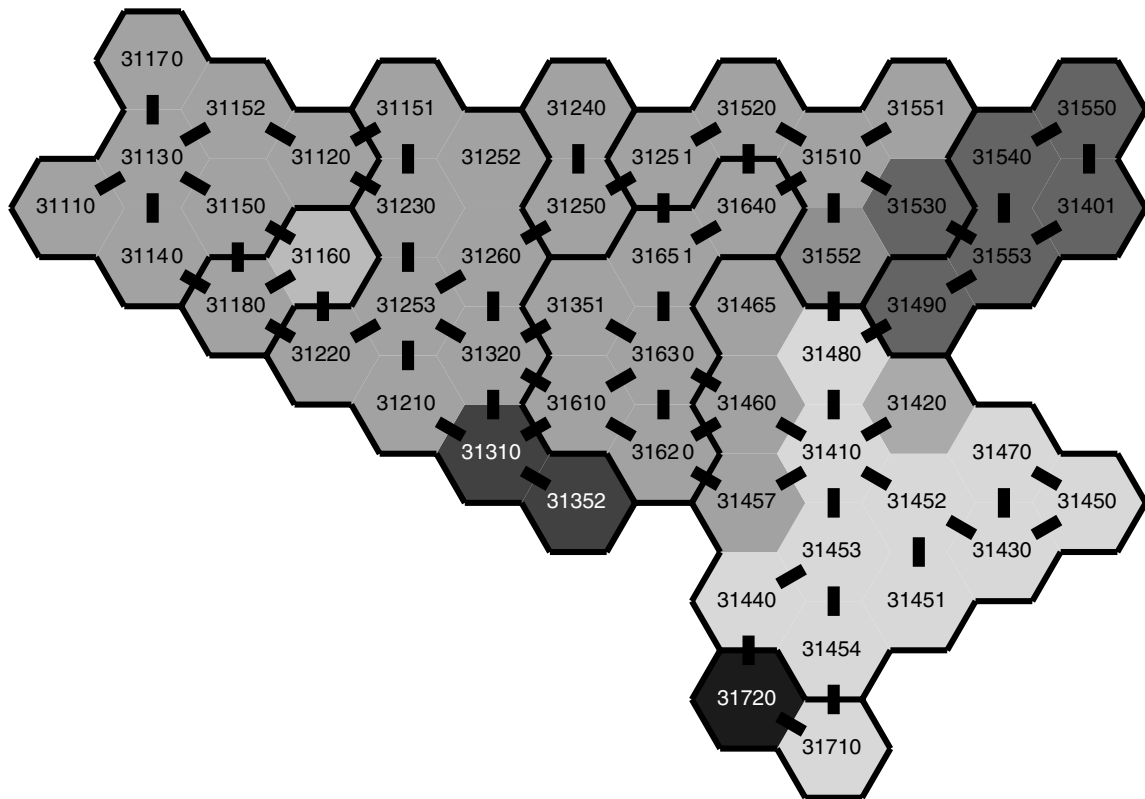
- The Deutsche Bahn (German Rail) still applies its distance tariff.

- There is also a public transportation company (serving the city of Saarland's capital, Saarbrücken) that uses a zone tariff with variable prices.

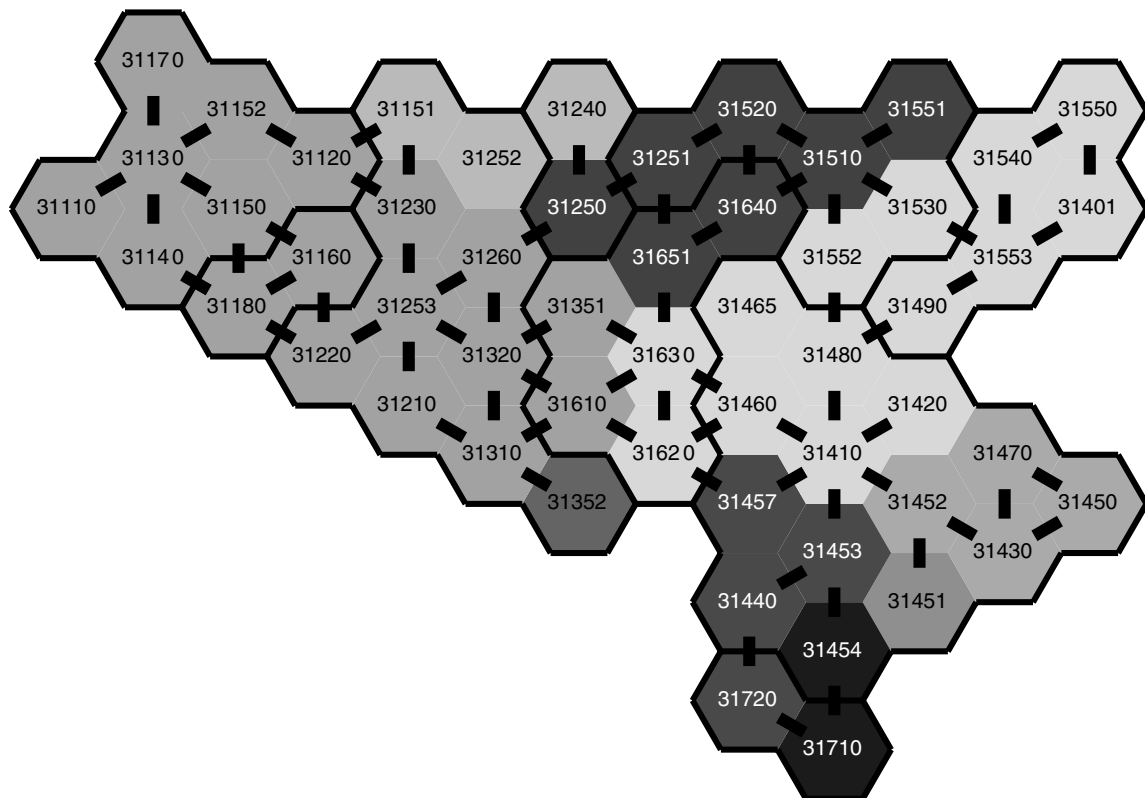
The traffic association of the Saarland is considering the introduction of one common counting zone tariff system that would be applied by all the public transportation companies. The public transportation network in Saarland consists of roughly 4,000 stations, where a preclustering into 600 minizones is given. The goal is to design about 100 zones and install a counting zone tariff system in such a way that the differences between the current and the new fares are as small as possible. It is also important that the new income of each of the public transportation companies not differ too much from its current income. Consequently, the reference prices in this application are the current prices for traveling. While the current fare structure is known and therefore relatively easy to get, it is usually hard to get

**Figure 6.** Comparison of the heuristics:  $b_{\max}$  graphed for any number  $L$  of planned zones.**Figure 7.** Political suggestion;  $b_{\max} = 5.15$  DM.

**Figure 8.** Solution of the single linkage algorithm;  $b_{\max} = 5.00$  DM.



**Figure 9.** Solution of the greedy algorithm;  $b_{\max} = 3.75$  DM.



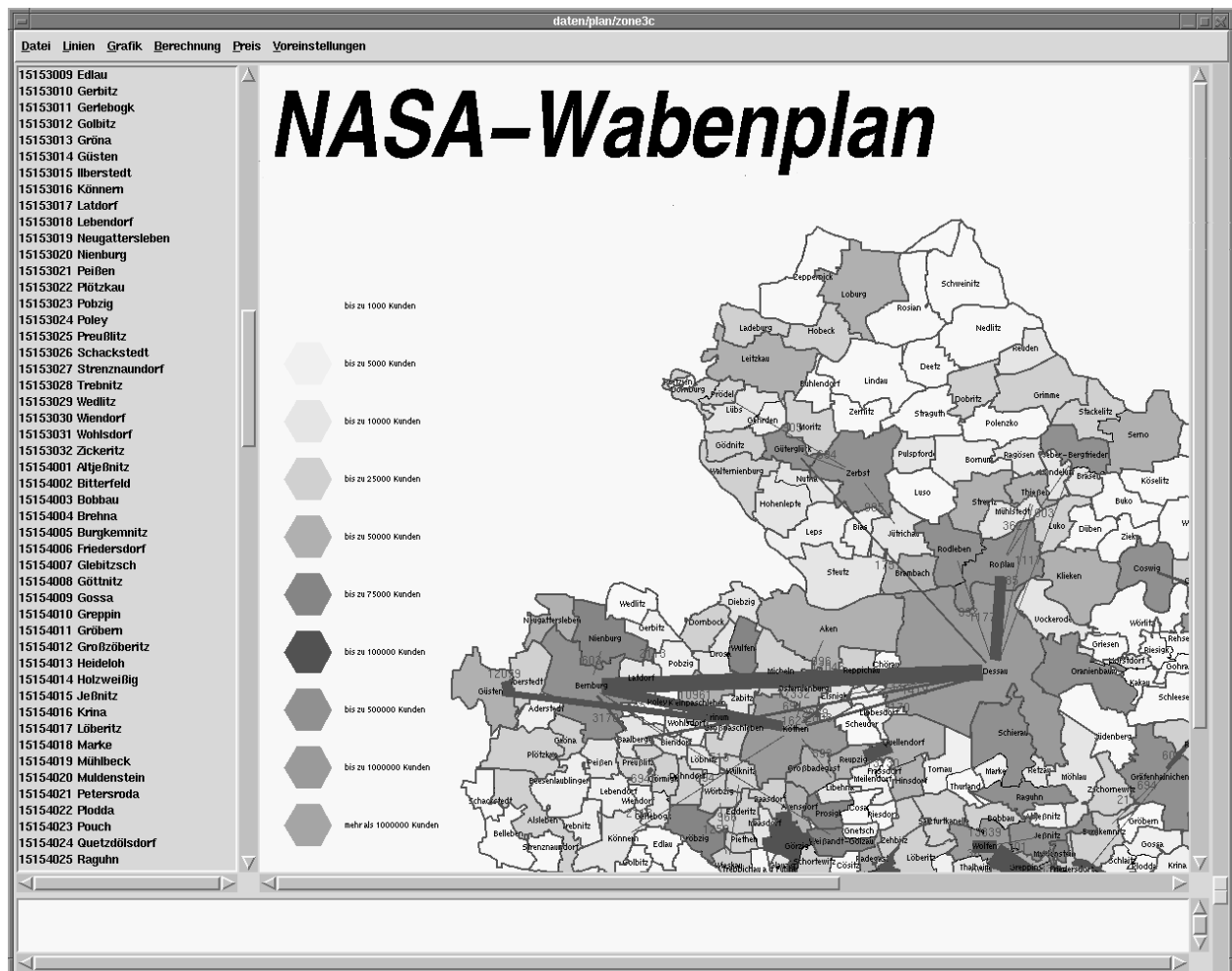
realistic data about the customers' behavior. In our project in the Saarland this was solved by using the income data of each of the transportation companies and dividing the income with the help of available statistics over the origin-destination pairs used by the customers.

We tested our algorithms on the data described above. The results of Algorithms 1, 2, and 3 are shown in Figure 6. This figure shows the objective value  $b_{\max}$  for any number of possible zones from 1 to 600. All objective values refer to a single-trip ticket for an adult, given in German Marks (DM). On the one hand, it turns out that in this practical application the greedy heuristic (Algorithm 2) is the clear winner in terms of the objective value: It generated the best results for any number of desired zones. On the other hand, the running time for Algorithm 2 for all possible numbers of zones, i.e., from  $L = 1, \dots, 600$ , was nearly two weeks altogether in our first implementation (on an AixJ90). The spanning tree approach (Algorithm 3) and the single linkage algorithm (Algorithm 1) both needed

only a few hours, but the results are much less convincing regarding the objective value  $b_{\max}$ , again. For a small number  $L$  of desired zones, single linkage did better than the spanning tree approach, whereas for a higher number of planned zones it was the other way around. This is due to the fact that the spanning tree approach starts with only one zone, while single linkage starts with 600 zones.

On a subset consisting of only 400 stations (or 54 mini-zones) the heuristics have also been tested. In this smaller setting the running times of Algorithms 1 and 3 were within seconds, and Algorithm 2 needed only two minutes to obtain again the clearly best results. The results for nine zones are shown graphically in Figures 7, 8, and 9. Figure 7 shows a suggestion for a zone partition that is due to the political districts in this area of the Saarland. The objective value for this zone partition is  $b_{\max} = 5.15$  DM, i.e., there exists a customer who will have a difference of 5.15 DM between his current fare and the new one. The result of the single linkage algorithm for nine zones is shown in

**Figure 10.** Zone planning with the software *WabPlan* in the state of Sachsen-Anhalt.



Source. The lines show the (fictional) customers who will have a change of more than 5% in their ticket prices.

Figure 8. Single linkage tends to form one large zone with several smaller zones surrounding it (see, e.g., Duran and Odell 1974). This behavior is also shown in Figure 8. The objective value of the graphed zone partition is 5.00 DM. The objective value in the spanning tree approach also was 5.00 DM for nine zones, but without these big differences in the zone sizes. The best results, however, were obtained by the greedy approach with an objective value of only 3.75 DM. The corresponding zone partition is shown in Figure 9.

For evaluating tariff zones in more detail, we use the software package WabPlan (Schöbel and Schöbel 1999). A graphical front-end provides a detailed analysis of all trips for which the fare will increase or decrease dramatically (see Figure 10). Furthermore, the expected income for each of the transportation companies in each ticket category is compared with its current income.

For practical purposes, many special rules for using fare zones are common. Several of these rules have also been implemented in our algorithms and tested on the Saarland data.

**Empty Zones.** First, in most zone tariff systems empty zones are used to increase the fare on some special trips without affecting all other relations. This seems to make sense in practice and can easily be incorporated in the algorithms presented in §4. In this way, given reference prices can be approximated arbitrarily close if the number  $L$  of zones is large enough. In our model this means that the optimal objective value goes to zero for  $b_{\max}$ ,  $b_1$ , and  $b_2$  in this case.

**Border Stations.** To avoid injustice, stations can be located on zone borders, meaning that they belong to more than one zone, and the cheapest choice for determining the fare applies. Because the zone tariff system should be clear and understandable, we tried to avoid this in the Saarland. In most cases it turned out that border stations can be avoided without losing anything in the objective values by changing only the zone design.

**Special Rules for Large Zones.** Also, some zones might be so large that they have to be counted twice when crossing them, in which case a special fare structure would have to be implemented.

In the Saarland, the tariff system proposed by using our methods is now in the implementation process.

## 6. Conclusion

In this paper, we have presented OR models for the counting zone tariff problem. For fixed zones we have shown that closed-form solutions can be provided. In contrast, the zone problem, i.e., the design of zones, is NP-hard. Three heuristics have been proposed and compared with respect to their numerical behavior. The practical usefulness of the approach is shown by its actual implementation in the state of Saarland, Germany. Other German states are currently

using our system to evaluate their tariff structures. Due to the importance of this problem, we hope that this contribution motivates further research.

## Appendix

**PROOF OF THEOREM 2.** We use a reduction to the problem partition into cliques for  $L = 3$  cliques, which is NP-hard (see problem GT15 in Garey and Johnson 1979). Let a graph  $G = (V, E)$  be given. We want to find out if there exists a partition of  $V$  into three node sets  $V_1, V_2$ , and  $V_3$  such that the induced subgraphs  $G_1, G_2$ , and  $G_3$  are complete. To reduce this problem to a zone design problem, we consider the complete graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , defined by

$$\mathcal{V} = V \cup \{a_1, a_2, a_3, b_1, b_2, b_3\} \quad \text{and}$$

$$\mathcal{E} = \{e = (k, l) : k, l \in \mathcal{V}, k \neq l\}.$$

Furthermore, let  $w_{ij} = 1$  for all  $i, j \in \mathcal{V}$  and define the reference prices as the following edge weights:

$$d_{kl} = \begin{cases} 1 & \text{if } (k, l) \in E, \\ 1 & \text{if } k \in V, l \notin V, \\ \frac{1}{2} & \text{if there exists } i = 1, 2, 3 \\ & \text{such that } (k, l) = (a_i, b_i), \\ 2 & \text{if } k, l \in V, (k, l) \notin E, \\ 2 & \text{if } k, l \notin V, (k, l) \neq (a_i, b_i) \\ & \text{for all } i = 1, 2, 3. \end{cases}$$

Now we show that  $G$  can be partitioned into three cliques if and only if the zone design problem in  $G$  has a solution with three zones and with  $b_{\max} < 3/4$ .

$\Rightarrow$ : Let  $V = V_1 \cup V_2 \cup V_3$  be the partition of  $G$  into cliques. Define  $C_i = V_i \cup \{a_i, b_i\}$  for  $i = 1, 2, 3$ . Using (5) we calculate

$$K_{\max}(0) = \frac{1}{2} \left(1 - \frac{1}{2}\right) = \frac{1}{4},$$

$$K_{\max}(1) = \frac{1}{2} (2 - 1) = \frac{1}{2},$$

such that we get  $b_{\max} < 3/4$ .

$\Leftarrow$ : Let  $C_1, C_2, C_3$  be a partition of  $\mathcal{V}$  with  $b_{\max} < 3/4$ . Define  $V_i = C_i \cap V$  for  $i = 1, 2, 3$ . First, we note that by renaming  $C_i$  we can assume that  $a_i, b_i \in C_i$  for  $i = 1, 2, 3$ . This can be proved by a simple case analysis (see Penner 1997) that verifies the following.

1. If  $a_i$  and  $b_i$  do not belong to the same zone, this yields  $b_{\max} \geq 3/4$ .

2. If  $a_i, b_i$  and another  $a_j$  or  $b_j$ ,  $j \neq i$ , belong to a single common zone, then  $b_{\max} \geq 3/4$ .

Now let  $k, l \in V_i$ . We have to show that the edge  $(k, l) \in E$ . Assume the contrary, i.e.,  $d_{kl} = 2$ . According to (5), this yields

$$K_{\max}(0) \geq \frac{1}{2} \left(2 - \frac{1}{2}\right) = \frac{3}{4},$$

a contradiction. For more than three zones, the proof can be done analogously with a reduction to partition into  $L > 3$  cliques. Q.E.D.

## Acknowledgments

The authors thank two anonymous referees for very helpful comments regarding the first version of this paper. This project was partially supported by the Deutsche Forschungsgemeinschaft under HA1795/7.

## References

- Babel, L., H. Kellerer. 2001. Design of tariff zones in public transportation systems: Theoretical and practical results. Technical report, Faculty of Economics, University of Graz, Graz, Austria.
- Duran, B. S., P. L. Odell. 1974. *Cluster Analysis—A Survey*. Lecture Notes in Economics and Mathematical Systems, no. 100. Springer, Berlin-Heidelberg-New York.
- Floyd, R. W. 1962. Algorithm 97: Shortest path. *Comm. ACM* **5** 345.
- Garey, M. R., D. S. Johnson. 1979. *Computers and Intractability—A Guide to the Theory of NP-Completeness*. H.W. Freeman and Co., San Francisco, CA.
- Hamacher, H. W. 1995. *Mathematische Lösungsverfahren für planare Standortprobleme* (Mathematical Solution Algorithms for Planar Location Problems). Vieweg, Braunschweig, Germany.
- Hamacher, H. W., A. Schöbel. 1995. On fair zone design in public transportation. *Computer-Aided Transit Scheduling*, no. 430. Lecture Notes in Economics and Mathematical Systems. Springer, Berlin, Germany and Heidelberg, Germany, 8–22.
- Hays, W. L. 1981. *Statistics*, 3rd ed. Holt, Rinehart and Winston, New York.
- Love, R. F., J. G. Morris, G. O. Wesolowsky. 1988. *Facilities Location*, Chapter 3.3. North-Holland, Amsterdam, The Netherlands, 51–60.
- Penner, K. 1997. Fahrpreisgestaltung im öffentlichen Verkehr. Master's thesis, Universität Kaiserslautern, Kaiserslautern, Germany.
- Plastria, F. 1995. Continuous location problems. Z. Drezner, ed. *Facility Location: A Survey of Applications and Methods*, Chapter 11. Springer, New York, 225–262.
- Sarkadi, K., I. Vincze. 1974. *Mathematical Methods of Statistical Quality Control*. Academic Press, New York and London, U.K.
- Schöbel, A. 1994a. Fair zone design in public transportation networks. *Operations Research Proceedings 1994*. Springer Verlag, Berlin, Germany, 191–196.
- Schöbel, A. 1994b. Methoden der kombinatorischen Optimierung in der Tarifplanung im öffentlichen Personennahverkehr (Methods of combinatorial optimization in tariff planning in public transportation). Master's thesis, Universität Kaiserslautern, Kaiserslautern, Germany.
- Schöbel, A. 1996. Zone planning in public transportation. L. Bianco, P. Toth, eds. *Advanced Methods in Transportation Analysis*. Springer Verlag, Berlin, Germany, 117–134.
- Schöbel, G., A. Schöbel. 1999. WabPlan—A software tool for design and evaluation of tariff systems.
- Warshall, S. 1962. A theorem on Boolean matrices. *J. ACM* **9** 11–12.