*Electronic Notes in Theoretical Computer Science 50 No. 1 (2001) { Proc. ATMOS 2001 URL:* [*http://www.elsevier.nl/locate/entcs/volume50.html*](http://www.elsevier.nl/locate/entcs/volume50.html) *11 pages*

*Locating New Stops in a Railway Network ?*

# *Horst W. Hamacher a;1, Annegret Liebers b;2, Anita Schobel a;3,*

*Dorothea Wagner b;4, Frank Wagner c;5*

*a University of Kaiserslautern, Department of Mathematics, and Fraunhofer Institute ITWM, 67653 Kaiserslautern, Germany*

*b University of Konstanz, Department of Computer and Information Science, 78457 Konstanz, Germany*

*c Deutsche Bahn AG, Konzernentwicklung GSV, Potsdamer Platz 2, 10785 Berlin, Germany*

*Abstract*

*Given a railway network together with information on the population and their use of the railway infrastructure, we are considering the e ects of introducing new train stops in the existing railway network. One e ect concerns the accessibility of the railway infrastructure to the population, measured in how far people live from their nearest train stop. The second e ect we study is the change in travel time for the railway customers that is induced by new train stops. Based on these two models, we introduce two combinatorial optimization problems and give NP-hardness results for them. We suggest an algorithmic approach for the model based on travel time and present a real-world application with its rst experimental results.*

# *1 Introduction*

*The motivation for the problems studied in this paper is to increase the at-* tractiveness of train travel in an existing railway network for local traÆc, for example to sports and leisure activities or for the daily commuter trip to work, in not too costly a manner. So assume we are given an existing railway network as well as information on people (potentially) using this network. A relatively cheap way of increasing the attractiveness of train travel is then to open up

*? This work was partially supported by the Human Potential Programme of the European Union under contract no. HPRN-CT-1999-00104 (AMORE).*

*1* [*hamacher@mathematik.uni-kl.de*](mailto:hamacher@mathematik.uni-kl.de)

*2* [*Annegret.Liebers@uni-konstanz.de,*](mailto:Annegret.Liebers@uni-konstanz.de) *supported by DFG grant Wa 654/10-2.*

*3* [*schoebel@mathematik.uni-kl.de*](mailto:schoebel@mathematik.uni-kl.de)

*4* [*Dorothea.Wagner@uni-konstanz.de*](mailto:Dorothea.Wagner@uni-konstanz.de)

*5* [*Frank.H.Wagner@bku.db.de*](mailto:Frank.H.Wagner@bku.db.de)

*c 2001 Published by Elsevier Science B. V. Open access under* [*CC BY-NC-ND license.*](http://creativecommons.org/licenses/by-nc-nd/3.0/)

*new train stops along the already existing railway lines, where a new train* stop is meant to be a small stop that is cheap to build, and that will typically be served only by the local commuter trains.

*We will consider two e ects of such new train stops. The rst one concerns* the accessibility of the railway network: Assuming that living closer to a train stop increases the attractiveness of the railway network (and ultimately increases the tendency to use a train instead of other means of transport), we study the problem of locating (few) new train stops so that (many) people live within a short distance from a train stop. This accessibility model will be speci ed more precisely in Section 2, and a corresponding optimization problem will be shown to be NP-complete.

*The second e ect of new train stops we will study concerns the travel* time for the customers already using the railway network. On the one hand, passengers sitting in a train that now stops at the newly opened train stops are slowed down by the additional stops. On the other hand, customers living closer to a newly opened train stop than to any previously existing train stop need less time to reach their closest train stop. Summing up each of these two contrary e ects over all railway customers and subtracting the sums, we obtain a measurement for the e ect in overall travel time that a set of new train stops is having for the existing railway customers. Section 3 will specify this travel time model and also give an NP-completeness result for the corresponding optimization problem. Section 4 will suggest an algorithmic approach for this optimization problem based on a genetic algorithm, and Section 5 will give experimental results for the complete railway network of the Deutsche Bahn AG in Germany.

# *2 The Accessibility Model*

*Let P denote the set of coordinate points, each of them representing a settle-* ment, and let L be the set of feasible points along the tracks of the railway company. We assume that the set of tracks is piecewise linear, i.e. it is given as the union of a nite number of line segments Si 2 IR2, i = 1;::: ; L:

*L = fx 2 IR2 : x 2 Si for some i = 1;::: ; Lg:*

*A solution X to the problem is given by a set of points in the plane,* representing the stops that should be installed. X is feasible, if X L, i.e., if each x 2 X satis es x 2 L.

*To evaluate a feasible solution X, we are interested in how many settle-* ments are covered by X. A settlement p 2 P is covered by a point x 2 L, if

*l2(p; x) r;*

*where l2 refers to the Euclidean distance (or any other norm), and r is some*

*given radius. A settlement p is covered by a set X L if*

*l2(p; x) r for some x 2 X:*

*For a set X we de ne*

*cover(X) := fp 2 P : l2(p; X) rg;*

*where l2(p; X) := minx2X l2(x; p) for jXj 1 and l2(p; ;) := 1. Figures 1 and*

*2 illustrate how di erent sets of train stops cover di erent sets of settlements,*

*using a radius of r = 2 km.*

*Furthermore, we are interested in the number of new stops: We would like* to cover as many settlements with as few stops as possible. This goal leads to the de nition of the following optimization problem, which unfortunately is NP-complete:

*De nition 2.1 [Covering Along Lines] Given a set P of integer-coordinate* points in the plane, a connected set L which is given as the union of a nite number of line segments in the plane, and positive integers d and K < jPj, can the points of P be covered by at most K discs of diameter d, all with center points in L?

*Theorem 2.2 Covering Along Lines is NP-complete.*

*Proof. We reduce Geometric Covering by Discs to the above problem. Geo-* metric Covering by Discs has been shown to be NP-complete [3] and can be stated as follows: Given a set P of integer-coordinate points in the plane and positive integers d and K < jPj, can the points of P be covered by at most K discs of diameter d?

*Given an instance of Geometric Covering by Discs, use the same set P* of points and the same numbers d and K and de ne the set L as follows to construct an instance of Covering Along Lines: For each unordered pair of points p1 and p2 from P,

*(i) add the points forming the line segment between p1 and p2 to L and*

*(ii) add the points forming a suÆciently large piece of the bisector of p1 and* p2 to L.

*Claim: P can be covered by at most K discs of diameter d, if and only if* P can be covered by at most K discs of diameter d which all have their center points in L.

*To see this, rst assume that P can be covered by some collection C,* consisting of at most K discs of diameter d. Then, for each disc c 2 C:

*If c contains no points of P, then disregard c.*

*If c contains only one point p of P, then replace c by the disc with center* point p (p is in L, since all line segments from p to any other point are in L) and diameter d.

*If c contains a set of points A of P with jAj 2 (i.e., c contains more than* one point of P), then replace c by a disc with center point q and diameter d, where q is the center point of the smallest enclosing circle of A.

*Since c covers A, the diameter of the smallest enclosing circle is smaller*

*or equal to d, and hence the disc with radius d and center point q also* covers A. Note that nding q is a well-known problem of location theory and can be done in linear time [4]. Moreover it is known that q always lies on at least one bisector of points in A [2], such that q satis es q 2 L.

*Together, P can be covered by at most K discs of diameter d, all with center* points in L.

*The other direction is immediate. 2*

*Note that in the special case of only one line segment in L, Covering Along* Lines turns out to be a set covering problem whose constraint matrix is an interval matrix. Hence, in this case the problem is polynomially solvable by linear programming. This result can be extended to sets of line segments, if no settlement can be covered by two stops from di erent line segments.

# *3 The Travel Time Model*

*Some more notation is needed to state the travel time model. We suppose* that L is given as a network G = (V; E) embedded in the plane. Note that all the already existing train stops are contained in V . For a point x 2 L let e(x) 2 E denote the edge of G on which x is located. Furthermore, for each edge e the number of customers traveling along e is denoted by we. Then for a set X of new stops,

*td(X) = X s we(x)*

*x2X*

*gives the amount of additional travel time for the customers, which is caused* by the additional stopping activities of the trains. (We assume that a constant time delay s is caused by any additional stop of a train.)

*On the other hand, some of the customers will save travel time, since a new* train stop may reduce the distance to their closest train stop and hence the time they need to get on a train. In our model, the reduction of the distance for a point p 2 P is calculated by

*l2(p; S) l2(p; S [ X);*

*where S V denotes the set of already existing train stops, and l2(p; Y ) is* the closest Euclidean distance from p to any point in Y as it has been used in the accessibility model before.

*To transform the possible reduction of distance into an amount of saved* access time we introduce a piecewise linear function g : IR IR ! IR in two variables, assigning an amount of saved time to each reduction of the distance,

*given as a pair consisting of old and new distance of a settlement from the* nearest train stop. For example, g can be de ned as

*g(x; y) := (x y)=5;* assuming an average speed of 5 km/h, or as

*8 x y*

*if x 1 (customer walks)*

*4*

*>*

*<*

*g(x; y) = x y*

*7*

*x y*

*>*

*if 1 < x 5 (customer uses a bike)*

*if 5 < x (customer uses a bus or a car)*

*: 20*

*Note that this de nition would assume that the customer stays with the means*

*of transport used for distance x.*

*Denoting for each p 2 P the number of customers from the corresponding* settlement by vp, the positive e ect to the travel time through saved access time can be calculated by

*ta(X) = Xvp g(l2(p; S); l2(p; S [ X)):*

*p2P*

*The travel time model can now be summarized:*

*such that*

*max ta(X) td(X)*

*x 2 L for all x 2 X:*

*Note that we neglect the change in train riding time that is caused by starting* or ending the trip at a di erent train stop, assuming that these gains and losses in train riding time roughly even out.

*We can now de ne the corresponding problem Saved Travel Time and show* its NP-completeness:

*De nition 3.1 [Saved Travel Time] Given a set P of integer-coordinate points* in the plane, each of them with a nonnegative weight vp, an embedded network G = (V; E) in the plane, each of its arcs with a nonnegative weight we, a sub- set S V , a positive number s and a real number K, and given a piecewise linear function in two variables g : IR IR ! IR, is there a set of points X on the edges of G such that

*f (X) := Xvp g(l2(p; S); l2(p; S [ X)) Xs we(x) K?*

*p2P x2X*

*Theorem 3.2 Saved Travel Time is NP-complete.*

*Proof. We reduce Covering Along Lines to Saved Travel Time. Given an in-* stance of Covering Along Lines with P, L, d and K < jPj as in De nition 2.1, construct a network G = (V; E) out of the line segments in L in the natural

*way (i.e. with V being the end points and the intersection points of the line* segments, and with E being the parts of the line segments obtained by subdi- viding them with their intersection points), use the same set of points P, and de ne

*Furthermore, de ne*

*r := d=2;* S := ;;

*vp := M for all p 2 P with M > jPj;* we := 1 for all e 2 E; and

*s := 1:*

*g(x; y) := <8*

*1 if y r*

*With these de nitions, we get that*

*:0 otherwise*

*f (X) = M jcover(X)j jXj* where the cover is de ned with respect to the set P.

*Claim: There exists a solution X to Saved Travel Time with f (X) MjPj K* if and only if there exists a solution to Covering Along Lines with no more than K discs.

*To see this, assume rst that X is a solution to Saved Travel Time with* f (X) MjPj K. Then cover(X) = P, for otherwise jcover(X)j jPj 1 and hence

*f (X) = M jcover(X)j jXj*

*M jcover(X)j*

*MjPj M*

*< MjPj jPj*

*< MjPj K*

*f (X);*

*a contradiction. Furthermore, cover(X) = P implies jXj K.*

*For the other direction, let X be a solution to Covering Along Lines with* cover(X) = P and jXj K. Hence,

*f (X) = M jcover(X)j jXj*

*= MjPj jXj*

*MjPj K:*

*2*

*The travel time model can be viewed as a location problem in which the* new facilities must be located on a given network, while the existing facilities are given as points in the Euclidean plane. (For a recent survey of theory and applications in the eld of location theory, see [1].) This means that neither planar location models nor network location theory really ts. Moreover, even in the case of locating only one new stop, the objective function of Saved Travel Time is not convex. For both reasons, the classical approaches of location theory are not applicable, and new solution algorithms have to be developed. Our rst approach | a genetic algorithm | is presented in the next section.

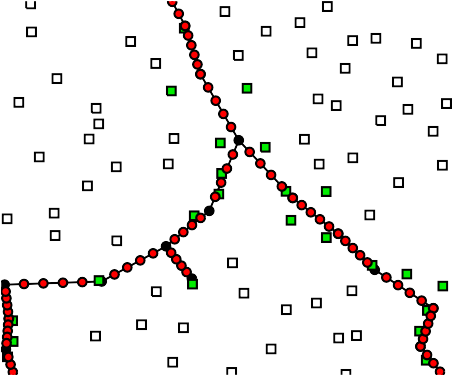
# *4 Algorithmic Approach*

*We propose to use a set of discrete candidates for new train stops, and to then* employ a genetic algorithm [5] for the maximization problem Saved Travel Time. Consider a simple set of candidates as illustrated on the left side of Fig. 2: Place four evenly spaced candidates on each line segment of the railroad network, and place a candidate at each end of such a line segment that is not an existing train stop. Clearly, if one of these candidates, c, does not by itself yield a positive saved travel time, then the saved travel time of a set X of new train stops containing c does not decrease if we drop c from X. So such a candidate c is never useful for a solution X, and we reduce the set of candidates by dropping every candidate c for which td(fcg) > ta(fcg). See the right side of Fig. 2 for an illustration. With this reduced set of candidates C for the travel time model, the natural application of a genetic algorithm works with bit vectors of length jCj where each bit stands for exactly one candidate. So 11...1 describes the whole set C, 00...0 describes the empty set, and every other bit vector of length jCj describes a set of new train stops consisting of some members of C.

*To generate a starting population of i individuals for a genetic algorithm,* we either use a xed probability p to determine for each of the i jCj bits whether they are set to 1, or, for each bit b, we rst determine a number pb 2 f0:001; 0:002;::: ; 0:999g randomly, and then use pb as the probability to determine whether b is set to 1.

*In order to generate the next generation from a given population, we divide* the given population in half, where one half contains the individuals with the highest saved travel time. Then among these \better" individuals, we repeat- edly choose two di erent individuals at random to \mate" them. We mate two individuals by rst randomly choosing a crossover point j 2 f1; 2;::: ; jCj 1g and by then applying the crossover by swapping the tails consisting of jCj j bits. For example, if two individuals 101100100 and 110110010 with jCj = 9 are mated with crossover point j = 4, then the two resulting individuals are 101110010 and 1101 00100. Finally, we ip some of the bits of the new indi- viduals with a very low probability pm (\mutation"). So the next generation consists of the better half of the old generation, plus the results of mating

*Fig. 1. Left:The situation with the existing train stops. Right: The solution found after 1000 generations as shown in Fig. 3. Black circles depict existing train stops, red (shaded) circles depict new train stops, and green (shaded) squares depict cov- ered settlements.*



*Fig. 2. Left:The raw set of candidates. Right: The reduced set of candidates. individuals from this better half with each other as just described.*

# *5 Application and Experimental Results*

*We applied the genetic algorithm described in Section 4 to a preliminary set* of input data describing the situation in Germany. There are 6 900 existing stops with polygonal edges consisting of 8 700 line segments, and there are

*30 600 settlements with their population and their estimated number of rail-*

*way customers. This status quo is depicted for a small excerpt in Southern* Germany on the left side of Fig. 1. The set of candidates used for the genetic algorithm consists of 6 700 potential new stops. The raw set of candidates and the reduced set of candidates are illustrated in Fig. 2. Assuming a time delay s of 2 minutes per additional stop and the function g(x; y) := (x y)=5 to describe the gain in access time when the distance to the nearest train stop is reduced from x to y km, we started the genetic algorithm with this

*set of candidates and with xed probabilities p 2 f0:25; 0:5; 0:75g three times* each for a population size of 20. The probability pm for mutation of bits after crossover was set to 0.0001. After 100 generations each, we let the result- ing population with the highest saved travel time so far evolve for another 900 generations. The development of this population over the course of its 1000 generations is shown in Fig. 3. For comparison, we started the genetic algorithm with varying initial probabilities pb for each bit b of each individ- ual with population size 100, and it turned out that the genetic algorithm

*nds the right \density" of the solutions very quickly, i.e. we actually do not*

*need to rst determine a good probability p for the creation of the starting* population: While the starting population with varying initial probabilities pb contained individuals with almost no candidates as well as individuals with almost all of the 6 700 candidates, the di erence between the lowest and the highest number of candidates in the individuals of the 10th generation had already shrunk to less than 1 000, and after 100 generations this di erence was a mere 22.

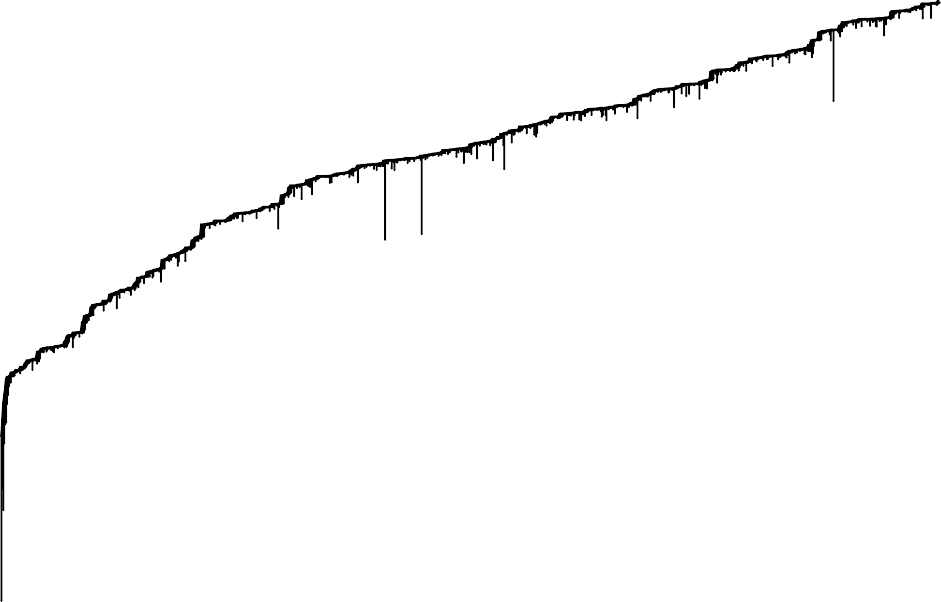
*The solution with the highest saved travel time found so far by our genetic* approach is illustrated on the right side of Fig. 1. It shows several candidates from the right side of Fig. 2 actually becoming new stops, and in this excerpt, the placement of new train stops looks reasonable and the development of this solution as depicted in Fig. 3 is promising. But when inspecting this solution more closely in other parts of Germany, we noticed that there are in some parts unrealistically many new stops in the solution. It turns out that such e ects are caused by inaccuracies in the input data: For example, unrealistically low numbers of customers traveling along edges make placing new stops along these edges too \cheap", resulting in too many new train stops.

*These and other problems mean that the solutions found by the genetic* algorithm are not to be taken as nal suggestions for placing new train stops in order to save travel time, and the amounts of saved travel time indicated on the vertical axis of Fig. 3 have to be read with caution. Rather, we have presented an algorithmic approach, that, if current problems with the input data are overcome, can yield useful results for the placement of new train stops when the goal is to maximize the saved travel time.

# *6 Conclusion*

*Studying the e ects of introducing new train stops in existing railway net-* works, we have considered two models, one based on the closeness of the train stops to the population, and the other based on maximizing the overall saved travel time by all customers. For both models, we have shown the NP-hardness of a naturally arising optimization problem. For the travel time model, we have applied a genetic algorithm and we have shown that it can yield useful solutions despite the NP-hardness results. For the rst model, however, an algorithmic approach still needs to be developed and applied. Here, the goal

12



11

10

saved travel time in mmin

9

8

7

0 200 400 600 800 1000

generations

*Fig. 3. The development of the objective function saved travel time over 1000*

*generations with population size 20.*

*would be to nd pareto-optimal solutions for the bicriterial problem with ob-* jective functions the percentage of population covered by a set of new stops (to be maximized) and the number of new stops (to be minimized).

*Furthermore, for both models, a re nement of the e ects under consid-* eration is also desirable: Instead of a settlement being either covered or not depending on its distance from a train stop being at most or more than a given radius r, one could consider a settlement being partially covered with the percentage of its covering depending in some more re ned way on its dis- tance from a train stop. Similarly, a more realistic function for converting a decrease in distance from a train stop into an amount of saved access time than simply assuming an average speed of 5 km/h is needed.

*Finally, other e ects of new train stops not considered in our two models* include building and operating costs. And it would be of interest to consider a dynamic model where a change in the numbers of customers due to new train stops is also taken into account.

# *Acknowledgement*

*We would like to thank Ste en Mecke for the implementations needed for this* work.

# *References*

*[1] Drezner, Z. and H. Hamacher, editors, \Location Theory: Applications and Theory," Springer, 2001.*

*[2] Elzinga, J. and D. Hearn, Geometrical solutions for some minimax location problems, Transportation Science 6 (1972), pp. 379{394.*

*[3] Johnson, D. S., The NP-Completeness Column: An Ongoing Guide,*

*J. Algorithms 3 (1982), pp. 182{195.*

*[4] Megiddo, N., Linear time algorithms for linear programming in IR3 and related problems, SIAM J. on Computing 12 (1983), pp. 759{776.*

*[5] Reeves, C. R., Genetic Algorithms, in: C. R. Reeves, editor, Modern Heuristic Techniques for Combinatorial Problems, McGraw Hill, 1995 pp. 151{196.*